CANADIAN THESES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THÈSES CANADIENNES

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

Canada

THE UNIVERSITY OF ALBERTA

PEPTIDE RACEMIZATION KINETICS

by

MICHAEL EDWARD MOIR

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

EDMONTON, ALBERTA

FALL 1985

ABSTRACT

The process of racemization in fossil proteins is modelled by heating some small peptides and peptide derivatives at 148.5° in pH 6.8 phosphate buffer. In addition, racemization of peptides containing one or two chiral centres is modelled by computer simulations of kinetic schemes of various levels of sophistication. In this way, the effects of peptide decomposition and asymmetric induction can be explored.

Asymmetric induction is observed during the heating of L-prolyl-L-leucylglycylglycine as excess D-proline is found in the total hydrolysis mixture. The D/L ratio for proline attains a maximum value of approximately 2.1 after 90 minutes. The D/L ratio then slowly approaches unity. A similar observation is recorded when L-prolyl-L-leucine is heated under the same conditions. The excess D-proline is attributed to the formation of a 2.3:1 mixture of the diketopiperazines cyclo-(D-prolyl-L-leucyl) and cyclo-(L-prolyl-L-leucyl).

These two species account for most of the leucine and proline in the final mixture after the peptides are no longer detectable. Only small amounts of prolylleucine can be detected after 50 hours. It is suggested that the above peptides undergo internal aminolysis.

The peptide derivatives blocked at the nitrogen

terminus show a decreased rate of racemization. This is interpreted to mean that the protonated nitrogen terminus promotes racemization. There is also some indication that the carboxylate anion at the C-terminus inhibits racemization.

Leucine in glycyl-L-leucylglycylglycine racemizes three times faster than in L-prolyl-L-leucylglycyl-glycine. This shows that an amino acid residue in a peptide chain does have an effect on the racemization of a neighbouring amino acid residue. Upon heating, glycyl-L-leucylglycylglycine gives cyclo-(glycylleucyl). As in the other cases where diketopiperazines were isolated, leucine racemizes fastest when in this form. Moreover, formation of diketopiperazines is competitive with racemization. It is questionable whether peptide racemization can be measured using the total hydrolysis mixture.

A discussion of the geochemical implications of the results is included.

ACKNOWLEDGEMENTS

I would like to thank Professor R.J. Crawford most sincerely for his patience and generosity while guiding me through this project. Many thanks are due to Professor Dennis D. Tanner for allowing me to use instrumentation in his laboratory. I am also indebted to the non-academic staff of the Department of Chemistry whose expertise in all fields made this work possible. Special thanks are to be given to Miss Annabelle Wiseman for her word-processing magic. To my wife also, for invaluable aid during the preparation of the thesis and for sharing both times of frustration and elation. Finally, I wish to thank the Department of Chemistry and the University of Alberta for financial support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

One of the primary concerns of geologists and archeologists is the reliability of the techniques used to determine the age of fossilized remains. What is needed is a method of dating which requires only a small amount of material, is simple to perform so that any reasonably equipped laboratory could do the analysis, has a fast turn around of samples and is inexpensive. The discovery of amino acids in fossils by Abelson[1] seemed to be most providential as the analysis of low concentrations of amino acids was well developed. The process of amino acid racemization (slow at ambient temperature) was seen to have promise as a potential tool for the dating of fossils.[2] Up to this point, radiocarbon dating* was the single most important dating method available but is generally considered to be accurate only to between 30000 and 50000 y.B.P.[3]

---

*Dating using tree-rings is also important. In areas where sufficient samples are available this technique can be extended to 8000 y.B.P. U-series dating is also available for a similar age range as radiocarbon dating and suffers from the same drawbacks.

The activation energy of the epimerization of amino acids is such that this limit could be extended. Until recently, the radiocarbon method required up to ten grams of sample. Consequently, it was impossible to date objects of small size or of great archeological value. Recent developments in radiocarbon dating methodology allow the sample size to be lowered to the milligram level.[3] Nevertheless radiocarbon dating is an expensive technique requiring dedicated instrumentation and thus increases the attraction of the amino acid method.

Geological dating by amino acid racemization is based on the conversion of a naturally occurring L amino acid to its D isomer. The interconversion depicted in Figure 1 which begins after the death of the organism (except in those tissues which are not repaired) has a half-life at ambient temperature varying from about 20000 to 300000 years depending on the amino acid used for dating.[4] Equation 1 describes amino acid racemization exactly.

$$\ln\left\{\frac{1 - K'\frac{[D]}{[L]}}{1 + \frac{[D]}{[L]}}\right\} - \ln\left\{\frac{1 - K'\frac{[D]}{[L]}}{1 + \frac{[D]}{[L]}}\right\}_{t=0} = -(k_1+k_{-1})t \qquad (1)$$

$$\text{where } K' = \frac{k_{-1}}{k_1}$$

A common procedure for extracting amino acids from fossils and measuring the D to L ratio of the individual amino

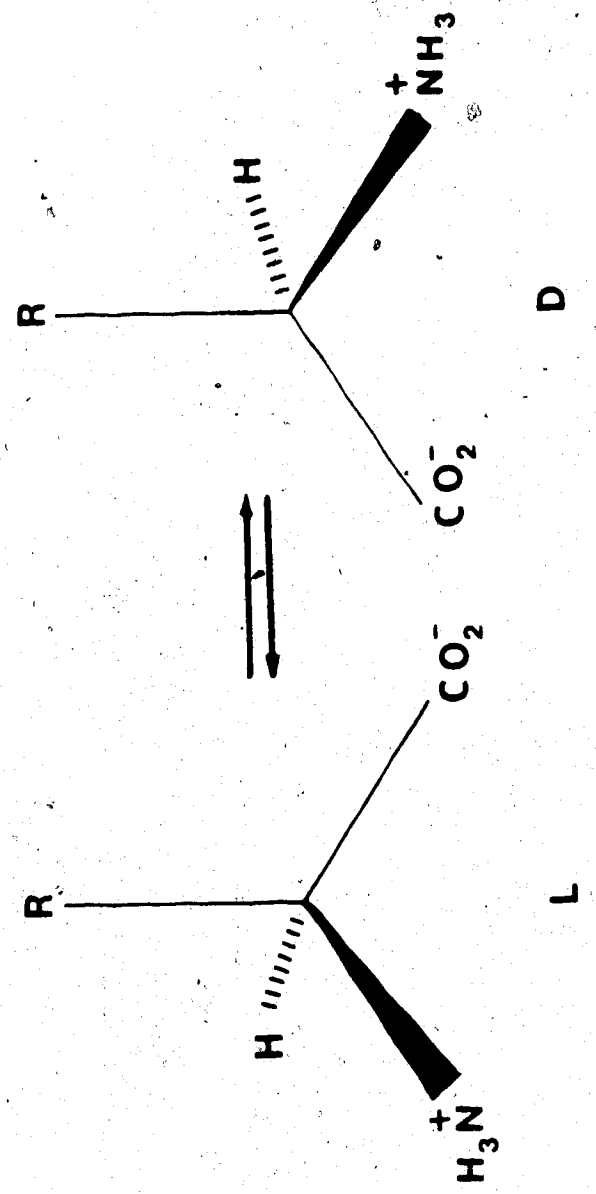


Figure 1. Racemization of an amino acid.

acids is as follows:[5] a small amount (100 mg) of the fossil is ground and then cleaned by sonication in 2N HCl and sonication in water to remove surface contamination including amino acids which leaching may have deposited on the sample. The sample is then treated with 5.5N HCl and heated in a sealed tube to decompose the fossil matrix and hydrolyze the remaining protein residue. Inorganic salts from the fossil matrix are removed from the solution by precipitation with HF or by ion exchange. The extract is evaporated and volatile derivatives of the amino acids are prepared for GC analysis. If a GC column with a chiral stationary phase is used, the amino acids are first esterified by heating with acidified 2-propanol. If a non-chiral column is used, the esterification is performed with (+)-2-butanol in order to obtain the diastereomers. In either case, the resulting esterified amino acids are subsequently acylated with pentafluoropropionic anhydride. The derivatives are then analyzed by GC.

Ultimately, the useful range of amino acid dating depends on two major factors. First, the temperature to which the fossil has been exposed is important since the rate of epimerization depends on temperature according to the Arrhenius equation. The second factor concerns the concentration of peptides in the fossil. The process of hydrolysis of the protein contained in the fossil is competitive with epimerization.[6] Eventually, the amino

acid containing species in the fossil will consist of amino acids* and small peptides which can be more readily exchanged with the environment. This factor clearly place upper limit on the age of fossils which the amino acid technique can measure.

Other practical problems exist with the amino acid dating technique. It was hoped that degree of racemization found for one species of fossil would be the same for another, but it has been found that this, in general, is not the case. For example, in a cave containing material dated independently to be 11000 years old, preserved plant matter of different species were found to have quite different amino acid D/L ratios for the same amino acid.[7] A similar species dependence has been shown for foraminiferas at the same geographical location.[8] Despite this problem, the use of different species to establish geological dates is still widespread. This is due in part to a partly successful technique of calibration introduced by Bada.[9,10] In this method, a single fossil from a particular site is dated by an independent method (usually radiocarbon). The D/L ratio for the amino acid of choice is measured and an

*In the thesis, an amino acid bound in a peptide or protein will be denoted by "peptide unit". A free amino acid, i.e. not bound in a peptide or protein, will be denoted by "amino acid".

average rate constant for the site is extracted from the reversible first-order rate expression (Eq. 1). This rate constant is then used for all other samples taken from the same location in different strata. It is assumed that any environmental factors which the site has experienced will have been homogeneous. Unfortunately there appear to be many examples where this approach fails.[11] One of the most spectacular was in the dating of the Dead Sea Scrolls. It was found that the aspartic acid D/L ratios varied considerably for different samples on the same piece of parchment.[12]

A further problem with the amino acid dating technique concerns the composition of the fossil. It is not guaranteed that the kinetics of epimerization for the protein will follow the simple rate law for amino acids (Eq. 1). Most of the protein in fossils is from collagen*

---

*Collagen is a class of glycoproteins consisting of three protein chains organized in a coiled coil having overall dimensions of approximately 15 × 3000 Angstrom units and containing between 0.4% to 5.8% carbohydrate linked to the alpha chains through hydroxylysine. Two of the alpha chains are in general identical and the entire length of the chains has a unique non-repeating sequence. The proline and hydroxyproline composition is quite variable (2 to 200 per 1000 residues) but is typically about 100. There is cross linking of the alpha chains of the N-terminus where there is also no coiling of the peptide chains. After the fifteenth to twentieth residue the coiling of the peptide chains begins, the repeating sequence -gly-X-Y- being a necessary requirement for the coiled structure. The coiled region of the peptide always begins with the sequence -gly-pro-met-.[13]

since this class of proteins is the most abundant in living tissue.

Much research has been done on the analysis of collagens in fossilized remains.[6] There is a considerable difference in the amino acid composition of fossil versus modern bones of the same species and likewise there is a plethora of peptide fragments present in fossilized protein. This result indicates a large amount of hydrolysis of the protein. It has been stated that the kinetics of epimerization of amino acids in protein will necessarily be complex and the rate constants which are extracted will be some composite rate constant with contributions from peptide units in different positions in each of the various peptides in the fossilized remains.[14,15] The hydrolysis of the collagen requires water and it is not likely that water will ever be excluded. Hence, the composition of fossil proteins (and the kinetics of epimerization) is always complex. Moreover, since the peptide composition of the fossil changes with time, the rate constant for the racemization of the mixture will be changing. This complication makes interpretation of the kinetics in terms of amino acid racemization somewhat suspect.

The best candidate for "preserved proteins" and limited contamination seems to be the inner portion of bones which have been less exposed to the environment.

Another approach is to identify a peptide in the fossil matrix which has some resistance to hydrolysis and could be isolated. This would circumvent the problem of unknown composition. Osteocalcin* has been suggested as a suitable peptide to be used for dating of fossil bones but as yet has not been isolated from fossils.[16] Even if such a peptide were found, there would still remain the problem of unknown kinetics.

Mechanism of Amino Acid and Peptide Racemization

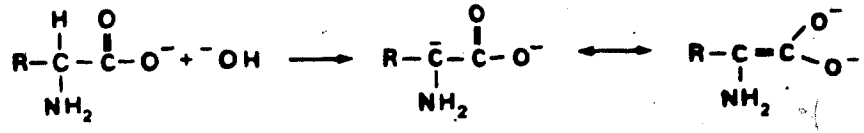
The aim of this thesis is to provide insight into the mechanism of peptide and protein racemization. The mechanism of amino acid racemization is understood and provides a good starting point for a discussion of studies dealing with the racemization of peptides and proteins.

Neuberger was the first to review the mechanism of amino acid racemization.[17] He suggested that the proposed intermediate (**1**) in base catalyzed racemization was a carbanion since high concentrations of base were necessary to cause significant racemization. Since the formation of the intermediate requires the formation of a dianion, any structural change to the amino acid which removes the

---

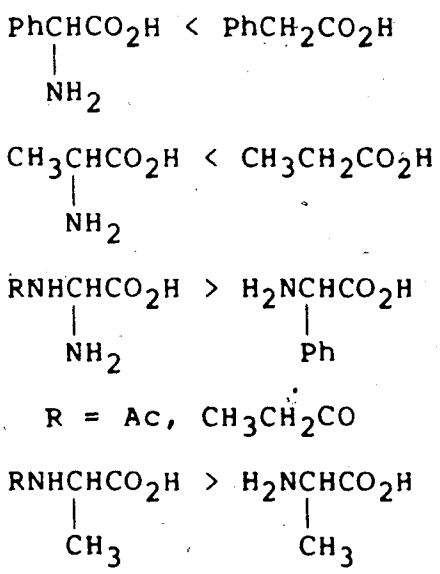
*Osteocalcin is a small protein (M.W. 6000) found in bone. It is characterized by a high content of the calcium binding amino acid γ-carboxyglutamic acid. The structure of osteocalcin varies with species.

carboxylate function would enhance the rate. Substitution on the alpha carbon or on the amino function with electronegative groups should likewise increase the rate of racemization.



1

It has been found that the rate of deuterium exchange is the same as the rate of racemization and establishes that removal of the alpha proton is likely the first step in amino acid racemization.[18,19] Studies on substituted amino acids measuring the rate of exchange in basic solution agree with Neuberger's prediction. Matsuo found that for the following pairs of derivatives:[20]



the order of reactivity of the derivatives was found to be:

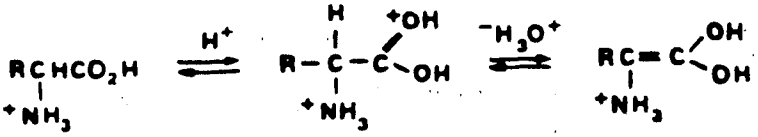
$$R'CONHCHRCONH_2 >> R'CONHCHRCO_2H > H_2NCHRCO_2R' > H_2NCHRCO_2H$$

The result of substitution on the alpha carbon was tested with the N-benzoyl anilide derivatives of amino acids and again was found to be consistent with the predicted behaviour:[21]

$$PhCONHCHRCONHPh$$

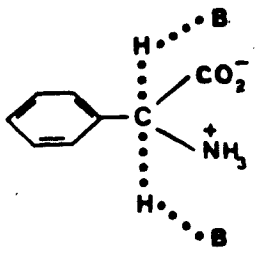
$$R = Ar >> ArCH_2 > Me > iBu > Et >> iPr$$

In the acid catalyzed case, Neuberger proposed that racemization proceeds via protonation of the carbonyl group:[17]



Smith has studied the kinetics of racemization for aliphatic amino acids and for arylglycines.[18,19,22] In both cases racemization follows pseudo reversible first order kinetics. In the case of arylglycines, it was shown

that the rate of exchange as determined by tritium incorporation was equal to the rate of racemization. This would appear to eliminate a mechanism involving a "push-pull" intermediate 2.[18]



2

A Hammett plot of log (k/k_o) (k_o = rate constant for phenylglycine) versus σ yielded a ρ value of 1.6 (calculated at 25° from a value of 1.15 at 110.5°). This was considered to be slightly low for a fully developed negative charge in a transition state (2.11 for the ionization of phenols at 25°). It was suggested that the protonated amino group was attenuating the effect of the substituents by an electrostatic stabilization like that proposed by Neuberger.

Smith also measured the activation parameters for the racemization of the series of arylglycines and found the enthalpy of activation was effectively constant for the entire series. The entropy of activation appears to provide the necessary variation to account for the substituent effect. Smith proposes that for arylglycines,

the change in entropy of activation arises from solvation differences. Aliphatic amino acids racemize about 60 times slower than arylglycines. This is caused by differences in both the enthalpy and entropy of activation. The enthalpy term is about 9 kcal/mol higher in the aliphatic case. If this were the only difference, the arylglycines would racemize about four orders of magnitude faster than the aliphatic amino acids. The difference in the entropy of activation (8 - 10 e.u. more negative) for the arylglycines attenuates the effect of the enthalpy change and may be indicative of the ordering of the medium in the case of the bulkier arylglycines.

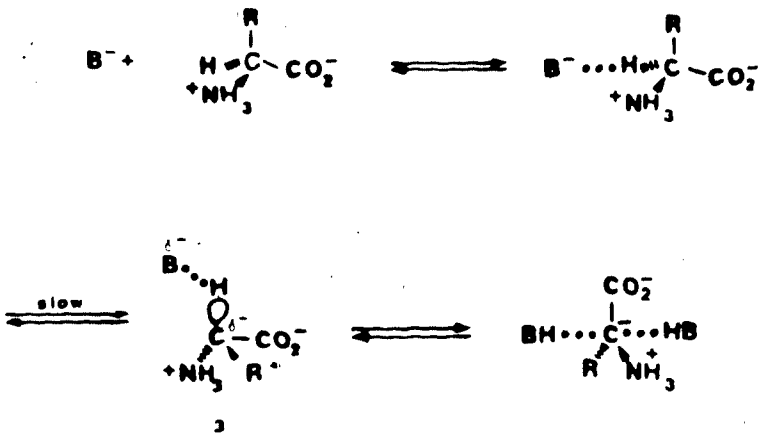
The effect of pH was studied for L-alanine and D-phenylglycine. It was found that there is less than an order of magnitude change in rate when the acid concentration is changed by fourteen orders of magnitude with a relatively level portion around the isoelectric point of the amino acid. This behaviour is similar to that of other amino acids which have been studied.[23] Assuming that the observed rate consisted of components of specific acid and base catalysis involving the zwitterionic, anionic and cationic species, Smith was then able to calculate that the most reactive species in solution is the zwitterion. However, Smith observed general base catalysis for racemization of L-alanine and phenylglycine. An increase in the rate of racemization was

observed when the concentration of phospate buffer was increased at constant ionic strength and constant pH. Therefore, the conclusion that the zwitterion is the most reactive species may not be correct.

The effect of ionic strength was also examined. At pH 7, the effect is modest with a doubling of the ionic strength resulting in only a 10% increase in rate. At pH 10, the same change in ionic strength results in about a 20% change. This concurs with the above modest effect of ionic strength since reactions involving zwitterions would be expected to show less dependence on ionic strength than those with anions.[18]

A Taft plot using the rates of racemization for the aliphatic amino acids and the order of entropies of activation (ala > val > ile) indicated to Smith the presence of a steric effect, but the order of rate (ala > leu > ile > val) must be considered to arise from electronic factors as well.

Smith proposes that the mechanism of racemization involves a polarized transition state **3**. The mechanism is consistent with the experimental facts and is patterned after the mechanism determined for the ionization of nitroalkanes.[24] The similarity of the ρ values provides Smith with the justification to propose this pathway for amino acid racemization. A deuterium isotope effect of 2.0 indicating a rate determining deprotonation is

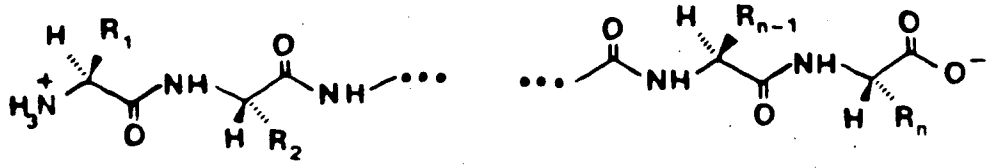


consistent with the above scheme.[19]

It should be pointed out that there are other mechanisms which are kinetically indistinguishable from Smith's mechanism. In particular, protonation of the carboxylate prior to the rate determining removal of the alpha proton could also lead to a flat pH profile and might be expected to show general base catalysis.

Racemization in Peptides and Proteins

In neutral aqueous medium a peptide is predominantly present as a zwitterion:



The rate of racemization of the C-terminal peptide unit is expected to be less than the rate of racemization of the interior peptide unit since the carboxylate anion would inhibit the formation of a carbanion. Similarly, the N-terminal peptide unit is expected to racemize faster than the interior peptide units since the positively charged N-terminus can inductively stabilize the incipient carbanion:[17,25]

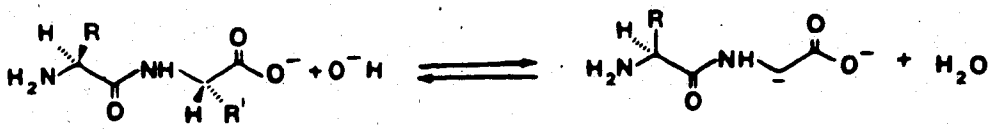
There is the possibility of a similar effect in the interior of the peptide chain. Since the bond linking the carbonyl group with the nitrogen atom in peptides is shorter than an isolated carbon-nitrogen single bond,[26] the peptide bond is considered to be a resonance hybrid:

This resonance form might contribute to the stability of the carbanion formed during racemization in the same way as the protonated N-terminus. Thus, the prediction that the N-terminal peptide unit will racemize fastest is not

firmly based. On the contrary, there may be no special preference for racemization to occur at the N-terminus for electrostatic reasons. This does not preclude the possibility that the N-terminus may racemize quickly by the action of another effect.

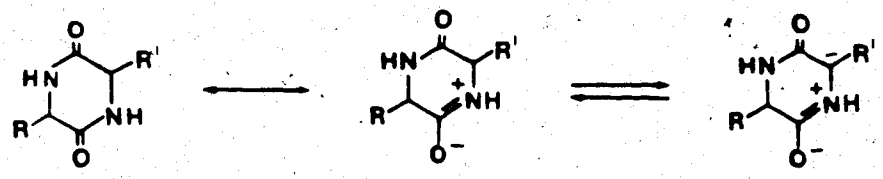
As in neutral media, racemization under basic conditions should be enhanced in the peptide chain. Early research by Levene on the racemization of alanine oligopeptides in alkali showed some indication that interior peptide units racemize more quickly since as the length of the peptide chain was increased, the rate of racemization also increased.[27]

Also, in basic solution several proteins (albumin, casein, edestin, fibrin and gelatin) show enhanced rates of racemization compared to amino acids not bound in a peptide chain.[28] Levene also studied the racemization of diketopiperazines in alkali and found that these compounds racemize more quickly than the corresponding dipeptides.[27] The relative stability of dipeptides was suggested to arise from the proximity of the carboxylate anion (at least for the C-terminal residue) in dipeptides.[17] As mentioned above, the negative charge would inhibit the formation of a carbanion at the asymmetric centre 4.

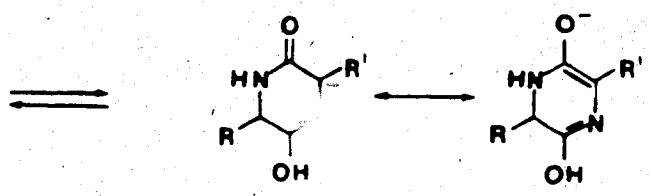


4

The relatively rapid epimerization of diketopiperazines was seen to be a result of the contribution from a dipolar resonance form having a positively charged nitrogen atom adjacent to the asymmetric centre (5).



5



6 7

It was discovered later by X-ray studies that the six-membered ring in diketopiperazines* is essentially flat as in c-(glycyl-glycyl)[30,31] and c-(D-alanyl-L-alanyl)[32,33]

---

*c- denotes the cyclic structure of the diketopiperazine.

or puckered as in c-(L-alanyl-L-alanyl).[33] This is presumably due to the contribution of the dipolar resonance form 5. The ease of racemization of diketopiperazines may be attributed to the relative stability of the conjugated system 7 in resonance with the carbanion 6 produced by removal of the asymmetric proton.

Amino acids isolated from fossilized materials were found by Hare[34,35,36] to be racemized to a greater degree than similarly treated free amino acids. This led some investigators to propose that racemization is concommitant with hydrolysis of the peptide chain.[34] However, it was earlier demonstrated that there seems to be no connection between the two processes. Levene showed for diketopiperazines[37] and proteins[38,39] and Darge for a dipeptide[40] that in strongly basic solution the rate of hydrolysis is faster than the rate of racemization. If the same substrate is then treated with dilute base, hydrolysis is found to be slower than racemization. Another proposal to explain the anomalously high degree of racemization of free amino acids found in fossils involves hydrolysis of the protein through some species (e.g. a diketopiperazine-like species from internal aminolysis, vide infra) which is highly susceptible to both racemization and removal from the fossil matrix either by leaching or by rapid hydrolysis to free amino acids.

Kriausakul and Mitterer were among the first to undertake a systematic investigation of peptide racemization using small isoleucine* containing peptides under mildly basic conditions (pH 8.0, 152°).[41] The elevated temperature is necessary for racemization to proceed at a measurable rate. In this study, Mitterer demonstrated that racemization occurred more rapidly in the peptide chain than in the free amino acid. By using a series of dipeptides, the tripeptides ile-gly-gly and gly-gly-ile, several natural peptides of known sequence and collagen from shell, Mitterer arrived at the conclusion that amino acids racemize depending on their position in the peptide chain: N-terminal > C-terminal >> interior > free. As with proteins, racemization of the peptides did not follow reversible first order kinetics. In all cases the initial rate of racemization was larger than the final rate which in turn was equal to the rate of racemization of the free amino acid (Figure 2)**. By assuming that there are only

---

*Epimerization in isoleucine produces the diastereomer alloisoleucine. It is therefore not necessary to use optically active materials for determining the degree of racemization of this amino acid and for this reason is widely used for the dating of fossils by the amino acid technique.

**Although Mitterer was the initiator of this type of analysis, he was not its sole user. Freidman and Masters, while studying the racemization of casein in dilute alkali, compounded the problem by attempting to use rate constants extracted in this manner in a linear free energy relationship.[42] Smith and Evans also (cont'd)

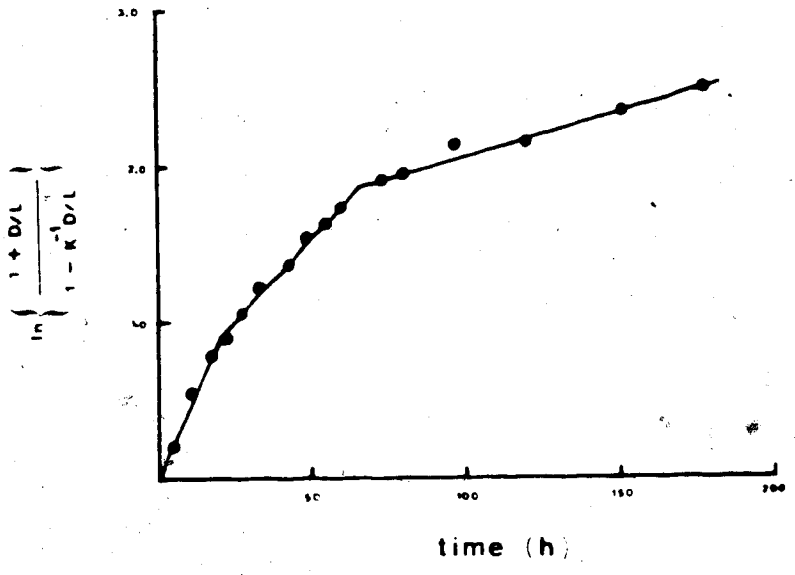


Figure 2. Kinetics of epimerization of isoleucine in heated Mercenaria at 152°C (Ref. 41).

two components in the hydrolysis mixture, Mitterer interpreted the initial slope of the canonical plot* as

> observed "two linear portions" in the racemization of collagen.[25] Lajoie, Wehmiller and Kennedy[43] and separately, Wehmiller[44] presented qualitative models in which the plots are considered as consisting of two linear portions. The authors recognized the limitations of the model and yet persisted in its use because of a failure to realize that hydrolysis can affect the magnitude of the rate constants extracted by this method. In order to further generalize, Wehmiller proposed a two component model in which there are two epimerizing groups of amino acids: 1) amino acids bound in a peptide chain and 2) free amino acids racemizing more slowly and being produced from group 1. This model would certainly predict a curved first order plot. Wehmiller used the model to attempt to produce an empirical method to compare the racemization behaviour (and thus the ages extracted by the amino acid method) of fossils from different species. There was no attempt to incorporate either of the models into a mechanism for protein or peptide racemization.

---

*The racemization rate constant is extracted from a plot of the left-hand side of Equation 1 versus time. If $k_1 = k_{-1}$ then the slope of this plot is equal to $-2k_1$. Otherwise the slope is equal to $-(k_1+k_{-1})$. It has become standard practice to extract rate constants from initial slopes of plots of this type even when the behaviour is seen not to obey Equation 1. For this reason, these plots will be called "canonical plots", Equation 1 will be called the "canonical equation" and the left-hand side of Equation 1 will be called the "canonical expression".

It is common for Equation 1 to be rearranged so that:

$$\ln\left(\frac{1+\frac{[D]}{[L]}}{1-K'\frac{[D]}{[L]}}\right) - \ln\left(\frac{1+\frac{[D]}{[L]}}{1-K'\frac{[D]}{[L]}}\right)_{t=0} = (k_1 + k_{-1})t$$

The slope of a plot of the left-hand of this equation versus time will be positive. Since racemization involves a decrease in optical activity, except for comparisons to the literature a convention will be adopted wherein the canonical plot will have negative slope. The two representations are equivalent and are related by reflection in the t-axis.

representing the rate of racemization of the starting material. However, in the cases where there is sufficient data to permit the measurement of the initial rate, there is no initial linear portion to the canonical plot (Figure 3). It is not clear that these are only two component systems and there is considerable evidence in the literature that the mixture from the hydrolysis of peptides will contain diketopiperazines as well as the dipeptides and free amino acids. Indeed, there is evidence in Mitterer's work that diketopiperazines are present. He showed that the rate of epimerization of ile-gly-gly is more than twenty times faster than that of gly-gly-ile. At the same time, the rate of production of free isoleucine is reversed with ile-gly-gly producing isoleucine some 30 times slower than gly-gly-ile. If hydrolysis of the peptides proceeds by the formation of the more easily racemized diketopiperazine by internal aminolysis (Figure 4) and subsequent hydrolysis to free amino acids then the above observations would be expected. Clearly, without separation of the hydrolysis mixture it is difficult to justify any of these conclusions.

In the following paper, Kriausakul and Mitterer[45] reported the activation energies for hydrolysis and epimerization for ile-gly to be 20.9 and 23.1 kcal/mol

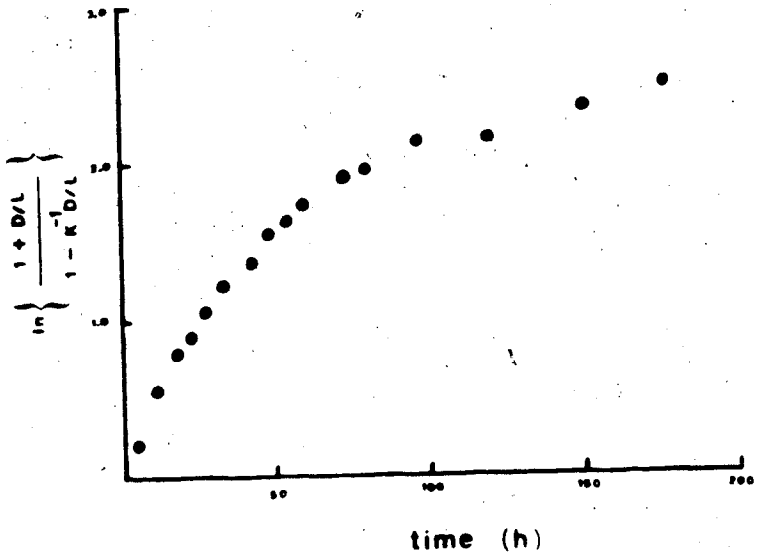


Figure 3. Kinetics of epimerization of isoleucine in heated Mercenaria at 152°C (Ref. 41). This is the same data shown in Figure 2 omitting the solid lines.

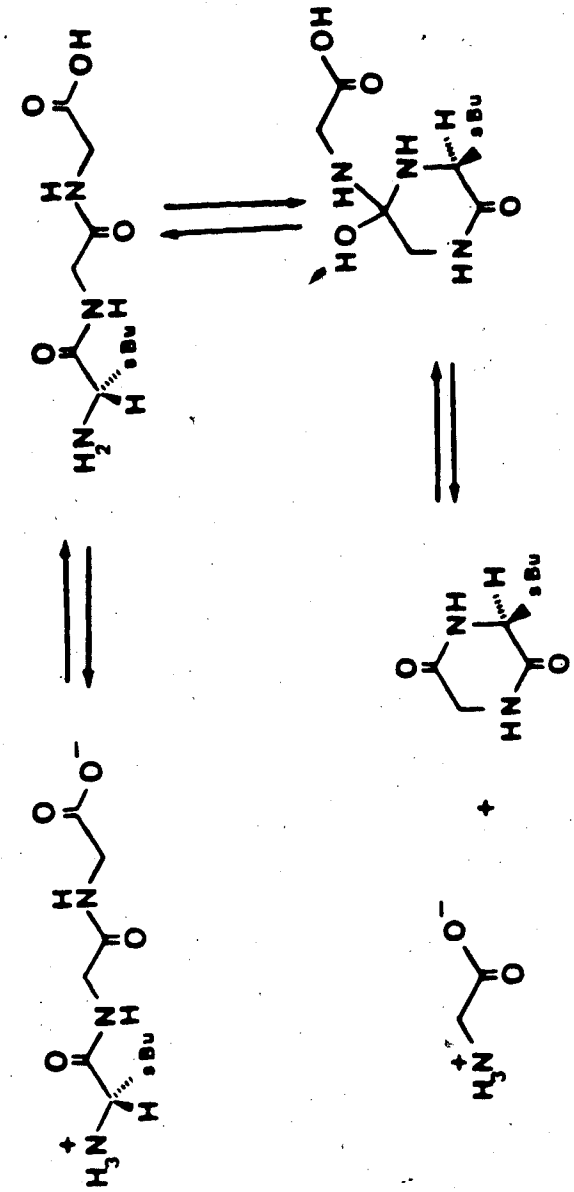


Figure 4. Diketopiperazine formation by internal aminolysis.

respectively.* The lower activation energy for epimerization compared to that of the free amino acid (29-30 kcal/mol) was attributed to the faster racemization of the N-terminal residue. Again Mitterer did not attempt to separate the contribution of diketopiperazine to the total racemization rate of the dipeptide. More interesting is the first attempt to separate a fossil protein hydrolysate according to molecular weight. When this was done, Mitterer discovered that the most highly racemized fraction (aside from the fraction containing free amino acids) was the one containing protein fragments having a molecular weight less than 500. Mitterer contended that the high degree of racemization observed in the fraction is due to the increased number of N-terminal residues. This observation is also consistent with the presence of diketopiperazines.

More recently Kriausakul and Mitterer described the measurement of the degree of racemization in the C-terminal position of dipeptides produced from the hydrolysis of fossil protein.[46] It was found that this position was the most racemized of all, quite at odds with

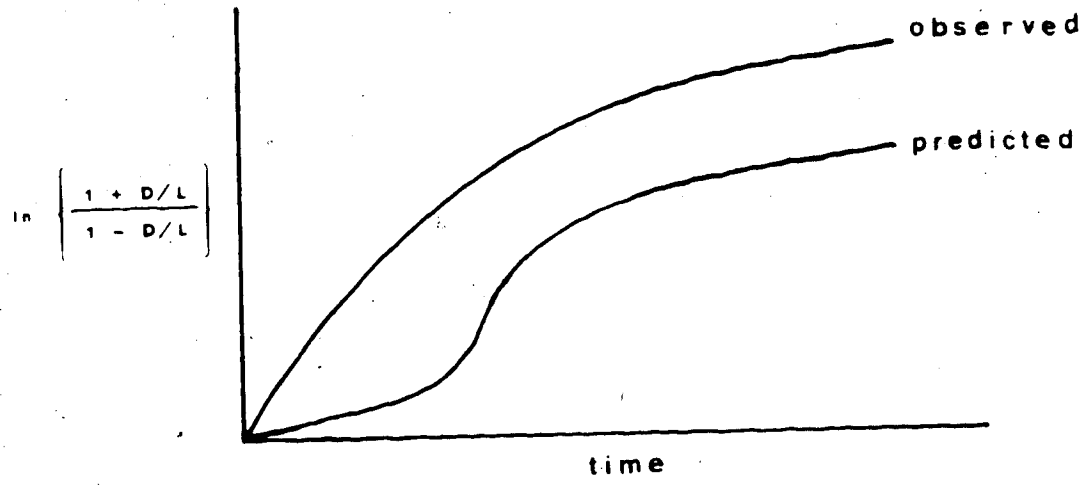
---

*For epimerization log A = 10.588; for hydrolysis log A = 8.263. At 15° epimerization is calculated to be about 4 times faster than hydrolysis while at 150° this increases to a factor of about 16. It should be noted that the rate of hydrolysis is measured by the production of ile, not from the disappearance of the dipeptide.

the observation that the N-terminal position in dipeptides racemizes fastest.* Mitterer attributed this observation to the formation of easily epimerized diketopiperazines

---

*There is an important inconsistency in Mitterer's reasoning which has not been commented upon previously. In the case of protein racemization with concommitant hydrolysis there will be few terminal residues near time zero. If the termini do indeed racemize faster than internal residues (as suggested by the model studies using dipeptides) in order to reflect the build-up of terminal residues there should be a portion of the curve near time zero where the slope is increasing. This has never been reported. Mitterer argues that the similarity of dipeptide and protein kinetics allows the extrapolation of dipeptide behaviour to proteins. The observation of a gradual decrease in rate constant is inconsistent with the predicted increase of the number of rapidly epimerizing terminal residues.



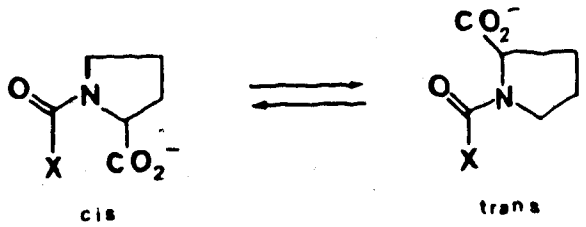
The lack of predicted behaviour could be due to the difficulty in obtaining data around time zero, but could equally reflect an important difference in kinetic behaviour between dipeptides and proteins. Mitterer observed similar behaviour for tripeptides indicating some degree of continuity between dipeptides and proteins. However, under the condition of high temperature it is likely that after a short period no tripeptide would be left due to rapid formation of dipeptides and diketopiperazines (vide infra).

and subsequent transfer of the epimerized peptide unit to a C-terminal position by hydrolysis. There is a difficulty with this rationale. If the hypothesized diketopiperazines are arising from dipeptides it seems reasonable that this enhanced rate of epimerization for the C-terminus would have been observed in the series of dipeptides studied previously. This was not observed for any of the dipeptides. In order to explain the protein result it may be necessary to propose the existence of additional species which are undergoing rapid epimerization.

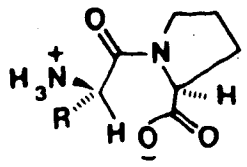
Smith and Sol have done a complementary study on dipeptide racemization but found that in the case of sterically unhindered dipeptides gly-X (X = ala, leu, phe, asp, met) and for alanine in val-ala, leucine in val-leu, phenylalanine in val-phe and aspartic acid in val-asp, the amino acid racemized fastest when in the C-terminal position.[47] The amino acids isoleucine and valine both with secondary carbons in the beta position show faster N-terminal racemization. Proline appears to cause the other amino acid residue in the dipeptide to show faster N-terminal racemization. Like Mitterer, Smith overlooked the possibility of diketopiperazine formation during the heating of dipeptides. This alone would be sufficient to cast doubt on the conclusions drawn from the study of these dipeptides, but it also is useful to scrutinize in

detail the mechanisms presented.

Smith attributed the affect of proline to the cis-trans equilibrium found for dipeptides of the form X-pro:



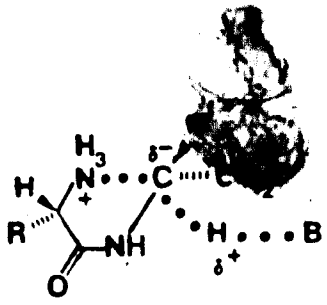
He argued that the enhanced rate for the N-terminal residue is due to the relatively close proximity of the carboxylate anion of proline:



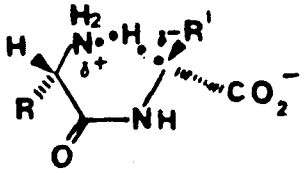
Smith cited the work of Dorman and Bovey[48] and that of Grathwohl and Wuthrich[44] where the preference of cis over trans was reported for phe-pro and ala-pro respectively in strongly basic solution. This preference was subsequently documented for the dipeptides X-pro (X = ala, leu, phe, trp, tyr).[50,51] At neutral pH, there is very little general preference for cis over trans with 78% cis for phe-pro, tyr-pro and trp-pro and from 37% to 51% for gly-pro, ala-pro, glu-pro and his-pro.[52] Therefore, the use of this conformational argument to explain the

racemization kinetics exaggerates the cis-trans equilibrium phenomenon.

Smith and Sol propose that the unexpected result (C-terminus > N-terminus) arises from some intramolecular interaction of the N-terminal amino group by either stabilization of the incipient carbanion by "solvation" by the protonated N-terminus, 8 or from the free amino group acting as a base to aid in the removal of the alpha proton, 9.[47]



8



9

Smith stated that the reverse trend in isoleucine and valine containing dipeptides is due to the removal of these possible effects for steric reasons as is found for S_N2 reactions. On the contrary, in intramolecular S_N2 reactions bulky groups can in fact speed up a reaction as in the formation of epoxides and anhydrides.[53] Long also uses the concept of "steric acceleration" to explain the

rapid ring closure to diketopiperazines in some sterically hindered dipeptides.[54] It appears that Smith's mechanisms need to be adjusted to account for these facts.

Previous work on the racemization of peptides has given results contrary to the work of Mitterer and Smith. In the treatment of the peptide hormone glucagon with 1 N sodium hydroxide (with little hydrolysis), Epand and Epand observed that with carboxypeptidase the treated peptide gave lower yields of the internal amino acids than for the untreated peptide.[55] The same result was obtained for leucine aminopeptidase. Since the enzymes are less effective at cleaving peptide bonds involving D isomers it was concluded that the interior of the peptide contained more of the biologically inactive amino acids residues. The conclusion to be drawn here is that the interior amino acid residues racemize faster than the termini. Similarly, when adrenocortigen was treated with 0.1 N sodium hydroxide at 100° (again with little hydrolysis), it was found that the phenylalanine residue at the C-terminus was unracemized whereas the internal phenylalanine residue was extensively racemized.[56,57] In a third study, the activity of melanocyte stimulating hormone was maintained after treatment with base (meaning that the active N-terminus still maintains the L configuration) while the internal residues phe and arg were found to be racemized.[58]

More direct evidence is available that internal residues racemize faster than the termini.[59] The amount of deuterium exchange was measured by NMR for a series of glycine containing peptides. At pH 11 to 13.5 after 21 hours there is no exchange in either glycine or glycylglycine. However, there is noticeable exchange in the middle methylene group of triglycine after only five hours. Likewise in the peptides acetyl-glycylglycine, tetraglycine, pentaglycine, diglycine-amide, acetyl-alanylglycine and acetyl-serylglycine, there is exchange only in the internal residues. The effect could be due to elastostatic repulsion between hydroxide ion and the carboxylate anion as predicted by Neuberger, since the methylene group in the residue adjacent to the N-terminus in tetraglycine exchanges faster than the methylene group in the residue adjacent to the C-terminus. The exchange is occurring most rapidly farther away from the carboxylate anion. The free amino group might also retard the rate of exchange possibly by an inductive interaction. The observation that in trifluoroacetyl-triglycine, the N-terminal methylene is exchanging at a much faster rate than the internal residue is consistent with this mechanism. The substituent also plays an important role. In glycylalanylglycylglycine, the alanine is only 5% exchanged when the adjacent glycine residue (towards the C-terminus) is 95% exchanged.[59]

There is convincing evidence that in certain cases the internal amino acid residues racemize faster than the termini. The above studies were performed under strongly basic conditions so extrapolation to neutral pH must be done with care. The observations may be applicable at least for the C-terminus since even at neutral pH the carboxylate anion is present in a predominant amount. With this caveat in mind, the above observations are completely opposite to the expectations derived from the dipeptide studies. The relevance of dipeptide racemization to that of proteins must be seriously questioned.

Factors Causing Deviation from Simple Behaviour Described by the Canonical Equation

It is already evident that some dipeptides and proteins do not obey the simple rate law given by Equation 1. The factors which result in this deviation will be the focus of this section. The factors special to peptides will be discussed later. The measurement of temperature is of particular interest to geologists. The measurement of fossil age by the amino acid technique is dependent on the knowledge of the temperature history of the fossil.[11] This problem of unknown temperature history can be overcome by use of the calibration method described previously.[9,10]

The amino acid technique in concert with some calibration dating method can also be used to measure geological temperature. Smith has pointed out that the temperature obtained by measuring rates is different from the time-averaged temperature.[60] Since the rate of reaction has an exponential dependence on temperature, the rate-averaged temperature will be somewhat higher than the time-averaged temperature. This idea can be extended in an analysis of the effect on rate of racemization from fluctuations in temperature. Geologically, the most obvious periodic change in temperature concerns the seasonal oscillation in the daily average temperature. The difference between the rate-averaged annual temperature and the true average annual temperature is certainly proportional to the amplitude of the seasonal oscillations. The size of the effect can be estimated.

If a sinusoidal time dependence of temperature is included, the Arrhenius expression becomes:

$$\bar{k} = A \times \exp(E_a/R(\Delta T \sin 2\pi t + T_{av})) \qquad (2)$$

where T_{av} is the true average temperature and ΔT is the amplitude of the temperature oscillation. The rate law in

equation (1) then becomes:*

$$dx/x = -2k_{av} \times \exp\left(\frac{T \times E_a \times \sin 2\pi t}{R \times (T_{av} \times (T_{av} + \Delta T \times \sin 2\pi t)}\right) dt \qquad (3)$$

The rate constant extracted from (1) as before is the effective rate constant k_r and the expression upon integration over one period yields:**

$$k_r/k_{av} = \exp\left(\frac{T \times E_a \times \sin 2\pi t}{R \times T_{av}(T_{av} + \Delta T \sin 2\pi t)}\right) dt \qquad (4)$$

The integral is best evaluated using Simpson's Rule and a short computer program was written to this end (Appendix 5). The difference between the effective average temperature and the true average temperature can be calculated by substitution into the expression:

$$T_r - T_{av} = (-R/E_a \times (\ln(k_r/k_{av}) + T_{av}) - T_{av}) \qquad (5)$$

---

* $x = \ln\left(\frac{1 - D/L}{1 + D/L}\right)$

** It is assumed in this model that one period (year) is the same as the next, as in the Bada/Smith model.[60] The averages will be unaffected if the integral is evaluated over only one period.

Examining the case for T_{av} = 283.15 K, ΔT = 10°C, E_a = 29.4 kcal/mol (Table 1) shows that the temperature that is extracted by the amino acid method gives a result which is approximately 3.7° higher than the true average temperature.*,†

It may be argued that this potentially important temperature effect is not present for a large proportion of the existence of a fossil. For example, fossils in the ocean bed have been at a constant average temperature and it is claimed that fossils found in certain caves are also maintained at a constant temperature.

However, this argument does not eliminate the possibility of brief fluctuations in temperature or

---

*For Edmonton: min. Jan -22°C, max. July +22°C, av. +13°C.[61]

†An estimate can be made of the error in an age determination resulting from using the rate-averaged temperature. The ratio of the ages derived from the rate-averaged temperature and the true average temperature is calculated using the above data in the Arrhenius equation:

$$\frac{k_r}{k_{av}} = \exp\left(\frac{E_a}{R}\left(\frac{1}{T_{av}} - \frac{1}{T_r}\right)\right)$$

$$= \exp\left(\frac{29400}{1.9872}\left(\frac{1}{283.15} - \frac{1}{286.87}\right)\right)$$

$$= 1.97$$

The age determined from the rate-averaged temperature will be too high by a factor of two.

Table 1. Effect of Temperature Oscillations on Average Temperature Extracted from Racemization Kinetics (Equation 5)

True Average Temperature (K)	Oscillation Amplitude (°C)	Activation Energy (cal/mol)	Effective Average Temperature (K)	Error $(T_r-T_{av})^a$ (°C)
273.15	1	29400	273.197622	.0476219654
273.15	2	29400	273.339062	.189062357
273.15	5	29400	274.275238	1.1252377
273.15	10	29400	277.07152	3.92152023
273.15	15	29400	280.67517	7.52516985
273.15	20	29400	284.631236	11.4812357
278.15	1	29400	278.195901	.0459010601
278.15	2	29400	278.332324	.182323933
278.15	5	29400	279.23854	1.08853996
278.15	10	29400	281.970762	3.82076192
278.15	15	29400	285.524875	7.37487472
278.15	20	29400	289.446785	11.2967852
283.15	1	29400	283.19427	.0442703962
283.15	2	29400	283.325929	.175929189
283.15	5	29400	284.203409	1.0534091
283.15	10	29400	286.872504	3.72250414
283.15	15	29400	290.376397	7.22639692
283.15	20	29400	294.263298	11.1132978
288.15	1	29400	288.192723	.0427234173
288.15	2	29400	288.319855	.169855475
288.15	5	29400	289.169773	1.01977277
288.15	10	29400	291.776754	3.62675369
288.15	15	29400	295.229834	7.07983339
288.15	20	29400	299.080881	10.9308813
293.15	1	29400	293.191255	.0412549973
293.15	2	29400	293.314083	.164082646
293.15	5	29400	294.137562	.98756218
293.15	10	29400	296.683508	3.53350818
293.15	15	29400	300.085274	6.93527365
293.15	20	29400	303.899643	10.7496426
298.15	1	29400	298.18986	.0398597717
298.15	2	29400	298.308591	.15859127
298.15	5	29400	299.10671	.956709743
298.15	10	29400	301.592757	3.44275713
298.15	15	29400	304.9428	6.79280031
298.15	20	29400	308.719687	10.5696871

a) T_r = effective average temperature; T_{av} = true average temperature

oscillations of long period. Depending on the time-scale, the duration of the present "constant temperature" may be geologically insignificant. But as shown above, the period of the oscillation has no effect whereas the amplitude certainly does affect the effective average temperature.

Peptides; Deviations from Simple Behaviour

There appear to be two major conceptual differences between amino acid and peptide/protein racemization: first, as mentioned before, peptides can yield through hydrolysis a number of different empimerizing species. Second, the racemization of any particular amino acid in the peptide chain can be affected by a neighbouring amino acid. This effect may be manifested by a positional dependence as suggested by Mitterer and Smith (ut supra) or by asymmetric induction. The possibility of achieving a D/L ratio in excess of unity (except for the case of isoleucine) has not been explored. This would clearly represent a large deviation from the behaviour typical of amino acids. By analogy with isoleucine, the value of K' in Equation (1) will not be unity. In order to measure the rate of epimerization in proteins and peptides it will be necessary to determine a value of K' for each species produced by hydrolysis.

Diketopiperazines in Protein Hydrolysis

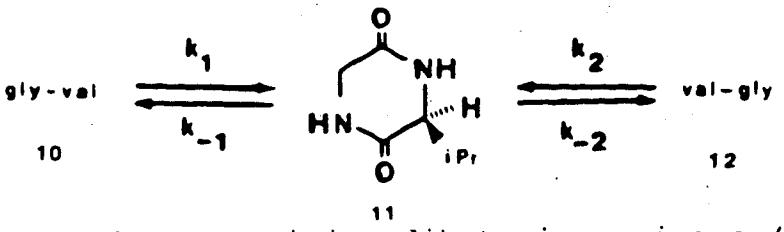
Of the possible fragments from the hydrolysis of proteins, diketopiperazines are especially interesting since they racemize significantly faster than the corresponding dipeptides.[27] A discussion of the formation of diketopiperazines may help to explain the observed racemization behaviour in the model systems discussed previously.

The incidence of diketopiperazines is well known in the partial hydrolysis of proteins in dilute (ca. 0.1N) HCl at elevated temperature.[62] It was of particular concern to Sanger in the determination of the amino acid sequence of insulin when N-terminal valine was observed from the heating of glycylvaline.[63,64] However, partial hydrolysis of proteins with 12N HCl at 37° does not in general produce inversions of this type.

At about the same time as Sanger's work, Hirohata and coworkers studied the acid hydrolysis of dipeptides in 1N H_2SO_4 at 100° in order to gain an insight into the relative stabilities of the peptide bond involving different amino acids.[65] It was discovered that diketopiperazines formed readily for dipeptides of the form gly-X (X = dl-ala, dl-nval, l-leu, dl-amino-butyric acid, dl-ileu, dl-nleu, dl-val, aminoisobutyric acid, dl-ival) as well as for dl-val-dl-val (25% H_2SO_4). Hirohata later did a similar study with a range of

dipeptides in 1.5N HCl at 100°.[66] In this work, it was found that peptides of the form X-val and X-ile gave diketopiperazines (X = ala, leu, phe, tyr, met, lys).

Diketopiperazine was not observed during the hydrolysis of X-gly. Hirohata explained this by using a steric argument. He reasoned that the two peptide bonds in the diketopiperazine (**11**) formed from dipeptides (**10**, **12** X = val) would be cleaved at different rates ($k_{-1} \neq k_{-2}$). The relative rate of ring opening ($k_{-2} > k_{-1}$) was explained by a more hindered approach of solvent to the carbonyl adjacent to the side chain. The difference could also be due to the relative steric requirements of the two tetrahedral intermediates. If also $k_{-2} >> k_2$ then no diketopiperazine would be observed. Evidence for this will be discussed further.



The proline containing diketopiperazine c-(pro-gly) has been observed frequently in protein hydrolysates.[67] It has also been produced in essentially quantitative yield in the catalytic hydrogenation of carbobenzoxy-glycylproline amide.[67] The formation of this diketopiperazine is so facile that a solution of glycylproline or glycylhydroxyproline (pH = 8.0 in pure water) standing at room temperature will gradually yield

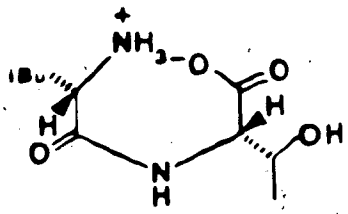
the corresponding diketopiperazine. Because of the relevance to protein sequencing most of the studies of dipeptide hydrolysis have been done in acidic aqueous solution.[68] Abderhalden and Komm were among the first to study the hydrolysis of tripeptides under acidic conditions.[69] They found that in 1% HCl (as well as dilute H_2SO_4) at 150° the diketopiperazine c-(gly-leu) was formed in good yield from l-leu-gly-l-leu, gly-dl-leu-gly, gly-dl-leu, and dl-leu-gly-gly. Of more relevance to racemization model studies at a neutral pH, Abderhalden also performed the hydrolysis in pure water at 150-160°.[70] He found that for the dipeptides gly-gly, gly-dl-leu, gly-l-leu, dl-leu-gly, gly-dl-phe, dl-leu-dl-leu and dl-ala-gly, the corresponding diketopiperazines were formed in greater than 90% yield. It is therefore not surprising that Bada and Steinberg observed predominant diketopiperazine formation during the hydrolysis of the tripeptides leu-gly-gly and gly-leu-gly under neutral conditions (0.1 M phosphate buffer, ionic strength 0.5, 100° or 130°)[71] or for the dipeptides ile-gly and gly-ile (pH 5-8).[72] The hydrolysis of the hexapeptide phe-gly-leu-gly-val-gly was inferred to proceed stepwise via the diketopiperazines c-(phe-gly), c-(leu-gly) and c-(val-gly) (pH = 7.26, 100°).[71] A noteworthy observation was the reversible isomerization of leu-gly-gly to gly-leu-gly presumably through a cyclic intermediate.[71]

There is unequivocal evidence that diketopiperazines are formed during the hydrolysis of proteins. However, it is unclear whether they are formed upon direct hydrolysis of the protein or peptide (Figure 4) or via an intermediate dipeptide.

The presence of diketopiperazines in the hydrolysis of peptides and their effect on the rate of hydrolysis is well documented. Long and coworkers published a series of careful kinetic studies on the hydrolysis of tripeptides in 2N HCl at temperatures less than 114° and found that the cleavage of the dipeptide from the N-terminus was faster than that from the C-terminus by a factor of about four.[73] The tripeptides gly-leu-gly and leu-gly-leu hydrolyze to a mixture of the dipeptides gly-leu and leu-gly along with the free amino-acids. There was a mass balance over the entire course of the hydrolysis indicating that if diketopiperazines were formed, they are present in a minor amount. Analysis was performed by ion exchange chromatography using the ninhydrin reaction[74] to detect the components in the mixture. Hence, any diketopiperazines in the mixture would not be observed.

The possibility of diketopiperazine formation was not addressed until the observation of sequence inversion in the hydrolysis of glycylphenylalanine.[75] The formation of inverted phe-gly was assumed to result from diketopiperazine formation followed by cleavage of the

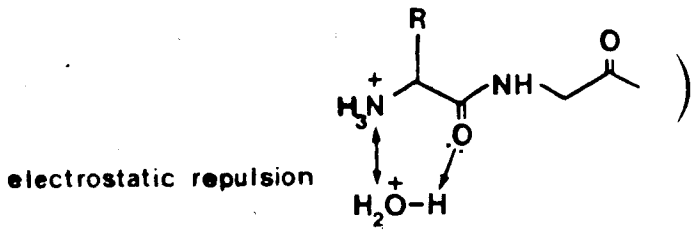
original peptide bond although in this case diketopiperazine was not detected. Long then examined the effect of diketopiperazine formation in the earlier work and discovered that this secondary hydrolytic pathway was significantly affecting the extracted rate constants. Long pursued the problem of sequence inversion and found that the phenomenon occurs in the dipeptides leu-thr, gly-leu, ala-gly, gly-val and their sequence inverted isomers.[54] The dipeptides were divided into two groups with gly-ala, ala-gly, gly-leu and leu-gly in a group having hydrolysis faster than sequence inversion and with gly-val, val-gly, leu-thr and thr-leu in a group where the rate of hydrolysis is less than or equal to the rate of sequence inversion. In the former group, the enthalpy of activation is 6 to 10 kcal/mol greater than in the latter group while the entropy of activation is about 20 e.u. less negative. Both these facts were rationalized by examination of the structure of the dipeptides in each group. Long reasoned that the bulky alkyl side-chains in the latter group of dipeptides were responsible for the lower value of $\Delta H^{\neq}$. The N- and C-termini are then forced together aiding ring closure producing, in Long's words, "steric acceleration":



The large negative entropy of activation (-40 e.u.) indicated to Long a highly ordered transition state. In the former group of dipeptides, the steric restrictions are less important.

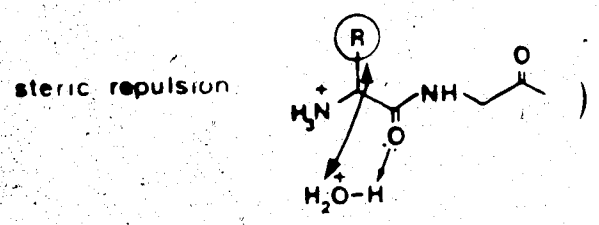
At higher pH Long observed a stoichiometric deficiency of material in the hydrolysis of dipeptides when the above analysis technique was used.[76] From pH 3.8 to 6.6 there was a measurable amount of diketopiperazine present in the hydrolysate.

None of the studies on the hydrolysis of dipeptides serve to explain the behaviour of larger peptides or proteins. It is clear that some of the diketopiperazine could arise from dipeptides but it is equally clear that some of the diketopiperazine could be formed directly from the peptide or protein. The stability of dipeptides and consequently their common appearance in acid protein hydrolysates is considered to be a result of the protonated N-terminus. For electrostatic reasons, the positive charge on the nitrogen atom inhibits the protonation of the carbonyl of the N-terminal residue, a step necessary for hydrolysis.[62] As the distance from the N-terminus increases, the rate of peptide bond cleavage should likewise increase:



The observation of increasing rate of hydrolysis in the homologous series of diglycine to hexaglycine would seem to support the electrostatic effect.[77,78] However, Long and coworkers found that the rate of cleavage of the C-terminal glycine residue in this same series of peptides was the same for gly-gly-gly to pentaglycine.[79] Also, the rate of cleavage for the interior peptide bonds was about the same except for the central amide bond in hexaglycine which was observed to cleave at about twice the rate of the other interior peptide bonds. This and the more negative entropy of activation for hydrolysis of the N-terminal residue was seen to be a result of the coiling of the peptide chain, a phenomenon which is observed in peptides of this length.

Steric effects are also evident in the hydrolysis of dipeptides. The resistance to hydrolysis of dipeptides containing N-terminal valine or leucine is well known and has been attributed to the steric restriction of the side-chain on the approach of a hydronium ion.[62] Long observed



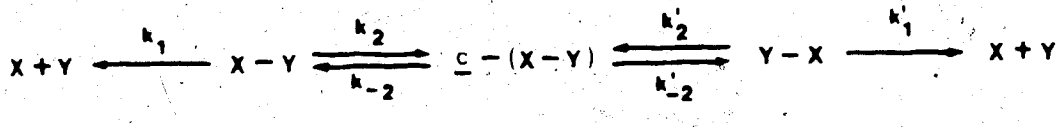
a similar effect in the hydrolysis of gly-leu-gly and leu-gly-leu.[73] The C-terminal peptide [redacted] in the former peptide hydrolyzes at about one-half the rate of the latter. As in gly-leu, the isobutyl group in gly-leu-gly again is considered to be interfering with the approach of the protonating species. These observations can also be explained by initial internal aminolysis to yield diketopiperazine. The steric restrictions of the side-chains could also be expected to influence diketopiperazine formation. The formation of dipeptides may be the result of hydrolysis of initially formed diketopiperazine favoured by the thermodynamic stability of the six-membered piperazine ring. The retarded production of free amino-acids from cleavage of the N-terminus, relative to the production of dipeptide, can still be rationalized by the electrostatic argument used to explain the stability of dipeptides. However, the extrapolation of dipeptide hydrolysis behaviour to that of higher peptides is questionable as the evidence is uncertain for the operation of the electrostatic effect further than the peptide bond adjacent to the N-terminus.

Alternative Explanation for Dipeptide Racemization Behaviour

The competitive processes of hydrolysis and diketopiperazine formation can effect the apparent rate of racemization, particularly when the total mixture of

peptides, diketopiperazines and hydrolysis fragments* is used. It is possible to provide an alternative explanation for the racemization behaviour of dipeptides studied by Smith and Mitterer.

Consider the hypothetical situation where a dipeptide X-Y and its sequence inverted isomer Y-X are undergoing competitive cyclization to the diketopiperazine and hydrolysis to the amino acids X and Y (Scheme 1).



SCHEME 1

It is assumed for this system that the amino acids racemize slowest and the diketopiperazine fastest in accord with the observations of Mitterer[45,46] and Smith[25,47] for dipeptides and Bada[72] for diketopiperazines. It will also be assumed that the rate of racemization of X or Y in the dipeptide is independent of position. Suppose that X-Y hydrolyzes faster than Y-X, $k_1 > k'_1$. Ignoring the effect of cyclization, the effect on the shape of the canonical plot can be determined. Since X-Y is yielding more slowly racemizing amino acids faster than Y-X, the observed rate constant for either X or Y

*For dipeptides, the hydrolysis fragments are amino acids.

centres found in the total mixture taken from a canonical plot for the total hydrolysis mixture, will approach the values for the amino acids faster for X-Y than for Y-X (Figure 5).

It will appear that X is racemizing faster at the C-terminus than at the N-terminus and that Y is racemizing faster at the N-terminus than at the C-terminus.

A similar analysis can be done for the cyclization to diketopiperazine. Suppose that X-Y gives diketopiperazine faster than Y-X. In this situation (ignoring the different rates of hydrolysis of X-Y and Y-X), the canonical plot will appear as in Figure 6.

The result of the effect of diketopiperazine formation is reversed in this situation. It will now appear that X is racemizing faster at the N-terminus than at the C-terminus and that Y is racemizing faster at the C-terminus than at the N-terminus since X-Y produces the more rapidly racemizing diketopiperazine faster than Y-X. These scenarios should be compared with the result from the literature (Figure 7).

Two things are apparent from this analysis. First, if X racemizes faster at one terminus than at the other terminus, the opposite will be true for Y. Second, the relative magnitude of the rate constants for cyclization and hydrolysis will determine which terminus will racemize faster.

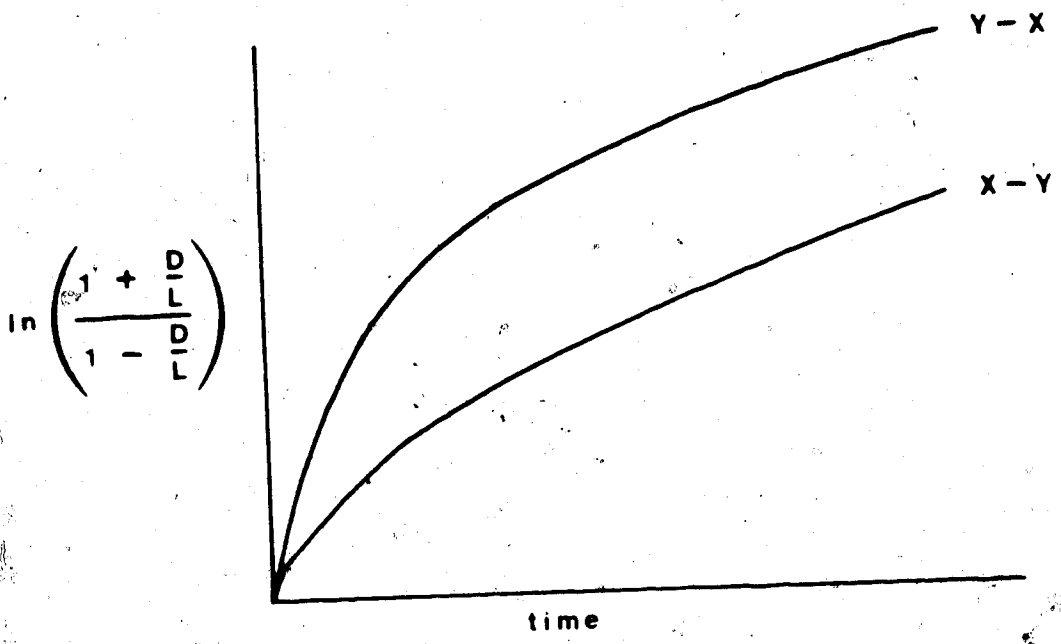


Figure 5. Predicted canonical plot, $k_1 > k_1'$, Scheme 1, showing the effect of hydrolysis of the dipeptide.

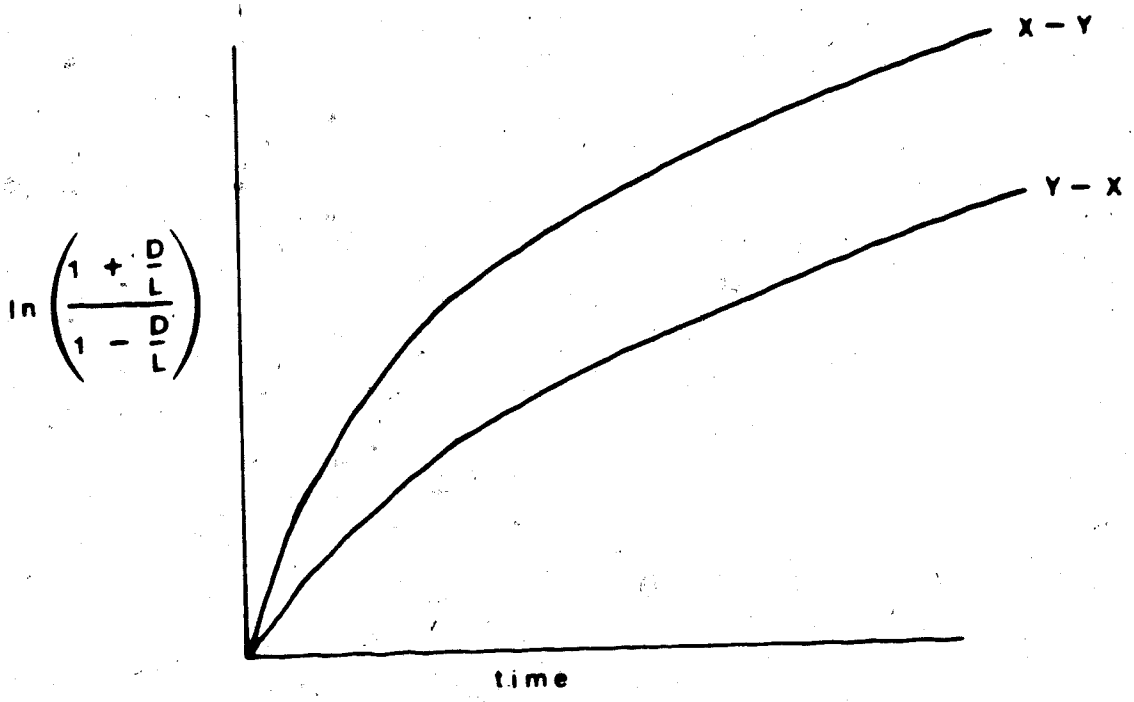


Figure 6. Predicted canonical plot, $k_2 > k_2'$, Scheme 1, showing the effect of cyclization of the dipeptide to the diketopiperazine.

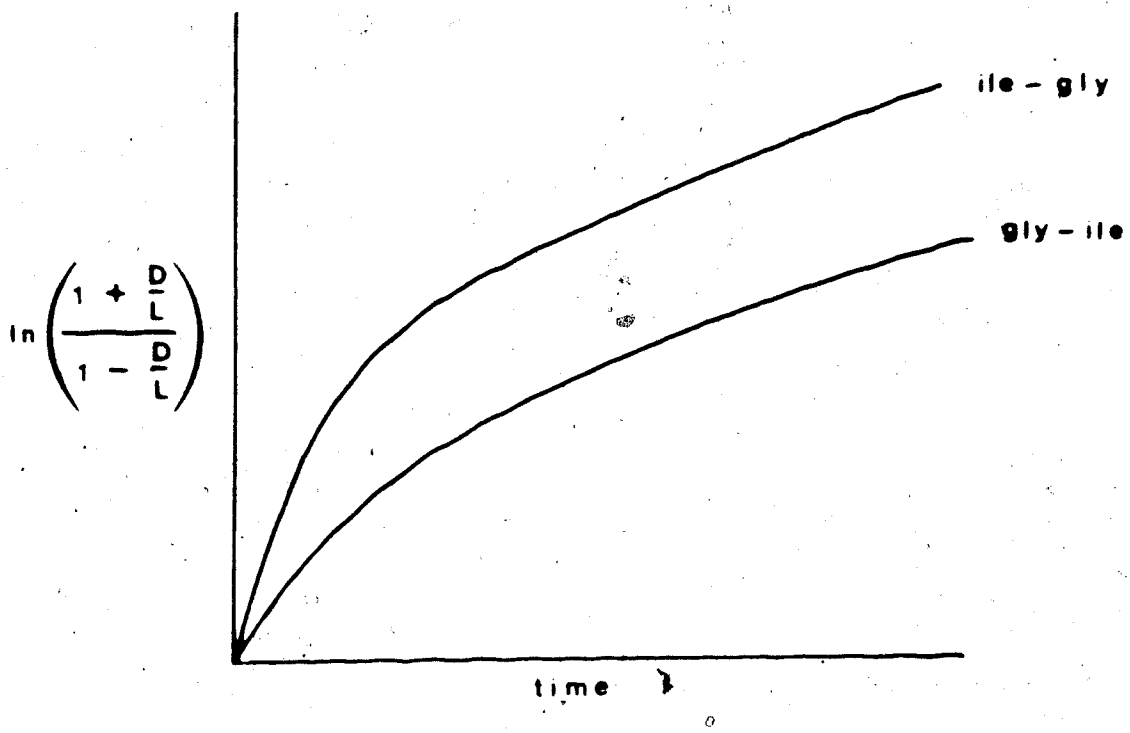
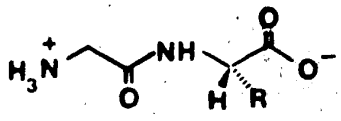
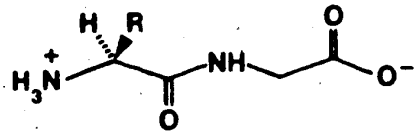


Figure 7. Experimental canonical plot from the work of Kriausakul and Mutterer (Ref. 41).

The rates of acid hydrolysis of dipeptides provide information which can be used to predict the outcome of dipeptide racemization. In general, gly-X hydrolyzes faster than X-gly. This result is interpreted to arise from steric limitations imposed on the approaching solvent molecule by the side-chain of X. In gly-X (**13**), the side-chain is beta to the carbonyl which is to be attacked by the solvent during hydrolysis.



13



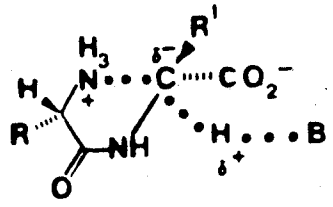
14

In X-gly (**14**), the side-chain is alpha to the carbonyl. The carbonyl is more hindered in this case, resulting in a lower rate of hydrolysis.

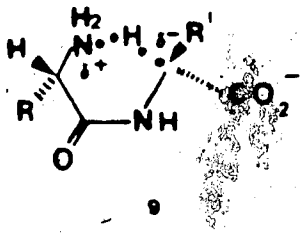
The importance of the size of X as well as the proximity to the carbonyl for the acid hydrolysis of gly-X dipeptides (1.5N HCl, 100°) has been demonstrated. The rate follows the order X = pro > ala > leu > val > ile.[64,54,80,81] In Smith's study,[25,47] the dipeptides X*-pro (X = val, phe, leu, glu, ala), val*-X (X = asp, phe, leu, ala) and X*-gly (X = val, ile, ser) showed N-terminal racemization to be faster than C-terminal racemization for the indicated (*) peptide units (e.g. for racemization of phe in phe-pro and pro-phe, phe-pro > pro-phe). The

dipeptides X-val (X = ala, leu, phe, asp) showed C-terminal racemization faster than N-terminal racemization for peptide unit X. These results may be seen as a consequence of competing hydrolysis. It is expected from the hypothetical case that X-val and pro-X will hydrolyze faster than their sequence inverted isomers. For example ala-val is predicted to hydrolyze faster than val-ala. Since ala-val will be producing the more slowly racemizing amino acids faster than val-ala, both the val and ala peptide units in val-ala should be observed to racemize faster than the same peptide units in ala-val. This is the observed behaviour for this pair of dipeptides as well as for the two series of dipeptides above.

The dipeptides X-gly (X = ala, leu, phe, asp, met) show C-terminal racemization to be faster than N-terminal racemization. Smith suggests that this behaviour results from intramolecular interaction of the N-terminal amino group with the alpha proton of the C-terminal peptide unit as in 8 and 9.



8



9

Since glycine is the least sterically demanding amino acid, this interaction is especially noticeable in gly-X

dipeptides. Gly-X produces sequence inverted X-gly faster (and by implication, diketopiperazine) than X-gly produces gly-X (X = ala, leu, val).[54] Indications of this tendency are evident in a study by Bada[72] where X = ile.

The alternative proposal to Smith's hypothesis involves preferential cyclization of gly-X to diketopiperazine compared with X-gly. As in the hypothetical dipeptide X-Y, the observed rate of racemization for the C-terminal peptide gly-X is perceived to be faster than for X-gly.

In this analysis the rate of epimerization of a peptide unit in dipeptide has been assumed to be independent of the position in the dipeptide. The preference for faster racemization of a peptide unit at the N- or C-terminus is swamped by the competing processes of hydrolysis and cyclization. This analysis fails for only one pair of dipeptides, val-ile and ile-val. Here, N-terminal val and ile racemize faster than C-terminal val and ile. It is noteworthy that first, this is the only change where both of the peptide units have branching at the beta carbon and second, that racemization is slower for this pair of dipeptides than for any other.[54] Both hydrolysis and cyclization may be retarded allowing for a third factor to become dominant.

The hydrolysis behaviour at neutral pH and their tendency to cyclize is largely unknown. In particular,

there appears to be no information available on the relative magnitudes of the rates of hydrolysis and cyclization. However, it is known that diketopiperazines are observed in greater amounts as the acidity of the solvent is reduced.[76] Knowledge of the relative magnitudes of the rates of hydrolysis and cyclization will determine the viability of the alternative explanation proposed here.

Summary

Peptide hydrolysis, cyclization to diketopiperazines and asymmetric induction are processes which can have a large effect on the observed rate of racemization for an amino acid in a peptide or protein. Tests for the effect of the primary structure of peptides on the racemization of the peptide units have given conflicting results.

Objectives

The use of amino acid racemization as a tool for geological dating has created the need to better understand racemization in peptides and proteins. Racemization of a peptide has been treated as if it were occurring in discrete amino acids. For the purposes of geological dating and stratigraphy, the first order (or pseudo first order) rate law has been assumed to be operating for amino acids, peptides, proteins and protein

hydrolysates of changing composition. This simplification ignores both the asymmetry of peptides and competing cleavage of peptide bonds. It has never been clearly demonstrated that the process of epimerization in peptides obeys the simple first order rate law. Furthermore, previous investigations have not involved the separation of the substrate peptide from the hydrolysis mixture. There have been no time studies carried out on a model peptide where both hydrolysis and racemization of the peptide are followed. The effect of the presence of more than one epimerizing centre has not been adequately addressed, particularly with respect to the potential for asymmetric induction.

In order to model the protein situation, the racemization of tetrapeptides and their N-terminal and C-terminal derivatives were studied.

By studying model peptides some of the properties of peptides were studied which can complicate the interpretation of peptide racemization measurements used for dating fossils. By studying peptides which differ by only one peptide unit it was possible to establish the effect of neighbouring peptide units on the rate of racemization. By examining peptides having more than one epimerizing chiral centre it was possible to observe asymmetric induction. This provided definite indication that the pseudo first order rate law for amino acid does

not apply in the case of peptides.

In previous model studies the rate constant extracted from the protein hydrolysate has been taken to be representative of the protein. These rate constants have then been used in structure reactivity arguments. The validity of this reasoning was tested by comparing the epimerization rate constants of the isolated tetrapeptides with those of the hydrolysate. The species which cause the change in the rate constant of the hydrolysate were identified.

In general, an attempt to extract meaningful rate constants for racemization of proteins from plots using the simple pseudo first order rate law will fail. Models were devised to include the influence of asymmetric induction and hydrolysis. In light of the agreement between the theoretically predicted and the experimentally observed behaviours of the tetrapeptides and the observed similarity between proteins and tetrapeptides, the racemization of proteins can be better understood.

CHAPTER 2

MATHEMATICAL MODELS FOR PEPTIDE RACEMIZATION

From the discussion in the Introduction it is evident that peptide and protein racemization will be complicated by hydrolysis and the intermediacy of diketopiperazines and intermediates formed by internal aminolysis. The failure of Equation 1 to handle these situations is intuitively obvious. A precise mathematical treatment of the possible effects of hydrolysis and diketopiperazine formation has not been done. The purpose of the following discussion is to explore the scope of these effects using mathematical models of varying degrees of sophistication.

The most general mathematical treatment of peptide racemization would involve many rate equations including the rate constants for the epimerization of each of the different chiral centres in the peptide. For a peptide with n chiral centres there will be 2^n diastereomers (except in the case of accidental symmetry). These must be solved for from 2^n differential equations. In addition the hydrolysis of the peptide must be accounted for as this process leads to more species which like the diastereomers of the original peptide, epimerize at different rates. Even for relatively small peptides

containing a few amino acid monomer units, the prospect of solving such a set of differential equations is intimidating.

The purpose of this work is to model the racemization of amino acids present in protein with small peptides containing one or two chiral centres. Restricting the model in this way has the advantage of reducing the number of hydrolysis products and diastereomers so that a complete experimental picture of the system is at least feasible. Likewise, the mathematical solution is within grasp. An accurate mathematical treatment is essential to a thorough understanding of the situation. Any mathematical model may or may not fit the experimental data and thus give an idea of the value of the theory, but the real importance of this mathematical analysis is in the potential for discovering cases where the experimental data may be interpreted incorrectly. For example, the slope of the line produced by plotting the canonical expression versus time may not necessarily contain only the rate constants for epimerization, even if it appears to be straight.

The above criticism is most readily apparent in the work done by Mitterer and Kriausakul[41,45,46] in which they are concerned with the comparison of the rates of epimerization of proteins and model dipeptides. Without

exception the canonical plots are curved, having an initial slope greater than that of the free amino acid and the final slope equal to that of the free amino acid. The authors quite reasonably take the initial slope as being equal to twice the rate of epimerization for the initial peptide and state that the curvature in the canonical plot is due to the contribution of the various components in solution which eventually all hydrolyze to free amino acids. Of course the problem with this approach is mainly with the difficulty in obtaining data near time zero. Therefore, the rates of epimerization which are quoted in studies using the initial slope method should always be taken as approximations of the actual rates. Further, it can be shown by mathematical simulations of these exact situations, that the canonical plots are in general always curved. The deviation from linearity and hence the degree to which the slope of the canonical plot does not reflect the rate of epimerization depends upon the values of the rates of epimerization of the various species in solution as well as the rates of hydrolysis. Clearly, the meaning of the value of the initial slope in studies of this nature is unknown without an independent measurement of the rate of hydrolysis and a check on the degree of reversibility. Mitterer and Kriausakul do try to measure the rate of hydrolysis following the production of free

amino acid. However, this method must be criticized for giving a lower limit for the rate of hydrolysis of proteins as it does not account for initial hydrolysis to species other than the free amino acid nor does it allow for the reversibility of any of the hydrolysis steps.

Kinetic Models

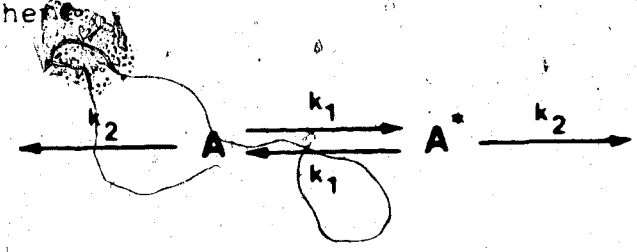
In deriving the rate expressions (Equation 9 and Appendices 1 through 4) two important assumptions have been made. Firstly, the effect of the chiral environment present in living systems is completely ignored. As racemization of the amino acids proceeds, the environment "racemizes" as well. Any preference for one enantiomer over the other produced by the chiral environment is gradually diminished with time. However, it appears that such an effect, which could be present even in the case of amino acids containing only one chiral centre, is not observed at the concentrations employed in the racemization studies.

Secondly, it is assumed that the starting material is homogeneous. Here "homogeneous" means that the starting material consists of only one conformer with a single rate of epimerization or that these conformers are equilibrating at a rate much faster than epimerization or hydrolysis. This assumption allows the substitution of this mixture of conformers by a single species.

Peptides with One Epimerizing Centre

The derivation of the simplest case will be detailed here and the more complex cases relegated to the appendices.

The L and D isomers of a peptide containing one chiral center are denoted by A and A* respectively. The enantiomers can undergo reversible epimerization as well as irreversible hydrolysis to shorter peptides, amino acids and diketopiperazines. This is represented in the following scheme:



SCHEME 2

Here and in the appendices concentration brackets are left off for clarity. The rate of change with respect to time of the enantiomers can be written as:

$$\frac{dA}{dt} = -(k_1 + k_2)A + k_1A^* \qquad (6)$$

$$\frac{dA^*}{dt} = k_1A - (k_1 + k_2)A^*$$

It follows that:

$$\frac{d}{dt}(A+A^*) = -k_2(A+A^*) \tag{7}$$

and

$$\frac{d}{dt}(A-A^*) = -(2k_1 + k_2)(A-A^*) \tag{8}$$

These differential equations are easily solved:

$$(A+A^*) = (A+A^*)_0 \, e^{-k_2 t}$$

$$(A-A^*) = (A-A^*)_0 \, e^{-(2k_1+k_2)t} \tag{9}$$

where $t = 0$, $A+A^* = (A+A^*)_0$ and $A-A^* = (A-A^*)_0$.

and the canonical Equation 11 is obtained by dividing the second solution by the first:

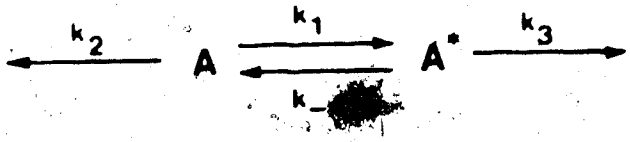
$$\frac{A-A^*}{A+A^*} = \frac{(A-A^*)_0}{(A+A^*)_0} \, e^{-2k_1 t} \tag{10}$$

$$\ln\left(\frac{1 - A^*/A}{1 + A^*/A}\right) = \ln\left(\frac{1 - A^*/A}{1 + A^*/A}\right)_{t=0} - 2k_1 t \tag{11}$$

A plot of the left-hand side of Equation 11 (the canonical expression) versus time will yield a straight line with the magnitude of the slope exactly equal to twice the

epimerization rate constant. More than one irreversible hydrolysis pathway could be included with an identical result, the canonical plot (Figure 8) will always yield a straight line independent of hydrolysis.

The presence of another chiral centre which does not epimerize (for example isoleucine), considerably complicates the situation. The equilibrium value of A*/A is no longer unity (1.3-1.4 for isoleucine). Similarly the rates of hydrolysis for the D and L isomers will not necessarily be the same. Appendix 1 contains the derivation of the equation pertaining to this kinetic scheme:



SCHEME 3

The solution to this system expressed in a form most closely resembling the canonical equation is:

$$\ln\left(\frac{A - xA^*}{A + zA^*}\right) = \ln y - I \cdot t \qquad (12)$$

where x, y and z are functions of all four rate constants. The exact expressions are in Appendix 1. The degenerate case where $k_2 = k_3$ does have a simple solution. The above equation simplifies to:

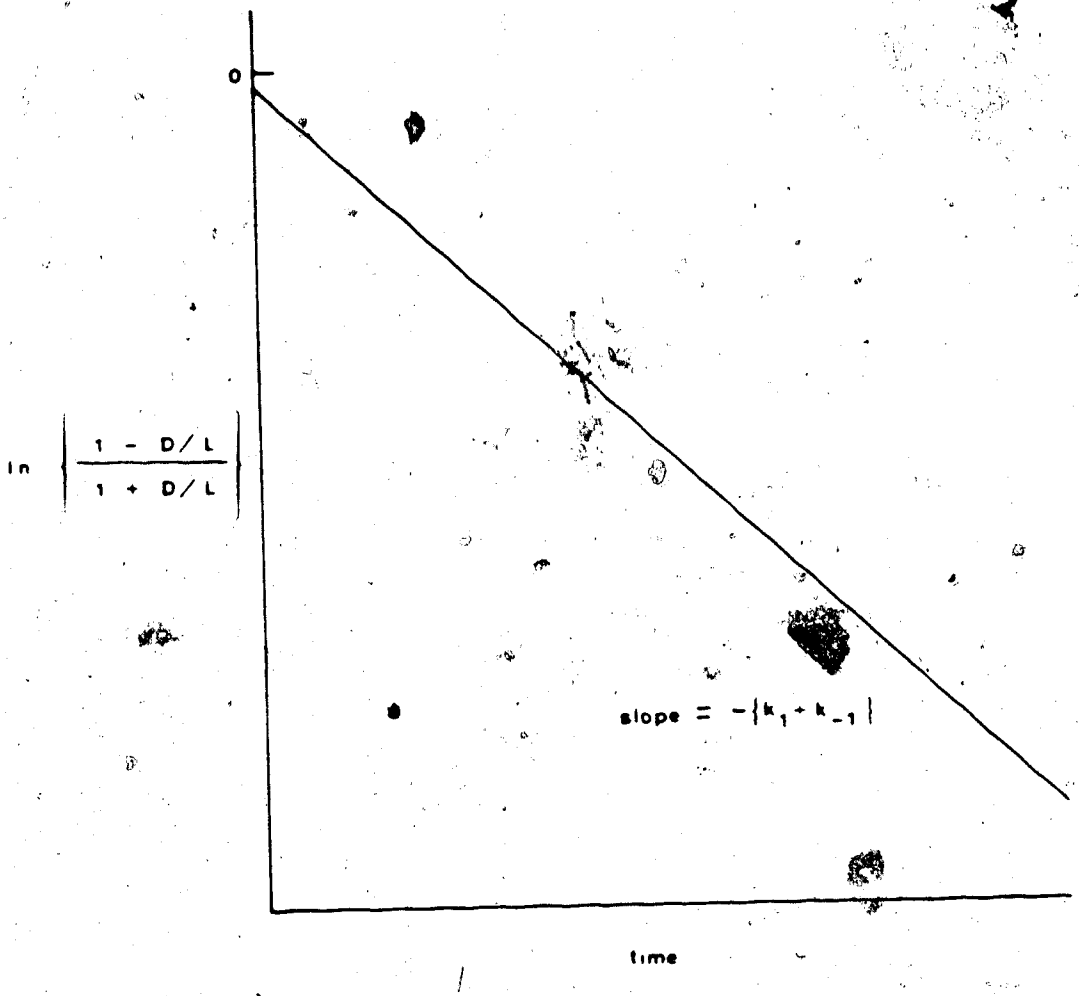


Figure 8. The canonical plot. For the interpretation of the rate constants see Scheme 2.

$$\ln\left(\frac{1 - K' A^*/A}{1 + A^*/A}\right)_t = \ln\left(\frac{1 - K'A^*/A}{1 - A^*/A}\right)_{t=0} - (k_1 + k_{-1})t \qquad (13)$$

where $K' = k_{-1}/k_1$.

Again the slope of the canonical plot is independent of hydrolysis. However, if $k_2 \neq k_3$ the situation is entirely different. The effect of hydrolysis in a number of simulations is quite apparent. In Figure 9, the canonical plot is the expected straight line when $k_2 = k_3$. When $k_2 \neq k_3$ but still with the values quite similar, there is a pronounced curvature. If the canonical expression (11) is extended to its more general form (12), the curvature is eliminated (Figure 10) although a contribution from hydrolysis to the slope is clearly present.

The curvature produced by hydrolysis in the canonical plot (and also the effect on the observed rate constant) was previously thought to arise from the presence of more than one epimerizing species. The above simulations show a completely different sort of effect which can occur in a system which is more complex than the system in which there is no hydrolysis. The magnitude of the curvature in the canonical plot for a single species is dependent on the size of the hydrolysis rate constants but will not be

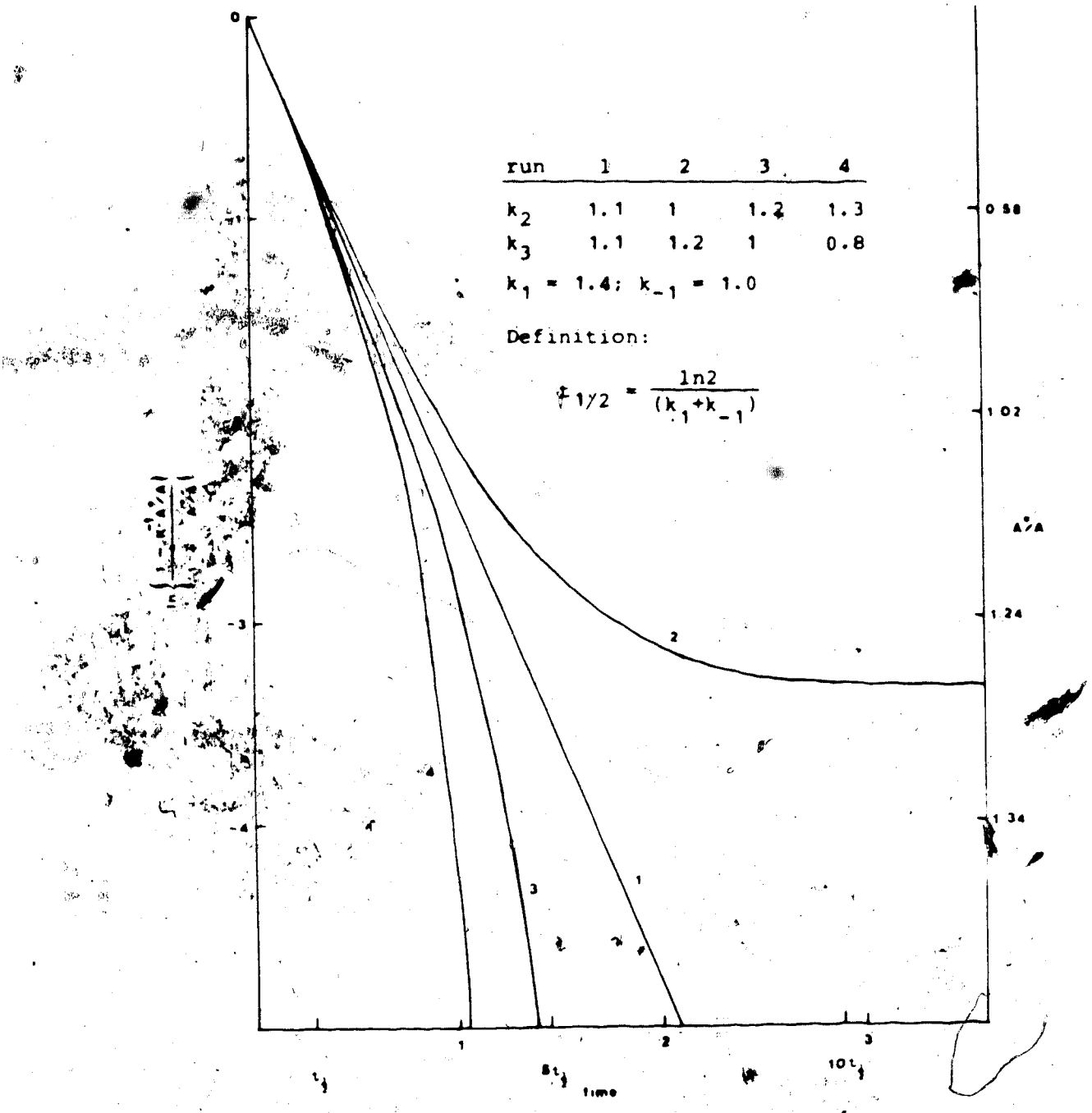


Figure 9. The effect of hydrolysis on the canonical plot, Scheme 3, Equation 13. In Figures 9 through 16, the units of time and the units of the the rate constants are not given. The units of the rate constants are the reciprocal of those chosen for time.

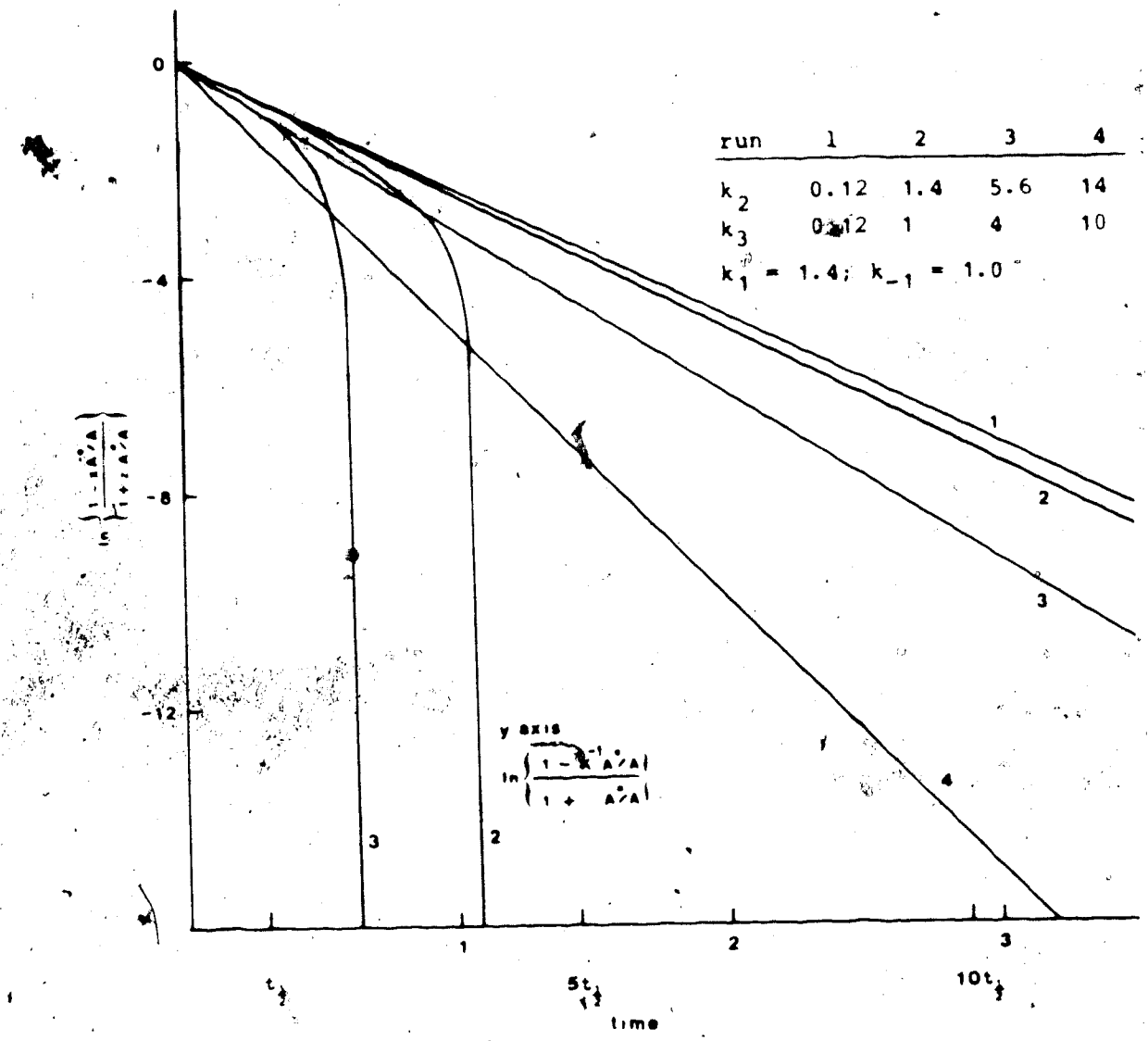
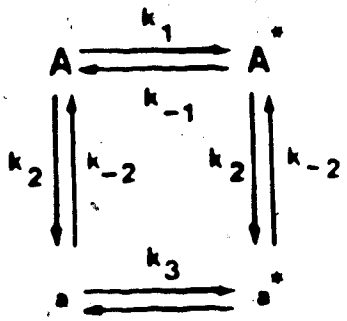


Figure 10. The effect of hydrolysis on the canonical plot, Scheme 3, Equation 12.

apparent unless there are diastereomers and the rates of hydrolysis for the diastereomers are different.

For the simple tetrapeptide gly-leu-gly-gly, the above effect will not be observed since there are no diastereomers. This does not preclude the possibility of other kinetic schemes which can lead to curvature of the canonical plot. Specifically, a mechanism which involves more than one epimerizing pair of enantiomers can produce curvature in canonical plots of the total hydrolysate. This is the same phenomenon which Mitterer (and others) described and is depicted in Figure 11. This simulation was produced using the scheme below.



SCHEME 4

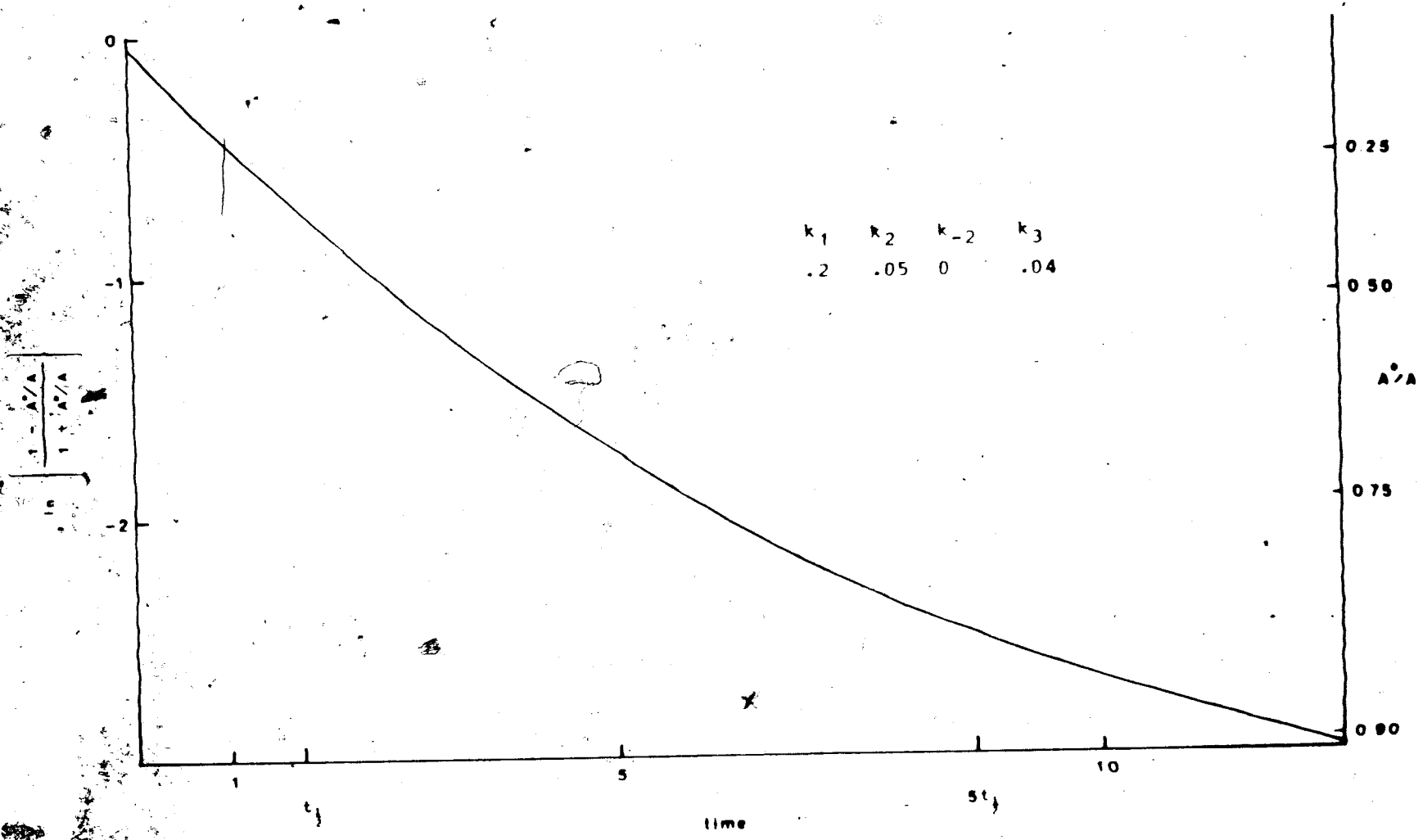
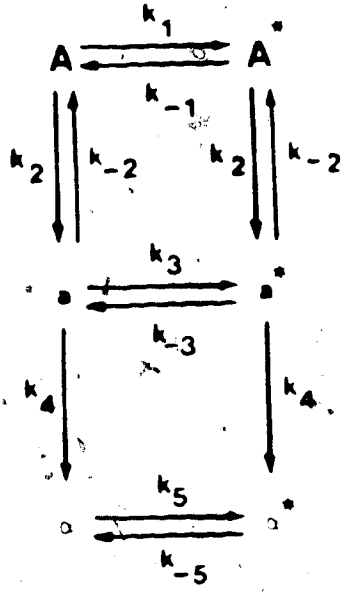


Figure 11. Simulated canonical plot for a fossil total hydrolysate, Scheme 4.

In the example in Figure 11, the curvature in the canonical plot for the total hydrolysis mixture is produced as a result of the inequality of k_2 and k_3. With $k_{-2} = 0$ this example represents a simple model of the typical fossil situation where there is a gradual irreversible change in the composition of the fossil peptide mixture with a concomitant change in the [illegible] racemization rate constant. A slightly more sophisticated model which consists of three pairs of enantiomers is shown below:



SCHEME 5

From a geochemical point of view, the pair A and A* might represent the original protein while pair a and a* represents the set of components produced by partial hydrolysis of the substrate protein. This mixture eventually hydrolyzes to the free amino acid pair α and α^*. As will be demonstrated below, curvature in the canonical plot of A and A* results only when the reversible step (k_{-2}) is included. As in the simpler models, curvature of the canonical plot for the total hydrolysate is possible without the reversible step. Appendix 2 contains the derivation of the rate expressions for this kinetic scheme and the computer program which generated the simulations.

The presence of curvature in the canonical plots presents an obvious problem for the extraction of accurate rate constants and can be stated as a theorem:

> Curvature in a canonical plot indicates that the rate constant extracted from the slope is in error.

A corollary to this theorem might be expressed as:

> The slope of a canonical plot exhibiting straight-line behaviour is equal to twice the epimerization rate constant.

The corollary is tested by simulation in Figure 12. A worker who produced these plots after separation of the hydrolysis mixture would be ecstatic with his good fortune. The canonical plot of the starting material (A

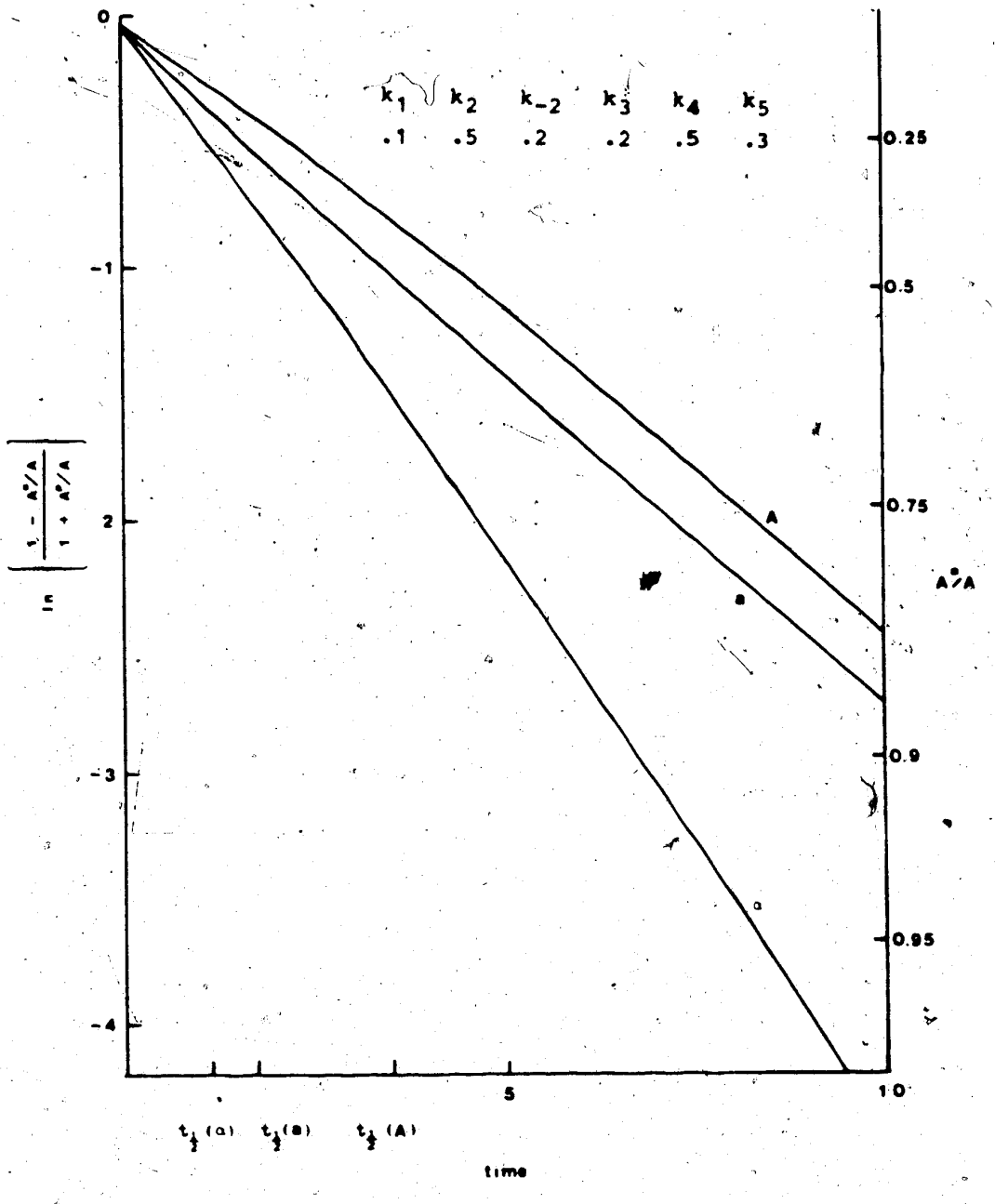


Figure 12. Simulated canonical plot, Scheme 5.

in Figure 12) shows some initial curvature. Linear regression of the data produces correlation coefficients in excess of 0.999 in all three cases. The "rate constants" extracted from the slope of the canonical plots are shown in Table 2 along with the true values used in the simulations. It is apparent that the corollary stated above is false. It was this kind of reasoning which was used in some reports on peptide racemization.

It should be noted that the rate constants extracted for intermediate hydrolysis fragments can be significantly different from the true values. This is also true for the free amino acid if the rate constant is extracted from data taken when there are even small amounts of the precursor species present in the hydrolysis mixture. The epimerization rate constant extracted for the substrate approximates the true rate constant when k_{-2} is nearly equal to zero (Table 3). It would be useful to have some indicators of the inaccuracy of the rate constant extracted for the substrate. From analysis of simulated kinetic runs it appears that these indicators are: (1) curvature of the canonical plot for the substrate near time zero and (2) regression y-intercept of the canonical plot not equal to those of the intermediates and free amino acid.

Table 2. Comparison of True and Observed Rate Constants from Simulated Kinetic Runs - Scheme 5

	k_1	k_3	k_5
true	0.100	0.200	0.300
obs	0.121	0.132	0.223
r	0.999	1.00	1.00

$k_2 = 0.5$; $k_{-2} = 0.2$; $k_4 = 0.5$

Table 3. Value of Extracted Rate Constant (k_1) in Scheme 5

	k_2	
k_{-2}	0.2	0.5
0.0	0.100	0.100
0.2	0.107	0.122
0.5	0.108	0.121

$k_1 = 0.1$; $k_3 = 0.2$; $k_4 = 0.5$; $k_5 = 0.3$

The steady-state situation is also capable of producing straight-line canonical plots with the slope approximating the true epimerization rate constant. Extracted rate constants for this degenerate case with $k_4 >> k_{-2}$ are shown in Table 4. The case is of particular interest when considering the possible existence of intermediates isomeric with the substrate peptide.

The accuracy with which a rate constant can be extracted from the canonical plot of the total hydrolysis mixture is shown (Figures 11-13). The initial slopes (Table 5) appear to have little in common with the true values of the rate constants of any of the species in the hydrolysis mixture. In theory, it should be possible to extract the true rate constant from the initial slope. The unknown values of the hydrolysis rate constants coupled with the difficulty of obtaining data near time zero make the chances for the success of this method very slim.

In the Introduction it was mentioned that the canonical plot of the total hydrolysis mixture should have a portion near time zero where the slope is increasing if the intermediate hydrolysis fragments epimerize at a rate faster than the substrate. This effect is visible only to a modest degree in most cases (Figures 11-13) but is quite apparent in situations where there is some reversibility

Table 4. Effect of Steady-State in a + a* on Extracted Rate Constant k_1 - Scheme 5

	k_4		
k_{-2}	0.1	1	10
0.0	0.100	0.100	0.100
0.1	0.143	0.106	0.1001
0.5	0.135	0.112	0.1002
1.0	0.126	0.117	0.1004

$k_1 = 0.1$; $k_2 = 0.5$; $k_3 = 0.2$; $k_5 = 0.3$

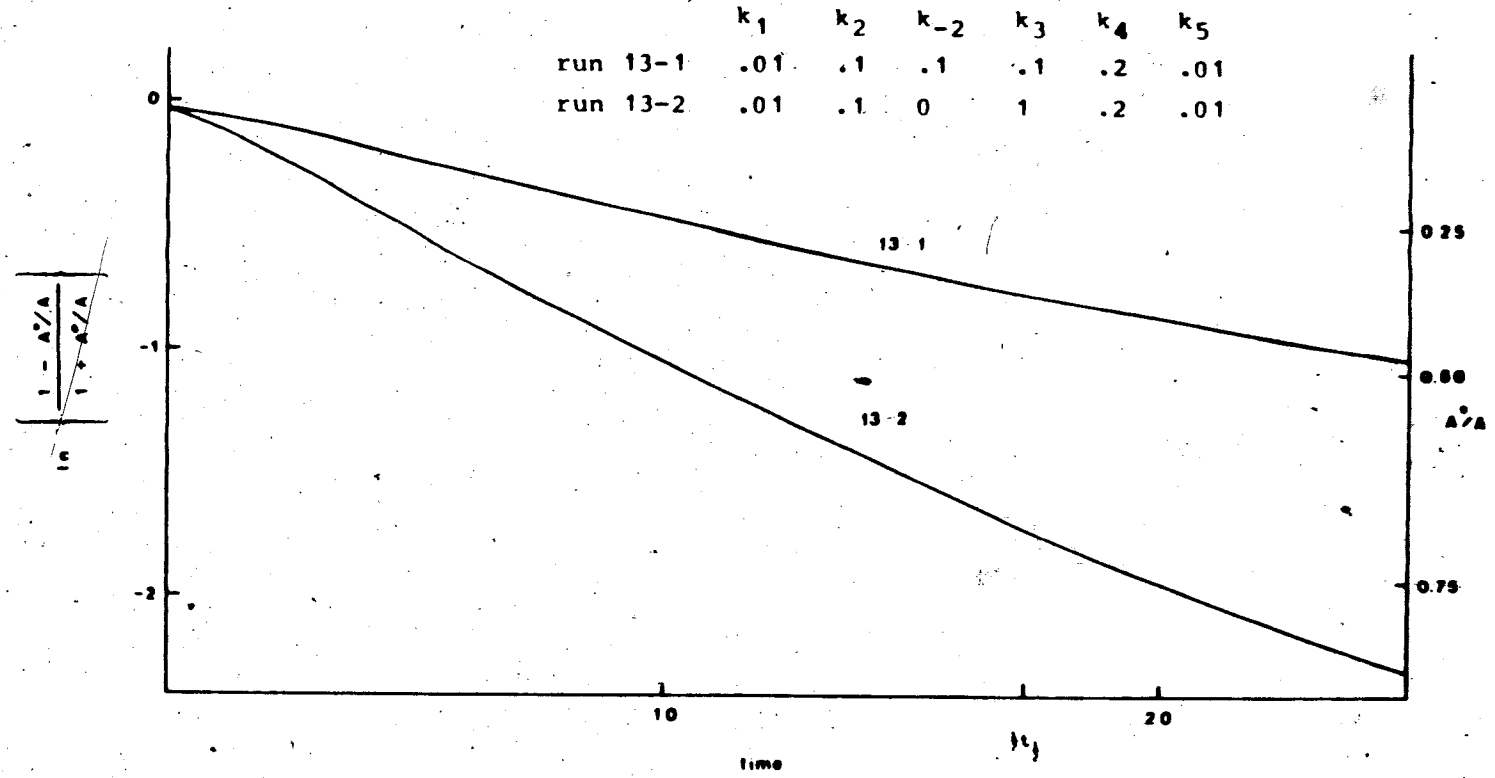


Figure 13. Simulated canonical plot for the total hydrolysate, Scheme 5, showing the difficulty in obtaining an accurate initial slope.

Table 5. Comparison of True Rate Constants with Extracted Initial Values from Total Hydrolysis Canonical Plots

Figure	k_{true}	k_{obs}
11	0.200	0.185
14	0.010	0.091
13-1	0.010	0.023
13-2	0.010	0.052

in the initial hydrolysis step and a large difference in the epimerization rate constants (Figure 14). This relatively rare event underscores the serious limitations of extracting rate constants from initial rate data since it will not always be apparent that this procedure will be incorrect.

From these computer generated simulations the prospect for measuring epimerization rate constants free of contributions from hydrolysis seems quite bleak. There appear to be only two chances that true rate constants can be measured. Firstly, if there is no reversibility in the initial hydrolysis step, the epimerization rate constant may be obtained. Alternatively the rate constants for all the process could be measured by time studies of all the components in the hydrolysis mixture. This is clearly a formidable undertaking even more complex than the studies undertaken by Long and coworkers on the hydrolysis of peptides.[54]

Peptides with Two Epimerizing Centres

The kinetic scheme for the case including diastereomers is shown in Scheme 6. This situation can be considered to be an extension of the models discussed in the previous section. Since the single-centred case is a degenerate two-centred case, it will be taken as proven that all the problems associated with the proper

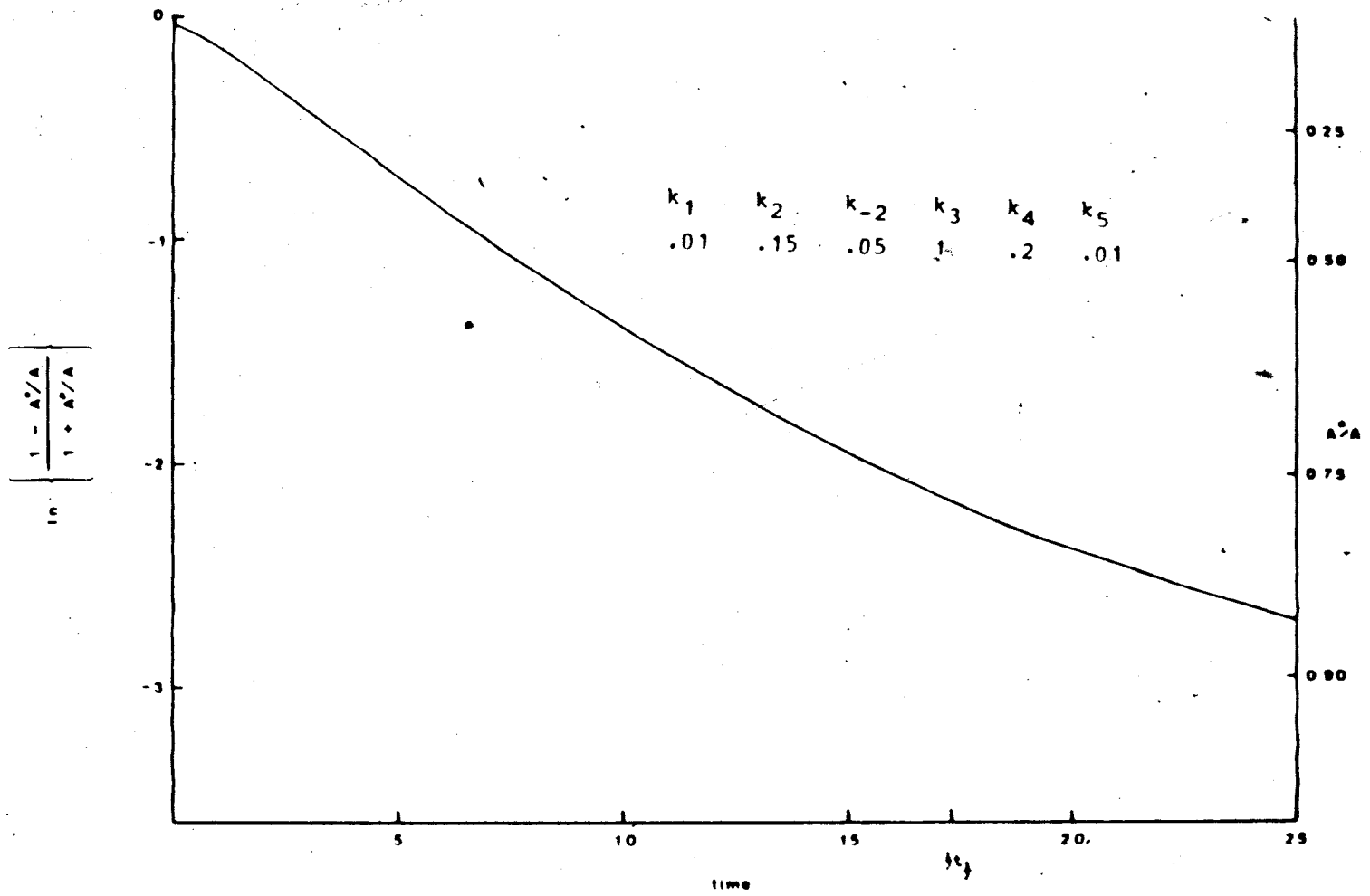
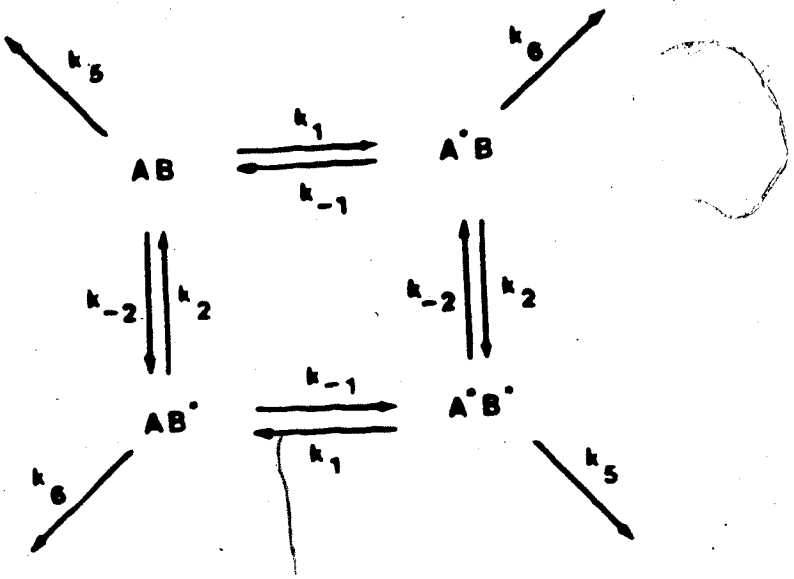


Figure 14. Simulation of epimerization in the total hydrolysate from a fossil protein, Scheme 5. The initial downward curvature of the canonical plot is demonstrated.



SCHEME 6

interpretation of experimental results are applicable here as well.

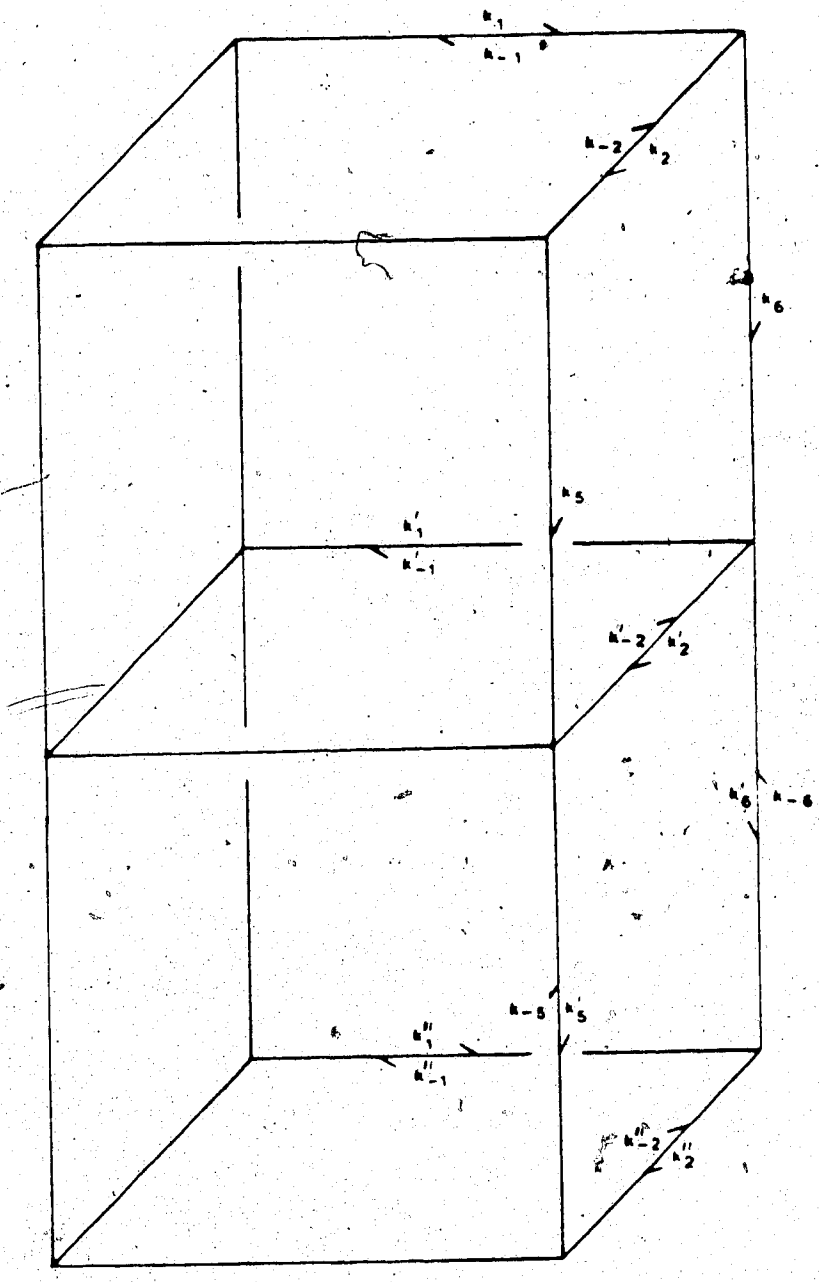
What remains to be discussed for the two-centred case is the possibility of asymmetric induction. The algebraic solution of Scheme 6 is presented in Appendix 3 where it is shown that asymmetric induction in either of the chiral centres is possible under certain circumstances. From the information in Table 6 it is evident that it is not possible to observe asymmetric induction at both centres at the same time.

A computer program was developed to model the more complex situations shown in Schemes 7 and 8. Here, the additional epimerizing species and the possibility of reversal of the first hydrolysis step makes the associated algebra unwieldy. Therefore, iterative techniques were used to solve the associated differential equations. The details of these solutions are given in Appendix 4.

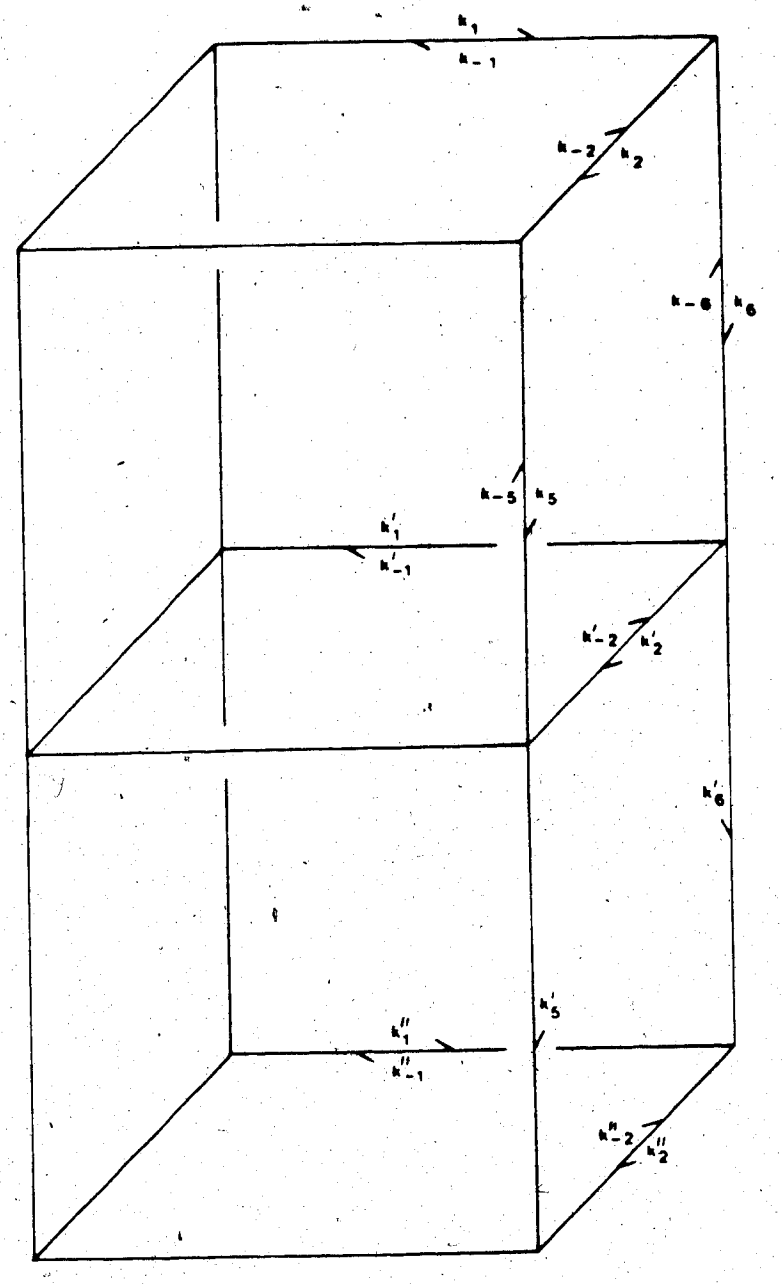
Each horizontal plane in Scheme 7 contains a set of four equilibrating diastereomers. The planes are connected by hydrolytic pathways. Only the lower hydrolysis pathway is reversible, a condition set up to model the irreversible formation of diketopiperazine from the substrate peptide. The diketopiperazines in turn are interconverting with the corresponding dipeptides. The accompanying figures (Figures 15 and 16) were generated by

Table 6. Conditions for Asymmetric Induction

Centre	Conditions
A	$k_1 > k_{-2}$ and $(k_5 - k_6) > 2(k_{-1} - k_1)$
B	$k_{-2} > k_1$ and $(k_5 - k_6) > 2(k_2 - k_{-2})$



SCHEME 7



SCHEME 8

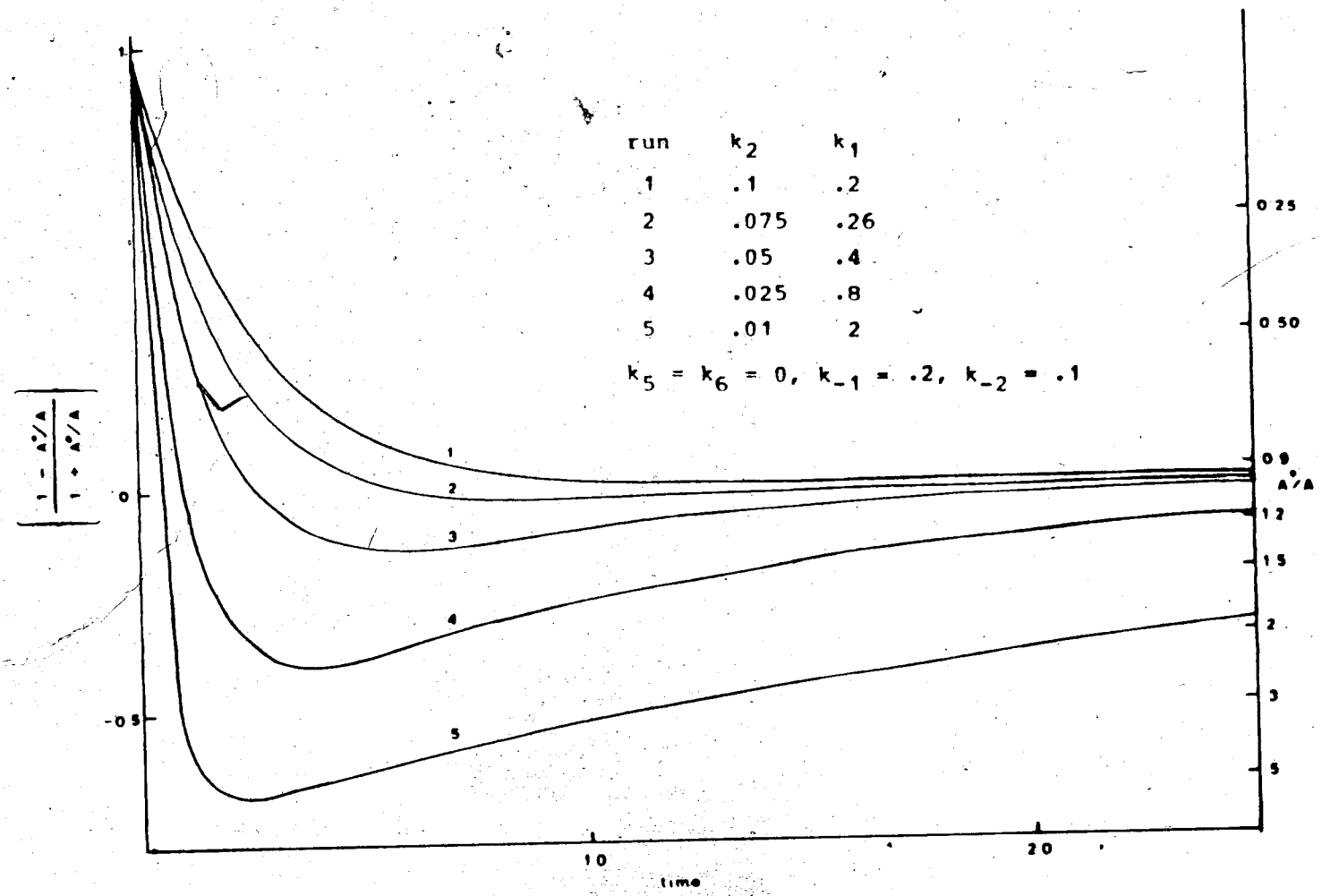


Figure 15. Asymmetric induction in peptides with two epimerizing chiral centres: A centre, Schemes 6 and 7.

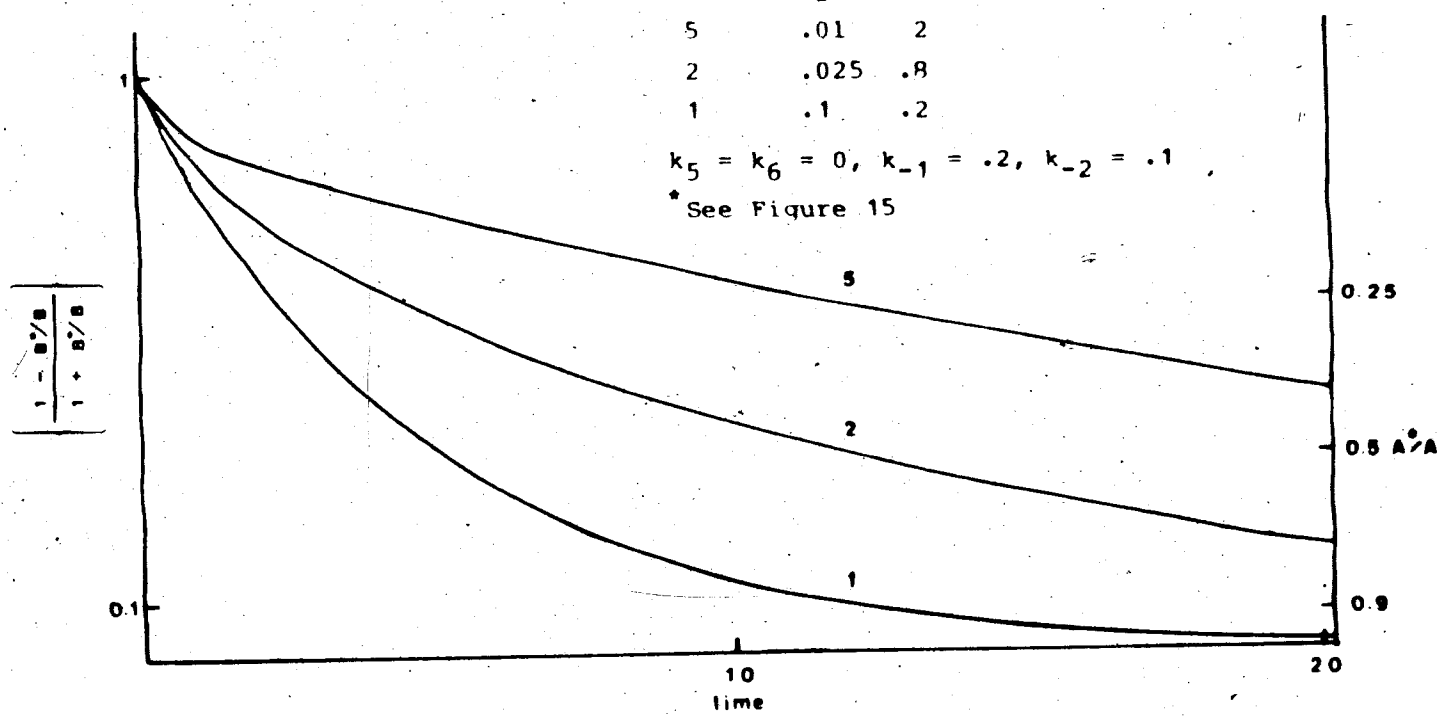


Figure 16. Asymmetric induction in peptides with two epimerizing chiral centres: B centre, Schemes 6 and 7.

this program for the situation identical to the simpler algebraic model shown at the beginning of this section. The results are in accord with those obtained for the algebraic solution.

Scheme 8 shows a situation in which only the upper hydrolysis pathway is reversible. This model was designed to simulate the reversible formation of an intermediate, isomeric with the substrate peptide. The intermediate can hydrolyze irreversibly to a set of interconverting diastereomers in the bottom plane of Scheme 8.

The figures show the range of possible curves for the canonical plots for any horizontal level in Scheme 7 (and for the total hydrolysate as well). The proof that these are the only forms possible with this model is given in Appendix 3.

Summary of Analysis of Kinetic Models

I. Accuracy of Rate Constants Derived from Canonical Plots

Substrate

1. Any reversibility of the initial hydrolysis step will be reflected by a deviation in the slope from its "true" value.
2. The rate of the reversible step and the difference between the epimerization rate constants of the substrate and the intermediate will determine the amount of deviation.

Total Hydrolysate

The extrapolated rate constant at time zero will not be accurate if hydrolysis is competitive with epimerization.

All Components

A straight-line canonical plot is not a guarantee that the true rate constant for epimerization can be extracted. It is in general a linear combination of the rate constants for all processes occurring in the mixture.

II. Asymmetric Induction

This phenomenon can occur at only one centre at a time in the same experiment. Asymmetric induction is another way in which curved canonical plots can be obtained.

CHAPTER 3

EXPERIMENTAL

Materials

Commercial peptides were found to be pure by electrophoresis (3000 V, pH 1.8, 45 min) and by HPLC (conditions below). Benzyl chloroformate, DMSO, N-methylmorpholine, isobutyl chloroformate and aniline were purified by distillation under vacuum. Ethyl acetate was purified by washing with sodium bicarbonate, water and saturated sodium chloride, drying with anhydrous potassium carbonate and distillation from 3A molecular sieves. Water used in lyophilization was deionized, doubly distilled, and deionized again.

Derivatives of peptides were prepared according to Bodanszky.[82,83] Melting points and yields are given in Table 7. Examples of the synthetic preparations follow.

N-Benzyloxycarbonyl L-pro-L-leu-gly-gly

L-pro-L-leu-gly-gly (Bachem, Torrance CA) (146.2 mg, 0.427 mmol) was dissolved in 20 mL water containing excess (18.2 mmol, 1.53 g) sodium bicarbonate. Benzyl chloroformate (14 mmol, 2.0 mL) was added and the mixture mechanically stirred for five hours at room temperature.

Table 7. Melting Points and Yields of Peptide Derivatives

Peptide	m.p. (°C) (uncorrected)	Purified Yield (%)
CBZ-plgg	204-205	62
CBZ-glgg	140 (d)	58
CBZ-plgg-NHPh	115-118	68
CBZ-glgg-NHPh	160-162	79
CBZ-gplg-NHPh	191-193	78
plgg-NHPh	160 (d)	87
glgg-NHPh	180 (d)	85
gplg-NHPh	160 (d)	36
gplg	183-185	54

CBZ = $C_6H_5CH_2OCO-$; Ph = C_6H_5-; g = glycyl, p = prolyl, l = leucyl

Reaction times in excess of this caused significant hydrolysis of the starting material. After this period, the mixture was extracted with two 20 mL portions of diethyl ether to remove the excess benzyl chloroformate. 5 M HCl was added dropwise until the solution reached pH 4. The solution was then extracted with three 25 mL portions of ethyl acetate. After drying the organic layer over magnesium sulfate, the solvent was removed on a rotary evaporator. The resulting white crystalline solid was recrystallized from ethyl acetate/petroleum ether to give 125.5 mg (62%) of material. The product, N-benzyloxycarbonyl-L-prolyl-L-leucyl-glycyl-glycine was characterized by its proton NMR (Tables 8 and 9) and FAB-MS spectra (Table 10).

N-Benzyloxycarbonyl L-prolyl-L-leucyl-glycyl-glycine anilide

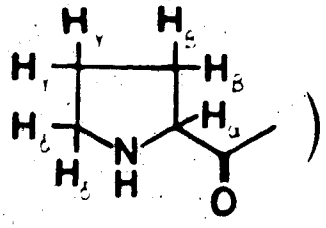
N-Benzyloxycarbonyl-L-prolyl-L-leucyl-glycyl-glycine (0.120 mmol, 57.3 mg) was partially dissolved in ethylacetate (25 mL) and cooled to -15°. Enough dimethyl sulfoxide was added to dissolve all of the starting material. N-methylmorpholine (0.144 mmol, 15 µL) was added by syringe followed immediately by isobutyl chloroformate (0.144 mmol, 20 µL). After five minutes "activation" time aniline (0.12 mmol, 14 µL) was added. The mixture was mechanically stirred at -15° for one hour. Cloudiness developed in the solution and a solid

Table 8. 200 MHz ^1H NMR Spectral Data for CBZ-L-pro-L-leu-gly-gly in DMSO-d_6, Bruker WP-200, TMS standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
8.06	3	m	NH
7.31	5	s	C_6H_5
7.26		s	
5.04	2	q[c]	benzylic CH_2
4.98		q[d]	
4.24	2	m	pro α + leu α
3.72	4	dd	gly CH_2
3.36	2	m	proδ
2.11	1	m	proγ
1.81	3	m	proγ + proβ
1.50	3	m	leuCHCH$_2$
0.84	6	dd	CH_3

a) s = singlet; d = doublet; t = triplet; q = AB quartet; dd = doublet of doublets; b = broad singlet; TT = triplet of triplets

b) α = proton on asymmetric carbon; gly_n = n^{th} glycine peptide unit from N-terminus; proα, proβ, proγ, proδ refer to the structure:



c) J_{AB} = 13; δ_A = 5.10; δ_B = 4.98

d) J_{AB} = 13; δ_A = 5.00; δ_B = 4.96

Table 9. 200 MHz ^1H NMR Spectral Data for CBZ-L-pro-L-leu-gly-gly in 50% DMSO-d_6, 50% D_2O, Bruker WP-200

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
8.54	3	b	NH
7.55	5	s	C_6H_5
5.23	2	q[c] + s[d]	benzylic CH_2
4.46	2	m	leuα + proα
4.03	4	m	gly CH_2
3.69	2	m	proδ
2.42	1	m	proβ
2.05	3	m	proβ + proγ
1.79	1	m	leuCH
1.66	2	m	leuCH_2
1.00	6	m	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) J_{AB} = 12.5 Hz; δ_A = 5.26; δ_B = 5.20

d) at 5.28 δ

Table 10. FAB-MS of Peptides and Peptide Derivatives (MS-9, glycerol matrix, Xenon gas)

Peptide	Fragment[a]			
	M+1-B[b]			
glgg[c]	16.6			
plgg[c]	2.0			
gplg	25.3			
	M+1-CO_2H	M+1-G[d]	M+1-CBZ	
CBZ-glgg	13.9	49.6	90.5	
CBZ-plgg	25.5	24.5	72.0	
CBZ-gplg[e]	55.6	33.3	116.7	
	M+1-CH_4	M+1-C_3H_8	M+1-NH_2Ph	M+1-CBZ
CBZ-glgg-NHPh	6.5	6.7	38.5	73.5
CBZ-plgg-NHPh	4.0	18.2	34.6	89.6
CBZ-gplg-NHPh	5.8	17.4	37.6	87.2
	M+1-B	M+1-NH_2Ph		
glgg-NHPh	7.4	9.7		
plgg-NHPh	0.7	1.9		
gplg-NHPh	15.6	23.6		

a) M+1 = 100

b) B = $(CH_3)_2CHCH_2$; Mass = 57

c) Commercial material (Bachem, Torrance, CA)

d) G = $NH_2CH_2CO_2H$; Mass = 75

e) Commercial material (Peptide Research Foundation, Japan)

appeared on the surface of the solvent. On warming to room temperature, the solid and the cloudiness disappeared. The mixture was mechanically stirred at room temperature for an additional six hours followed by washing with 25 mL each of 5% sodium bicarbonate, water, 1 M HCl and again water. The organic layer was dried over magnesium sulfate and evaporated to give a white solid which was recrystallized from ethylacetate/heptane to give 35.4 mg (56%) of N-benzyloxycarbonyl L-prolyl-L-leucyl-glycyl-glycine anilide. The product was characterized by its proton NMR (Table 11) and FAB-MS spectra (Table 10).

L-prolyl-L-leucyl-glycyl-glycine anilide

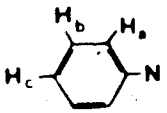
N-Benzyloxycarbonyl L-prolyl-L-leucyl-glycyl-glycine anilide (0.0919 mmol, 50.8 mg) was dissolved in 45% HBr in acetic acid (2 mL) and stirred for 30 minutes. The solution was evaporated to dryness and lyophilized several times with water. These derivatives proved the most difficult to purify but adequate material was obtained by the following procedure. After the above treatment, the solid was dissolved in 1 mL of ethanol. The derivative was precipitated with methylene chloride and the solid filtered and washed with methylene chloride. This procedure was repeated until the material obtained was white. L-prolyl-L-leucyl-glycyl-glycine anilide was obtained in 87% yield (33.4 mg). The product was

Table 11. 200 MHz ^1H NMR Spectral Data for CBZ-L-pro-L-leu-gly-gly Anilide in Acetone-d_6

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b,c]
8.82	1	b	NH
8.13	1	b	NH
7.90	1	b	NH
7.78	2	d[d]	H_a
7.33	5	s	benzylic C_6H_5
7.25	2	tt[d]	H_b
7.00	1	tt[d]	H_c
5.10	2	q[e]	benzylic CH_2
5.02	1	b	NH
4.26	2	m	leuα + proα
3.84	4	m	gly CH_2
3.45	2	m	proδ
1.69	3	m	leuCHCH_2
0.90	6	d	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) 

d) J_{ab} = 7.5 Hz; J_{ac} = 1.2 Hz

e) J_{AB} = 13 Hz; δ_A = 5.12; δ_B = 5.07

characterized by its proton NMR (Tables 12 and 13) and FAB-MS spectra (Table 10).

The peptide gly-L-pro-L-leu-gly was obtained by the same procedure.

Derivatization Procedure[5]

The peptide hydrolysate, peptide or amino acid standard mixtures (approximately 0.1 mg) were dissolved in 2 mL 5.5 M HCl and heated at 110° for 20 h followed by removal of the acid solution by using the centrifugal biodryer. The resulting mixture of amino acids was then esterified by treatment with 0.5 mL 3 M HCl in 2-propanol at 110° for 20 minutes. The 2-propanol was removed under a stream of nitrogen and the residue treated with 0.1 mL pentafluoropropionic anhydride in .1 mL of methylene chloride at 110° for five minutes. The solvent was evaporated. The residue was evaporated to dryness three times with 2 mL methylene chloride. The derivatized amino acid enantiomers were separated on a 0.2 mm I.D. glass capillary column coated with Chirasil-Val (Applied Science), an optically active stationary phase. The column was mounted in a Hewlett-Packard gas chromatograph model HP5840A. Figure 17 shows the chromatogram for a mixture of seven racemic amino acid derivatives. Typical conditions are given in Figure 17 as well.

Table 12. 200 MHz ^1H NMR Spectral Data for L-pro-L-leu-gly-gly Anilide in D_2O

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b,c]
7.47	4	s	H_a
7.44		s	H_b
7.31	1	m	H_c
4.41	2	m	proα + leuα
4.14	2	d[d]	gly CH_2
4.04	2	d[e]	gly CH_2
3.37	2	t	proδ
2.41	1	m	proβ
1.99	3	m	proβ + 2 proγ
1.72	3	m	leuCHCH$_2$
0.98	6	t	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) See footnote c) Table 11

d) J = 1.8 Hz

e) J = 3.8 Hz

Table 13. 200 MHz ^1H NMR Spectral Data for L-pro-L-leu-gly-gly Anilide in Methanol-d_4

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b,c]
7.59	2	d[d]	H_a
7.19	2	t[d]	H_b
7.09	1	t[d]	H_c
4.03	3	m[e]	gly_2CH_2+αleu
3.91	2	q[f]	gly_1CH_2
3.07	2	m	proδ
2.20	1	m	proβ
1.81	3	m	proβ + proγ
1.63	3	m	$leuCHCH_2$
0.97	6	t	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) See footnote c) Table 11

d) J_{ab} = 8.4 Hz; J_{ac} = 1.2 Hz

e) J_{AB} = 16.8 Hz; δ_A = 4.07; δ_B = 3.98

f) J_{AB} = 16.8 Hz; δ_A = 3.98; δ_B = 3.84

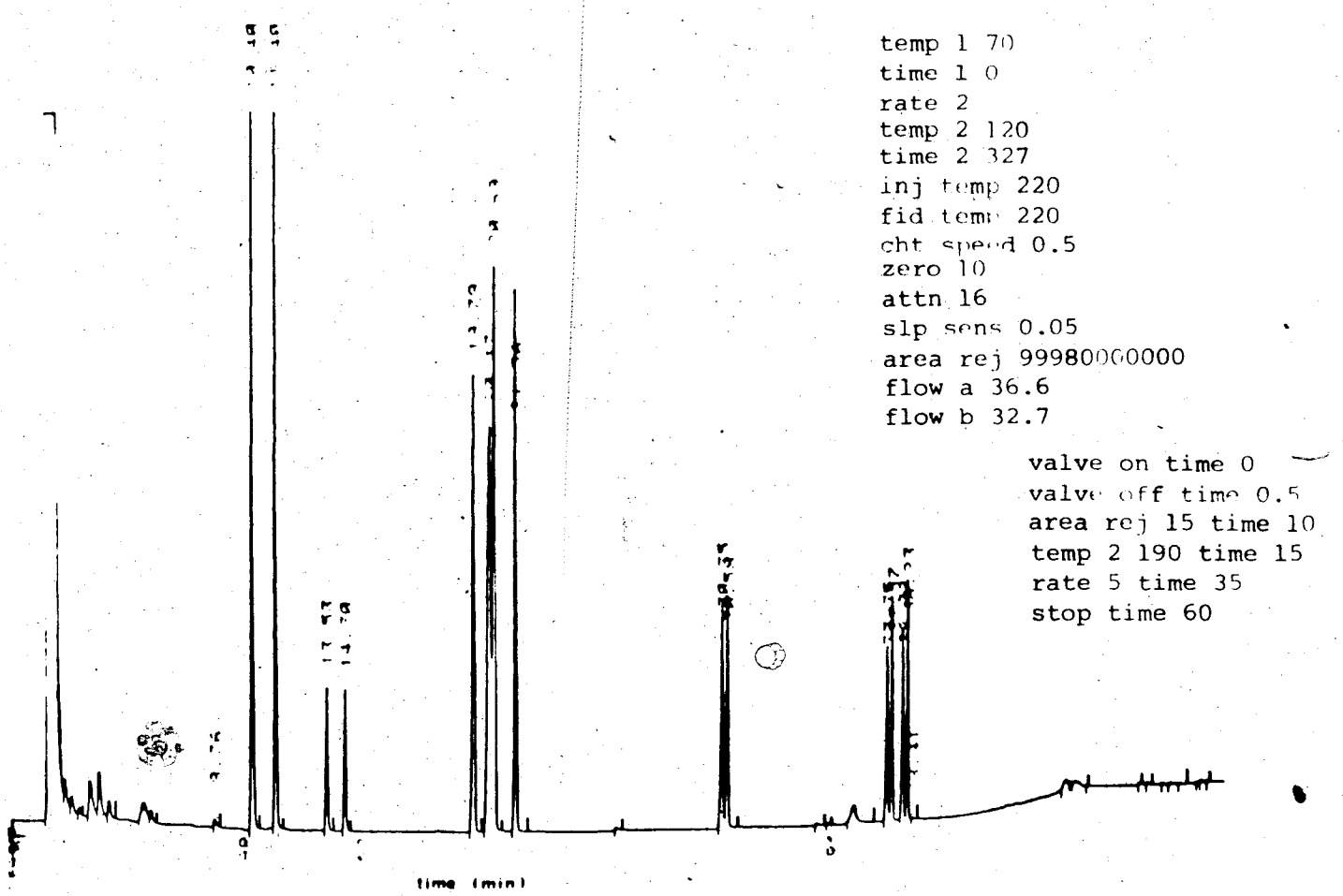


Figure 17. Mixture of racemic amino acid N-pentafluoropropionyl isopropyl esters. The major peaks are, in order of increasing retention time, CH_2Cl_2, the derivatives of D-ala, L-ala, D-val, L-val, D-leu, D-pro, L-pro, L-leu, D-asp, L-asp, D-phe, D-glu, L-phe and L-glu. The time program is shown at the upper right.

Racemization of Synthetic Peptides

The optical purity of the commercial peptides and the synthesized derivatives was determined using the procedure outlined above. The results are tabulated in Table 14.

The peptide derivatives were characterized by their mass (Table 10) and proton NMR spectra (see Table 13).

Mass Spectra

Owing to the low volatility of the peptides and derivatives, the mass spectra of these compounds were obtained by Fast Atom Bombardment Mass Spectroscopy (FAB-MS) in a glycerol matrix (Table 10). This method yielded strong M+1 peaks. The small number of additional peaks can be rationalized by a reasonable fragmentation pattern.

The similarity in fragment pattern among the same derivative of the three different peptides argues favourably for the purity of the compounds.

Purification of the products arising from removal of the carbobenzoxy protecting group was difficult. The mass spectrum of these products (anilides and gplg) shows no evidence of the appropriate precursors.

Table 14. Optical Purity of Peptide Derivatives

	D/L		Optical Purity (%)		
Compound	(pro)	(leu)	(pro)	(leu)	NMR Table
plgg	.0091	.0107	98.2	97.9	15
CBZ-plgg	.0106	.0203	97.9	96.0	8, 9
CPZ-plgg-NHPh	.0080	.0167	98.4	96.7	11
plgg-NHPh	.0099	.0175	98.0	96.6	12, 13
gplg	.0083	.0126	98.4	97.5	16
CBZ-gplg	.0094	.0130	98.1	97.4	17
CBZ-gplg-NHPh	.0116	.0188	97.7	96.3	18
gplg-NHPh	.0112	.0180	97.8	96.5	19
glgg	-	.0166	-	96.7	20, 21
CBZ-glgg	-	.0158	-	96.9	22
CBZ-glgg-NHPh	-	.0215	-	95.8	23
glgg-NHPh	-	.0196	-	96.2	24

CBZ = $C_6H_5CH_2OCO$; Ph = C_6H_5; p = prolyl; l = leucyl; g = glycyl

Proton NMR Spectra

The purity and the identity of the peptide derivatives was confirmed by 200 MHz proton NMR spectra (Tables 8,9,11-24). The spectral assignments are based on selective decoupling experiments performed on L-pro-L-leu-gly-gly (Figure 18) and L-pro-L-leu (Figure 19) as well as the pH behaviour in aqueous solution of these two compounds. This latter experiment is described below in the discussion concerning the identification of the components present in the hydrolysis mixture.

Finding a suitable solvent for NMR of the peptide derivatives was difficult as amides are generally sparingly soluble in most organic solvents. D_2O was found to be a good choice for the underivatized peptides. The anilides and carbobenzoxy derivatives were only slightly soluble in water at room temperature whereas the derivatives with both termini blocked were essentially insoluble. A variety of solvents were tried ($CDCl_3$, DMSO-d_6, methanol-d_4, acetone-d_6) with varying degrees of success. To overcome the solubility problem, the number of scans was increased (1024 was not unusual). Consequently, any impurity in the NMR solvent appeared in the spectrum as a significant absorption. For example, the signal from the protons in the isotopic impurity of acetone-d_6 was by far the largest signal in samples using

Table 15. 200 MHz ^1H NMR Spectral Data for L-pro-L-leu-gly-gly in D_2O, DOH standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
4.43	1	m	proα
4.39	1	m	leuα
3.96	2	q[c]	gly_1CH_2
3.77	2	q[d]	gly_2CH_2
3.40	2	m	proδ
2.46	1	m	proβ
2.04	3	m	proγ + proβ
1.64	3	m	$leuCHCH_2$
0.93	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) J_{AB} = 17.0 Hz; δ_A = 4.01; δ_B = 3.91

d) J_{AB} = 17.0 Hz; δ_A = 3.80; δ_B = 3.74

Table 16. 200 MHz ^1H NMR Spectral Data for Glycyl-L-prolyl-L-leucyl-glycine in D_2O, DOH standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
4.46	2	m	leuα proα
4.03	2	q[c]	gly_1CH_2
3.78	2	q[d]	gly_2CH_2
3.62	2	m	proδ
2.50	1	m	proβ
2.04	3	m	2proγ + proβ
1.67	3	m	$leuCHCH_2$
0.92	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) J_{AB} = 16.5 Hz; δ_A = 4.06; δ_B = 4.00

d) J_{AB} = 17.3 Hz; δ_A = 3.84; δ_B = 3.72

Table 17. 200 MHz ^1H NMR Spectral Data for CBZ-gly-L-pro-L-leu-gly in DMSO-d_6, TMS standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
8.28	1	b	NH
7.96	2	b	NH
7.92		b	NH
7.32	6	s	NH + C_6H_5
5.02	2	s	benzylic CH_2
4.36	2	m	proα + leuα
3.83	2	m	gly CH_2
3.71	2	m	gly CH_2
3.44	2[c]	m	proδ
2.18	1	m	proβ
1.88	3	m	proγ + proβ
1.51	3	m	leuCHCH_2
0.86	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) Partially hidden by HOD

Table 18. 200 MHz ^1H NMR Spectral Data for CBZ-gly-L-pro-L-leu-gly Anilide in Acetone-d_6, TMS standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b,c]
8.87	1	b	NH
7.86	1	b	NH
7.79	2	dd	H_a
7.74	1	b	NH
7.32	5	s	benzylic C_6H_5
7.25	2	tt	H_b
7.01	1	tt	H_c
6.55	1	b	NH
5.04	2	q[d]	benzylic CH_2
4.39	1	m	leuα
4.22	1	m	proα
3.90	4	t	gly CH_2
3.68	2	m	proδ
1.68	3	m	leuCHCH$_2$
0.88	6	t	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) See footnote c) Table 11

d) J_{AB} = 12.5 Hz; δ_A = 5.06; δ_B = 5.02

Table 19. 200 MHz ^1H NMR Spectral Data for Gly-L-pro-L-leu-gly Amslide in D_2O, DOH standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b,c]
7.44	2	d	H_a
7.42	2	s	H_b
7.30	1	m	H_c
4.50	1	m	leuα
4.41	1	m	proα
4.09	2	d	gly_2CH_2
3.97	2	s	gly_1CH_2
3.52	2	m	proδ
2.12	1	m	proβ
1.96	3	m	proβ + proγ
1.72	3	m	$leuCHCH_2$
0.99	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) See footnote c) Table 11

Table 20. 200 MHz ^1H NMR Spectral Data for Gly-L-leu-gly-gly in DMSO-d_6, TMS standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
8.79	1	d	NH
8.62	1	dd	NH
7.20	1	dd	NH
4.23	1	m	leuα
3.59	2	d	gly_3CH_2
3.50	2	s	gly_2CH_2
3.39	2	m	gly_2CH_2
1.49	3	m	$leuCHCH_2$
0.85	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

Table 21. 200 MHz ^1H NMR Spectral Data for Glycyl-L-leucyl-glycyl-glycine in D_2O, DOH standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
4.40	1	t	leuα
3.98	2	q[c]	gly_2CH_2
3.86	2	s	gly_1CH_2
3.80	2	s	gly_3CH_2
1.68	3	m	$leuCHCH_2$
0.95	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) J_{AB} = 17.5 Hz; δ_A = 4.01; δ_B = 3.95

Table 22. 200 MHz ^1H NMR Spectral Data for CBZ-gly-L-leu-gly-gly in D_2O, DOH standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
7.40	5	s	C_6H_5
5.12	2	s	benzylic CH_2
4.35	1	m	leuα
3.93	2	d[c]	gly_1CH_2
3.87	4	d[d]	$gly_{2,3}CH_2$
1.62	3	m	$leuCHCH_2$
0.90	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) J = 2.6 Hz

d) J = 2.4 Hz

Table 23. 200 MHz ^1H NMR Spectral Data for CBZ-gly-L-leu-gly-gly Anilide in Acetone-d_6, TMS standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b,c]
8.84	1	b	NH
8.10	1	b	NH
7.76	3	d,b[d]	H_a + NH
7.30	6	m	benzylic C_6H_5 + NH
7.24	2	tt[d]	H_b
7.00	1	tt[d]	H_c
6.64	1	b	NH
5.02	2	s	benzylic CH_2
4.32	1	m	leuα
3.89	2	m	gly_1CH_2
3.79	4	m	$gly_{2,3}CH_2$
1.64	3	m	$leuCHCH_2$
0.87	6	t	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) See footnote c) Table 11

d) J_{ab} = 7.5 Hz; J_{ac} = 1.5 Hz

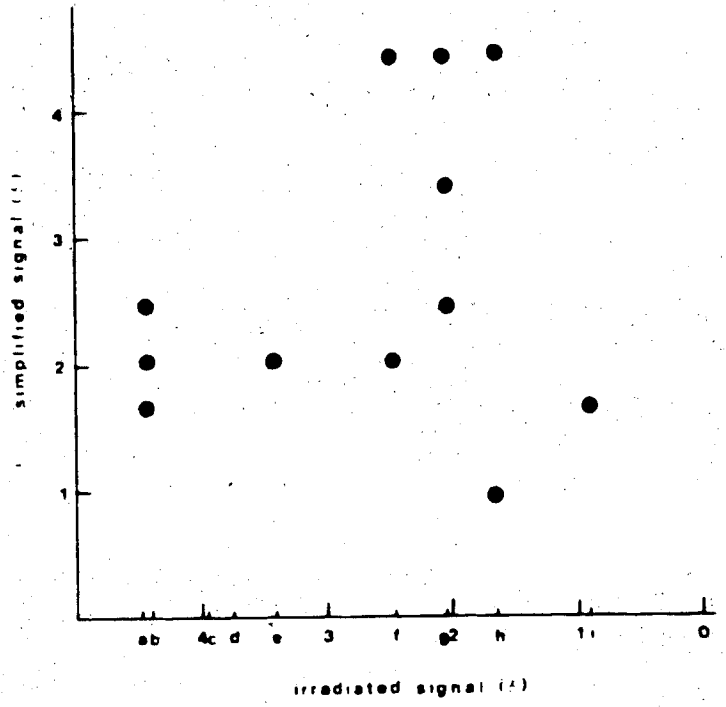
Table 24. 200 MHz ^1H NMR Spectral Data for Gly-L-leu-gly-gly Anilide in D_2O, DOH standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
7.46	2	s	H_a
7.44	2	s	H_b
7.32	1	m	H_c
4.42	1	t	leuα
1.66	3	m	leuCHCH$_2$
0.92	b	t	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) See footnote c) Table 11



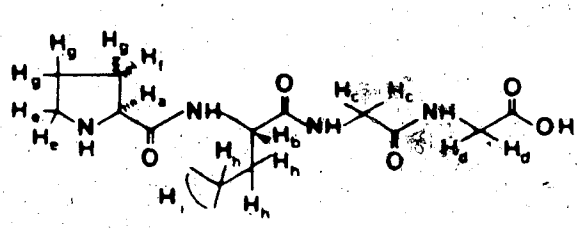
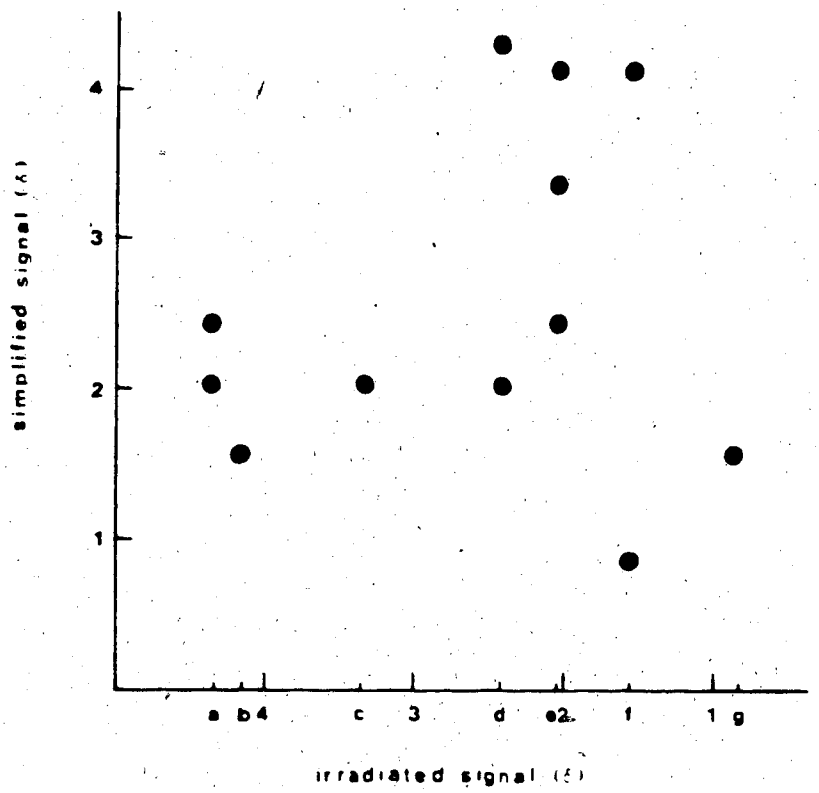


Figure 18. Assignment of 200 MHz ^1H NMR spectrum of L-pro-L-leu-gly-gly by selective decoupling in D_2O, pH 6.40, DOH standard



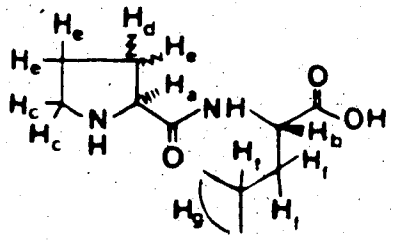


Figure 19. Assignment of 200 MHz ^1H NMR spectrum of L-pro-L-leu by selective decoupling in D_2O, pH 7.33, DOH standard

this solvent. Smaller, but still significant signals are obtained from HOD and HOH and from another unidentified peak in the solvent.

These problems were minimized by obtaining the spectrum in more than one solvent. A change in solvent did affect the chemical shift of some of the protons. The protons on the alpha carbons appeared to be most sensitive to solvent, an effect which can be rationalized by the effect of solvation of the adjacent amide functional groups.

Synthesis of Diketopiperazines

C-(L-prolyl-L-leucyl) and c(-glycyl-L-leucyl) were synthesized by heating the corresponding dipeptide in phenol (purified by sublimation) under nitrogen at 145° for 1 h. The phenol was removed by sublimation. The resulting white solid was recrystallized from ethanol-water. C-(L-prolyl-L-leucyl) yielded hexagonal plates upon crystallization while c-(glycyl-L-leucyl) produced a white powder. The compounds were identified by their characteristic proton NMR (Tables 25 and 26) and chemical ionization mass spectra (Tables 27 and 28). The melting points agreed well with the values reported in the literature (Table 29).[84,85]

Table 25. 400 MHz ^1H NMR Spectral Data for c-(L-pro-L-leu) in $CDCl_3$, Bruker WP-400, TMS standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
7.02	1	s	NH
4.13	1	t	proα
4.02	1	dd	leuα
3.58	2	m	proδ
2.34	1	m	proβ
2.08	3	m	proγ + proβ
1.88	2	m	leuCH_2CH
1.54	1	m	leuCH_2
0.97	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

Table 26. 400 MHz ^1H NMR Spectral Data for c-(gly-L-leu) in $CDCl_3$, TMS standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
6.19	1	b	NH
6.08	1	b	NH
4.02	2	s	gly CH_2
3.98	1	m	leuα
1.72	3	m	leu$CHCH_2$
0.98	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

Table 27. Chemical Ionization (NH_4Cl) Mass Spectral Data for c-(pro-leu)

Compound	Mass[a] 211 (%)	228 (%)	229 (%)
c-(L-prolyl-L-leucyl)[b]	100	25.7	3.4
c-(D-prolyl-L-leucyl)[c]	100	41.0	5.3
c-(L-pro-L-leu)[c]	100	55.1	5.2
c-(pro-leu)[d]	100	85.4	11.5
c-(pro-leu)[e]	100	14.9	2.8
L-pro-L-leu[f]	2.8	0.7	100
D-pro-L-leu[g]	32	0	100

a) c-(pro-leu): M+1 = 211; M+18 = 228

pro-leu: M+1 = 229

b) Synthesized from L-pro-L-leu

c) Isolated from the decomposition mixture of plgg; 148.5°, pH 6.8, 10 min by HPLC separation

d) Isolated for the decomposition mixture of plgg; 148.5°, pH 6.8, 300 min by extraction with $CHCl_3$

e) Produced by heating of L-pro-L-leu in methanol at 110° for 22 h followed by HPLC separation

f) Commercial material (Serva)

g) Isolated by HPLC from decomposition of L-pro-L-leu; 148.5°, pH 6.8, 60 min

Table 28. Chemical Ionization (NH_4Cl) Mass Spectral Data for c-(gly-leu)

	Mass[a]			
Compound	171	188	189	206
c-(gly-L-leu)[b]	2.4	100	10.0	0.6
c-(gly-leu)[c]	7.7	100	11.1	0
L-leu-gly[d]	0.3	0	100	3.0
gly-L-leu[d]	0	0	100	25.8

a) c-([redacted]L-leu): M+1, 171; M+18, 188

gly-leu: M+1, 189; M+18, 206

b) Synthesized from gly-L-leu

c) Isolated by HPLC from the decomposition mixture of glgg; 148.5°, pH 6.8, 90 min

d) Commercial material (Serva)

Table 29. Melting Points of Synthetic Diketopiperazines

Compound	m.p. (°C)[a] obs	lit
c-(L-prolyl-L-leucyl)	157-158	157-158[b]
c-(glycyl-L-leucyl)	251-252 (d)	250-251[c]

a) Uncorrected

b) Reference 84

c) Reference 85

Kinetic Procedure

Kinetic runs were carried out in phosphate buffer (pH 6.8, 25°) with a total phosphate concentration of 0.050 M. Two millilitres of ~10 mM peptide solution were placed in a glass tube and sealed after five freeze-pump-thaw cycles. The tubes were placed in a thermostated oil bath at a temperature of 148.50 ± 0.05°. The bath was found to be stable to this degree of precision for a period (24 h) exceeding the maximum time used for the kinetic runs. After the glass tubes were removed from the oil bath, the hydrolysis mixture of partially racemized material was separated by TLC or HPLC.

The hydrolysis mixture was separated on silica gel TLC plates (Whatman LK6DF, 250 μ) developed with 4:1:1 n-butanol:acetic acid:water. The components in the mixture were visualized by spraying with ninhydrin reagent (90 mL, 0.3 wt% aqueous solution of ninhydrin and 10 mL of a solution of 3 g cadmium acetate in 300 mL acetic acid and 150 mL water) or with a UV lamp. The components in the mixture were isolated by scraping off the appropriate section of silica gel followed by extraction with 5.5 M HCl. After filtering the acid extracts (2 μm Metricel) the samples were derivatized and the extent of racemization measured as described above).

Separation of the hydrolysis mixture by HPLC was

achieved by using a C-18 reversed phase radial compression column (Waters Associates) connected to a Perkin-Elmer Series 2 liquid chromatograph solvent delivery system. The components of the hydrolysis mixture were detected by UV absorption at 230 nm. The analogue output of the UV detector (Perkin-Elmer LC-55B) was interfaced to a Varian Vista terminal for data collection or to a strip-chart recorder. The separation was performed isocratically using 0.1 M ammonium acetate (pH 7) containing 10% acetonitrile. Ammonium acetate was purified by sublimation. The water used for the separation was deionized and distilled from potassium permanganate. HPLC grade acetonitrile (Caledon) was used untreated. The solvent was degassed by bubbling helium through the mixture for two minutes.

Typically, one millilitre of the reaction mixture was separated by repetitive injections of 100 µL. The hydrolysis fragments were collected manually as they eluted from the column after passing through the UV detector. Six fractions were collected for the hydrolysis and racemization studies of pro-leu-gly-gly and pro-leu:

1. LL* + DD peptide
2. DL = LD peptide
3. LL + DD + DL + LD peptide
4. DL + LD DKP
5. LL + DD DKP
6. DL + LD + LL + DD DKP

In the case of gly-L-leu-gly-gly, three fractions were collected:

1. initial fraction containing dipeptides
2. substrate enantiomers
3. c-(gly-L-leu) + c-(gly-D-leu)

For all three peptides, a portion of the total hydrolysis mixture was saved to obtain a measurement of racemization of the total hydrolysate. The HPLC solvent was removed from the fractions in a centrifugal biodryer. The identity of the hydrolysis components was determined by FAB-MS and by NMR spectroscopy (vide infra). The extent of racemization of the separated components was subsequently determined by the GC method outlined above.

---

*e.g. DL refers to the sequence D-pro-L-leu in the tetrapeptide or dipeptide. DKP = diketopiperazine.

Sample chromatograms and conditions are shown for pro-leu-gly-gly (Figure 20), pro-leu (Figure 21) and gly-leu-gly-gly (Figure 22).

Identification of Hydrolysis Products

The dried eluate from chromatography was analyzed by FAB-MS in the case of the peptides or by chemical ionization mass spectroscopy (CI) in the case of the diketopiperazines (DKP). The absolute configuration of the alpha carbons of the hydrolysis products was determined by the GC method.

Besides the substrate peptide, the hydrolysis of pro-leu-gly-gly yielded three major components. The component having the shortest retention time after the substrate was found to have an M+1 peak at mass 343. GC analysis showed that it consisted of D-proline, L-leucine and glycine. Although it is tempting to assign the structure D-pro-L-leu-gly-gly to this component, there is a potential complication. The formation of a cyclic tetrapeptide[71] from the substrate followed by hydrolysis could in principle yield four isomeric tetrapeptides: plgg, lggp, ggpl or gplg.* The sequence of the peptide units was

---

*It is assumed that the peptide pglg is not possible. The formation of this peptide would require the breaking of (cont'd)

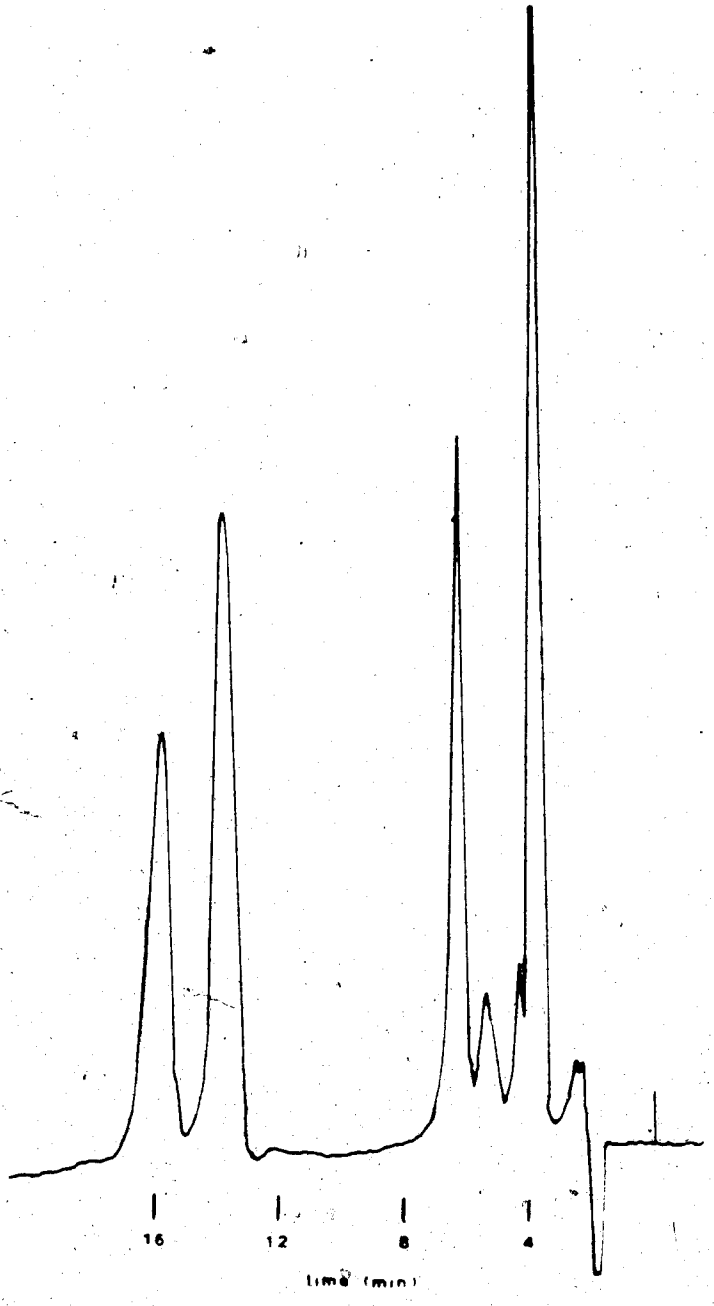


Figure 2. HPLC trace of the mixture produced after heating L-pro-L-leu-gly-gly for 30 min at 148.5°, pH 6.8. Chromatographic conditions: reversed phase C-18 column (Waters), 0.1 M NH_4OCOCH_3, 10% CH_3CN, 1.2 mL/min. The major peaks in order of increasing retention time are L-pro-L-leu-gly-gly, D-pro-L-leu-gly-gly, c-(D-pro-L-leu) and c-(L-pro-L-leu).

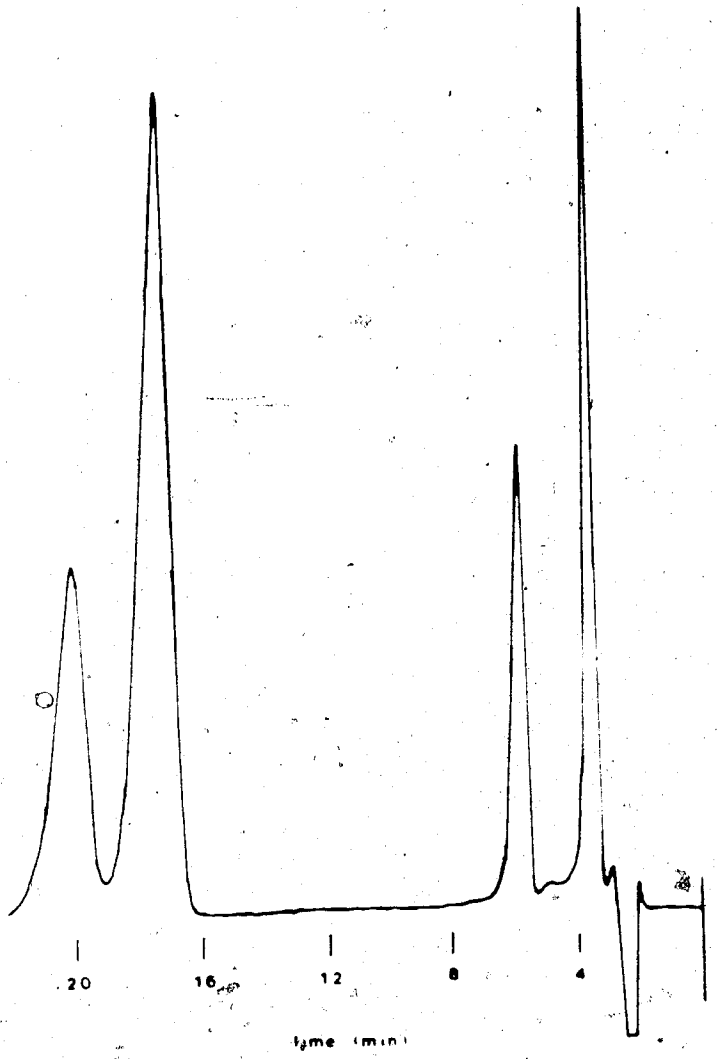


Figure 21. HPLC trace of the mixture produced after heating L-pro-L-leu for 60 min at 148.5°, pH 6.8. Chromatographic conditions: reversed phase C-18 column (Waters), 0.1 M NH_4OCOCH_3, 10% CH_3CN, 1.0 mL/min. The peaks in order of increasing retention time are L-pro-L-leu, D-pro-L-leu, c-(D-pro-L-leu) and c(L-pro-L-leu).

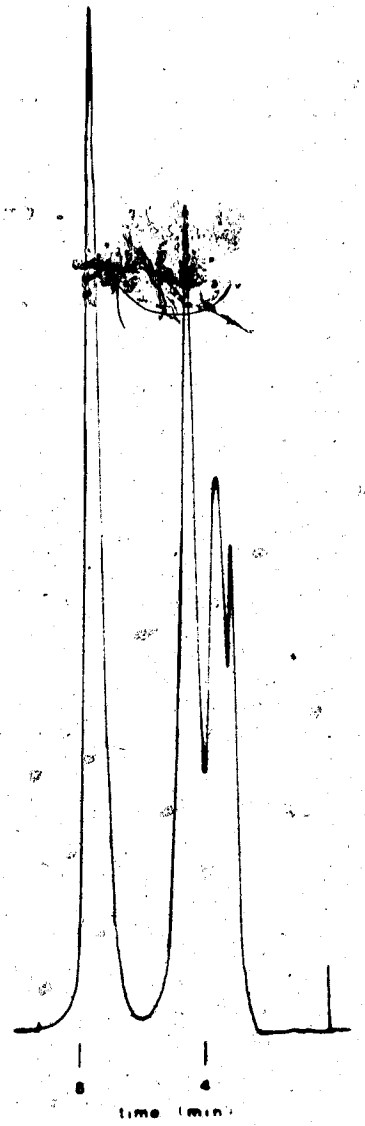


Figure 22. HPLC trace of the mixture produced after heating gly-L-leu-gly-gly for 300 min at 148.5°, pH 6.8. Chromatographic conditions: reversed phase C-18 column (Waters), 0.1 M NH_4OCOCH_3, 10% CH_3CN, 0.5 mL/min. The two largest peaks in order of increasing retention time are gly-leu-gly-gly and c-(gly-leu).

determined by exploiting the acid-base chemistry of peptides. At neutral pH, the peptide will exist predominantly in the zwitterionic form. When the medium is made strongly acidic, the carboxyl terminus will be protonated thus removing the negative charge. In strong base, the N-terminus, protonated and positively charged at neutral pH, is deprotonated and becomes uncharged. The removal of charge is predicted to have an effect on the chemical shift of those protons near to the centre of charge.[86-89] The predictions are summarized in Table 30.

The pH behaviour of the unknown peptide from the heating of L-pro-L-leu-gly-gly was compared to the behaviour of an authentic sample of L-pro-L-leu-gly-gly* (Table 31). A comparison of the change of chemical shift as a function of pH for the two compounds shows that the behaviour is identical (Table 32). With this evidence the structure D-pro-L-leu-gly-gly is assigned to this component from the heating experiment.

The CI mass spectra of the other two components in the hydrolysis mixture each showed an M+1 peak at mass 211, corresponding to c-(prolyl-leucyl) (Table 27). The

---

the pro-leu and the gly-gly amide bonds followed by the condensation of proline, glycine and leu-gly. The probability of this occurrence is likely prohibitive.

*Spectral assignments of ^1H spectra are based on selective decoupling experiments on L-pro-L-leu-gly-gly (Figure 21).

Table 30. Predicted Change in Chemical Shift of Protons Near Peptide Termini as pH is Changed from Neutrality

Terminus	Medium	
	Acid	Base
N	0	-
C	+	0

+ = downfield shift

- = upfield shift

0 = no change

Table 31. 200 MHz ^1H NMR Spectral Data[a] for Diastereomeric Tetrapeptides as a Function of pH

L-pro-L-leu-gly-gly

pH	proα[b]	leuα[b]	gly$_1$CH$_2$[c]	gly$_2$CH$_2$	proδ	proβ	proγ[b] proβ	leu[b] CHCH$_2$	CH$_3$[d]
1.00	4.36	4.39	3.97[e]	4.02[f]	3.40	2.45	2.03	1.64	0.89
6.40	4.43	4.39	3.96[g]	3.77[h]	3.40	2.46	2.04	1.64	0.93
9.18	3.97	4.37	3.97[k]	3.78[l]	3.06	2.24	1.84	1.66	0.92

D-pro-L-leu-gly-gly

1.05	4.15	4.40	4.00[m]	4.04[f]	3.45	p	p	1.66	0.92
5.75	4.40	4.35	3.90[n]	3.72[f]	3.36	p	p	1.59	0.85
8.80	4.22	4.40	3.97[n]	3.82[f]	3.26	p	p	p	0.92
[illegible]	3.96	4.37	3.96[o]	3.77[f]	2.84	p	p	1.64	0.90

a) [illegible]ndard, corrected for pH dependent HOD signal
b) Multiplet
c) AB quartet
d) Doublet of doublets
e) J_{AB} = 17.4 Hz; δ_A = 3.99; δ_B = 3.95
f) Singlet
g) J_{AB} = 17.0 Hz; δ_A = 4.01; δ_B = 3.91
h) J_{AB} = 17.0 Hz; δ_A = 3.80; δ_B = 3.74
k) J_{AB} = 17.0 Hz; δ_A = 4.00; δ_B = 3.94
l) J_{AB} = 17.3 Hz; δ_A = 3.80, δ_B = 3.76
m) Doublet, J = 1.25 Hz
n) Doublet, J = 2.3 Hz
o) Doublet, J = 2.0 Hz
p) Region obscured by CH_3CO_2H

Table 32. Change in Chemical Shift for Diastereomeric Tetrapeptides as a Function of pH

L-pro-L-leu-gly-gly [a]

pH	proα	leuα	gly_1CH_2	gly_2CH_2	proδ	proβ	proγ proβ	leu $CHCH_2$	CH_3
1.00	-0.06	0	0.01	0.25	0	-0.01	-0.01	0	-0.04
9.18	-0.48	-0.08	-0.05	-0.05	-0.40	-0.28	-0.26	-0.04	-0.07

D-pro-L-leu-gly-gly [b]

1.05	0.05	0.05	0.18	0.32	0.09	-	-	0.07	0.07
8.80	-0.18	0.05	0.07	0.10	-0.10	-	-	-	-
10.20	-0.44	0.02	0.06	0.05	-0.52	-	-	0.05	0.05

a) relative to pH 6.40

b) relative to pH 5.75

component with the shorter retention time was shown by GC analysis to consist of D-proline and L-leucine while the other consisted of L-proline and L-leucine. The proton NMR and mass spectra of authentic c-(L-prolyl-L-leucyl) and the latter component were identical (Table 25 and Table 33). The spectra of the former component was consistent with the structure c-(D-prolyl-L-leucyl) (Table 34).

There were also three major components from the heating of L-pro-L-leu, two of which correspond to the DKPs: c-(D-pro-L-leu) and c-(L-pro-L-leu) obtained from the heating of the tetrapeptide. The mass spectrum (CI) of the other component eluting after the substrate had an M+1 peak at mass 229 (Table 27). GC analysis of this compound showed that it consisted of D-proline and L-leucine. Again, the sequence of the peptide units is ambiguous and was determined from the pH dependence of chemical shift in the proton NMR (Tables 35 and 36). The assignment of the resonances is based on selective decoupling experiments on the authentic dipeptide L-pro-L-leu (Figure 22). The pH behaviour of the dipeptide and the unknown component from the heating experiment was the same, demonstrating that the structure was D-pro-L-leu.

The component from the heating experiment with gly-L-leu-gly-gly with largest retention time gave a mass

Table 33. 400 MHz ^1H NMR Spectral Data ($CDCl_3$) for c-(L-pro-L-leu) from Decomposition of L-pro-L-leu-gly-gly, TMS standard

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b,c]
5.74	1	b	NH
4.07	1	t	proα
3.97	1	m	leuα
3.52	2	m	proδ
2.28	1	m	proβ
0.92	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

c) Region from δ 2.2 to 1.2 contains signals due to components in HPLC solvent

Table 34. 400 MHz ^1H NMR Spectral Data[a] ($CDCl_3$) for c-(D-pro-L-leu) from Decomposition of L-pro-L-leu-gly-gly, TMS standard

Chemical Shift[b] (δ)	Intensity (number of protons)	Multiplicity	Assignment[c]
6.20	1	b	NH
4.03	1	dd	proα
3.89	1	m	leuα
3.58	2	m	proδ
2.45	1	m	proβ
0.91	6	dd	CH_3

a) see footnote a) Table 8

b) See footnote b) Table 8

c) Region from δ 2.2 to 1.2 contains signals due to components in HPLC solvent

Table 35. 200 MHz ^{1}H NMR Spectral Data[a] for Diastereomeric Dipeptides as a Function of pH

L-pro-L-leu

pH	proα[b]	leuα[b]	proδ[b]	proβ[b]	proγ proβ[b]	leuCHCH$_2$[b]	CH$_3$[c]
1.05	4.42	4.42	3.41	2.45	2.07	1.68	0.90
7.33	4.34	4.15	3.37	2.42	2.05	1.58	0.85
12.20	3.65	4.14	2.85	2.07	1.68	1.53	0.83

D-pro-L-leu

1.15	4.48	4.48	3.45	2.50	2.08	1.72	0.94
7.13	4.43	4.23	3.44	2.39	2.08	1.59	0.92
13.05	3.58	4.10	2.78	2.02	1.62	1.49	0.77

a) HOD standard, corrected for pH dependent HOD signal

b) Multiplet

c) Doublet of doublets

Table 36. Change in Chemical Shift for Diastereomeric Dipeptides as a Function of pH

L-pro-L-leu [a]

pH	proα	leuα	proδ	proβ	proγ proβ	leuCHCH$_2$	CH$_3$
1.05	0.08	0.27	0.04	0.03	0.02	0.10	0.05
12.20	-0.69	-0.01	-0.52	-0.35	-0.37	-0.05	-0.02

D-pro-L-leu [b]

1.05	0.05	0.25	0.01	0.11	0	0.13	0.02
13.05	-0.84	-0.13	-0.66	-0.37	-0.46	-0.10	-0.15

a) relative to pH 7.33

b) relative to pH 7.13

spectrum (CI) with an M+1 peak at mass 170 (Table 28). The proton NMR and mass spectra of this compound were identical to the spectra obtained for authentic c-(glycyl-L-leucyl) (Table 26 and Table 37). The proton NMR spectrum of the least retained component indicated that it was a mixture. The composition was tentatively assigned to a mixture of glycine and leucine containing dipeptides.

Calibration of UV Detector

Since authentic D-pro-L-leu-gly-gly, D-pro-L-leu and c-(D-pro-L-leu), were not available, the extinction coefficients of these compounds could not be measured directly. To accomplish this, the ratio of diastereomers recorded from the output of an UV detector was compared with the ratio obtained from the 400 MHz proton NMR spectrum of the mixture collected from the chromatograph.

Figure 23 shows part of the 400 MHz proton NMR spectrum of the mixture of L-pro-L-leu-gly-gly and D-pro-L-leu-gly-gly isolated during the decomposition of L-pro-L-leu-gly-gly. The protons on the glycine peptide units of the diastereomers were distinguished by spiking the sample with pure L-pro-L-leu-gly-gly (Figure 24). A similar experiment was performed on the mixture of L-pro-L-leu and D-pro-L-leu (Figures 25 and 26). Even at 400 MHz, the diastereomers could not be completely

Table 37. 400 MHz ^1H NMR Spectral Data ($CDCl_3$) for c-(gly-L-leu) from the Decomposition of gly-L-leu-gly-gly

Chemical Shift (δ)	Intensity (number of protons)	Multiplicity[a]	Assignment[b]
5.97	1	b	NH
5.82	1	b	NH
4.01	2	s	$glyCH_2$
3.96	1	m	leuα
1.68	3	m	$CHCH_2$
0.96	6	dd	CH_3

a) See footnote a) Table 8

b) See footnote b) Table 8

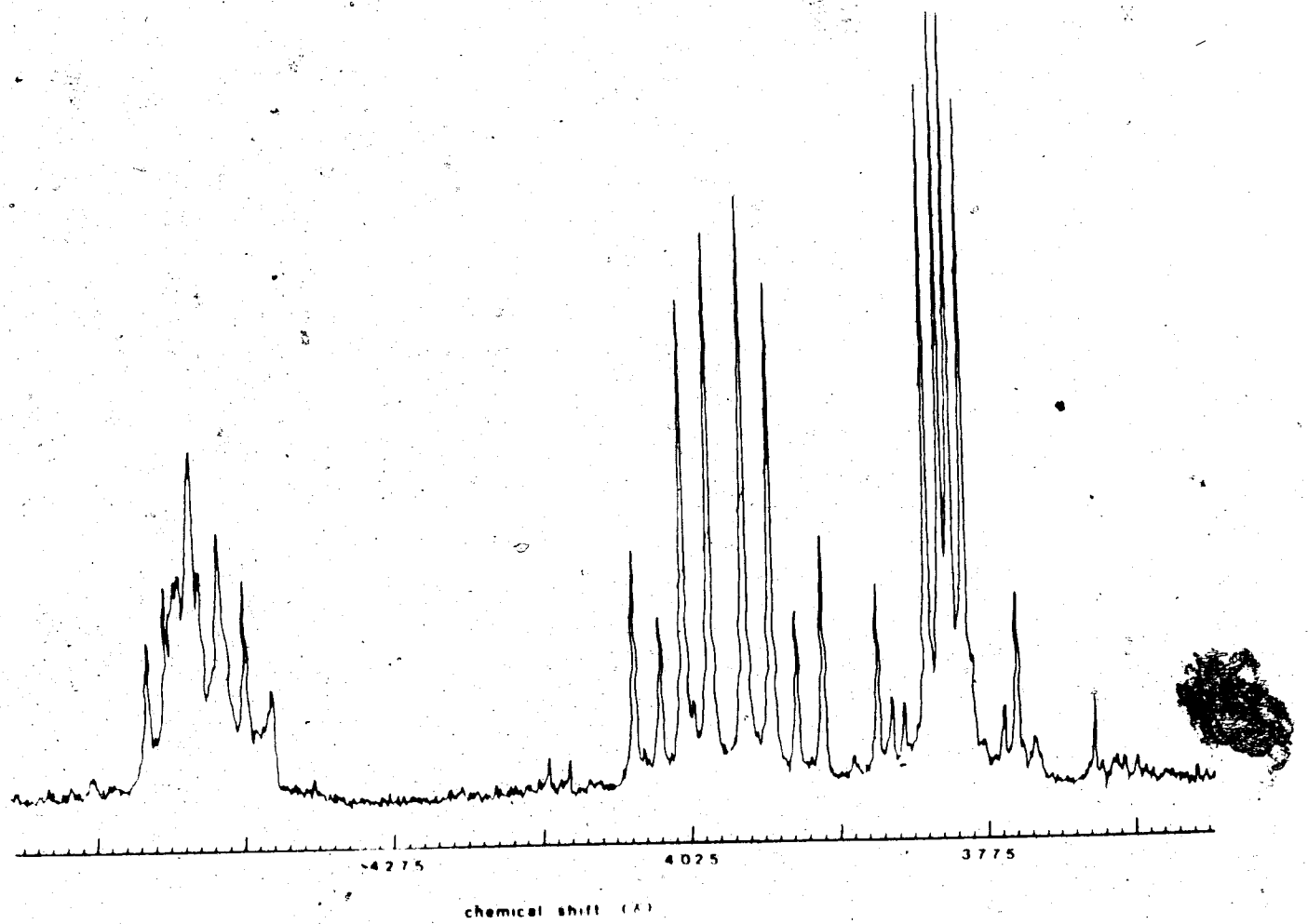


Figure 23. 400 MHz ^{1}H NMR of the mixture of L-pro-L-leu-gly-gly and D-pro-L-leu-gly-gly isolated during the hydrolysis of L-pro-L-leu-gly-gly, 148.5°, 30 min, pH 6.8, 8.45 mM. The expansion shows the α proton region. The signals arising from the α protons of the glycine peptide units are at higher field and appear as two pairs of overlapping AB quartets.

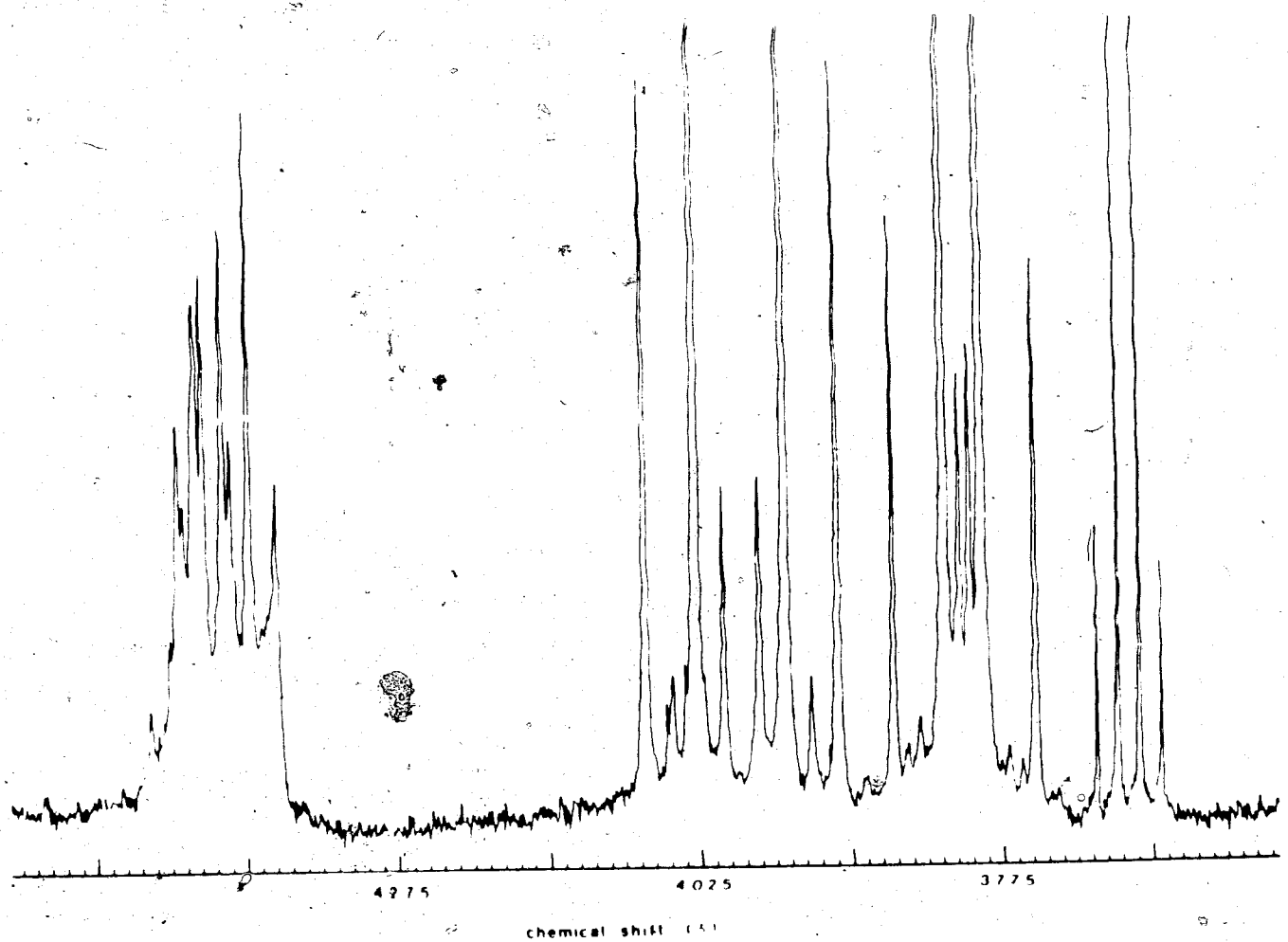


Figure 24. The same sample as in Figure 23 was spiked with authentic L-pro-L-leu-gly-gly. The protons of the glycine peptide units of the diastereomers can be distinguished. The ratio of diastereomers in the sample was calculated from the integral in Figure 23. The quartet at high field is due to a small amount of ether present in the sample of L-pro-L-leu-gly-gly.

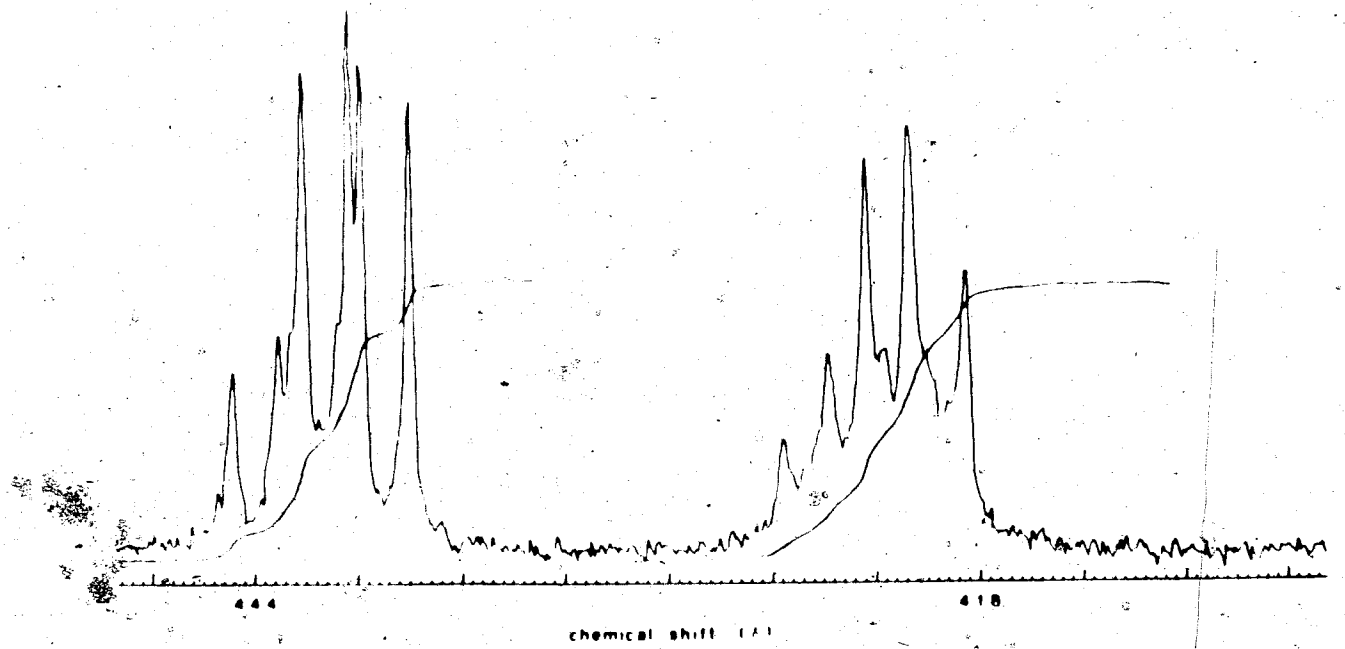


Figure 25. 400 MHz ^1H NMR spectrum of the mixture of D-pro-L-leu and L-pro-L-leu isolated during the decomposition of L-pro-L-leu after 30 minutes at 148.5°, pH 6.8. The signal at low field is due to the α protons. The other signal is due to the δ protons of the proline peptide unit. The solvent was $CDCl_3$.

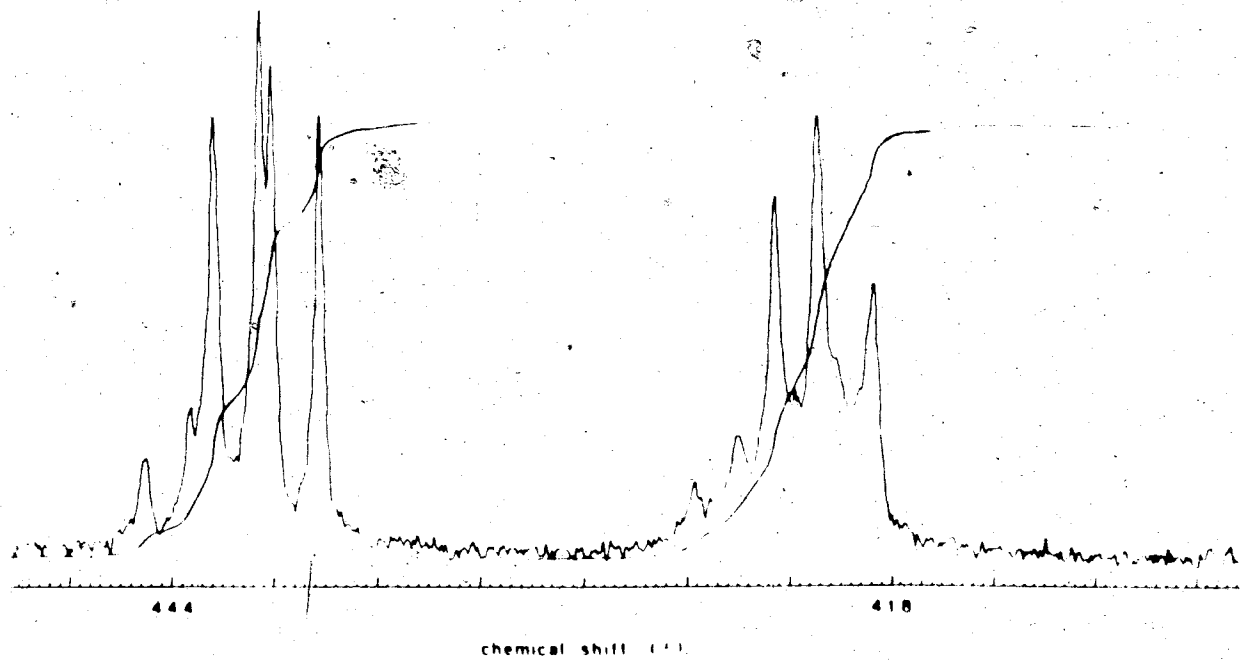


Figure 26. The same sample as in Figure 25 has been spiked with authentic L-pro-L-leu. This showed that the two signals on the high field side of each multiplet are due to the α protons of D-pro-L-leu.

separated. By using the signals from the alpha protons an approximate measurement of the ratio of diastereomers was obtained.

The diastereomers of the diketopiperazine c-(pro-leu) were distinguishable by their alpha protons (Figure 27), by their methyl protons (Figure 28) or by the imino proton. The amount of each diastereomer was measured by the integrated peak area in the liquid chromatogram. The relative extinction coefficients were measured by taking the NMR spectrum of the mixture (Table 38).

The relative extinction coefficients for the remainder of the compounds obtained from the heating experiments were measured by preparing samples of authentic compounds with a known composition and injecting these into the liquid chromatograph. The results of this calibration are given in Table 39.

Time Studies

The hydrolysis and racemization kinetics of L-prolyl-L-leucyl-glycyl-glycine, (plgg), glycyl-L-leucyl-glycyl-glycine, (glgg), and L-prolyl-L-leucine, (pl), were measured at 148.50 ± 0.05°. Eight to ten samples were analyzed for the disappearance of substrate peptide, the appearance of new species and the extent of racemization in these components. Samples were removed from the heating bath at times between 5 and 300 minutes.

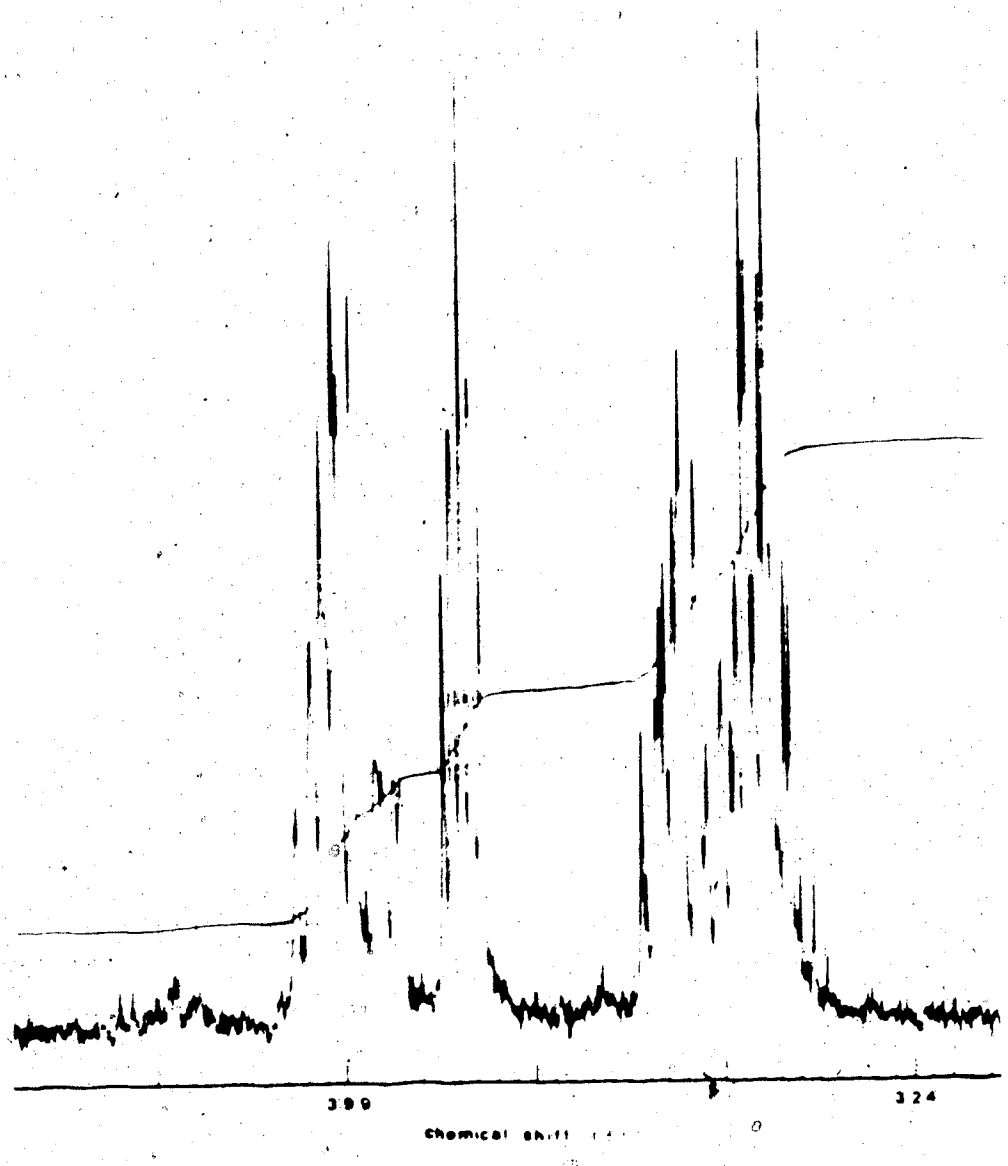


Figure 27. Part of the 400 MHz ^{1}H NMR spectrum of the mixture of c-(D-pro-L-leu) and c-(L-pro-L-leu) isolated during the decomposition of D-pro-L-leu-gly-gly, 148.5°, 30 min, pH 6.8, 9.01 mM. The expansion shows the signals due to the α protons. The central multiplet is due to the α proton of the leucine peptide unit in c-(D-pro-L-leu). The small multiplet at 3.94 ppm, is due to the same proton in c-(L-pro-L-leu).

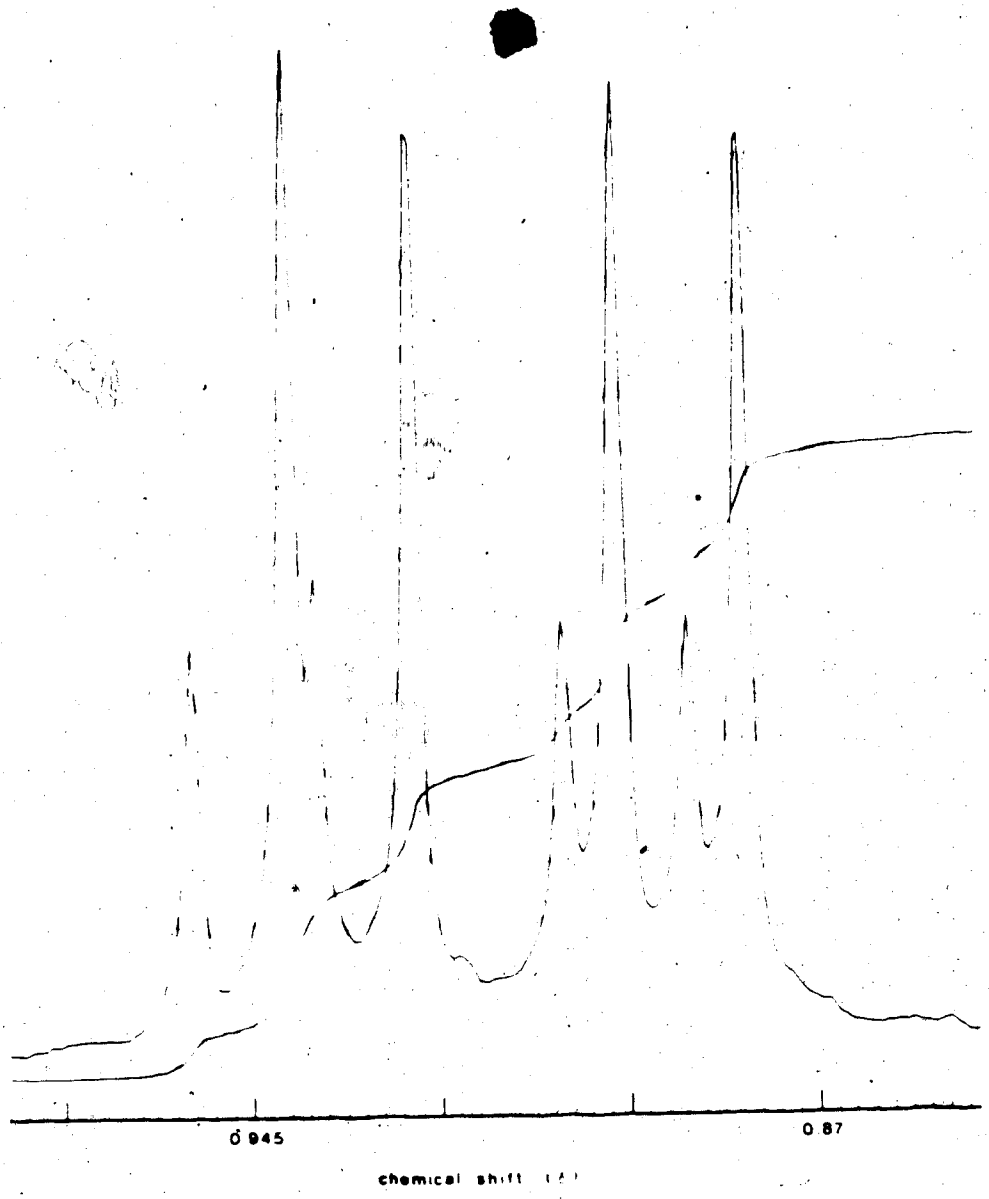


Figure 28. Part of the same spectrum as in Figure 27. The expansion shows the signals due to the methyl protons. The diastereomers are distinguishable by the slight difference in chemical shift. The relative amounts of the diastereomers was calculated by measuring the integral. The four smaller peaks are due to c-(L-pro-L-leu).

Table 38. Measurement of Relative Extinction Coefficients at 230 nm. After 30 minutes reaction time, the ratio of diastereomers was measured by NMR. By comparison to the ratio obtained by HPLC, the relative extinction coefficients were found.

$$\frac{DL + LD}{DD + LL}$$

Compound	measured by HPLC	measured by 400 MHz ^1H NMR
plgg	0.748	1.032[a]
cyclo-(pro-leu)	1.693	2.12[b]
pl	0.676	0.63[c]

a) From integration of gly_1CH_2

b) 2.18 from overlapping CH_3 signals
2.05 from overlapping α protons
1.96 from NH signals

c) 0.60 from leuα
0.66 from proα

Table 39. Relative Extinction Coefficients at 230 nm From Calibration Standards

Mixture	ε_{rel}
$\frac{glgg}{cyclo\text{-}gl}$	1.861
$\frac{gl}{glgg}$	1.105
$\frac{lg}{glgg}$	1.014
$\frac{plgg}{cyclo\text{-}pl}$	0.621
$\frac{pl}{plgg}$	0.877
$\frac{cyclo\text{-}pl}{cyclo\text{-}gl}$	3.69
$\frac{cyclo\text{-}pl}{plgg}$	1.610

The dipeptide pl and its diastereomer have similar retention times to the tetrapeptide plgg and its diastereomer. From the measurement of the rate of formation of the DKP's, it appeared that the tetrapeptide was hydrolyzing to the DKP with little or no incidence of the dipeptides. However, at long reaction times, the DKP's could be hydrolyzing to the dipeptides, thus complicating the kinetic picture. In order to test this idea and to determine if the DKP's were in equilibrium with each other, the DKP's were separated from the reaction mixture, heated under the same conditions and then separated again by HPLC. It was found that the DKP's were in equilibrium with each other and with the dipeptides. However, the dipeptides were found to be present in the equilibrating mixture to an amount not exceeding ~10% of the total. It was assumed that the DKP's were essentially stable to the reaction conditions and that the dipeptides were not significantly affecting the rate constant for hydrolysis derived from the measurement of the disappearance of the substrate.

Competitive Studies

The relative rates of racemization of the synthesized peptide derivatives was measured at one time. The reaction period was chosen to give a measurable amount of

racemization when the substrate was still visible in the chromatogram. This period was determined to be 90 minutes for plgg and 150 minutes for glgg and gplg.

The substrate was separated from the hydrolysis mixture in the same way as in the time studies. For the derivatives, the substrate eluted from the column at longer retention times than for the parent peptide. There was no evidence that the diastereomers of the derivatives eluted at different times so that only one peak from the chromatogram was collected. A qualitative measure of the rate of hydrolysis of the derivatives was obtained.

CHAPTER 4

RESULTS AND DISCUSSION

In previous model studies there was no attempt to isolate the substrate peptide from the hydrolysis mixture. In an exploratory experiment, thin layer chromatography was used to attempt to separate the components of the mixture after heating a buffered aqueous solution of the model peptides. The thin layer chromatograms showed numerous hydrolysis fragments illustrating the unavoidable experimental difficulties. First, the numerous interferring components makes it difficult to be sure that only one compound was being recovered from the TLC plate. Second, in order to achieve a significant degree of racemization, the peptide had to be allowed to decompose to the point where there was nearly none of the starting peptide remaining. This last point underscores the limitation of studies which focus on the racemization of the total hydrolysis mixture. The assignment of the experimentally determined rate constant to the substrate is not justified if the substrate accounts for only a small amount of the material in the hydrolysis mixture.

Separation of the hydrolysis mixture is not without hazards either. In the case of a peptide with several chiral centres, there could be several separable diastereomers of the original peptide. Analysis of the component having the same retention time as the original peptide will result in a rate of epimerization which will certainly be lower than the rate of epimerization for all of the diastereomers combined. Therefore, it is important to be sure that the substrate and its diastereomers are not separated from each other or that the positions in the chromatogram of the diastereomers be known. For experimental purposes, especially for determining epimerization rate constants for a protein, it is far better to try to achieve the former situation.

Despite the problems associated with separation, conclusions are possible from heating experiments on model peptides using TLC as the separation technique. The data in Table 40 was collected by isolating a fraction from the TLC plate having the same R_f as the substrate.

Aside from the starting material, the identity of each spot on the TLC is not known. The important observation is the large variation in the D/L ratios in the components from the hydrolysis of peptides containing more than one chiral centre. It is not known whether this is due to partial separation of diastereomers or due to

Table 40. D/L Ratios from Components Isolated from TLC[a] of Peptide Hydrolysate, 148.5°, pH 6.8, 270 min

Peptide	Component[b]		D/L val	leu	glu	phe	ala	pro
val-leu-ser-glu-gly	1	.36	0.50	0.40	0.03			
	2	.29	1.13	1.01	0.04			
	3*	.17		0.05	0.15			
	4	.14			1.27			
	5	.11	0.50	0.40	0.003			
glgg	1	.25		0.66				
	2*	.18		0.43				
ala-phe-pro-ala-β-napth	1	.38					0.20	
	2*	.35				0.03	1.15	0.085
	3	.28				0.04	0.90	0
	4	.18					0.49	
Z-gplg	1*	.27		0.02				0
	2	.21		0.34				0.78
	3	.18		0.22				
	4	.16		0.02				0.77
gly-gly-pro-ala	1	.061					0.020	
	2*	.036					0.07	0.09
pro-leu-gly-gly	1	.27		0.40				0.73
	2	.22		0.18				1.04
	3*	.12		0.04				2.03

a) Stationary Phase: silica gel (Whatman LK6DF 250 μ). Mobile phase: n-butanol/acetic acid/water, 4:1:1

b) Substrate marked by *

the different rates of racemization of the different components in the hydrolysis mixture.

The one peptide which cannot suffer from the problem of diastereomers is the tetrapeptide containing only one chiral centre, gly-L-leu-gly-gly. The peptide was heated at 148.5° at pH 6.8 and the hydrolysis mixture separated by TLC. Seven distinct components were detected, isolated and the D/L ratios measured at several times (Table 41). The canonical expression is appropriate for the substrate if there is no reversible process involving the substrate. This does not preclude the possibility of accidentally linear canonical plots. Straight line canonical plots were observed for the substrate and the other components isolated from the TLC plate (Figure 29). Using the gly/leu ratio from the GC analysis of the amino acid derivatives produced after acid hydrolysis of the components, the identity of the components was tentatively assigned.

The racemization rate constants given in Table 42 do not span a very large range. Nevertheless, these experiments showed the potential for error if the total hydrolysis mixtures are used to measure the rate constants for the substrates. A positive aspect is that it is possible to find a peptide whose kinetics are well behaved.

Table 41. Racemization[a] (D/L) of Leucine Peptide Unit in Components Isolated by TLC During Hydrolysis of Glycyl-L-leucyl-glycyl-glycine, 148.5°, pH 6.8, 7.71 mM

Time (min)	D/L Ratios of Component in Mixture[b]				
	1	2	3	4	5
65	0.205	0.213	0.173	0.101	0.118
150	0.205	0.418	0.437	0.244	0.255
210	0.234	0.556	0.610	0.350	0.349
270	0.302	0.440	0.729	0.444	0.430
330	0.344	0.806	0.799	0.522	0.522
390	0.375	0.844	0.859	0.654	0.576
450	0.401	0.849	0.895	0.582	0.553
600	0.476	0.784	0.960	0.665	-

a) Separation performed on Whatman LK6DF; mobile phase 4:1:1, n-butanol:acetic acid:water

b) Component 1, R_f = 0.40; component 2, R_f = 0.36; component 3, R_f = 0.33, component 4 (substrate), R_f = 0.25; component 5, R_f = 0.20

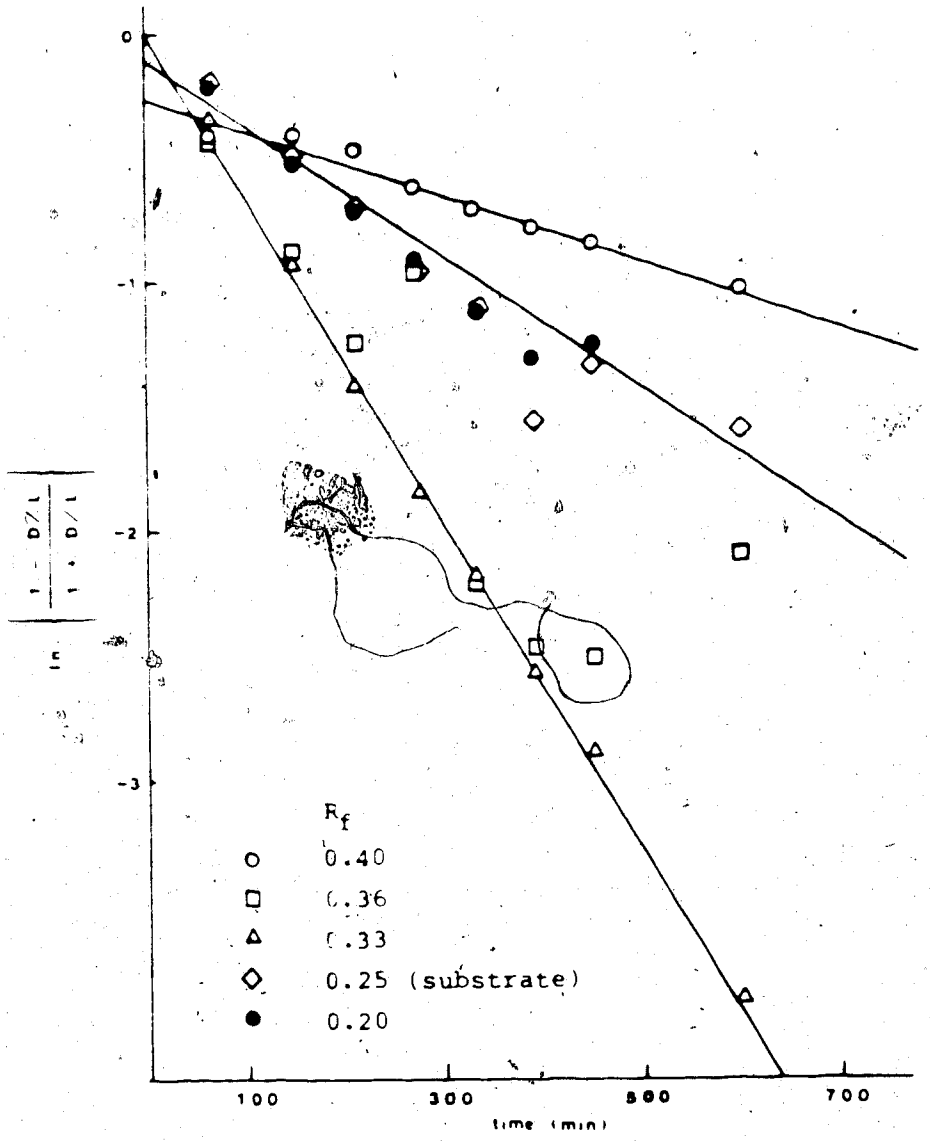


Figure 29. Racemization of the leucine peptide unit in components isolated by TLC during the decomposition of glycyl-L-leucyl-glycyl-glycine, 148.5°, pH 6.8, 7.71 mM. The solid lines were produced for the components having R_f values 0.40, 0.33 and 0.25. See Table 41.

Table 42. Leucine Epimerization Rate Constants for Components Isolated by TLC[a] During the Decomposition of Gly-L-leu-gly-gly, 148.5°, pH 6.8, 7.71 mM

R_f[b] Component	gly/leu	$10^5 k$ (s^{-1})	Component
0.40± 0.03	0.04± 0.01	1.10± 0.08	leu
0.36± 0.03	0.43± 0.15	4.93± 0.79[c]	gly-leu
0.33± 0.03	1.06± 0.12	5.45± 0.10	gly-leu-gly
0.25± 0.02	1.46± 0.12	2.30± 0.33	gly-leu-gly-gly
0.20± 0.03	1.56± 0.13	2.40± 0.15	

a) Stationary phase: Whatman LK6DF 250 μ silica gel. Mobile phase: 4:1:1, n-butanol:acetic acid:water

b) Other spots at R_f = 0.17± 0.02 and R_f = 0.10± 0.03 contained predominantly glycine

c) Only the first seven points (Table 41) were used to calculate the rate constant

A goal of this work was to model the racemization of protein by using peptides having more than one chiral centre. It appeared that separation by TLC was incapable of giving good results for peptides of this type. High Performance Liquid Chromatography was considered to be the analytical method of choice since the experimental parameters could be better controlled. For subsequent studies, HPLC was used exclusively for separation of the decomposition mixtures.

Racemization and Decomposition of L-pro-L-leu-gly-gly

The predominant feature of the racemization of pro-leu-gly-gly is the occurrence of an excess of D-proline in the total hydrolysate and in the diketopiperazine isolated during the decomposition of the substrate (Figure 30).

Excess D-proline has been observed in similar systems. For example, N-benzyloxycarbonyl-L-alanyl-L-phenylalanyl-L-proline treated with acetic anhydride and sodium acetate gives **15**, N-(N-benzyloxycarbonyl-L-alanyl)-c-(L-phenylalanyl-D-prolyl).[90] Also phenylacetyl-L-alanyl-L-phenylalanyl-L-prolyl-p-nitrophenyl ester treated with NaH in DMF at 0°C for 2 hours gives **16**, N-(N-phenylacetyl-L-alanyl)-c-L-phenylalanyl-D-prolyl).[91] Earlier, a 1:10 ratio of c-(L-phe-D-pro):c-(L-phe-L-pro) was formed by treating c-(L-phe-L-pro) with 0.5N sodium hydroxide at room temperature for 15 minutes.[92]

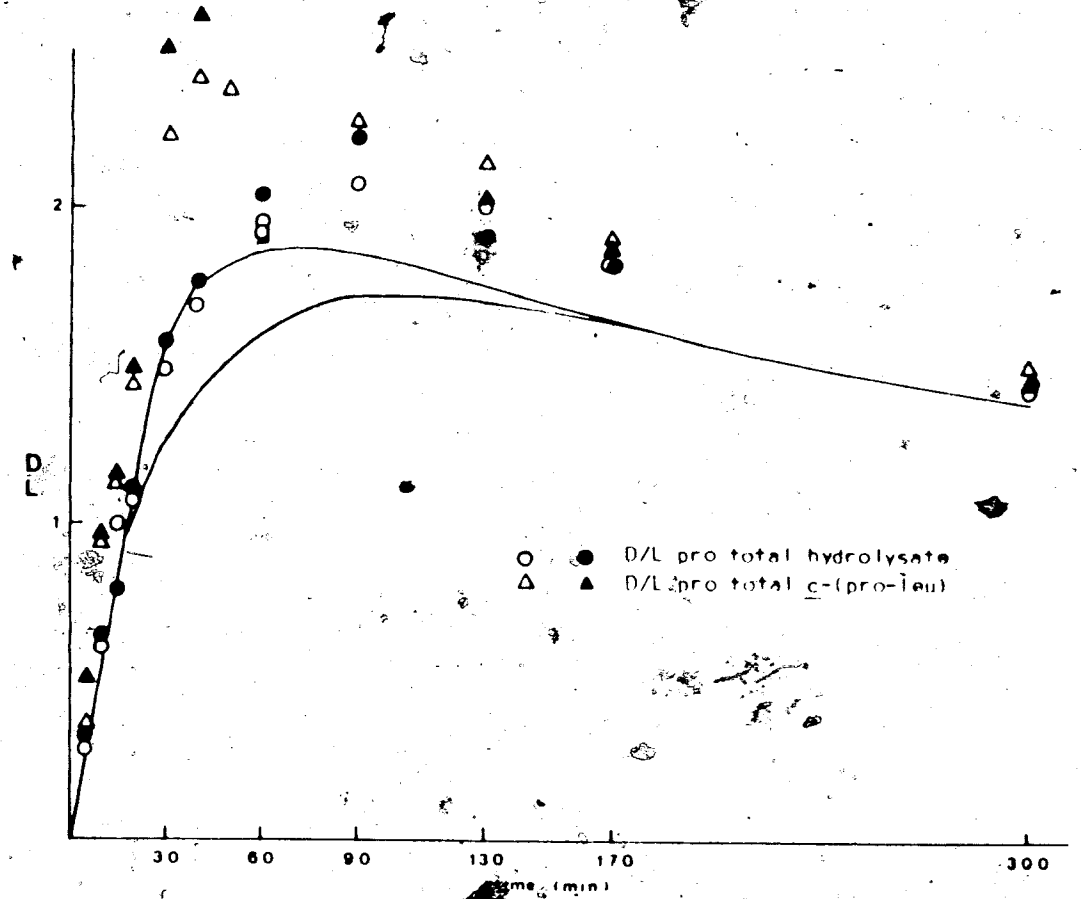
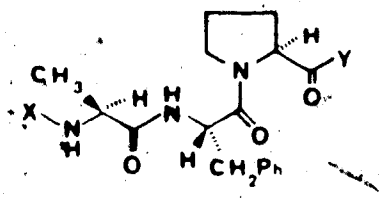
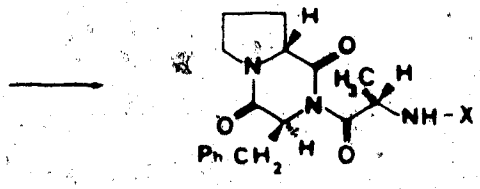


Figure 30. Racemization of proline peptide unit in total hydrolysate and total c-(pro-leu) isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.0, 8.45 mM. The N-pentafluoropropionyl isopropyl esters of D and L-proline could not be base-line resolved by GC. The measured D/L proline ratio was corrected in two ways. The solid symbols involving proline racemization represent D/L proline ratios corrected by use of a calibration curve (Appendix 6). The open symbols represent D/L proline ratios corrected by a second method. Immediately after a sample run, a standard D/L proline sample was injected into the GC. The observed D/L ratio for the sample was corrected by multiplying by the ratio $(D/L)_t/(D/L)_{obs}$. $(D/L)_t$ is the true D/L ratio of the standard produced by mixing appropriate amounts of solutions of pure D-proline and pure L-proline. $(D/L)_{obs}$ is the ratio observed in the GC for the standard. The upper and lower solid lines show the computer simulation produced for total c-(pro-leu) and the total hydrolysate respectively using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

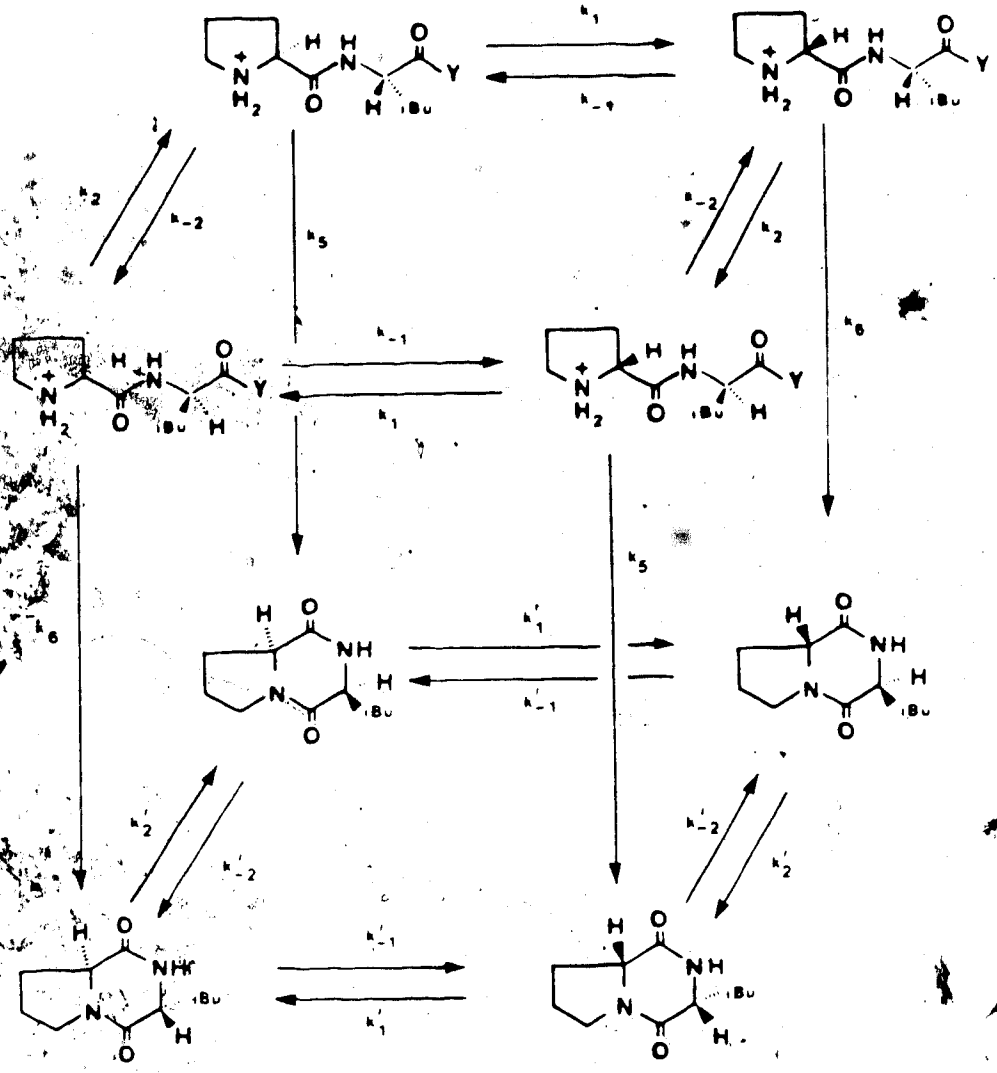


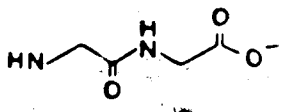


	X	Y
15	$Ph-CH_2-OCO$	OH
16	$Ph-CH_2-CO$	$PNO_2\ C_6H_4O$

In early studies of ergot alkaloids, D-proline was isolated from the alkaline hydrolysate in the form c-(D-prolyl-X) (X = leu, phe, val).[93] This caused confusion regarding the stereochemistry of proline in the alkaloids.[94] It was later determined that the proline was in the L-form. By using only 1 equivalent of base to hydrolyze the alkaloids, the diketopiperazine containing L-proline can be obtained. Addition of more base epimerizes the diketopiperazine to the thermodynamically more stable, c-(L-leucyl-D-prolyl).[95]

Scheme 9 shows the pathways assumed to be operating for the racemization and decomposition of L-pro-L-leu-



pro-leu-gly-gly ; Y = 

pro-leu ; Y = O^-

SCHEME 9

gly-gly (plgg).* There is sufficient information available in the experimental data to estimate the ten rate constants, of which eight are independent.

The disappearance of plgg and its diastereomer D-plgg** and the appearance of c-(L-pro-L-leu) (cpl) and its diastereomer c-(D-pro-L-leu) (cDpl) are shown in Figure 31. The data is given in Table 43. The rate constants k_5 and k_6 (Scheme 9) can be seen from Figure 31 to be equal, since after approximately twenty minutes, the diastereomers are disappearing at the same rate. The facile formation of diketopiperazine is not unexpected. Dipeptide esters and amides are known to cyclize especially when one of the peptide units is proline or N-methylated.[96]

Adding together the concentrations of the diastereomers of plgg (Figure 32) and plotting the natural logarithm of the concentration versus time should yield a straight line with slope $-k_5$ or $-k_6$ (Figure 33). A similar treatment of the data for the diketopiperazines should yield the same rate constant if there is no

---

*The processes of cyclization of plgg to give diketopiperazine cpl and gg and hydrolysis to pl and gg under pseudo first order conditions are kinetically equivalent. The generic term "decomposition" will be used to describe both hydrolysis and cyclization.

**Where not specified, the peptide unit is in the L form.

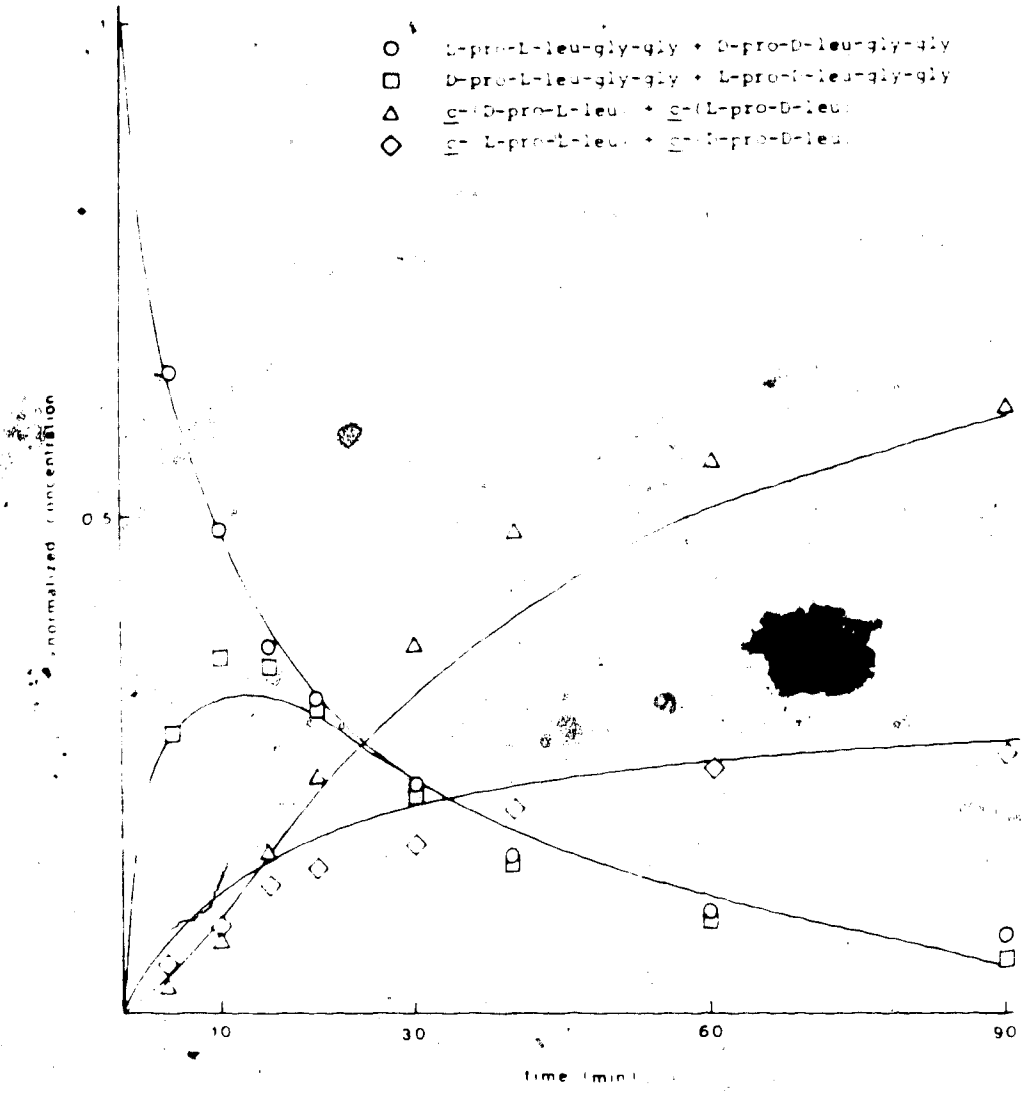


Figure 31. Decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The solid lines show the computer simulation produced using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in scheme 9.

Table 43. Decomposition[a] of Pro-leu-gly-gly, 148.5°, pH 6.8,[b] 8.45 mM

	Normalized Concentration[c]			
Time (min)	plgg	D plgg	cDpl	cpl
5	0.645	0.281	0.023	0.049
10	0.486	0.358	0.071	0.086
15	0.386	0.349	0.162	0.127
20	0.313	0.305	0.236	0.145
30	0.231	0.221	0.370	0.170
40	0.158	0.151	0.484	0.205
60	0.100	0.097	0.555	0.246
90	0.079	0.055	0.608	0.260

a) Measured by the UV absorption at 230 nm of the eluate from the HPLC separation

b) 0.025 M KH_2PO_4, 0.025 M Na_2HPO_4

c) Abbreviations: plgg = L-prolyl-L-leucyl-glycyl-glycine; Dplgg = D-prolyl-L-leucyl-glycyl-glycine; cDpl = cyclo-(D-prolyl-L-leucyl); cpl = cyclo-(L-prolyl-L-leucyl); glgg = glycyl-L-leucyl-glycyl-glycine; cgl = cyclo-(glycyl-L-leucyl); pl = L-prolyl-L-leucine; DPl = D-prolyl-L-leucine; DKP = diketopiperazine. Each abbreviation refers to the pair of enantiomers. In the hydrolysis data, the abbreviation denotes the sum of the concentrations of the enantiomers, while in the racemization data the abreviation denotes the ratio of the concentrations of the enantiomers.

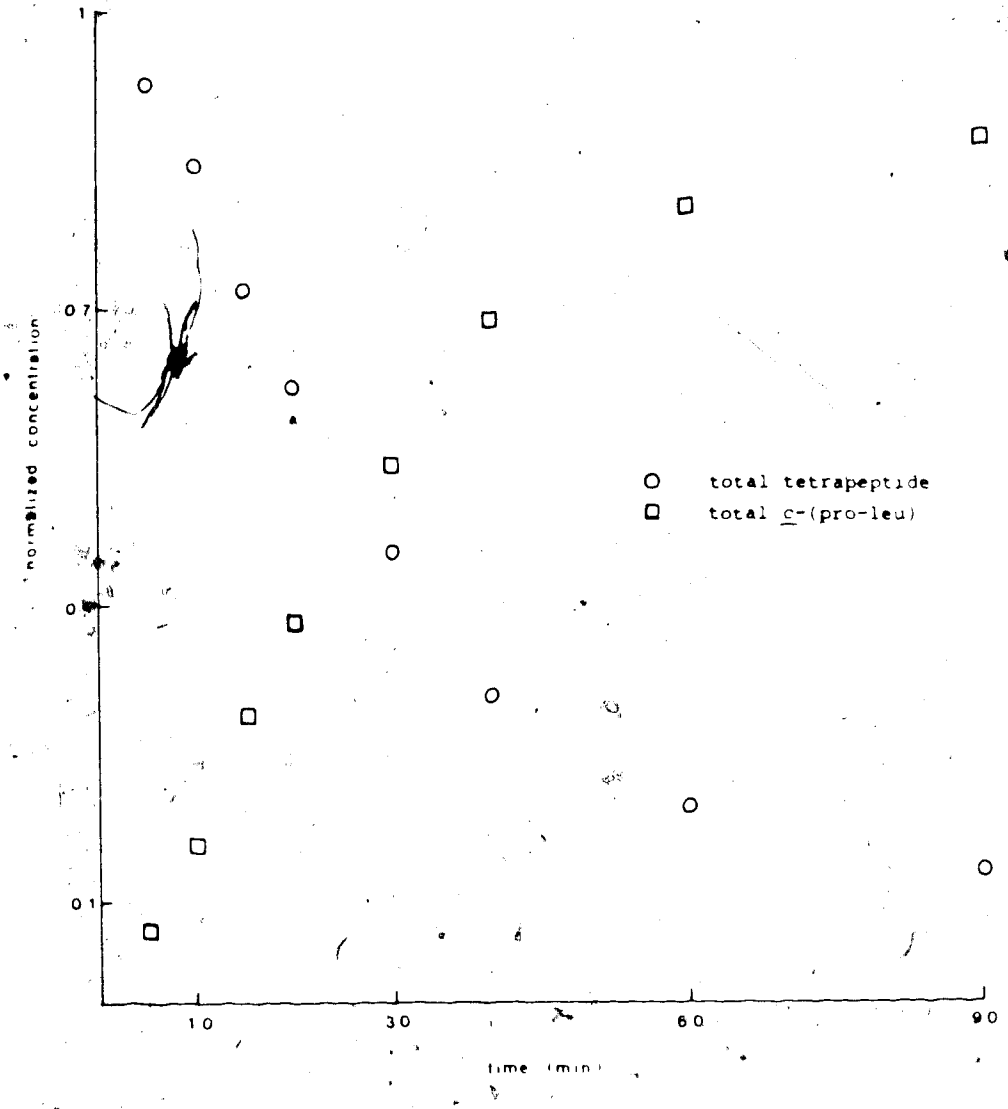


Figure 32. Decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM.

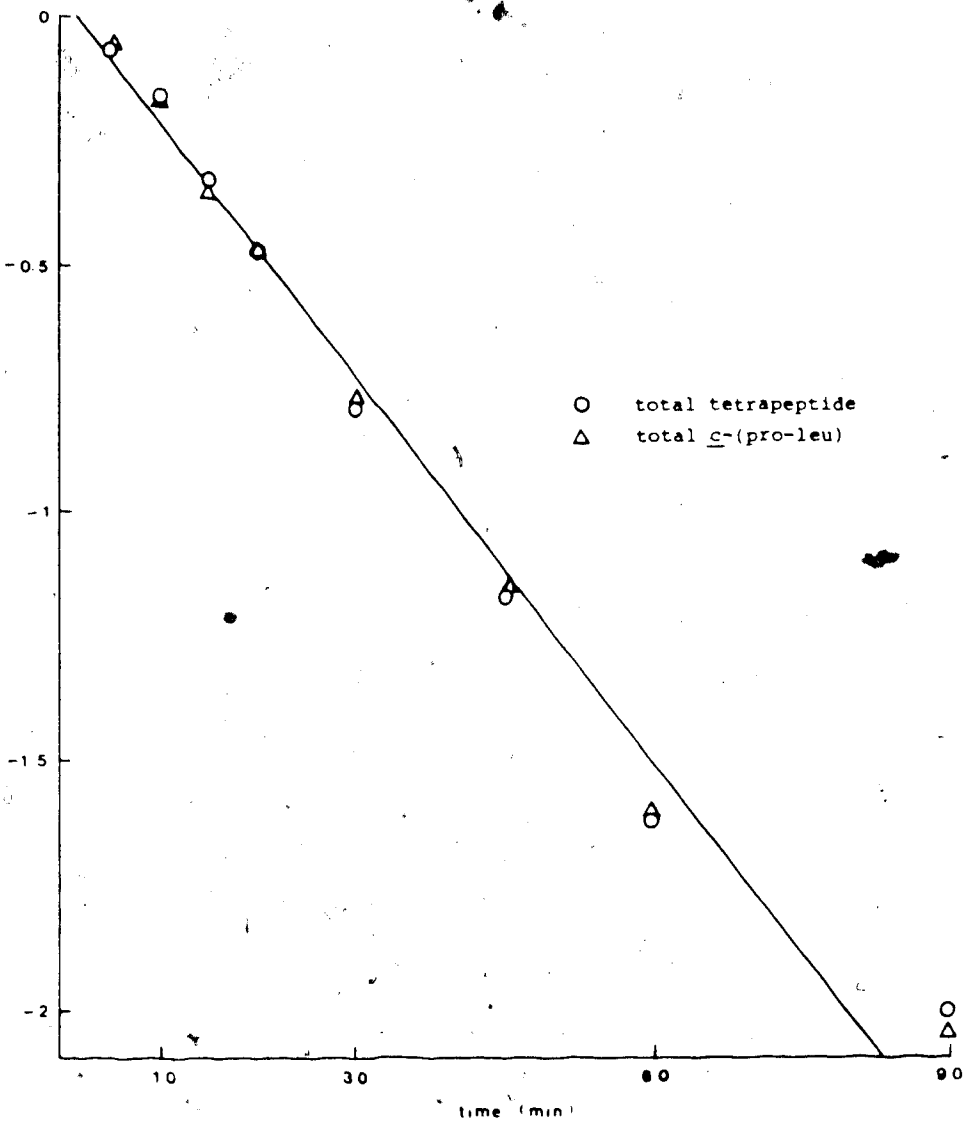


Figure 33. Decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. For pro-leu-gly-gly the ordinate is -ln(c) where c is the normalized concentration of total tetrapeptide. For c-(pro-leu) the ordinate is -ln(1-c) where c is the normalized concentration of diketopiperazine. The solid line was produced by linear regression analysis.

competing hydrolysis of the tetrapeptide.*

Figure 31 shows that the diastereomers plgg and D-plgg achieved the same concentration at 20 minutes. This can be expressed as:

$$\frac{DL + LD}{DD + LL} = 1 \qquad (14)$$

and requires that $k_{-1} = k_1$ and $k_2 = k_{-2}$. This agrees with the racemization kinetics of plgg. The racemization data for the proline and leucine peptide units in the components isolated during the decomposition of plgg are given in Tables 44 through 47. The D/L ratio for the proline tetrapeptide unit never exceeded unity (Figure 34). The fact that $k_5 = k_6$, $k_{-1} = k_1$ and $k_2 = k_{-2}$ significantly simplifies the extraction of the rate

---

*There are two pieces of evidence to support this idea:

1) A test sample of plgg (2.99 mg/mL) was extracted with 3 × 5 mL $CHCl_3$ after 60 minutes of heating at 148.50°. Material weighing 3.2 ± 0.2 mg was isolated having the same 200 MHz ^1H NMR spectrum as the DKP and with mass 210. (M+1 in CI mass spectrum). This constitutes an 86 ± 5% yield in good agreement with expected quantity in Figure 1.

2) The eluate from the HPLC up to the elution of plgg (the first major peak) was collected. By GC analysis this was found to contain mostly glycine, with a gly/pro/leu ratio of 1/.06/.01. The relative amount of glycine is likely larger than this since the derivatization procedure depletes the glycine derivative more than pro or leu.

Table 44. Racemization[a] (D/L) of the Leucine Peptide Unit in Components from the Decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM

Time (min)	Component[b] plgg	Dplgg	total tetrapeptide	cDpl	cpl	total DKP	total hydrolysate
5	0.010	0.016	0.013	0.065	0.069	0.028	0.014
10	0.017	0.022	0.018	0.039	0.037	0.056	0.019
15	0.017	0.027	0.019	0.077	0.026	0.045	0.029
20	0.019	0.032	0.030	0.032	0.030	0.040	0.029
30	0.024	0.024	0.023	0.052	0.086	0.051	0.039
40	0.029	0.029	0.028	0.064	0.047	0.066	0.060
60	0.046	0.044	0.046	0.088	0.098	0.083	0.096
60	0.043	0.037	0.036	0.090	0.105	0.129	0.087
90	0.055	0.063	0.060	0.179	0.177	0.193	0.137
130	0.096	0.081	0.103	0.199	0.222	0.195	0.218
170	0.185	0.160	0.160	0.268	0.285	0.280	0.283
300	0.077[c]	0.317[c]	0.416[c]	0.469	0.487	0.483	0.488

a) D/L of N-pentafluoropropionyl isopropyl esters of amino acids produced by acid hydrolysis (110°, 24 h) of the components from the decomposition. Stationary phase: Chirasil-Val

b) For abbreviations, see footnote c), Table 43

c) Essentially no substrate was left by this point

Table 45. Racemization[a,b] (D/L) of the Proline Peptide Unit in Components from the Decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM

Time (min)	Component[c] plgg	Dplgg	total tetrapeptide	cDpl	cpl	total DKP	total hydrolysate
5	0.009	20.53	0.282	8.63	0.010	0.356	0.296
10	0.014	28.36	0.545	11.24	0.039	0.896	0.593
15	0.018	14.64	0.767	1.81	0.051	1.041	0.727
20	0.024	12.7	0.729	16.1	0.057	1.328	1.023
30	0.027	18.56	0.92	9.85	-	2.192	1.396
40	0.028	19.21	0.892	11.20	0.100	2.287	1.560
60	0.041	12.56	0.897	9.51	0.157	2.229	1.796
60	0.035	14.25	0.755	8.41	0.202	2.223	1.667
90	0.064	10.46	0.821	1.663	0.103	2.014	1.946
130	0.110	2.11	0.631	3.540	0.263	1.766	1.664
170	0.231	2.82	0.547	3.652	0.332	1.664	1.603
300	0.626[d]	-	0.508[c]	2.176	0.507	1.289	1.284

a) See footnote a), Table 44

b) Values are not corrected for unresolved N-PFPA isopropyl esters of D-proline and L-proline

c) For abbreviations see footnote c), Table 43

d) See footnote c), Table 44

Table 46. Racemization[a,b] (D/L) of the Proline Peptide Unit in Components from the Decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM

Time (min)	Component[c] plgg	Dplgg	total tetrapeptide	cDpl	cpl	total DKP	total hydrolysate
5	0.014	32.4	0.292	13.6	0.010	0.364	0.283
10	0.014	36.7	0.564	13.7	0.039	0.936	0.604
15	0.026	23.5	0.778	1.92	0.054	1.106	0.969
20	0.026	20.4	0.974	21.3	0.072	1.433	1.066
30	0.028	24.5	0.965	15.36	-	2.221	1.480
40	0.029	19.7	0.944	13.42	0.100	2.399	1.681
60	0.045	14.80	0.932	10.95	0.148	2.367	1.919
60	0.056	30.78	0.913	14.37	0.246	2.361	1.942
90	0.065	14.01	0.860	1.896	0.142	1.663	2.064
130	0.174	3.643	0.697	4.595	0.372	2.136	1.954
170	0.289	3.089	0.744	5.205	0.445	1.885	1.815
300	0.942[d]	-	0.981[d]	2.45	0.546	1.484	1.409

a) See footnote a), Table 44

b) Values are corrected for unresolved N-PFPA isopropyl esters of D-proline and L-proline by comparison with a D/L proline standard with approximately the same D/L ratio as the sample. The standard was injected immediately upon completion of the chromatographic run of the sample.

c) For abbreviations see footnote c), Table 43

d) See footnote c), Table 44

Table 47. Racemization[a,b] (D/L) of the Proline Peptide Unit in Components from the Decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM

Time (min)	Component[c] plgg	Dplgg	total tetrapeptide	cDpl	cpl	total DKP	total Hydrolysate
5	0.011	27.4	0.305	11.72	0.014	0.385	0.320
10	0.016	37.7	0.588	15.15	0.036	0.967	0.640
15	0.020	19.6	0.828	2.76	0.044	1.150	0.784
20	0.027	17.1	0.787	21.54	0.056	1.489	1.129
30	0.030	24.8	0.993	13.33	-	2.505	1.57
40	0.031	25.6	0.962	15.1	0.109	2.619	1.76
60	0.045	16.9	0.968	12.88	0.170	2.551	2.03
60	0.039	19.1	0.815	11.44	0.218	2.280	1.89
90	0.070	14.4	0.886	2.57	0.112	1.473	2.21
130	0.119	2.41	0.681	4.10	0.284	2.01	1.90
170	0.250	3.25	0.590	4.23	0.356	1.86	1.81
300	0.676[d]	-	0.548[d]	2.49	0.551	1.44	1.43

a) See footnote a), Table 44

b) Values are corrected for unresolved N-PFPA isopropyl esters of D-proline and L-proline using calibration curve (Appendix 6)

c) For abbreviations see footnote c), Table 43

d) See footnote c), Table 44

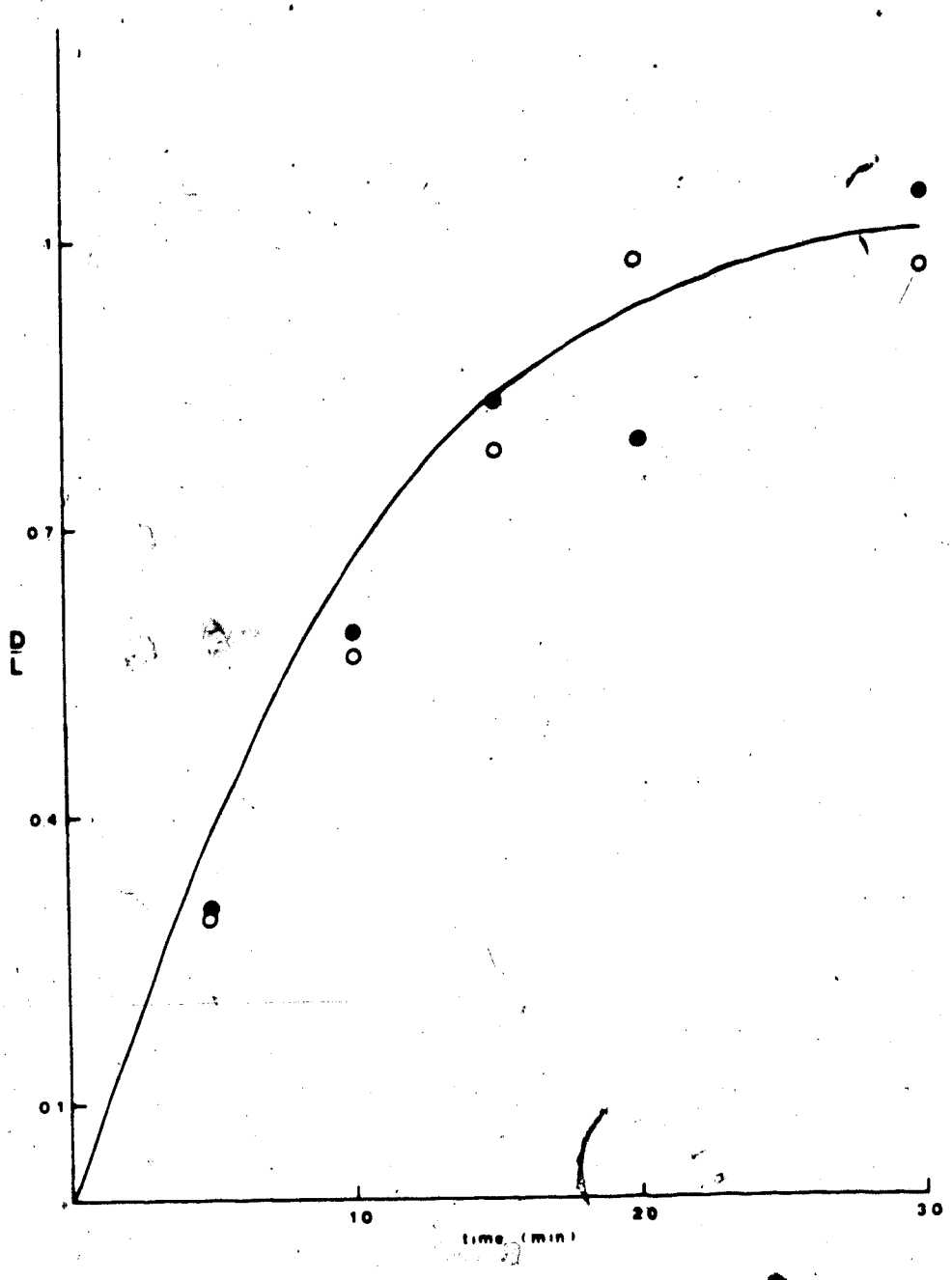


Figure 34. Racemization of the proline peptide unit in total pro-leu-gly-gly isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line shows the computer simulation produced using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

constants. From the algebraic solution of Scheme 6, equation A3-17 becomes:

$$A-A^* = c_1 \exp(-2k_2t) + c_3 \exp(-2k_1t). \qquad (15)$$

where A is the proline peptide unit. From equation A3-24:

$$c_1 = \frac{1}{2(k_1 - k_2)}\left(-2k_1 + \frac{2k_1 + 2k_2 + 2(k_1 - k_2)}{2}\right) = 0 \qquad (16)$$

Since $c_3 = 1 - c_1$, therefore:

$$A-A^* = (A-A^*)_0 \exp(-2k_1t) \qquad (17)$$

Thus a plot of

$$\ln\left(\frac{A-A^*}{A+A^*}\right)$$

equivalent to

$$\ln\left(\frac{1 - D/L}{1 + D/L}\right) \qquad (18)$$

for the proline peptide unit versus time will yield a straight line with slope $-2k_1$ (Figure 35). Similarly for $B-B^*$,

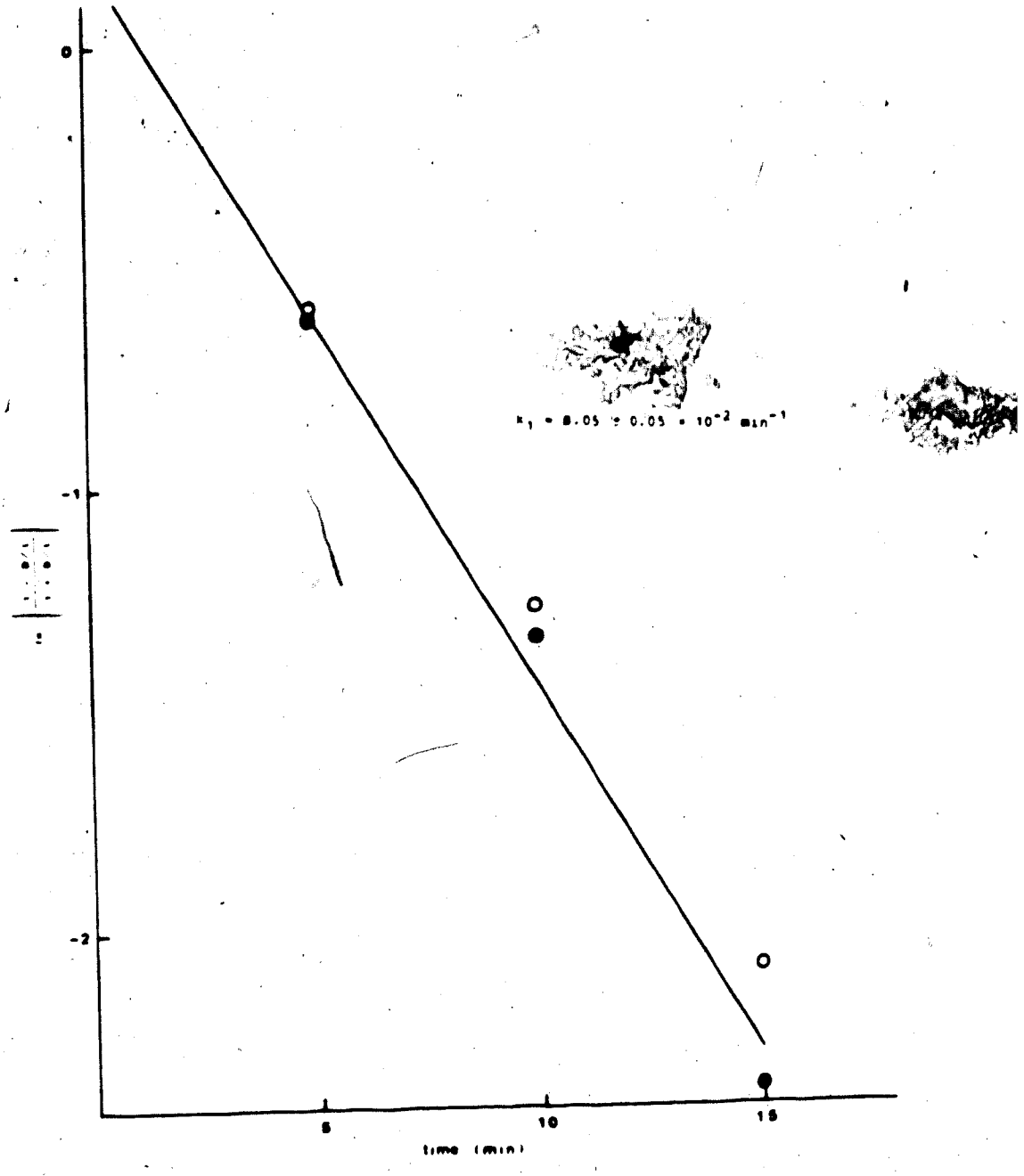


Figure 35. Racemization of the proline peptide unit in the total tetrapeptide isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line was produced by linear regression analysis.

$$B-B^* = (B-B^*)_0 \exp(-2k_2t) \qquad (19)$$

and a plot of

$$\ln\left(\frac{1 - D/L}{1 + D/L}\right) \qquad (20)$$

for the leucine peptide unit versus time will yield a straight line with slope $-2k_2$.

It is possible to measure two other ratios, namely DD/LL and LD/DL after separately isolating the two diastereomers. The racemization measured by proline and leucine gives two separate measurements of the two ratios. Again assuming the algebraic solution is applicable, equation A3-22 becomes:

$$AB + A^*B^* - AB^* - A^*B = (AB + A^*B^* - AB^* - A^*B)_0$$

$$e^{-2(k_1+k_2)t} \qquad (21)$$

Rewriting the above equations (17) and (19):

$$AB + AB^* - A^*B^* - A^*B = (A - A^*)_0 \exp(-2k_1t) \qquad (22)$$

$$AB + A^*B - A^*B^* - AB^* = (B - B^*)_0 \exp(-2k_2t) \qquad (23)$$

With

$$AB + AB^* + A^*B + A^*B^* = 1 \tag{24}$$

these equations can be combined so that:

$$\frac{AB - A^*B^*}{AB + A^*B^*} = \frac{(A - A^*)_0 \, e^{-2k_1 t} + (B - B^*)_0 \, e^{-2k_2 t}}{1 + D_0 \exp(-2(k_1 + k_2)t)} \tag{25}$$

and

$$\frac{AB^* - A^*B}{AB^* + A^*B} = \frac{(A - A^*)_0 \exp(-2k_1 t) - (B - B^*)_0 \exp(-2k_2 t)}{1 - D_0 \exp(-2(k_1 + k_2)t)} \tag{26}$$

where

$$D_0 = (AB + A^*B^* - AB^* - A^*B)_0$$

If $k_1 >> k_2$ then

$$\frac{1 - \frac{A^*B^*}{AB}}{1 + \frac{A^*B^*}{AB}} = (B - B^*_0) \exp(-2k_2 t) \tag{27}$$

and

$$\frac{1 - \frac{A^*B}{AB^*}}{1 + \frac{A^*B}{AB^*}} = -(B - B^*)_0 \exp(-2k_2 t) \tag{28}$$

It is noteworthy that k_2 can be obtained from five measurements whereas k_1 can be found from only one. Figures 36 through 39 show the plots used to extract k_2 for plgg. A further approximation was used here. For small D/L:

$$-\frac{1}{2}\ln\left(\frac{1 - D/L}{1 + D/L}\right) = D/L \qquad (29)$$

A plot of D/L versus time will yield a straight line with slope k_2.

The excess D-proline (asymmetric induction) observed in the total hydrolysate (Figure 30) arises from the production of excess diketopiperazine cDpl. According to the model (Scheme 6), this is occurring partly because $k'_1 > k'_{-1}$. Consequently, some of the symmetry in the algebraic solution is destroyed and a different approach must be taken to successfully extract the rate constants.

For asymmetric induction to occur, $k'_1 > k'_{-2}$. Hence $k'_{-1} > k'_2$ and $k'_1 + k'_{-1} > k'_2 + k'_{-2}$. From equation A3-18

$$I^2 = (1+K)^2(k'_{-1} + k'_2)^2 - 16K\,k'_{-1}k'_2 \qquad (30)$$

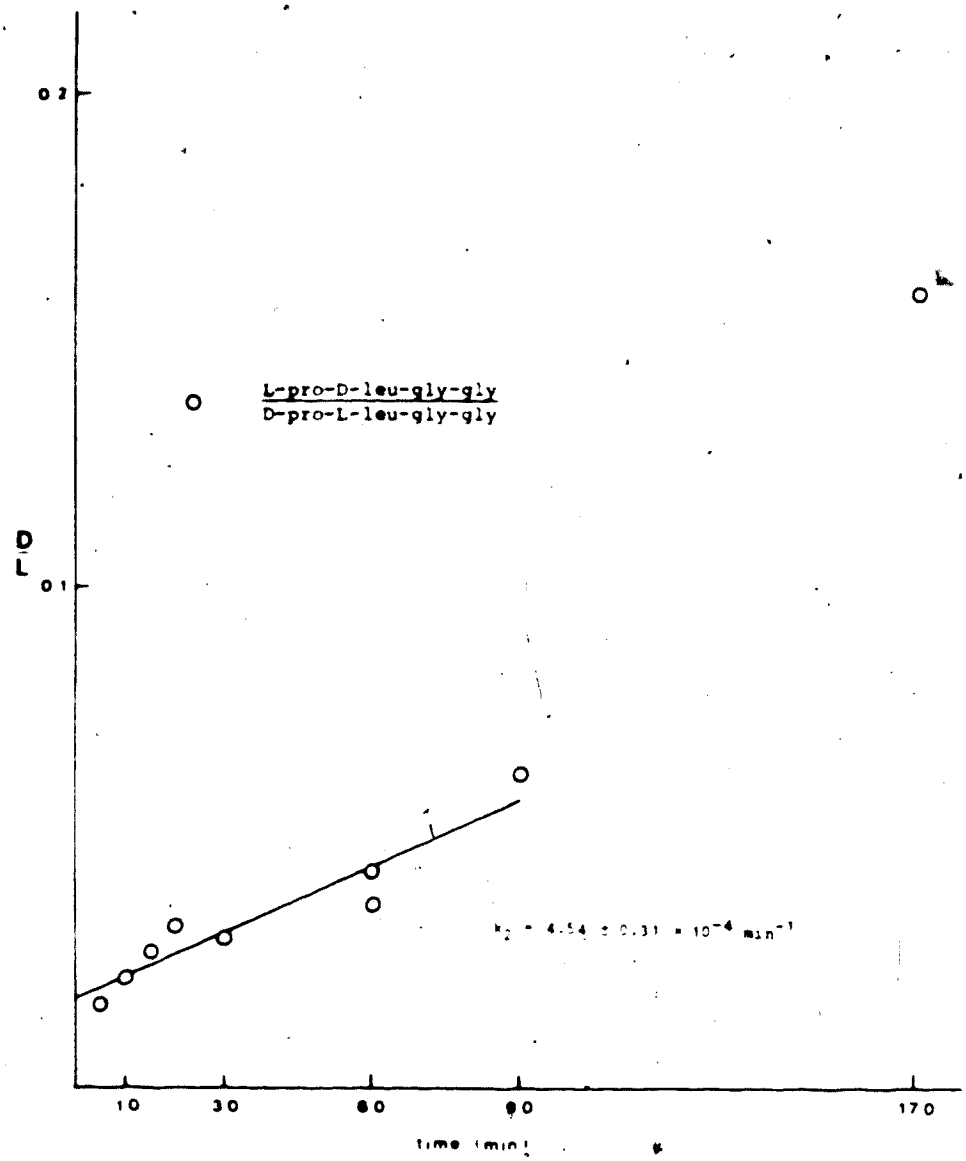


Figure 36. Racemization of the leucine peptide unit in D-pro-L-leu-gly-gly isolated during hydrolysis of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The solid line is produced by linear regression analysis. The point at 170 min was not used. Since there is very little substrate left at this point, there is considerable error associated with this measurement.

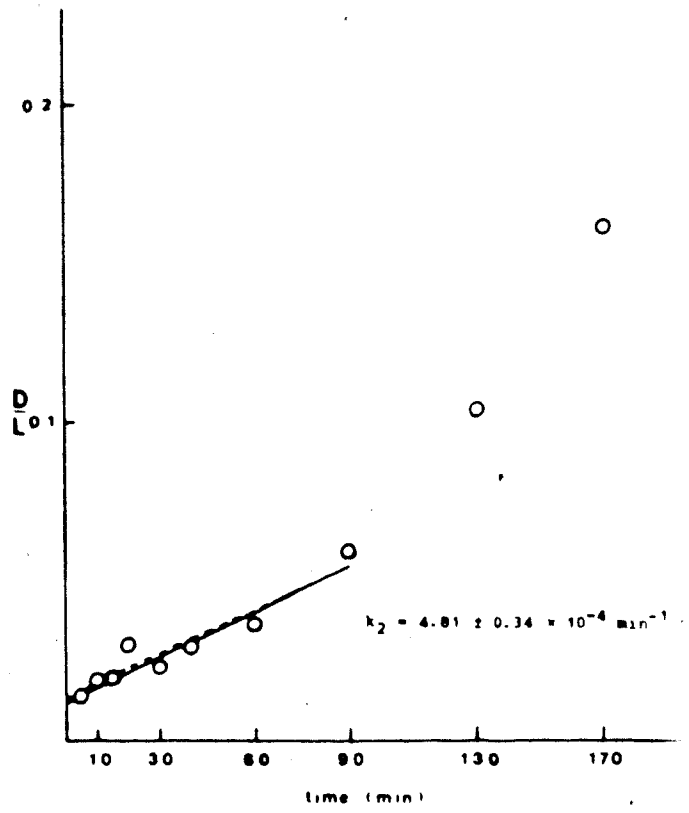


Figure 37. Racemization of the leucine peptide unit in total pro-leu-gly-gly isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The solid line was produced by linear regression analysis. The dashed line shows the computer simulation produced using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

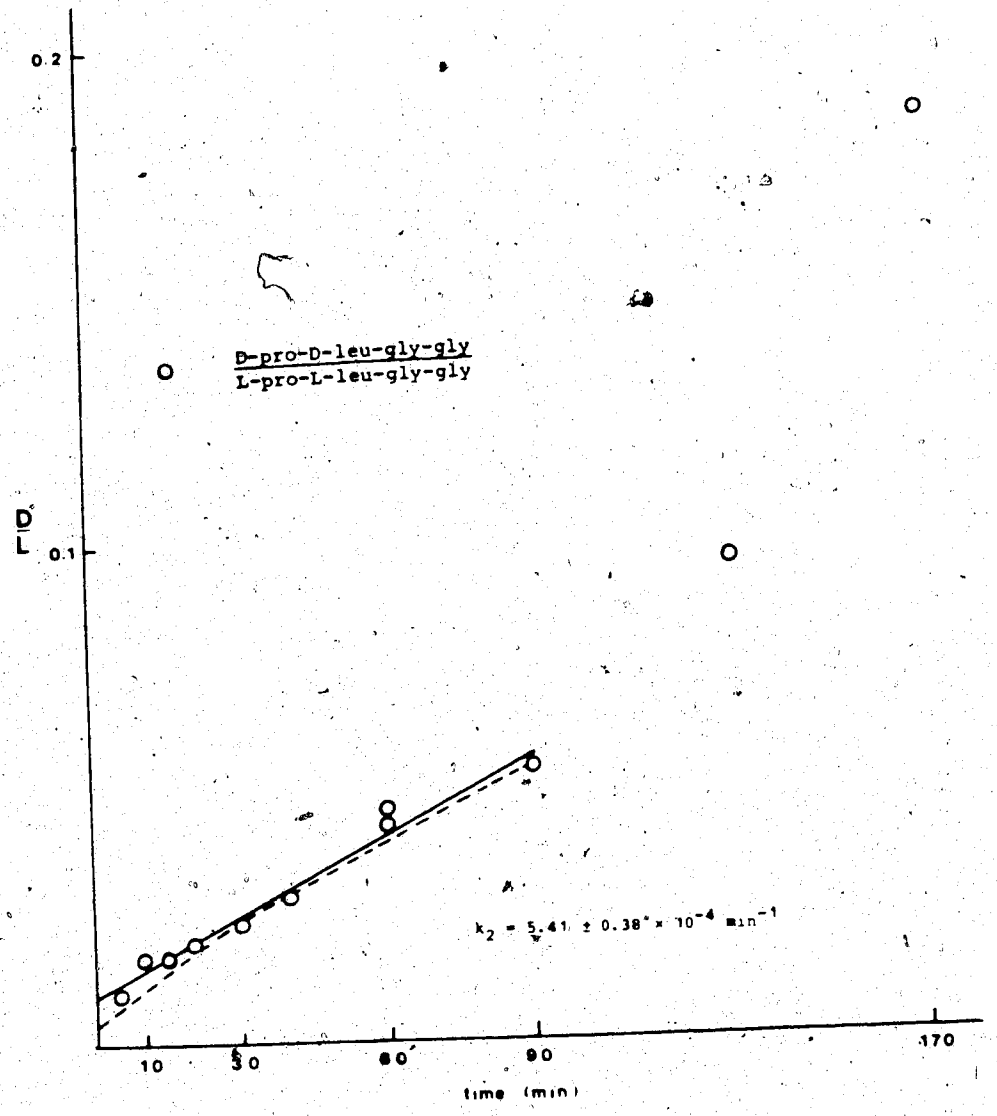


Figure 38. Racemization of the leucine peptide unit in pro-leu-gly-gly isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The solid line was produced by linear regression analysis of the data points before 90 minutes (see Figure 36). The dashed line shows the computer simulation produced using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

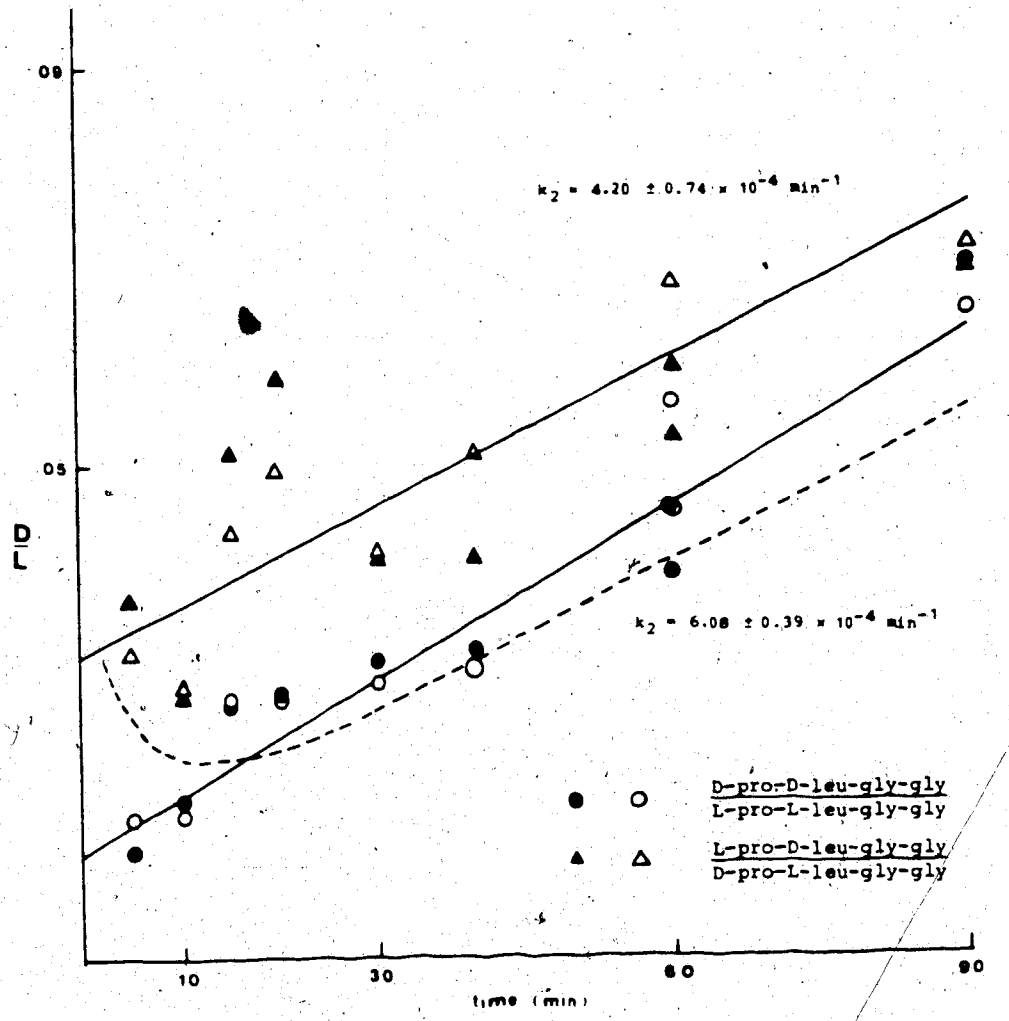


Figure 39. Racemization of the proline peptide unit in pro-leu-gly-gly isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid lines were produced by linear regression analysis. The dashed line shows the computer simulation produced for L-pro-D-leu-gly-gly/D-pro-L-leu-gly-gly using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

where because of the relative values of k'_{-1} and k'_2 the value of I will be slightly less than $k'_1 + k'_{-1}$. From A3-17

$$A-A^* = c_1 \exp\left(\frac{-\Sigma k + I}{2} \cdot t\right) + c_3 \exp\left(\frac{-\Sigma k - I}{2} \cdot t\right) \tag{31}$$

where $\Sigma k = k'_1 + k'_{-1} + k'_2 + k'_{-2}$ and $c_3 = 1 - c_1$.

Since $-\Sigma k + I \approx 0$ and $k'_1 + k'_{-1} > k'_2 + k'_{-2}$ this expression becomes

$$A-A^* = c_1 + (1 - c_1) \exp(-(k'_1 + k'_{-1})t) \tag{32}$$

From A3-24 and A3-25

$$c_1 = \frac{1}{I}\left\{-2k'_1 + \frac{\Sigma k + I}{2}\right\}$$

and with the inequality $k'_1 + k'_{-1} > k'_2 + k'_{-2}$

$$c_1 = \frac{-2k'_1 + (1 + K)k'_{-1}}{(1 + K)k'_{-1}} \tag{33}$$

$$= \frac{1 - K}{1 + K}$$

Equation 32 can be expressed as:

$$A - \left(\frac{1 + c_1}{1 - c_1}\right) A^* = \frac{2(K - 1)}{(K + 1)} \exp(-(k'_1 + k'_{-1})t) \quad (34)$$

so that

$$A - \frac{A^*}{K} = \frac{2(K - 1)}{(K + 1)} \exp(-(k'_1 + k'_{-1})t) \quad (35)$$

Thus a plot of

$$\ln \left(\frac{1 - \frac{D/L}{K}}{1 + D/L}\right)$$

versus time will give a straight line with slope $-(k'_1 + k'_{-1})$ (Figure 40).

The ratio K is obtained from the equilibrium value of (DL + LD)/(DD + LL) for the diketopiperazine at long reaction times. After twenty minutes (Figure 31), this ratio reaches a constant value of 2.3 (Table 48).

When t is large the expression for A-A* (Equation 31) can be simplified to:

$$A-A^* = c_1 \exp\left(\frac{-\Sigma k + I}{2} \cdot t\right) \quad (36)$$

since the term $\exp\left(\frac{-\Sigma k - I}{2} \cdot t\right)$ approaches zero much faster than $\exp\left(\frac{-\Sigma k + I}{2} \cdot t\right)$.

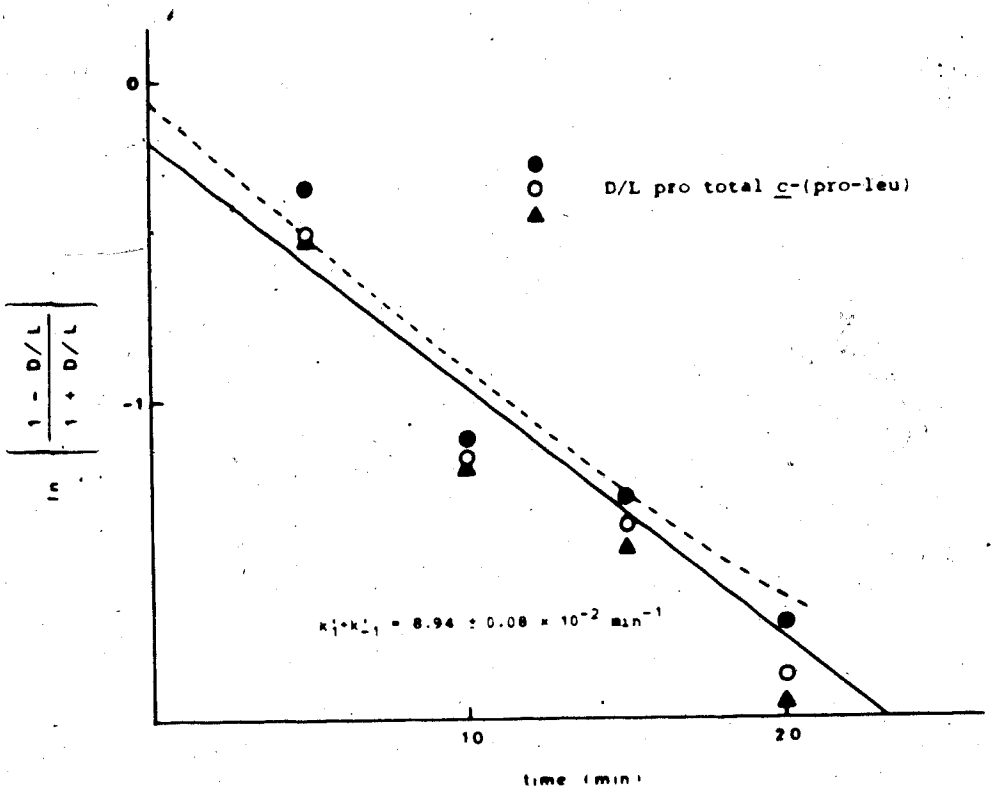


Figure 40. Racemization of the proline peptide unit in total c-(pro-leu) isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The closed circles represent the uncorrected data points. Otherwise, see Figure 30 for the interpretation of the solid and open symbols. The solid line was produced by linear regression analysis.

Table 48. Measurement of k_1'/k_{-1}' for c-(pro-leu) and c-(D-pro-leu) Isolated During Decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM

Time (min)	$\frac{DL + LD}{DD + LL}$[a]
20	1.63
30	2.17
40	2.36
60	2.25
90	2.34
1500[b]	2.23
3000[b]	2.36

a) Data after 40 minutes gives a value of 2.31 ± 0.08

b) From pl data 148.5°, pH 6.8, 9.01 mM

From the expression for I and with the slope of

$$\ln \left(\frac{1 - D/L}{1 + D/L} \right)$$

versus t at long reaction times (after the maximum asymmetric induction has occurred) is

$$m = \frac{-\Sigma k + I}{2}$$

then

$$k'_2 = \frac{-m(m + k'_1 + k'_{-1})}{1 + K + 4k'_1} \qquad (37)$$

k'_{-2} is obtained from:

$$k'_{-2} = k'_2 K \qquad (38)$$

The procedure for extracting the rate constants can be summarized:

(1) Obtain K by measuring (DL + LD)/(DD + LL) at long reaction times.

(2) Plot

$$\ln\left(\frac{1 - D/L}{1 + D/L}\right)$$

versus time for the peptide unit undergoing asymmetric induction to yield $k_1 + k_{-1}$

(3) Plot

$$\ln\left(\frac{1 - D/L}{1 + D/L}\right) \text{ versus time}$$

for the other racemizing centre or using the ratios DD/LL, DL/LD and obtain from the slope at long reaction times.

There are two criticisms of this method. The accuracy of $k'_1 + k'_{-1}$ is sensitive to the precision of the data and the number of points taken around time zero. Also, the determination of "around time zero" and "long reaction times" is somewhat arbitrary. A choice must be made between using a small number of data points and the precision of the individual data point. For example, using data at t = 5, 10 and 15 minutes will yield a value of $k'_1 + k'_{-1}$ which will be less than the true value and probably less than that obtained from the first two data points. It is also likely a better approximation of the true average slope.

k'_1 is calculated from $k'_1 + k'_{-1}$, K and the slope at long reaction times. k'_2 (and hence k'_{-2}) may be

sensitive to the value of $k'_1 + k'_{-1}$. Table 49 shows the effect of changing $k'_1 + k'_{-1}$ over an order of magnitude. Clearly the calculated value of k'_2 is not sensitive to $k'_1 + k'_{-1}$. This is a consequence of $k'_1 + k'_{-1}$ being much larger than k'_2. The expression for k'_2 can be simplified to

$$k'_2 = -m\frac{(1 + K)}{4K} \qquad (39)$$

This approach can be criticized for tacet assumption that the diketopiperazine appears only from the LL tetrapeptide. In reality, the tetrapeptide is approaching a 50:50 mixture of LL and DL diastereomers (assuming that leucine epimerization is slow). There is no detectable asymmetric induction in the tetrapeptide and proline is epimerizing quite rapidly. Consequently, the measured value of the slope near time zero must be considered to be giving a lower limit to the value of $k'_1 + k'_{-1}$.

Figures 41 and 42 show the racemization behaviour of the diketopiperazine isolated during the decomposition of plgg.

Figures 43 to 46 show the plots which yield the value of m (Equation 39). k'_2 and k'_{-2} are then calculated.

The upper straight lines in Figures 43 to 46 indicate the error in the extracted rate constants produced if the inappropriate plot is generated; i.e.

Table 49. Sensitivity of k_2' to $k_1' + k_{-1}'$ (Scheme 6)

k_1'	$k_1' + k_{-1}'$	k_2'[a,b]
0.0401	0.0575	0.001218
0.0802	0.1150	0.001211
0.1603	0.230	0.001209
0.401	0.575	0.001206

a) Calculated according to Equation 37, m = 3.86 × 10^{-3}

b) From Equation 39, $k_2' = 0.001205$

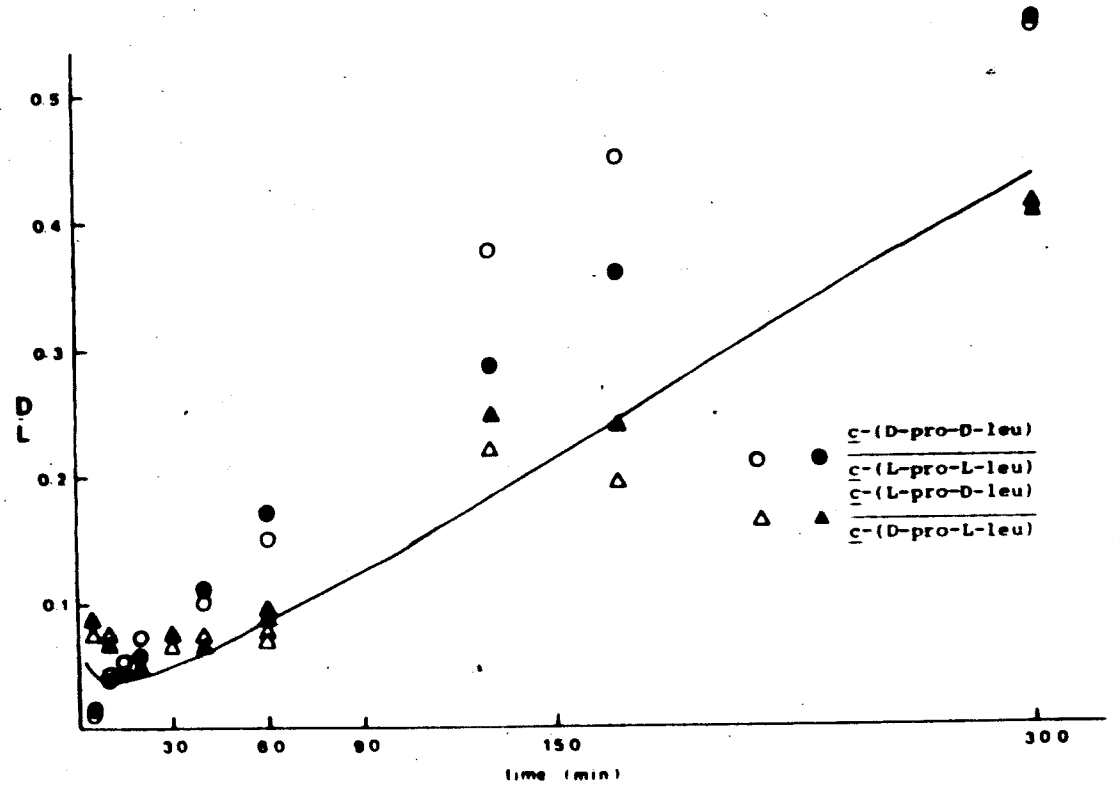


Figure 41. Racemization of the proline peptide unit in c-(pro-leu) isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line shows the computer simulation produced for c-(L-pro-D-leu)/c-(D-pro-L-leu) using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

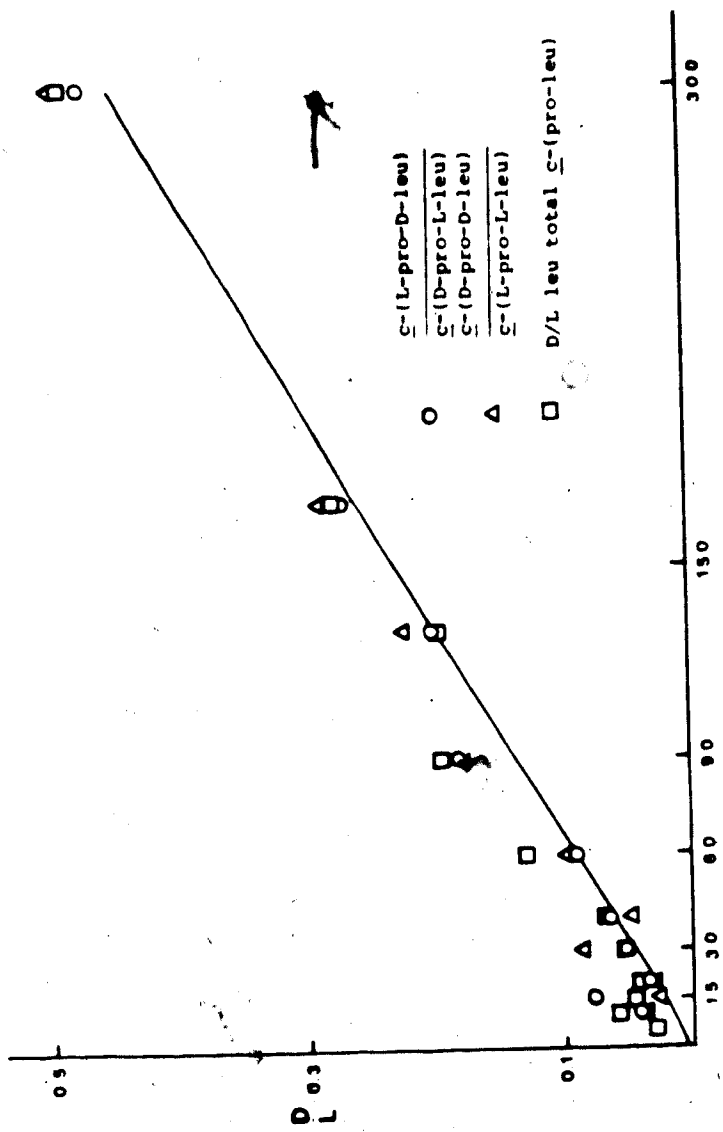


Figure 42. Racemization of the leucine peptide unit in c-(pro-leu) isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The solid line shows the computer simulation produced for c-(D-pro-D-leu)/c-(L-pro-L-leu) using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

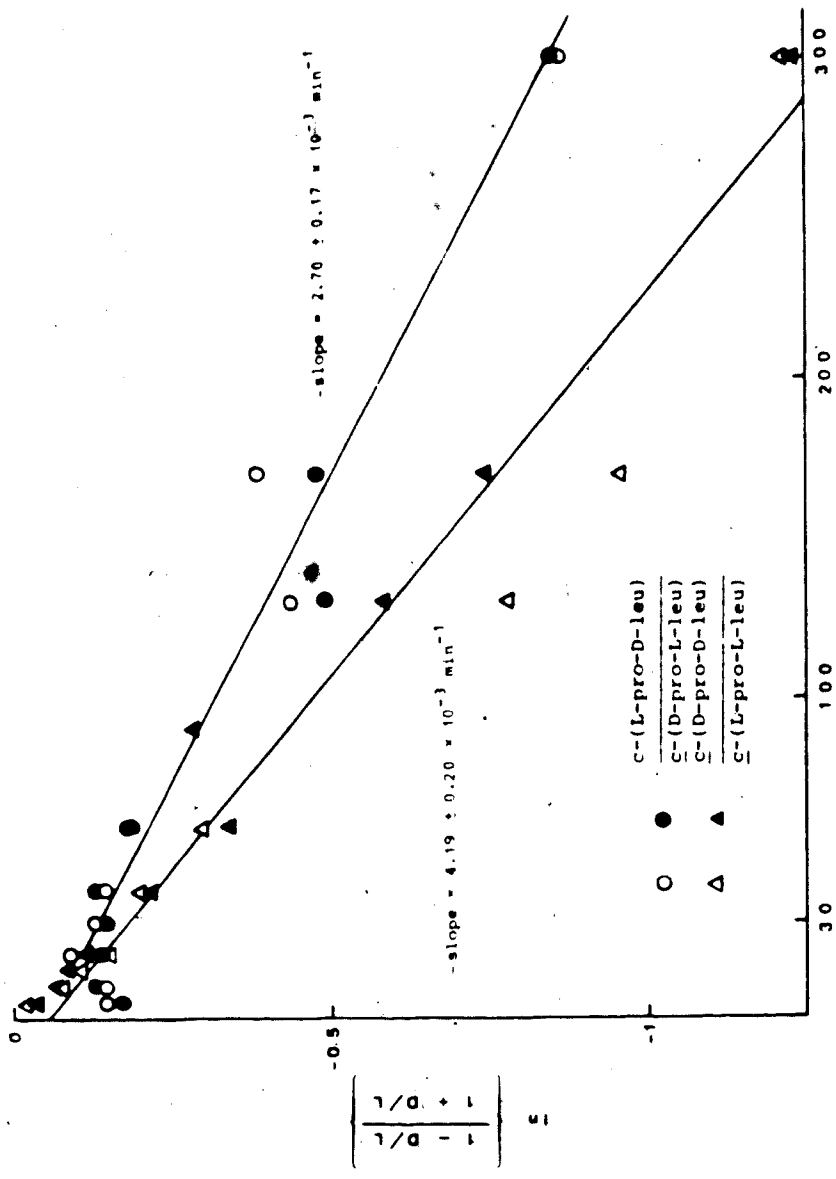


Figure 43. Racemization of the proline peptide unit in c-(pro-leu) isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid lines were produced by linear regression analysis. The upper solid line uses points after 20 minutes. For further information see text page 189.

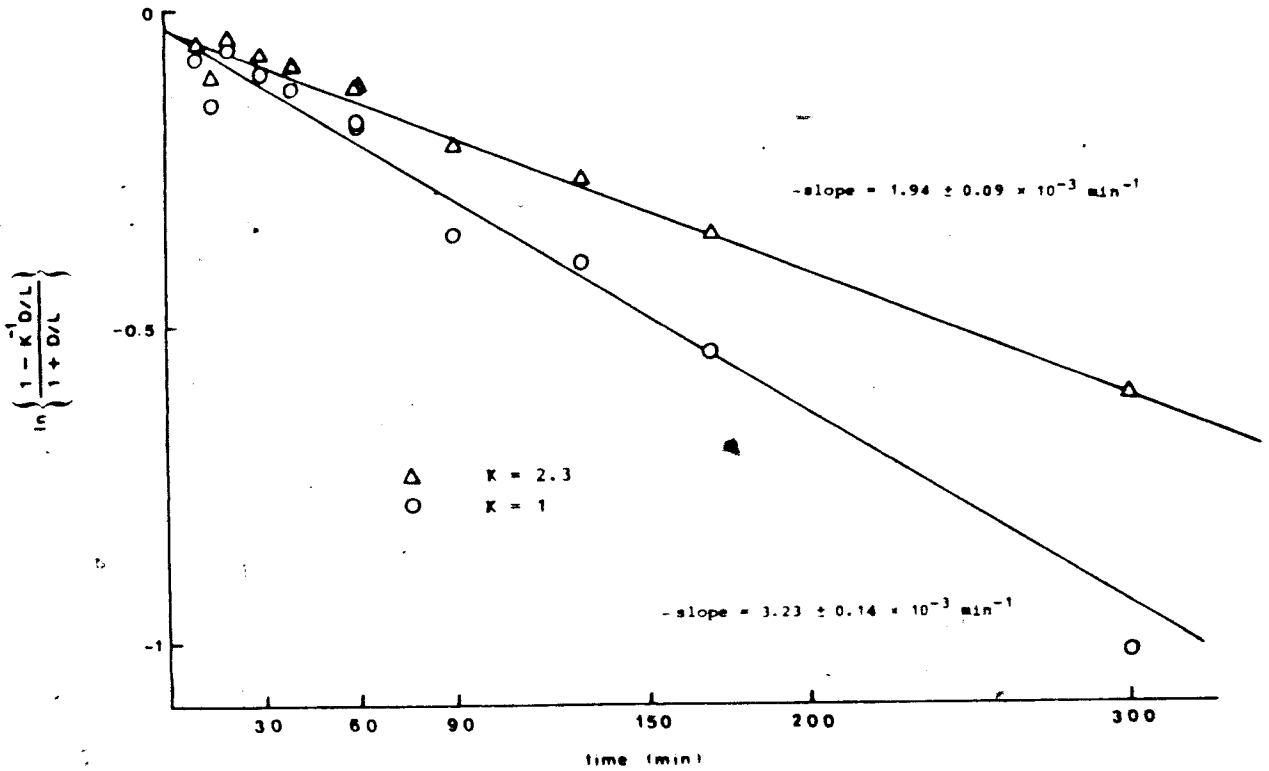


Figure 44. Racemization of the leucine in c-(pro-leu) isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The solid line was produced by linear regression analysis. This plot shows the inappropriate method (K = 2.3) and the appropriate method (K = 1) for extracting k_2'. For further information see text page 175-188.

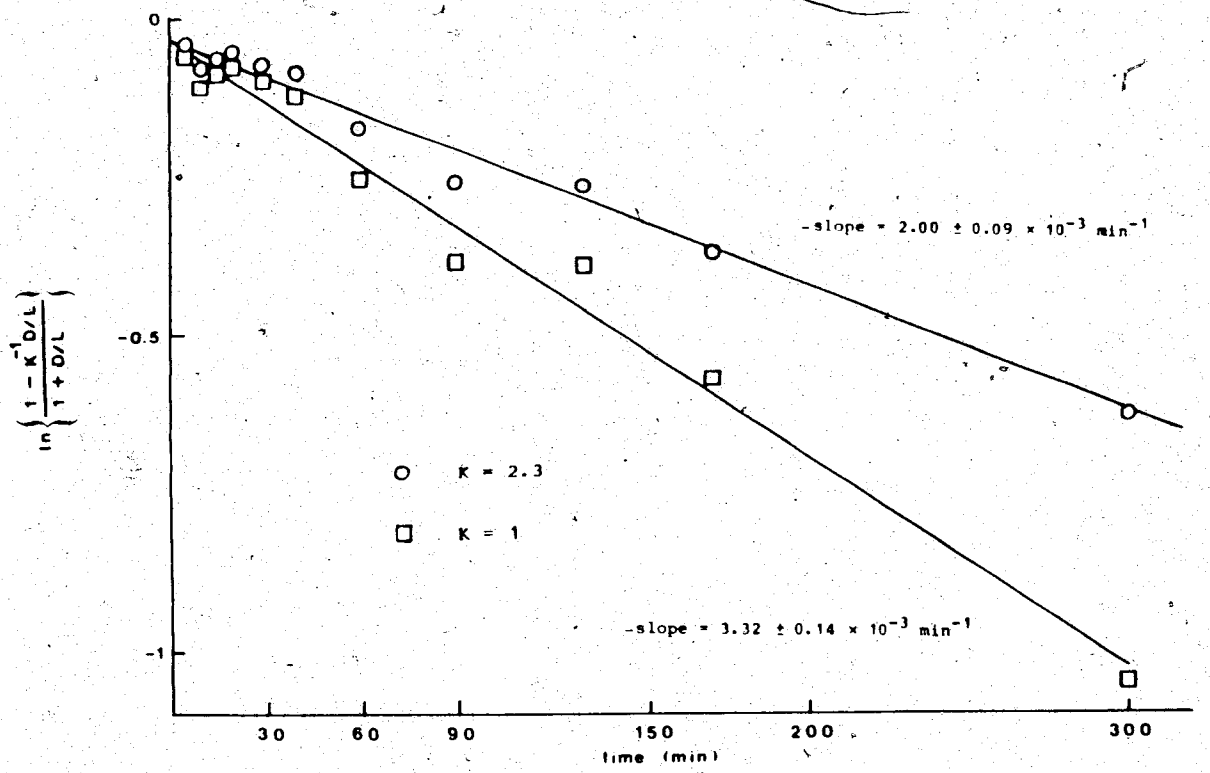


Figure 45. Racemization of the leucine peptide unit in the total c-(pro-leu) isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The solid lines were produced by linear regression analysis. This plot shows the inappropriate method (K = 2.3) and the appropriate method (K = 1) for extracting k_2'. For further information see text p. 175-188.

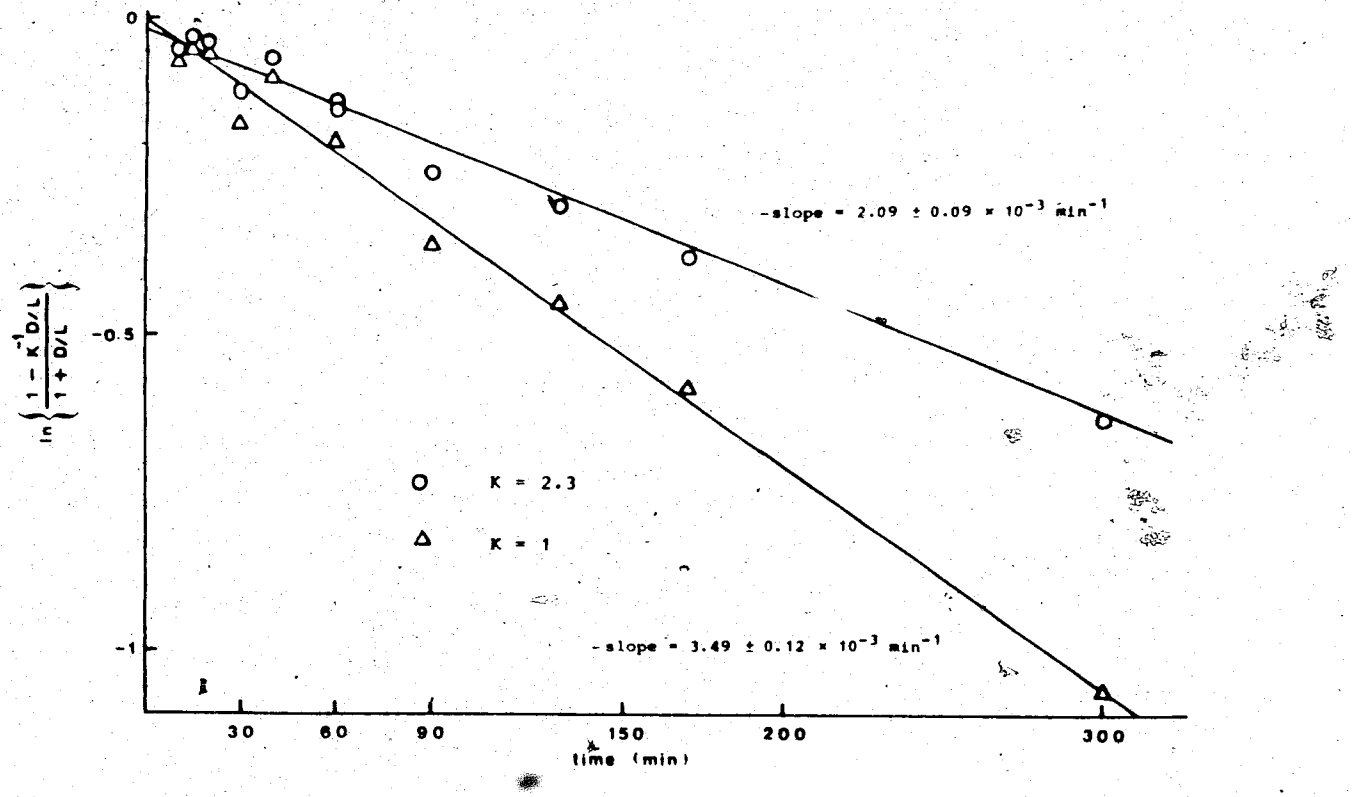


Figure 46. Racemization of the leucine peptide unit determined by measuring c-(D-pro-D-leu)/c-(L-pro-L-leu). This pair of enantiomers was isolated during the decomposition of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.45 mM. The solid line was produced by linear regression analysis. This plot shows the inappropriate method (K = 2.3) and the appropriate method (K = 1) for extracting k_2'. For further information see text p. 175-188.

$$\ln\left(\frac{1 - \frac{D/L}{K}}{1 + D/L}\right)$$ versus time.

This presumably should give slope $-(k'_2 + k'_{-2})$. According to the previous analysis, this is not the correct approach.

Figure 47 shows the racemization behaviour of the leucine peptide unit in the total hydrolysate. Data found in Table 47 and Table 50 was used to produce this plot. The rate constant extracted from the initial slope (<90 min) was $2.47 \times 10^{-5} s^{-1}$. The actual value of the rate constant for epimerization of leucine in the tetrapeptide is $8.35 \times 10^{-6}\ s^{-1}$. Clearly the rate constant from the initial slope is not representative of the tetrapeptide.

The experimentally determined rate constants for plgg are given in Table 51. The consistency of the kinetic model is tested by substituting the experimental values in the computer simulation program, TLSA (Appendix 4). Overall, the agreement is quite good. As expected, the major discrepancy is associated with rate constant for epimerization of the proline peptide unit ($k'_1 + k'_{-1}$) in the diketopiperazine. The computer simulations using the experimentally determined values are superimposed on the experimental data in Figures 31, 33, 37-41 and 47.

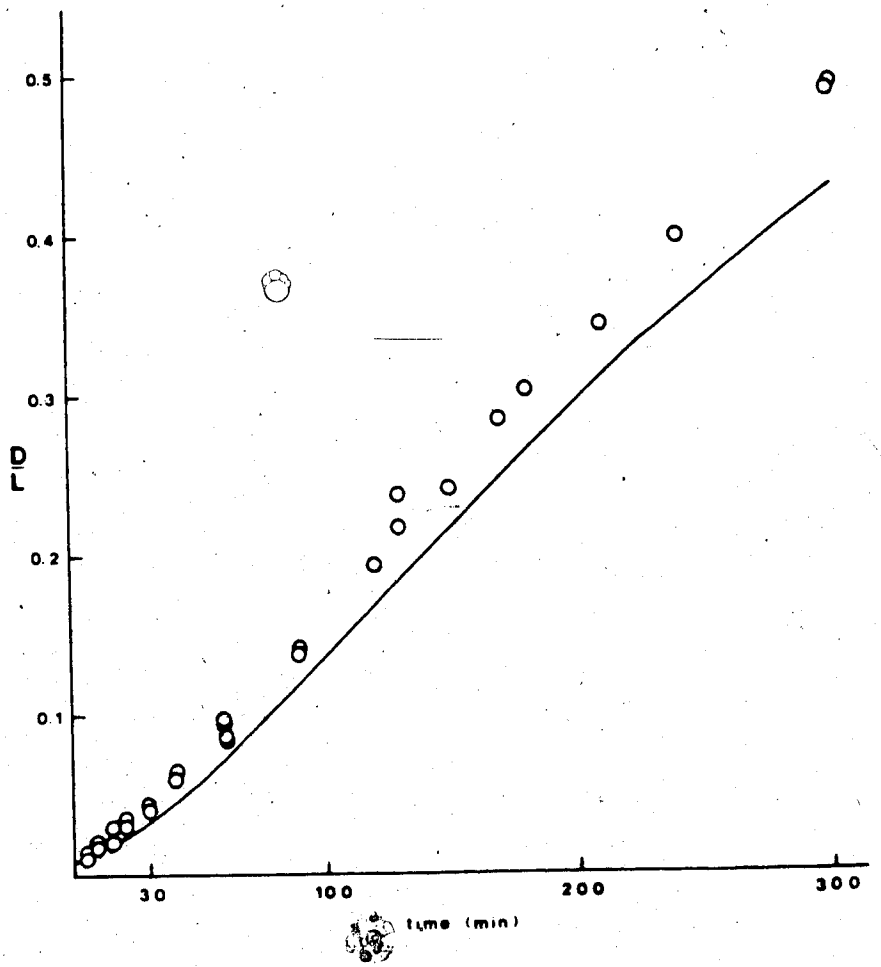


Figure 47. Racemization of the leucine peptide unit in the total hydrolysate of L-pro-L-leu-gly-gly, 148.5°, pH 6.8. The solid line shows the computer simulation produced using the experimentally derived rate constants (Table 51). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

Table 50. Racemization[a] of the Leucine Peptide Unit in the Total Hydrolysate of L-pro-L-leu-gly-gly, 148.5°, pH 6.8, 8.83 mM

Time	D/L
5	0.0096
10	0.015
15	0.020
20	0.026
30	0.043
40	0.063
60	0.095
60	0.086
90	0.142
120	0.194
150	0.238
150	0.241
180	0.302
210	0.342
240	0.397
300	0.489

a) See footnote a), Table 44

Table 51. Experimentally Determined Rate Constants (10^4 s^{-1}) for L-pro-L-leu-gly-gly, (Scheme 9), 148.5°, pH 6.8, 8.45 mM

$k_1 = k_{-1}$ [a]	$k_2 = k_{-2}$	$k_5 = k_6$	k'_1	k'_{-1}	k'_2 [b]	k'_{-2}
13.4 ± 0.8 [c]	0.0835 ± 0.0155	4.03 ± 0.30	10.4 ± 1.0	4.52 ± 0.450	0.202 ± 0.015	0.463 ± 0.036

a) k_1, k_{-1}, k_2 and k_{-2} refer to the tetrapeptide; k'_1, k'_{-1}, k'_2 and k'_{-2} refer to the diketopiperazine

b) From m = $(3.36 \pm 0.26) \times 10^{-3}$ min^{-1} from leucine epimerization, Equation 39

$(3.39 \pm 0.67) \times 10^{-3}$ min^{-1} including proline epimerization

c) The rate constants are quoted as averages of several measurements. The error is reported at the 95% confidence level.

Extraction of $k'_1 + k'_{-1}$ from simulated kinetic runs using rate constants similar to the experimentally determined values, show that the extracted value of $k_1 + k_{-1}$ will be lower than the actual value.

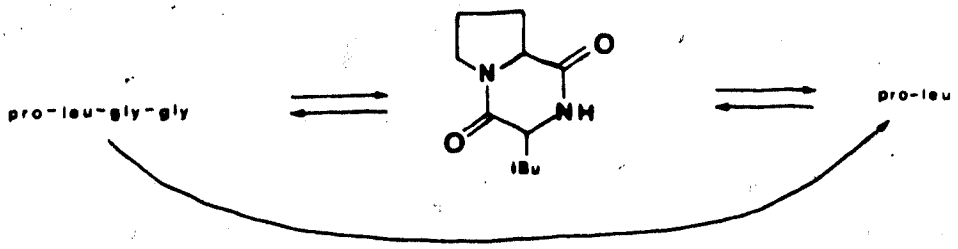
The size of this discrepancy depends on the magnitudes of the rate constants and the number of points used in the evaluation of the rate constant. Substituting the experimental rate constants into the simulation program and calculating the D/L ratio for the proline peptide unit in the diketopiperazine at 5, 10, 15 and 20 minutes gives a value of $k'_1 + k'_{-1}$ of 0.0746 min^{-1}. The input into the program was 0.0894 min^{-1}. Using only the first two points yields a value of 0.0856 min^{-1}. This shows that in principle it is possible to extract $k'_1 + k'_{-1}$ in the region where the leucine peptide unit is not yet significantly racemized (see also Figure 40).

The maximum D/L ratio predicted by theory using the experimentally determined rate constants is about 1.86 (at 40 minutes) (Figure 30). If the decomposition from the tetrapeptide was instantaneous, i.e. the starting condition was pure c-(pro-leu), the maximum D/L ratio predicted is 1.92 (at 50 minutes). Therefore an underestimate of the rate constant k'_1 or k'_{-1} cannot be completely responsible for the discrepancy in $k'_1 + k'_{-1}$.

There are two potential sources for the error. First, the data may not be sufficiently precise to allow a proper measurement of $k'_1 + k'_{-1}$. When the value of k'_1 is increased to 0.1869 min^{-1} (three times the experimental value), the predicted maximum D/L ratio is 2.11 (at 20 minutes). The scatter in the experimental data does not seem to be great enough to encompass such a large error in the extracted rate constant. An alternative explanation appears necessary.

The second potential source of error lies with the model used to extract the rate constant (Scheme 9). It does not account for more than the presence of the tetrapeptide and the diketopiperazine. Other species in the mixture which yield excess D-proline could account for the discrepancy in $k'_1 + k'_{-1}$ and maximum D/L ratio observed for the diketopiperazine and the total hydrolysate.

The dipeptide pro-leu was thought to be a logical candidate. It could result from hydrolysis of the tetrapeptide or from hydrolysis of the diketopiperazine (Scheme 10). This latter process might be reversible. To explore this potential complication, a time study of the racemization and hydrolysis behaviour of pro-leu was undertaken.



SCHEME 10

Racemization and Decomposition of L-pro-L-leu

Figure 48 shows the racemization of the proline peptide units in the total hydrolysate and in the diketopiperazine isolated from the heating of pro-leu (pl). As in the heating of plgg, excess D-proline is observed. The D/L ratios for both leucine and proline approach unity at long reaction times (Tables 52-55).*

The reversibility of the cyclization of pro-leu to c-(pro-leu) was established by subjecting the isolated diketopiperazine and dipeptide diastereomers to the

*At 3000 min, D/L(leu) = 0.94 and D/L(pro) = 1.04.

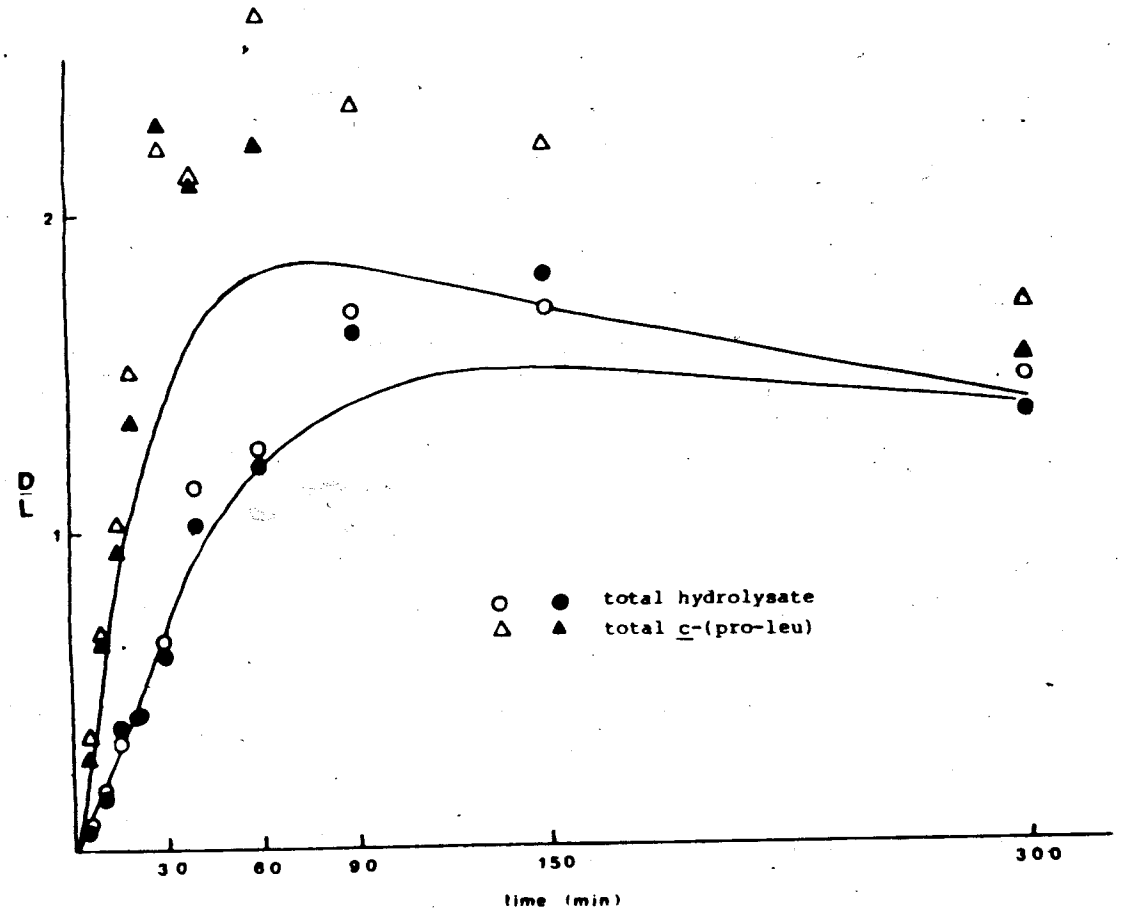


Figure 48. Racemization of the proline peptide unit in the total hydrolysate and total c-(pro-leu) isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. See Figure 30 for the interpretation of the solid and open symbols. The upper and lower lines show the computer simulation produced for c-(D-pro-L-leu) + c-(D-pro-D-leu)/c-(L-pro-D-leu) + c-(L-pro-L-leu) and the total hydrolysate respectively using the experimentally derived rate constants (Table 57). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

Table 52. Racemization[a] (D/L) of the Leucine Peptide Unit in Components from the Decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM

Time (min)	Component[b] pl	Dpl	total dipeptide	cDpl	cpl	total DKP	total hydrolysate
5	0.010	0.017	0.010	0.064	0.016	0.033	0.010
10	0.010	0.011	0.012	0.043	0.015	0.025	0.011
15	0.012	0.019	0.012	0.040	0.024	0.030	0.015
20	0.011	0.010	0.018	0.040	0.[illegible]	0.031	0.016
30	0.012	0.013	0.014	0.045	0.040	0.049	0.021
40	0.018	0.016	0.012	0.054	0.048	0.062	0.032
60	0.011	0.013	0.020	0.078	0.086	0.083	0.043
90	0.013	0.014	0.051	0.126	0.13	0.134	0.078
150	0.029	0.027	0.025	0.221	0.246	0.237	0.187
300	0.079	0.191	0.106	0.506	0.533	0.512	0.442
1500	0.376	0.813	0.312	0.997	0.984	0.994	0.908
3000	0.601	0.917	0.617	0.994	0.937	0.997	0.937

a) See footnote a), Table 44

b) For abbreviations see footnote c), Table 43

Table 53. Racemization[a,b] (D/L) of the Proline Peptide Unit in Components from the Decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM

	Component[c]						
Time (min)	pl	Dpl	total dipeptide	cDpl	cpl	total DKP	total hydrolysate
5	0.008	8.30	0.050	3.47	0.026	0.264	0.049
10	0.009	8.43	0.127	8.11	0.027	0.593	0.139
15	0.011	9.65	0.218	11.23	0.067	0.862	0.259
20	0.010	10.47	0.259	10.77	0.062	1.201	0.385
30	0.008	24.9	0.375	9.08	0.067	2.002	0.559
40	0.014	12.9	0.451	9.84	0.047	1.830	0.950
60	0.022	23.3	0.630	8.32	0.124	1.943	1.115
90	0.013	26.5	0.739	5.95	0.213	2.220	1.432
150	0.044	11.11	0.861	3.63	0.345	1.968	1.584
300	0.189	3.98	0.874	1.645	0.521	1.358	1.306
1500	0.578	0.921	0.602	0.814	0.801	0.964	0.865
3000	0.816	0.920	0.712	0.858	0.754	1.053	0.849

a) See footnote a), Table 44

b) Values are not corrected for nonresolved N-PFPA isopropyl esters of D-proline and L-proline

c) For abbreviations see footnote c), Table 43

Table 54. Racemization[a,b] (D/L) of the Proline Peptide Unit in Components from the Decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM

	Component[c]						
Time (min)	pl	Dpl	total dipeptide	cDpl	cpl	total DKP	total hydrolysate
5	0.012	10.10	0.064	3.612	0.030	0.353	0.076
10	0.010	11.7	0.184	13.462	0.036	0.675	0.177
15	0.011	17.88	0.246	16.716	0.082	1.024	0.329
20	0.014	16.52	0.357	19.959	0.079	1.499	0.411
30	0.016	30.88	0.518	13.692	0.094	2.214	0.654
40	0.024	24.64	0.617	15.508	0.070	2.126	1.139
60	0.011	40.52	0.853	13.005	0.146	2.622	1.254
90	0.024	37.33	1.000	8.191	0.283	2.339	1.686
150	0.062	21.00	1.092	4.776	0.403	2.229	1.683
300	0.235	4.115	1.101	1.862	0.557	1.695	1.343
1500	0.774	0.971	0.704	1.116	0.932	1.068	1.068
3000	0.856	0.996	0.902	1.176	0.901	1.116	1.039

a) See footnote a), Table 44

b) Values are corrected for nonresolved N-PFPA isopropyl esters of D-proline and L-proline by comparison with a D/L proline standard with approximately the same D/L ratio as the sample. The standard was injected immediately upon completion of the chromatographic run of the sample.

c) For abbreviations see footnote c), Table 43

Table 55. Racemization[a,b] (D/L) of the Proline Peptide Unit in Components from the Decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM

Time (min)	Component[c] pl	Dpl	total dipeptide	cDpl	cpl	total DKP	total hydrolysate
5	0.011	11.29	0.055	4.94	0.028	0.285	0.054
10	0.012	11.46	0.138	11.04	0.031	0.640	0.151
15	0.014	13.07	0.236	15.14	0.073	0.930	0.280
20	0.013	14.14	0.280	14.54	0.068	1.339	0.416
30	0.011	33.1	0.405	12.32	0.073	2.283	0.603
40	0.018	17.38	0.487	13.31	0.052	2.080	1.025
60	0.026	31.00	0.680	11.32	0.134	2.214	1.237
90	0.016	35.21	0.797	8.20	0.230	2.540	1.611
150	0.49	14.98	0.929	5.15	0.373	2.243	1.790
300	0.201	5.61	0.943	2.55	0.562	1.534	1.463

a) See footnote a), Table 44

b) Values are corrected for nonresolved N-PFPA isopropyl esters of D-proline and L-proline using D/L calibration curve (Appendix 6)

c) For abbreviations see footnote c), Table 43

reaction conditions.* However, the equilibrium constant favours the diketopiperazine:**

$$\frac{[cDpl]\ [cpl]}{[Dpl]\ \ [pl]} = 7.4$$

A first approximation of the eight independent rate constants for this system can be extracted as before assuming that the cyclization of pl to cpl and cDpl is irreversible (Scheme 9).

Since the diketopiperazine racemizes faster than the dipeptide and the cyclization is reversible, the apparent rate of racemization in the dipeptide will increase with time. This effect is most noticeable in the racemization of leucine in the dipeptide (Figure 49). This is a result of the large difference between the rates of leucine epimerization for the diketopiperazine and the dipeptide.

Figure 50 shows a plot of concentration versus time for the diketopiperazine and dipeptide diastereomers using data from Table 56. The rate constant for cyclization was determined in Figures 51 and 52 assuming irreversible formation of the diketopiperazine.

---

*After 30 min at 148.50°. When each of the four components is subjected to this treatment, the other three are observed.

**Calculated by averaging data at 1500 and 3000 minutes.

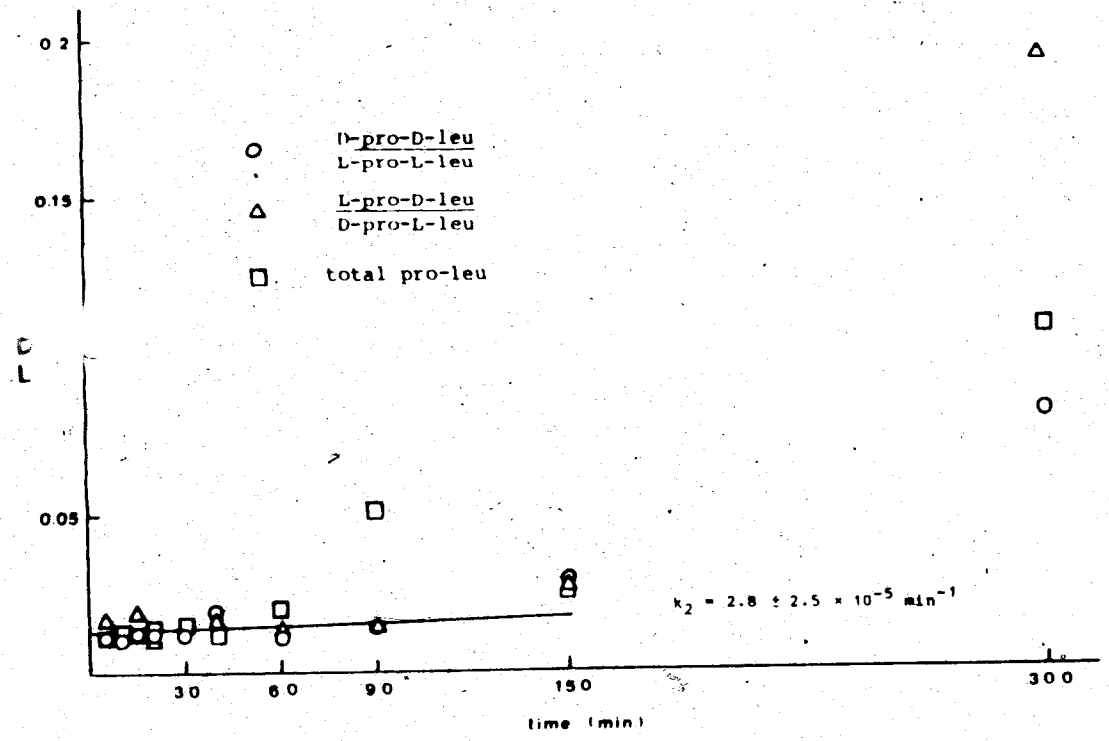


Figure 49. Racemization of the leucine peptide unit in pro-leu isolated during the decomposition of L-pro-L-Leu, 148.5°, pH 6.8, 9.01 mM. The solid line was produced by linear regression analysis using the data points from 5 to 150 minutes. The final points were not used because of the potential for ring opening of the diketopiperazine.

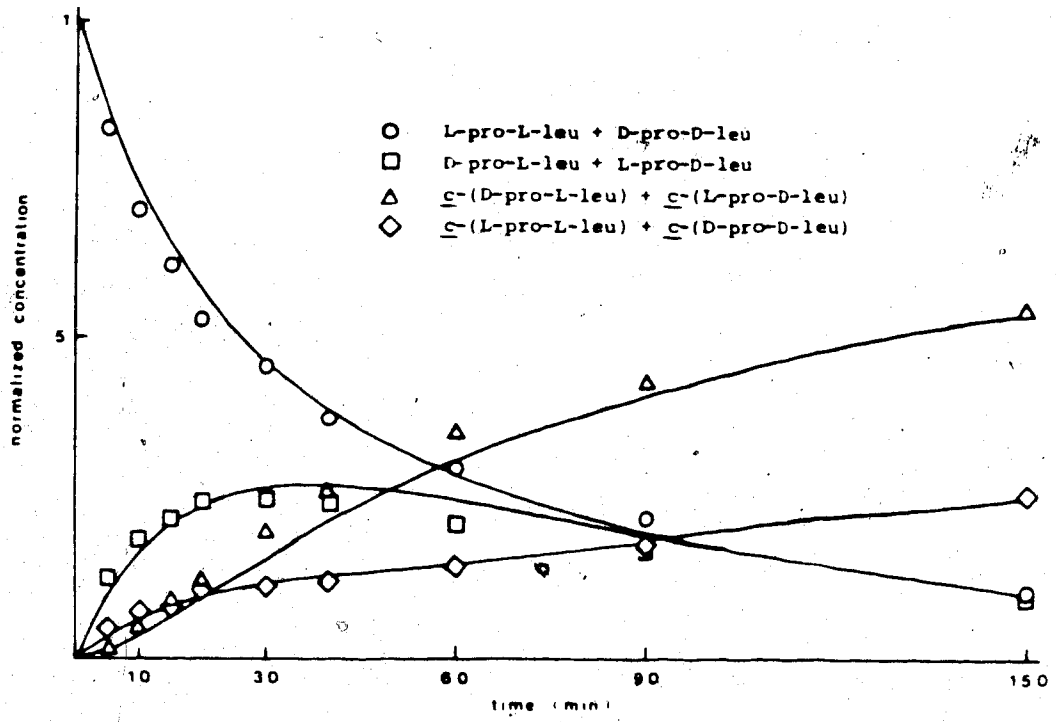


Figure 50. Decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. The solid lines show the computer simulation produced using the experimentally derived rate constants (Table 57). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

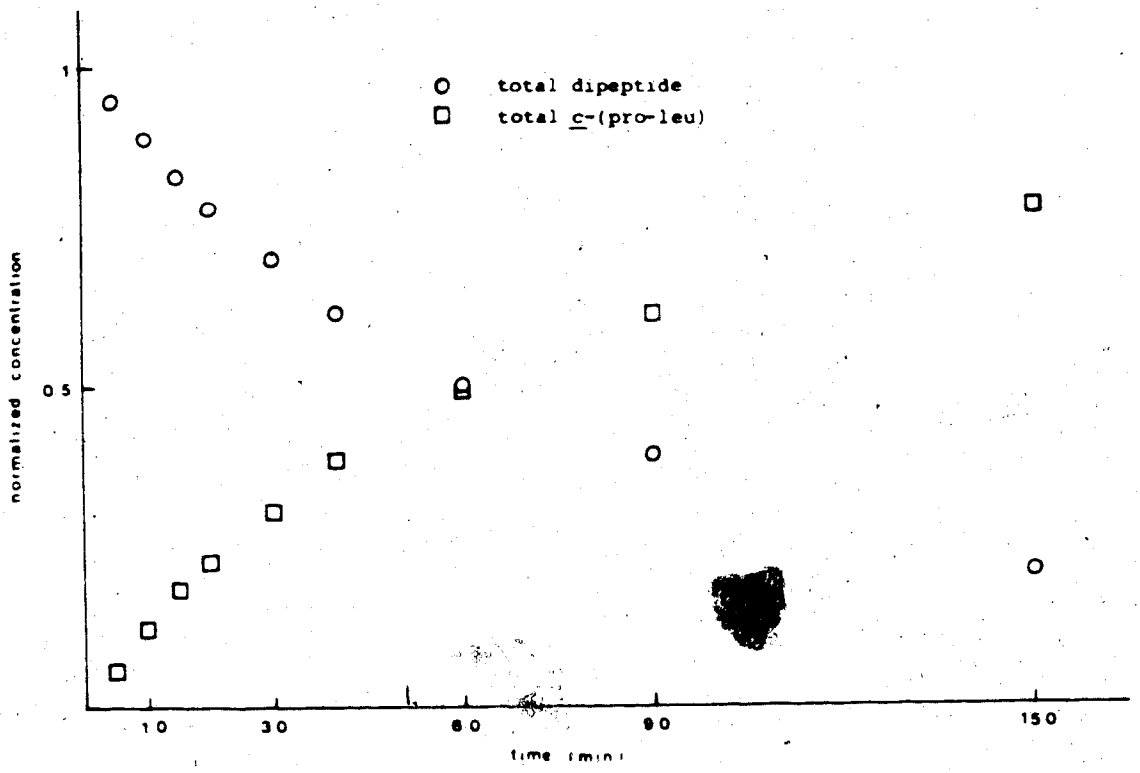


Figure 51. Decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM.

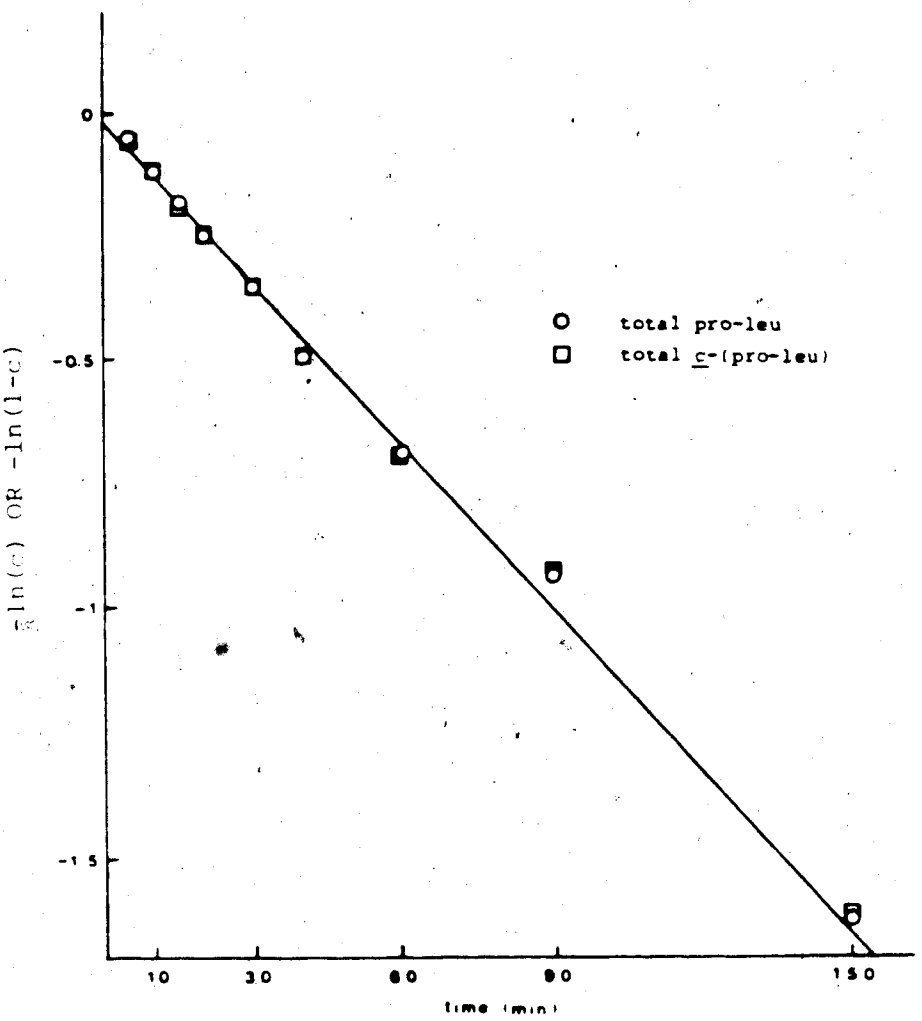


Figure 52. Decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. For pro-leu the ordinate is -ln(c) where c is the normalized concentration of total dipeptide. For c-(pro-leu) the ordinate is -ln(1-c) where c is the normalized concentration of total diketopiperazine. The solid line is produced by regression analysis.

Table 56. Decomposition[a] of L-pro-L-leu, 148.5°, pH 6.8[b], 9.01 mM

Time (min)	pl	Dpl	cDpl	cpl
5	0.828	0.120	0.012	0.044
10	0.699	0.184	0.047	0.072
15	0.613	0.216	0.089	0.089
20	0.527	0.247	0.121	0.105
30	0.456	0.242	0.196	0.107
40	0.374	0.240	0.266	0.116
60	0.295	0.207	0.352	0.144
90	0.219	0.173	0.428	0.178
150	0.105	0.091	0.546	0.255
300	0.032	0.037	0.640	0.291
1500	0.079	0.034	0.614	0.275
3000	0.072	0.044	0.622	0.262

a) Measured by the uv absorption at 230 nm of the eluate from the HPLC separation

b) 0.025 M KH_2PO_4, 0.025 M Na_2HPO_4

c) For abbreviations see footnote c), Table 43

Figures 49 and 53 show how k_1 and k_2 were extracted. The rate constant for epimerization of the leucine peptide unit in pro-leu (Figure 49) is obtained from the initial slope. This procedure produces an upper limit of the true value.

The epimerization rate constants for c-(pro-leu) were obtained by the same method used in the decomposition of pro-leu-gly-gly. The plot used to extract $k'_1 + k'_{-1}$ is shown in Figure 54. Figures 55, 56 and 57 show plots of D/L versus time used for extraction of k'_2. Figures 58, 59 and 60 show the canonical plots for Figures 55, 56 and 57, respectively. The experimentally determined rate constants are given in Table 57.

A comparison of the rate of racemization of the leucine peptide in the total hydrolysate (8.5×10^{-6} s^{-1}) with that of the dipeptide (4.7×10^{-7} s^{-1}) again reveals a lack of correlation. The evidence indicates that mechanistically meaningful rate constants cannot be extracted from the total hydrolysis mixture.

Besides the excess D-proline produced from the heating of pl, three other features stand out in Tables 51 and 57:

(1) The rate of cyclization (dehydration) of pl to the corresponding diketopiperazine is less than the rate of production of the diketopiperazine from the plgg.

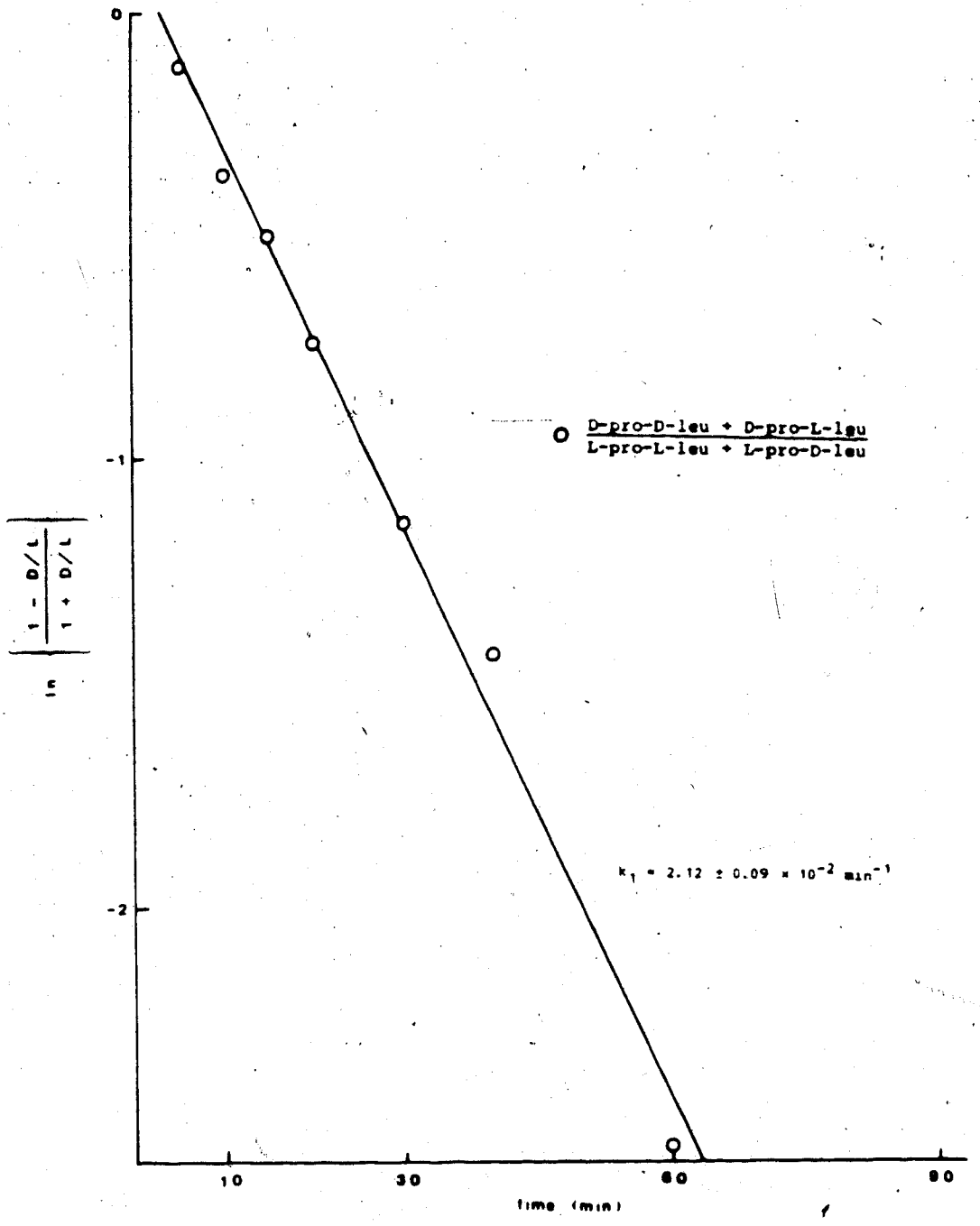


Figure 53. Racemization of the proline peptide unit in total pro-leu isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line was produced by linear regression analysis.

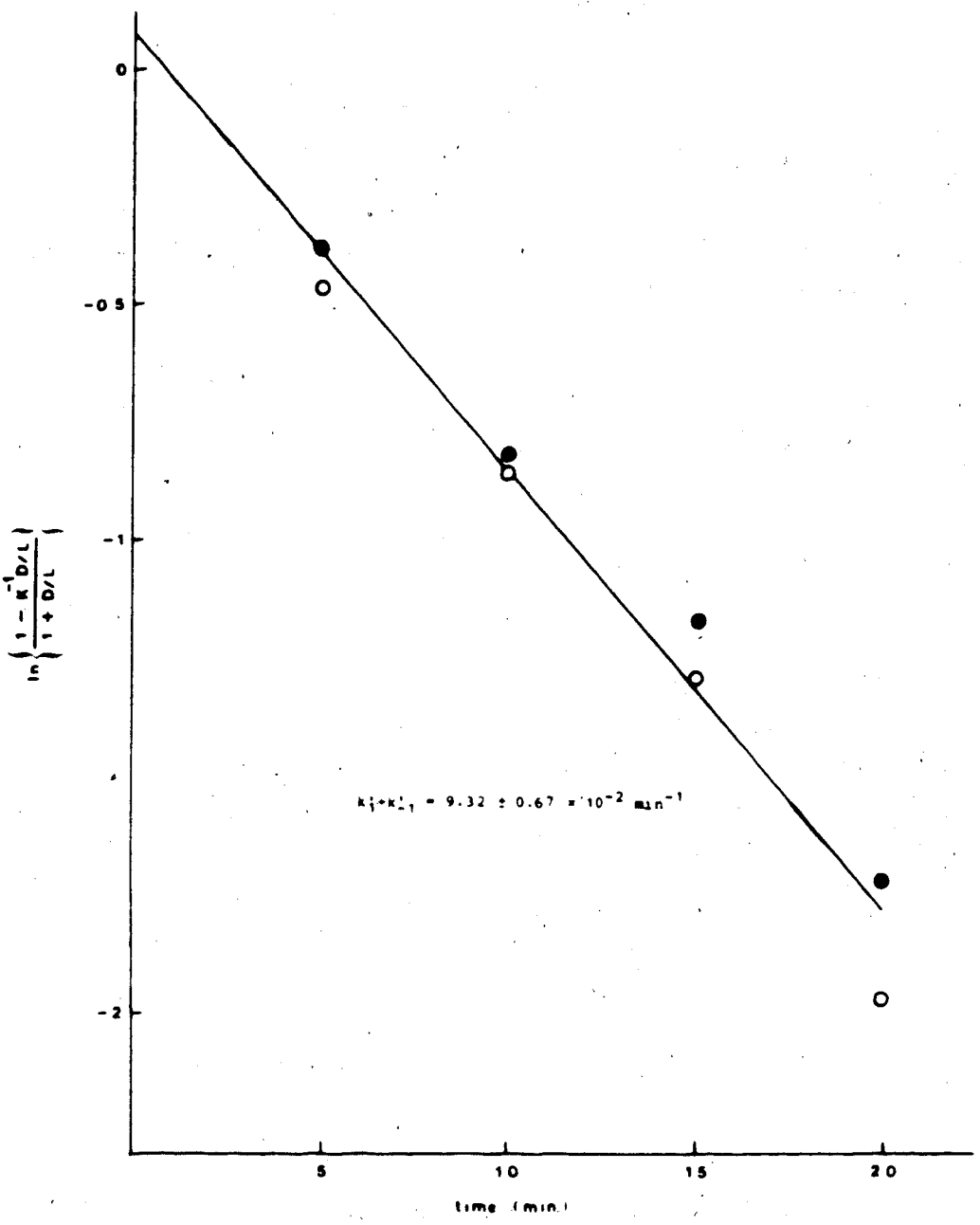


Figure 54. Racemization of the proline peptide unit in total diketopiperazine isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line was produced by linear regression analysis.

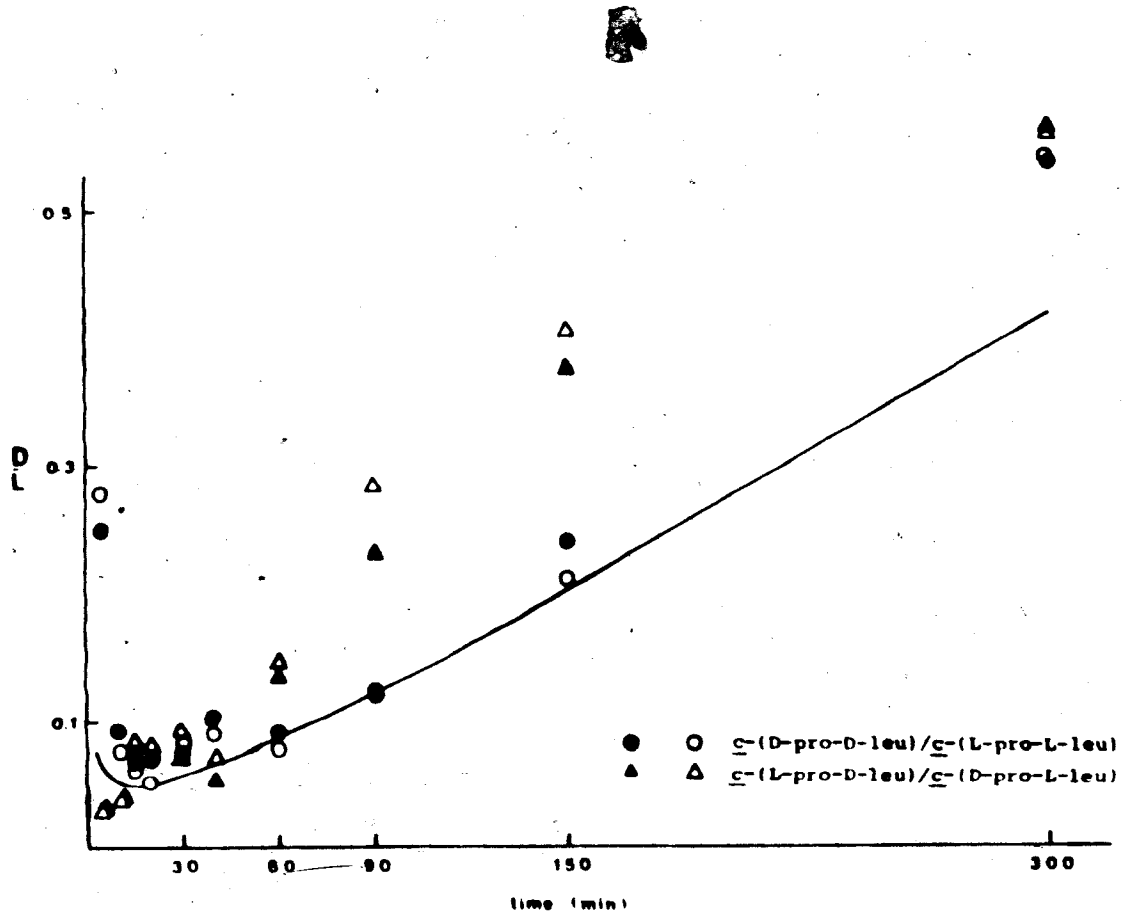


Figure 55. Racemization of the proline peptide unit in c-(pro-leu) isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line shows the computer simulation produced for c-(L-pro-D-leu)/c-(D-pro-L-leu) using the experimentally derived rate constants (Table 57). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9. The points marked by an asterisk arise from the initial amounts of the DL diastereomers.

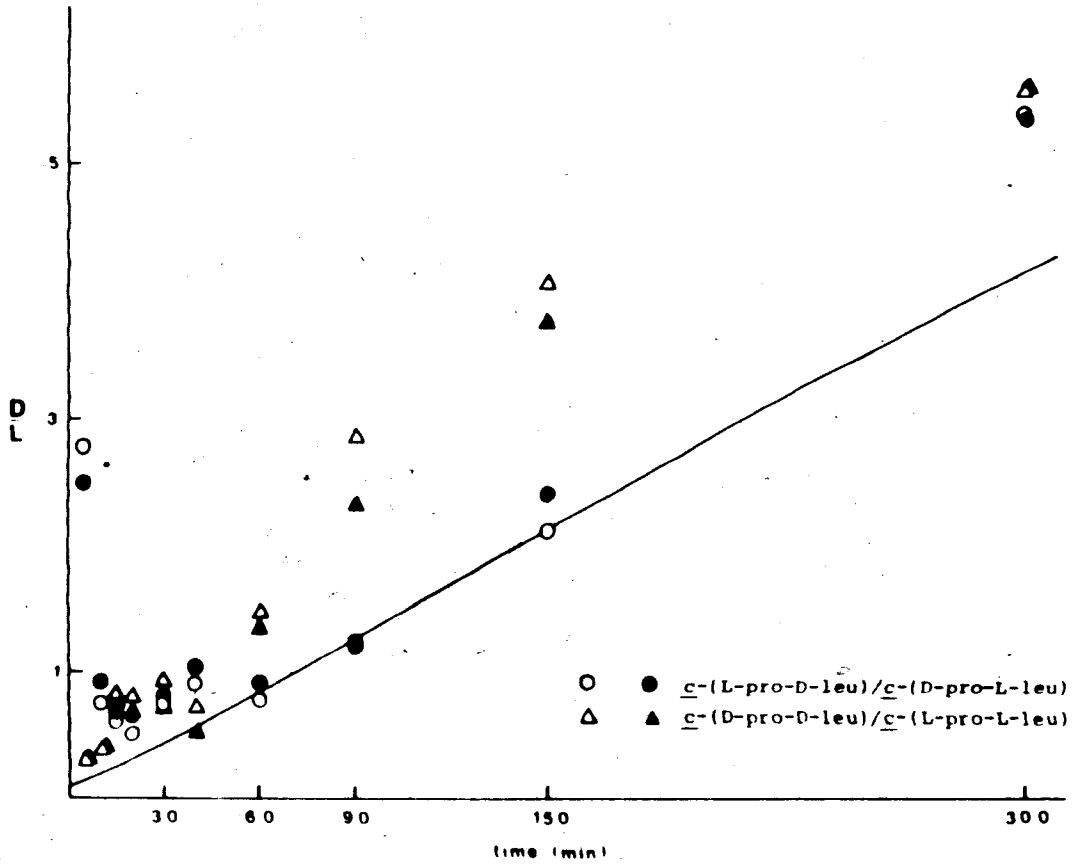


Figure 56. Racemization of the leucine peptide unit in c-(pro-leu) isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. The solid line shows the computer simulation produced for c-(D-pro-D-leu)/c-(L-pro-L-leu) using the experimentally derived rate constants (Table 57). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9.

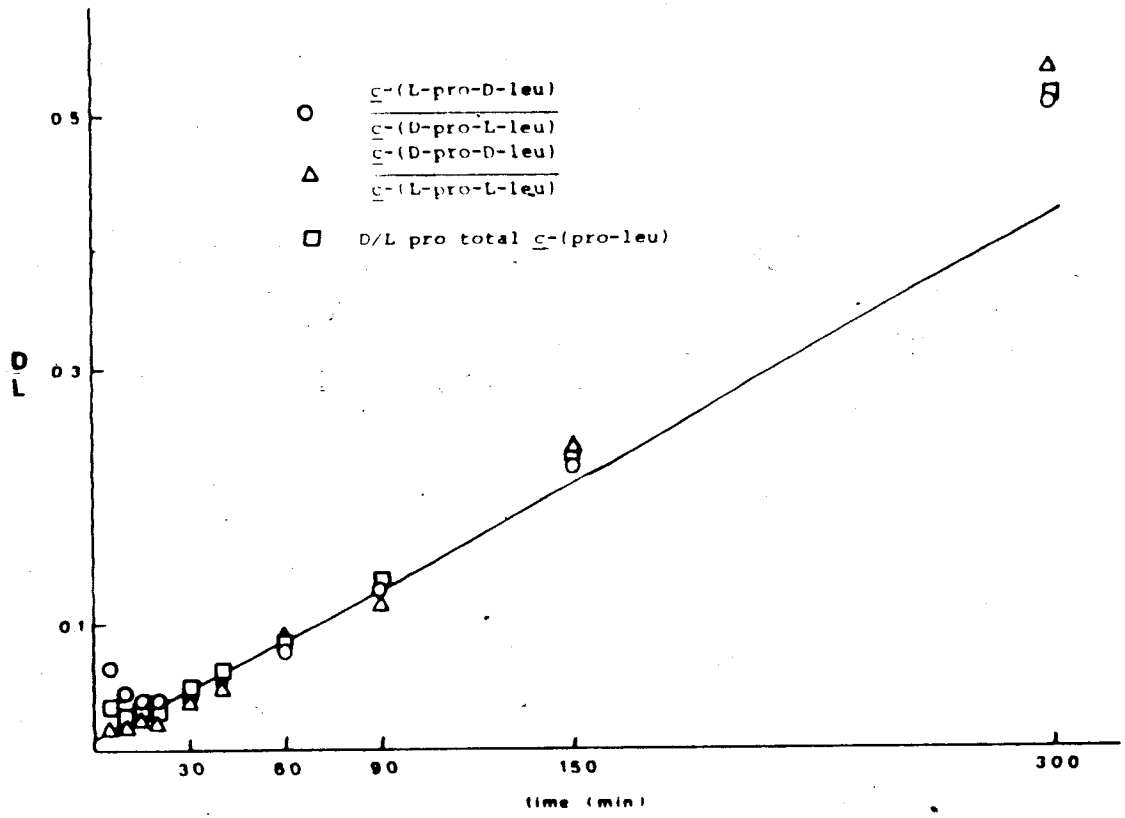


Figure 57. Racemization of the proline peptide unit in c-(pro-leu) isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line shows the computer simulation produced for c-(D-pro-D-leu)/c-(L-pro-L-leu) using the experimentally derived rate constants (Table 57). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9. For an explanation of the data points marked by an asterisk see Figure 55.

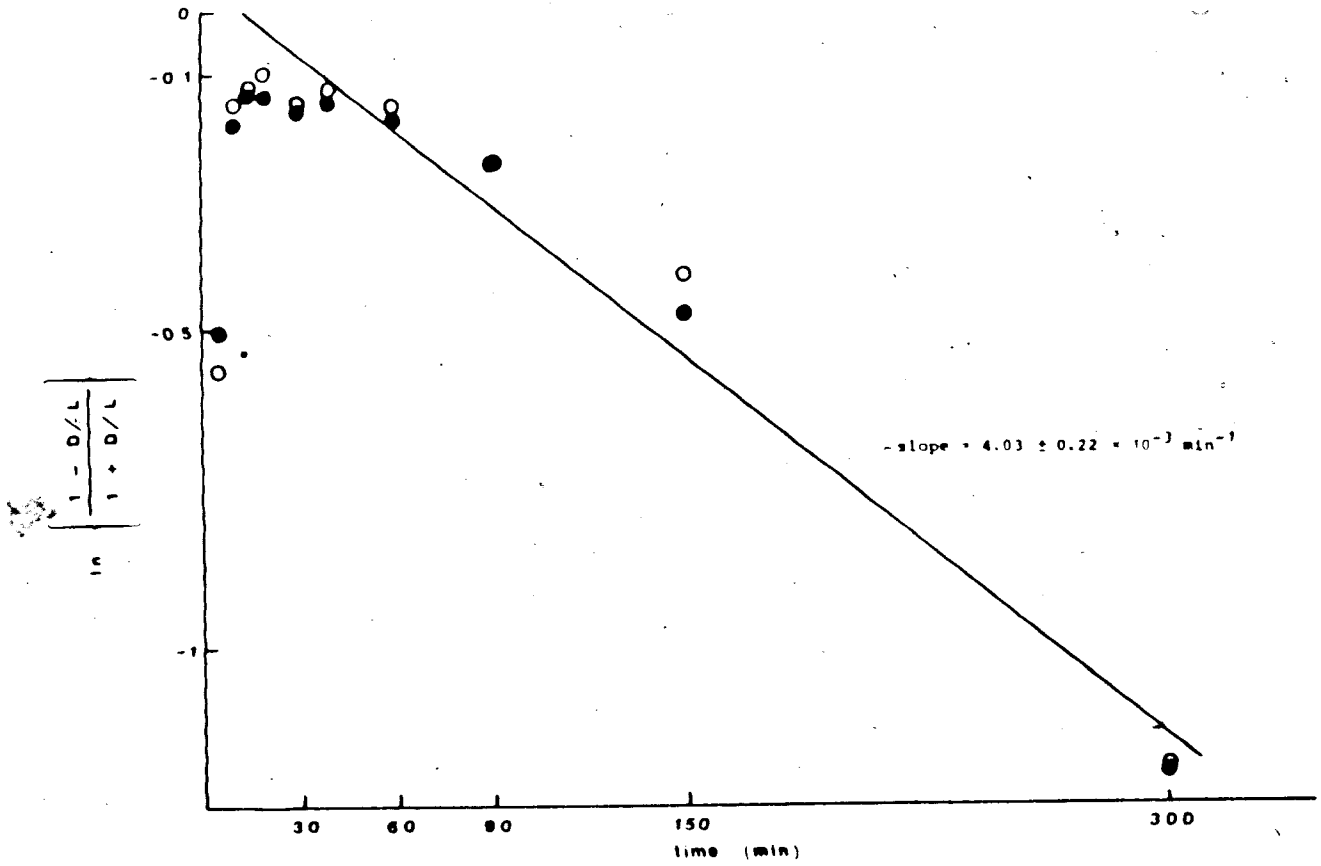


Figure 58. Racemization of the proline peptide unit determined by measuring c-(L-pro-D-leu)/c-(D-pro-L-leu). This pair of enantiomers was isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line was produced by linear regression analysis of the data points after 20 minutes.

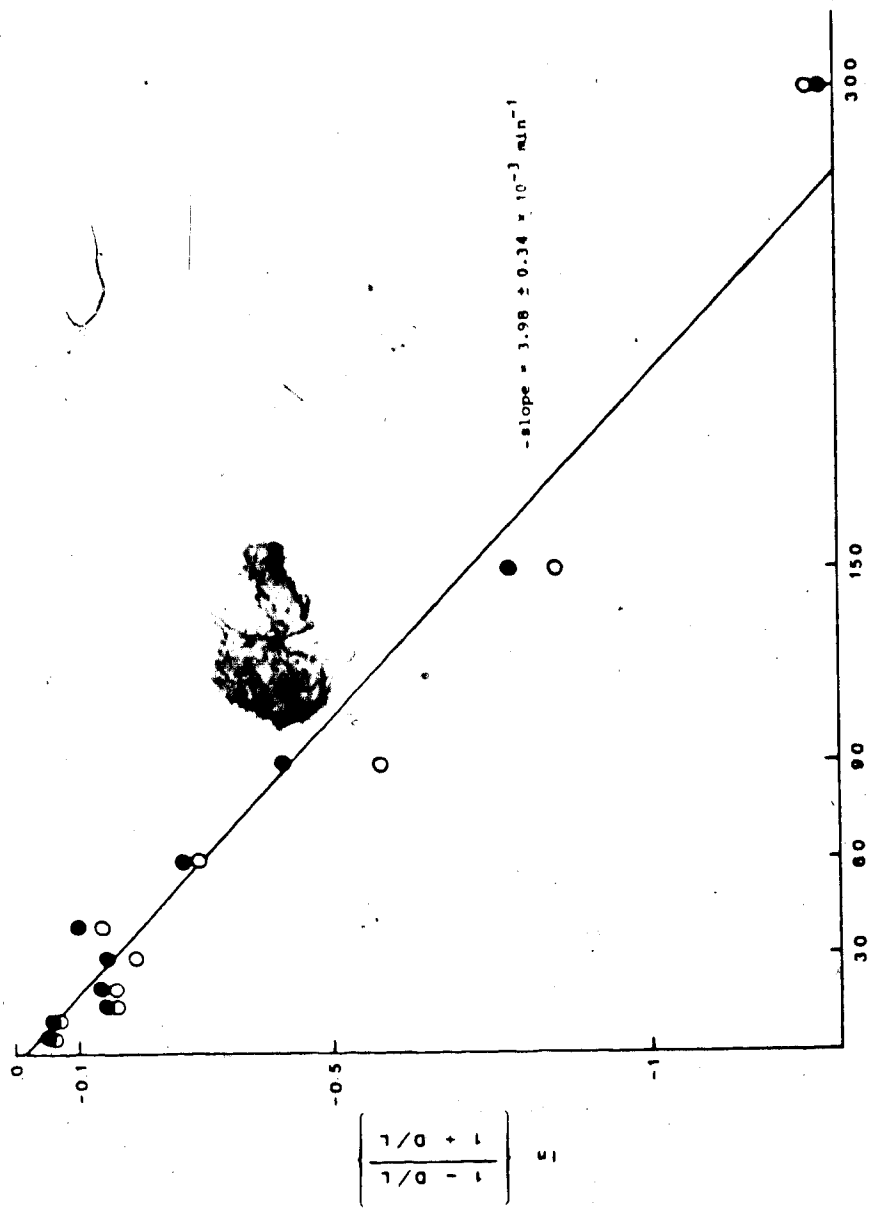


Figure 59. Racemization of the proline peptide unit determined by measuring c-(D-pro-D-leu)/c-(L-pro-L-leu). This pair of enantiomers was isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. See Figure 30 for the interpretation of the solid and open symbols. The solid line was produced by regression analysis using points after 20 minutes.

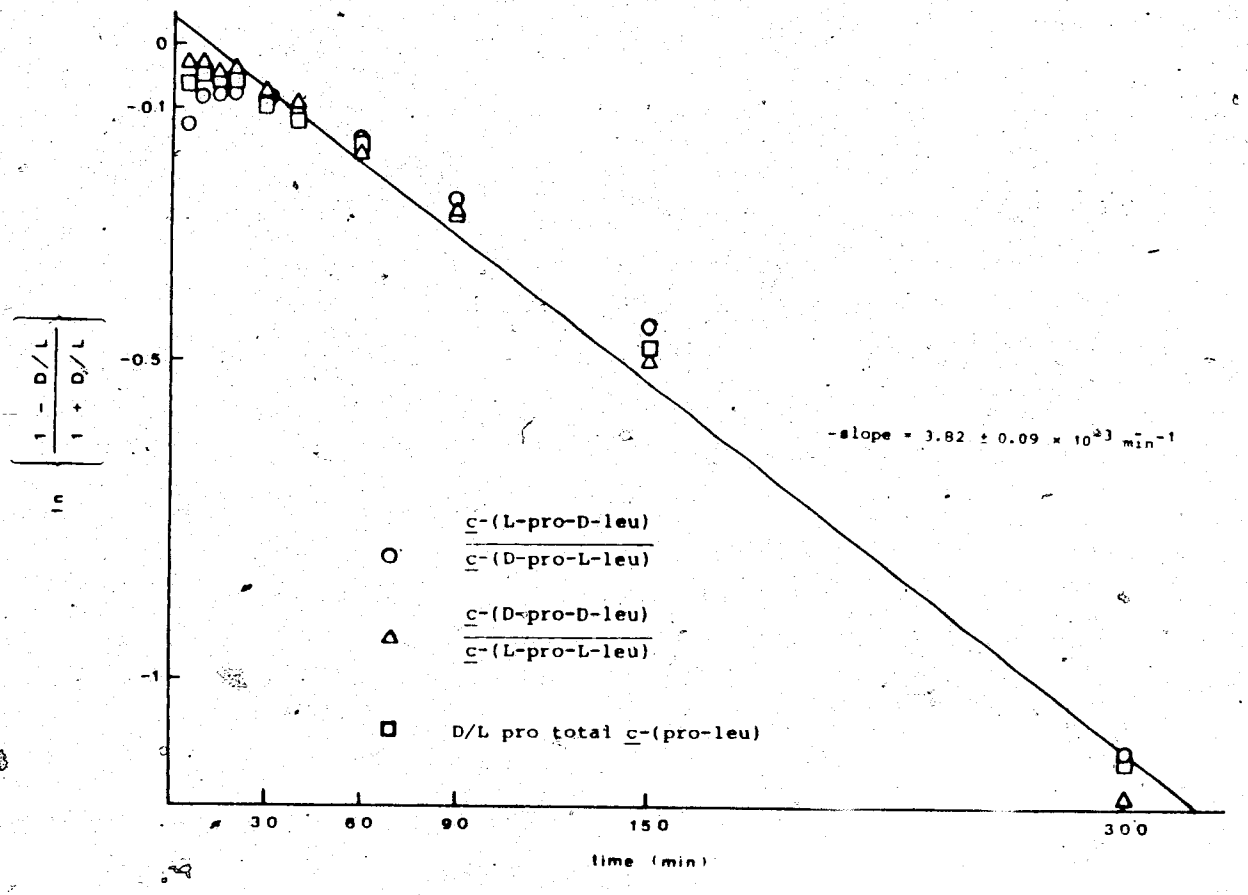


Figure 60. Racemization of the leucine peptide unit in c-(pro-leu) isolated during the decomposition of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. The solid line was produced by linear regression analysis of the data points after 20 minutes.

Table 57. Experimentally Determined Rate Constants (10^4 s^{-1}) for L-pro-L-leu, (Scheme 9), 148.5°, pH 6.8, 9.01 mM

$k_1 = k_{-1}$[a]	$k_2 = k_{-2}$	$k_5 = k_6$	k'_1	k'_{-1}	k'_2	k'_{-2}
3.58 ± 0.15[b]	0.0046 ± 0.042	1.72 ± 0.06	13.9 ± 1.5	6.07 ± 0.66	0.235 ± 0.030	0.540 ± 0.069

a) k_1, k_{-1}, k_2 and k_{-2} refer to the dipeptide; k'_1, k'_{-1}, k'_2 and k'_{-2} refer to the diketopiperazine

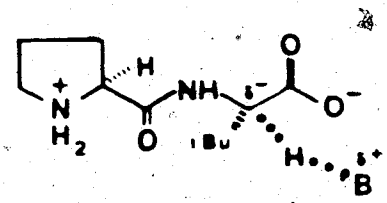
b) The rate constants are quoted as averages of several measurements. The error is reported at the 95% confidence level.

(2) Proline racemization is slower in the dipeptide than in the tetrapeptide.

Features 1 and 2 taken together indicate that the dipeptide cannot be giving rise to the underestimation of $k'_1 +, k'_{-1}$ in the diketopiperazine.

(3) Leucine racemization in the dipeptide is at least an order of magnitude slower than in the tetrapeptide.

Removal of the proton at the asymmetric centre may be inhibited by the electrostatic repulsion between the carboxylate anion and the developing negative charge in **17**.



17

The reduction by a factor of four in the rate of proline epimerization may also be due to this effect.

The experimentally determined rate constants were substituted into the computer simulation. Since the rate constant for leucine epimerization in the substrate (k_2) is small, two of the Eigenvalues produced in the numerical solution have nearly the same value. As a result of round-off error (see Appendix 4), a solution cannot be achieved. The minimum k_2 which can be used in the simulation program is 10 times larger than the experimental value. With the rest of the experimental rate constants entered into the simulation program without adjustment, the agreement between theory and experiment can be tested.

The data generated by the simulation are overlayed in Figures 48, 50, 55-57 and 61 where it can be seen that the agreement is similar to that obtained for the decomposition of pro-leu-gly-gly. Again, the value of $k'_1 + k'_{-1}$ appears to be underestimated.

To probe this possibility, pure c-(L-pro-L-leu) was epimerized under the same conditions. Since there is little ring opening to the dipeptide, the total mixture was used to obtain the D/L ratio of the diketopiperazine. Surprisingly, the rate constant extracted using points at 5 and 10 minutes was found to be 2.03×10^{-3} s^{-1}, essentially the same value obtained from plgg and pl. This result justifies the assumption that

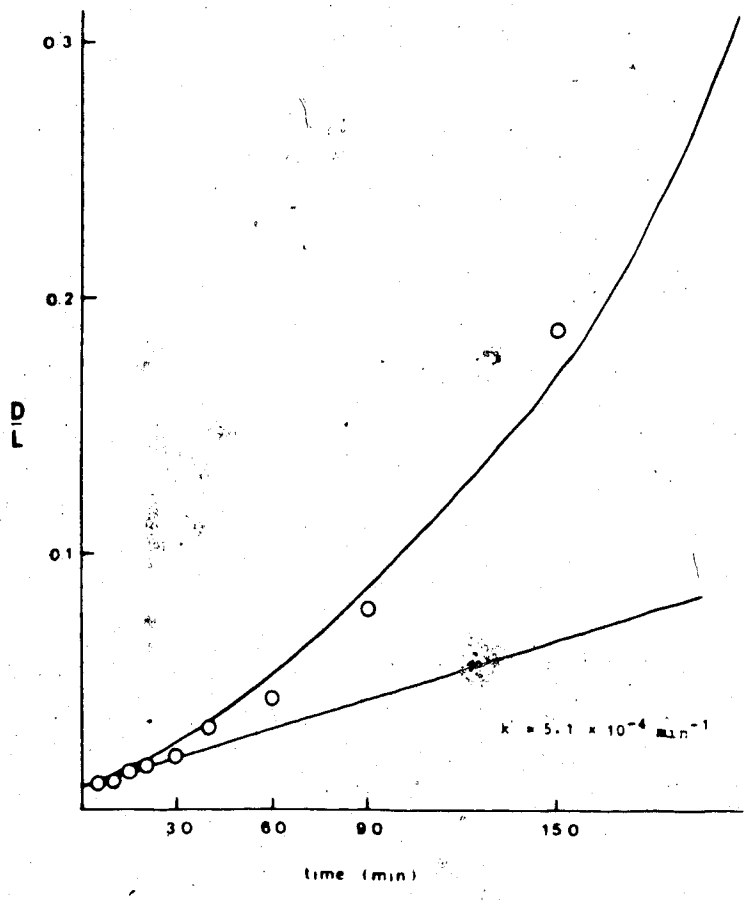
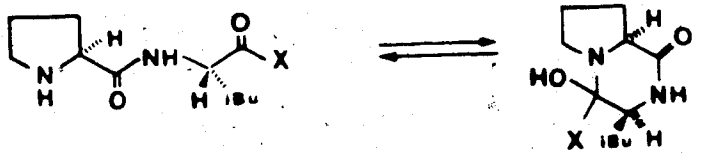


Figure 61. Racemization of the leucine peptide unit in the total hydrolysate of L-pro-L-leu, 148.5°, pH 6.8, 9.01 mM. The upper solid total hydrolysate shows the computer simulation produced using the experimentally derived rate constants (Table 57). The details of the assumptions used to obtain the rate constants are given in the text (p. 175-188). The computer simulation is based on the pathways given in Scheme 9. The lower line is the estimate of the "initial slope", 5.1×10^{-4} min^{-1}. The actual value of the rate constant for leucine racemization was 2.8×10^{-5} min^{-1} (Figure 49).

cyclization from plgg or pl can be ignored when obtaining epimerization rate constants for the diketopiperazines. The result is disappointing because the rate constant extracted in this manner is still too small to account for the degree of asymmetric induction observed for the diketopiperazine and the total hydrolysate.

Another possible source of asymmetric induction is the tetrahedral intermediate **18** resulting from internal aminolysis. An analogous tetrahedral intermediate was postulated to fit the kinetics of cyclization of dipeptide esters.[96]



18

X = OH, gly-gly

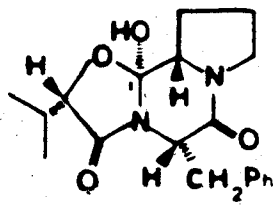
The intermediate must have been present in a sufficient quantity to cause the observed asymmetric induction (ca. 10%) and be undetected in the HPLC, perhaps as an unretained component.

The marked (') carbon in **19** would be expected to have a chemical shift of about 100 ppm in the ^{13}C NMR spectrum.* 0.5 mL of 94.7 mM L-pro-L-leu-gly-gly was heated for 30 min at 148.5° in pH 6.8 buffer.

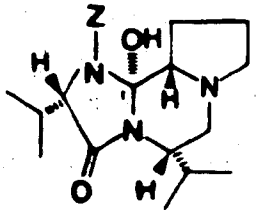
The ^{13}C NMR spectrum of the reaction mixture was measured at 100.61 MHz (broadband decoupled, 20000 scans, pulse width = 2.5 μs, no relaxation delay). No signals were detected in the region where this type of quaternary carbon would be expected to absorb.

---

*The marked (') carbon of the ergot alkaloid analogues **19** and **20** have ^{13}C chemical shifts of 104.624 ppm (Ref. 97) and 98.4 ppm (Ref. 98) respectively.



19



20

Z=PhCH$_2$-O-CO-

There are a number of possible reasons for the failure to observe the tetrahedral intermediate:

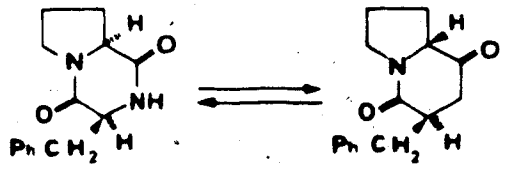
(1) It is not present in a large enough amount for observation by ^{13}C NMR.

(2) It is not sufficiently stable for the time necessary to measure the spectrum.

(3) The relaxation time of the quaternary carbon may be quite long.

Asymmetric Induction in Diketopiperazine Epimerization

The greater thermodynamic stability of c-(D-pro-L-phe) compared with c-(L-pro-L-phe) has been attributed to a steric effect.[92] Any crowding between the pyrrolidine ring and the benzyl substituent on the same side of the diketopiperazine ring in c-(L-pro-L-phe) **21** is absent in c-(D-pro-L-phe) **22**.



21 **22**

A similar effect is probably working in c-(pro-leu). This argument does not help reveal the reason for preferential epimerization of the proline asymmetric centre. As an amino acid, proline racemizes 2.8 times faster than leucine at 142°.[10] This difference might be considered to be a measure of the steric requirements of the isobutyl group of leucine and the pyrrolidine ring of proline. In c-(pro-leu), the proline peptide unit epimerizes about 20 times faster than the leucine peptide unit (Tables 51 and 57).

Although a difference in solvation around the asymmetric centres might account for this observation, ring strain may also play a role. This possibility was tested by measuring the ^{13}C-H coupling constants for the asymmetric centres in the diketopiperazines. Table 58 shows the results of these experiments. The coupling constants for the asymmetric centres of c-(L-pro-L-leu) seem to indicate that the proline peptide unit is more acidic than the leucine peptide unit predicted by the epimerization measurements. Contrariwise, the leucine peptide unit in c-(D-pro-L-leu) appears to be more acidic than the proline peptide unit. Moreover, the ^{13}C-H coupling constants for the asymmetric centres of proline and leucine are the same or larger than those for the diketopiperazines. This observation predicts that the

Table 58. ^{13}C-H Coupling Constants[a] for c-(L-pro-L-leu) and c-(D-pro-L-leu)

	$J_{^{13}C-H}$ (Hz)		%s[b]	
	proα	leuα	proα	leuα
c-(L-pro-L-leu)[c]	139.5	136.5	27.9	27.3
c-(D-pro-L-leu)[c]	139.2	144.9	27.8	29.0
proline[d]	149.0	-	29.8	-
leucine[d]	-	145.0	-	29.0

a) Measured at 100.61 MHz, gated broadband decoupled, NOE observed, D1 = 1 to 8 s

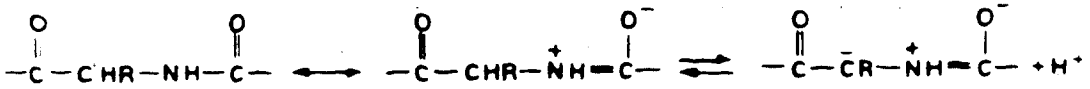
b) Calculated from %s = 0.2 J

c) $CDCl_3$

d) D_2O

rate of racemization of the amino acids should be larger than the rate of racemization for the diketopiperazine. The actual situation is the reverse: leucine racemizes* about 17 times slower than the leucine peptide unit in c-(pro-leu).

These observations indicate that the measurement of the coupling constant may not be a useful probe of the acidity of the alpha proton. Any effects which stabilize or destabilize the anion (solvation, steric crowding, electrostatic interaction with a nearby charge) appear to be swamping out any potential ground state acidity differences. The enhanced rate of racemization in the tetrapeptides and diketopiperazines is more consistent with the stabilizing effect of a partial positive charge adjacent to the anion at the asymmetric centre:



This stabilizing effect is likely partially offset by the adjacent carboxylate anion in amino acids. The lower rate of racemization of leucine compared with proline is

---

*Calculated from $\Delta S^{\ddagger}$ = -19.8 e.u., $\Delta H^{\ddagger}$ = 27.5 kcal/mol.[25]

probably indeed due to the greater solvation or steric problems of the isobutyl group compared with the pyrrolidine ring.

Racemization of Gly-L-leu-gly-gly

A time study of the racemization and hydrolysis behaviour of glgg was carried out for comparison with the study on plgg. The species in the hydrolysis mixture are the substrate glgg, the diketopiperazine c-(glycyl-leucyl) (cgl) and at least two other components. These unknown components are almost unretained in the liquid chromatogram. The eluant from the chromatogram before the substrate peak was collected and the amino acid composition estimated by the GC method. Recall that the derivatization procedure depletes the glycine derivative. The substrate gives a gly/leu ratio of 1.63 ± 0.16, the diketopiperazine cgl 0.59 ± 0.03 and the peaks at the beginning of the chromatogram 1.69 ± 0.11. The authentic dipeptides gly-leu and leu-gly have approximately the same retention times as the unknown components. Therefore, these components are likely the dipeptides gly-leu, leu-gly and gly-gly.

The data obtained for the decomposition of gly-leu-gly-gly is found in Table 59. In this experiment the source of the diketopiperazine is ambiguous since there is

Table 59. Decomposition[a] of Gly-L-leu-gly-gly, 148.5°, pH 6.8[b], 9.00 mM

Time (min)	Normalized Concentration[c] dipeptides	glgg	cgl
5	0.172	0.777	0.051
10	0.254	0.613	0.133
15	0.273	0.542	0.149
20	0.301	0.494	0.205
30	0.305	0.449	0.246
40	0.344	0.400	0.256
60	0.306	0.342	0.352
60	0.309	0.322	0.369
90	0.280	0.289	0.431
150	0.297	0.155	0.547
420	0.343	0.030	0.626

a) See footnote a), Table 56

b) See footnote b), Table 56

c) For abbreviations see footnote c), Table 43

initially more dipeptide than diketopiperazine.* Some of the diketopiperazine must be from dipeptides (gly-leu or leu-gly) arising from hydrolysis of the tetrapeptide.

There are three important findings from this experiment. First, a diketopiperazine (c-(gly-leu)) is formed during the heating of another peptide. Second, the racemization rate of the substrate tetrapeptide ($2.57 \pm 0.08 \times 10^{-5}$ s^{-1}) (Table 60, Figures 62 and 63) does not resemble that of the total hydrolysis mixture ($4.53 \pm 0.11 \times 10^{-5}$ s^{-1}). This further demonstrates that the epimerization rate constant of a peptide cannot be measured by the initial slope of the canonical plot. Third, the racemization of glgg is about three times faster than the rate of epimerization of the leucine peptide unit in plgg. This shows that neighbouring peptide units can have an effect on the racemization rates.

Racemization of Peptide Derivatives

There have been several model studies directed toward determining the way in which the position of the peptide unit along the peptide chain affects the rate of

*The tetrapeptide decomposes with a rate constant of $1.1 \pm 0.2 \times 10^{-2}$ min^{-1}.

Table 60. Racemization[a] (D/L) of the Leucine Peptide Unit in Components from the Decomposition of Gly-L-leu-gly-gly, 148.5°, pH 6.8, 9.00 mM

Time (min)	Component[b] dipeptides[c]	glgg	cgl	total hydrolysate
5	0.027	0.020	0.039	0.024
10	0.049	0.029	0.067	0.024
15	0.045	0.028	0.094	0.039
20	0.062	0.037	0.120	0.063
30	0.077	0.049	0.161	0.070
40	0.096	0.060	0.212	0.099
60	0.159	0.095	0.306	0.154
90	0.216	0.156	0.423	0.254
150	0.345	0.284	0.546	0.439
420	0.622	0.568	0.933	0.811

a) See footnote a), Table 44

b) For abbreviations see footnote c), Table 43

c) Peaks at the beginning of the chromatogram are tentatively assigned to the dipeptides gly-leu, leu-gly and gly-gly

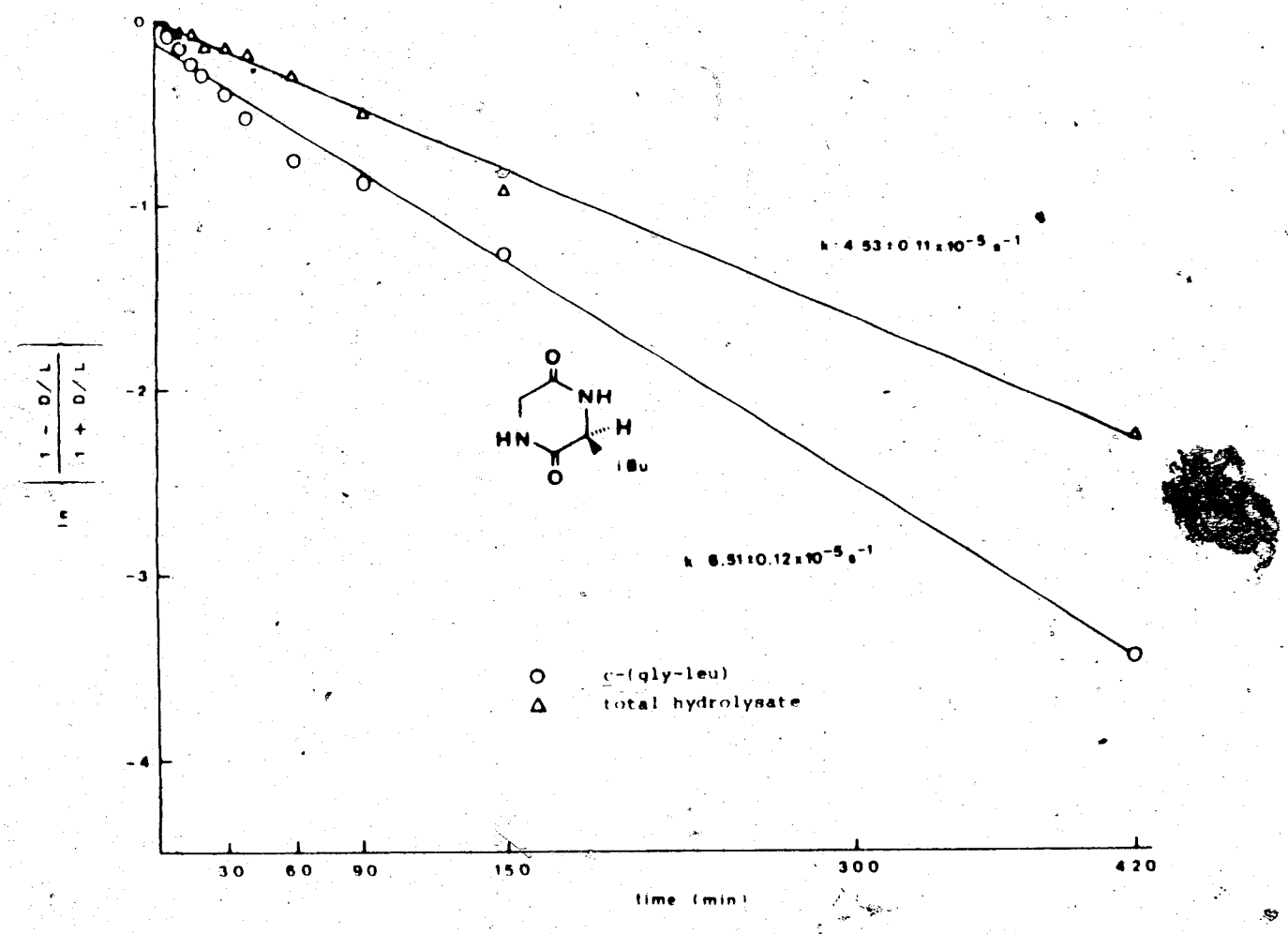


Figure 62. Racemization of the leucine peptide unit in c-(glycyl-leucyl) isolated during the decomposition of glycyl-L-leucyl-glycyl-glycine, 148.5°, pH 6.8. The solid lines were produced by linear regression analysis.

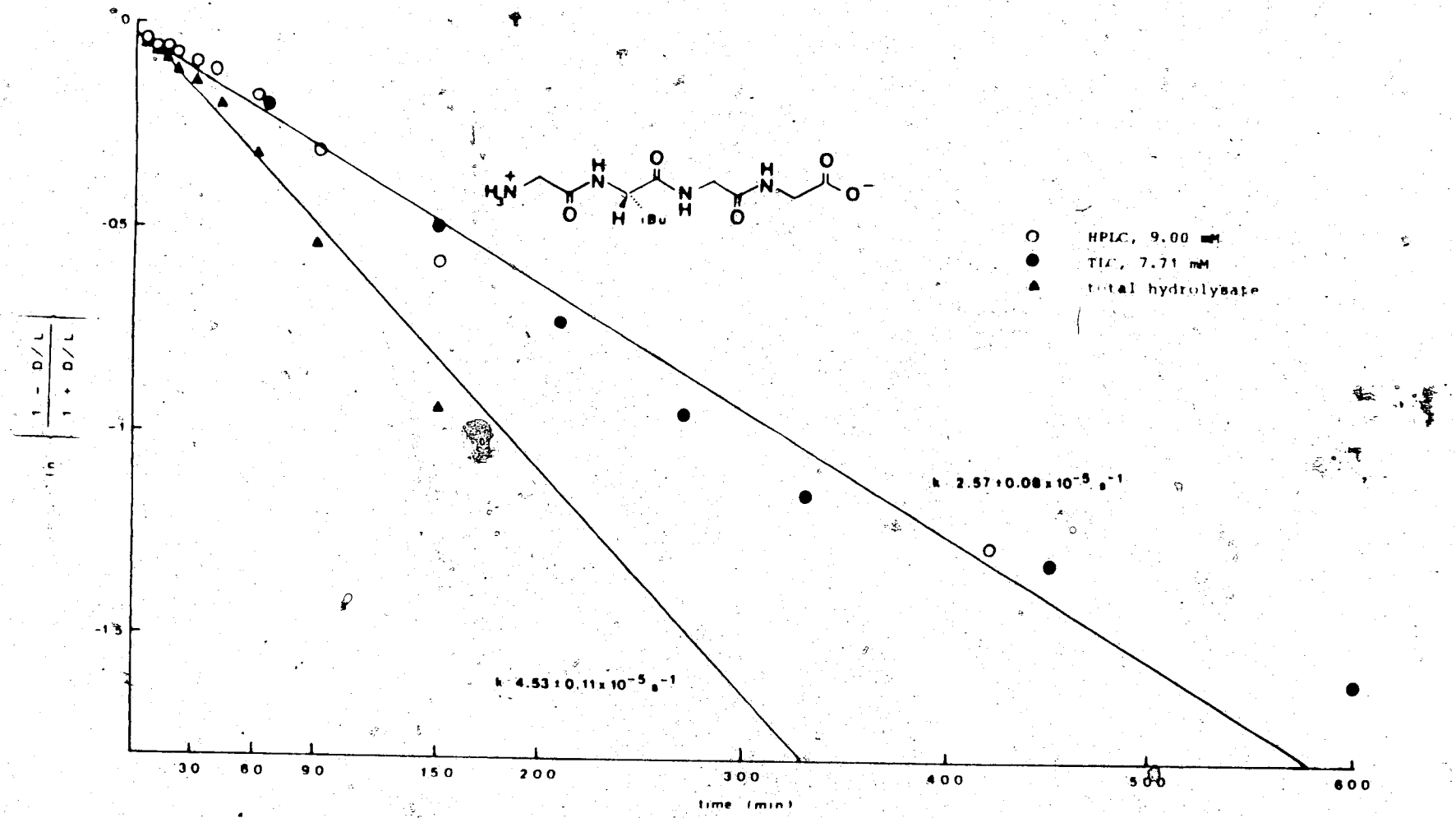


Figure 63. Racemization of the leucine peptide unit in gly-leu-gly-gly isolated during the decomposition of glycyl-L-leucyl-glycyl-glycine, 148.5°, pH 6.8, 9.00 mM. Data from the separation by TLC and HPLC are combined in this plot. The solid line was produced from the HPLC data by linear regression.

racemization. As mentioned in the introduction proximity to the termini has been suggested to increase the rate of racemization. By use of the N-terminal benzyloxycarbonyl and C-terminal anilide derivatives of the tetrapeptides plgg, gplg and glgg, this idea was tested. Tetrapeptides were considered to be better protein models than di- or tripeptides. Also, the N-, C- and N- and C blocked peptides were taken to model the C-terminal, N-terminal and interior portions of the peptide respectively.

The compounds (Table 61) were heated under the same conditions used in the time studies. Since hydrolysis (or cyclization) consumes the unblocked parent peptides quite rapidly, the degree of racemization in the compounds is in some cases quite small. The resulting large relative error in some of the rate constants is the unavoidable consequence of the competing processes which are depleting the substrate.

The data from this experiment is given in Table 61 for leucine epimerization in the isolated substrates. Assuming that the epimerization follows the canonical equation, the rate constants for epimerization can be calculated. Although the rate constants for leucine epimerization for the substrates plgg and glgg do not agree well with the values obtained from the time studies,

Table 61. Leucine Epimerization[a] in Recovered Substrate, 148.5°, pH 6.8[b]

Peptide[c]	(D/L)	$(D/L)_{t=0}$	10^5 k (s^{-1})[d]
plgg	0.08408	0.01049	1.4 ± 0.2
Z-plgg	0.2490	0.01097	0.25 ± 0.20
Z-plgg-NHPh	0.04051	0.01159	0.53 ± 0.18
plgg-NHPh	0.09437	0.01191	1.5 ± 0.2
gplg	0.1264	0.01759	1.2 ± 0.2
Z-gplg	0.02823	0.01871	0.1 ± 0.1
Z-glpg-NHPh	0.05317	0.02014	0.37 ± 0.12
gplg-NHPh	0.1836	0.0443	1.8 ± 0.1
glgg	0.3489	0.01939	3.8 ± 0.1
Z-glgg	0.05795	0.01888	0.43 ± 0.12
Z-glgg-NHPh	0.07356	0.01770	0.62 ± 0.12
glgg-NHPh	0.2697	0.01563	2.9 ± 0.1

a) See footnote a), Table 44

b) After 90 min for plgg and derivatives, 150 minutes for all others

c) p = prolyl, l = leucyl, g = glycyl, Z = $C_6H_5CH_2OCO-$, Ph = C_6H_5

d) $k = -\frac{1}{2t} \ln \frac{1-D/L}{1+D/L}_t - \ln \frac{1-D/L}{1+D/L}_{t=0}$

the relative rates are essentially the same. For the time study:

$$\frac{k_{glgg}}{k_{plgg}} = 2.87 \pm 0.29$$

and for the competitive study at only one point:

$$\frac{k_{glgg}}{k_{plgg}} = 2.79 \pm 0.55$$

Since these measurements agree within experimental error, the relative rates can be used with some security. Placing a CBZ group (carbobenzoxy- or benzyloxycarbonyl-) on the N-terminus decreases the rate of epimerization of the leucine peptide unit.* CBZ has the added effect of decreasing the rate of hydrolysis. This is illustrated qualitatively by the chromatograms of the reaction mixtures for CBZ-plgg-NHPh (Figure 64) and CBZ-plgg (Figure 65). These should be compared with the decomposition of the parent peptide, plgg (Figure 20).

Conversely, addition of a C-anilide group has little effect. It does not significantly change the rate of epimerization of leucine nor does it prevent hydrolysis (Figure 66).

*By 7.4 × for plgg, 6.5 × for gply, 6.4 × for glgg.

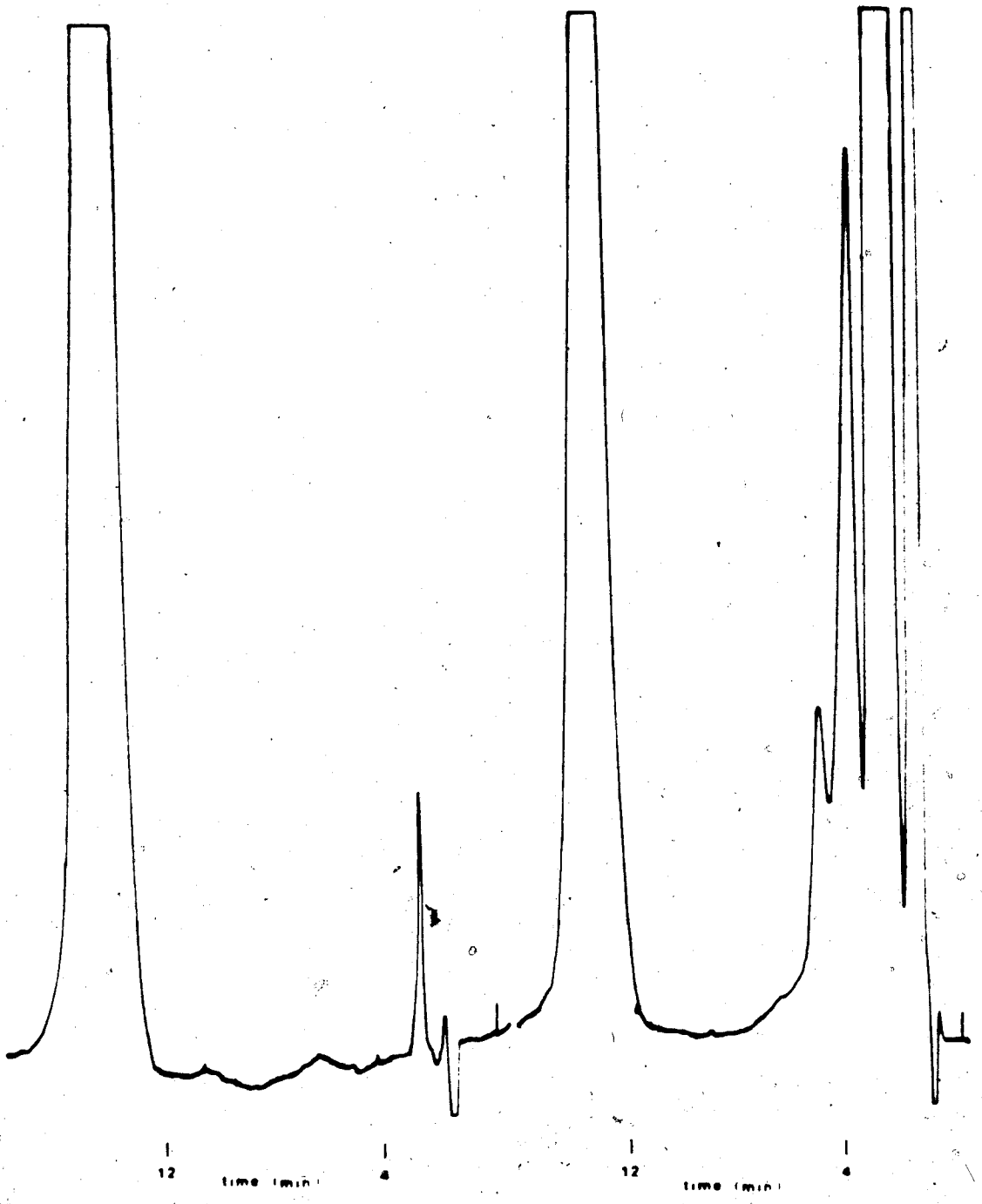


Figure 64. HPLC traces of CBZ-L-pro-L-leu-gly-gly anilide before and after heating for 90 min at 148.5°, pH 6.8. Chromatographic conditions: reversed phase C-18 column (Waters), mobile phase, 0.03 M NH_4OAc, 33% acetonitrile, 1.0 ml

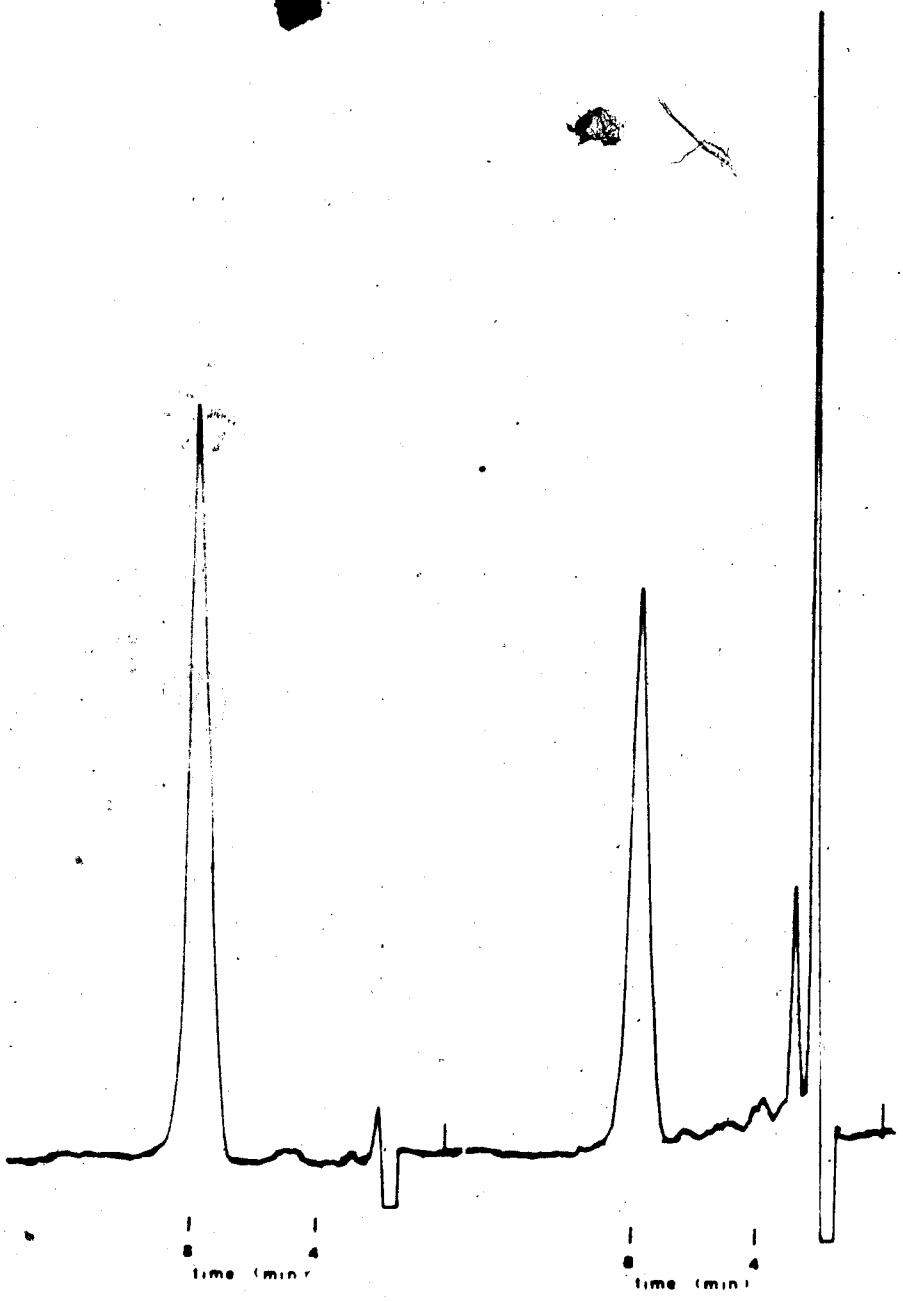


Figure 65. HPLC traces of CBZ-L-pro-L-leu-gly-gly before and after heating for 90 min at 148.5°, pH 6.8. Chromatographic conditions: reversed phase C-18 column (Waters); mobile phase, 0.065 M NH_4OAc, 22% acetonitrile, 1.0 mL/min.

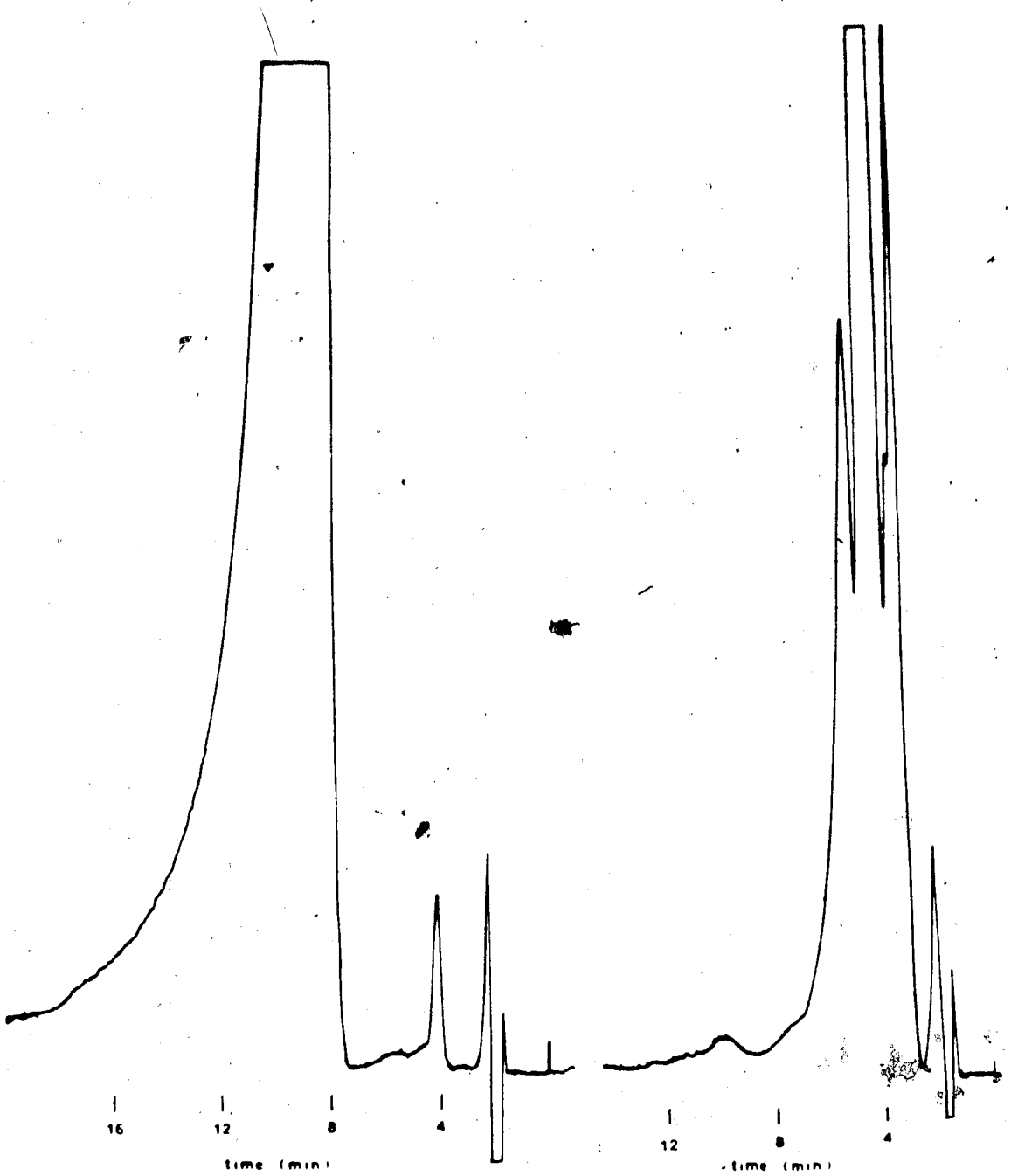


Figure 66. HPLC traces of L-pro-L-leu-gly-gly anilide before and after heating for 90 min at 148.5° & pH 6.8. Chromatographic conditions: reversed phase C-18 column (Waters); mobile phase, 0.065 M NH_4OAc, 22% acetonitrile, 1.0 mL/min.

The data for the epimerization of proline in the eight substrates are given in Table 62. The results for proline are essentially the same as for leucine. The presence of the CBZ group decreases the rate* while masking the C-terminus with an anilide group again has little effect.

The rate decrease caused by the CBZ group could be explained in three ways:

(1) Free NH_2 can act as a base **23.** When blocked with the CBZ group, the nitrogen is not so readily able to remove the proton from the asymmetric centre of either the adjacent peptide unit or a peptide unit of another peptide molecule.

Since peptides are present at neutral pH as zwitterions there would be very little free NH_2. It would be tempting to eliminate this pathway for this reason. However, since diketopiperazine formation requires an unprotonated amino function **24** and since diketopiperazine formation is so predominant, elimination of this pathway is not justified. Indeed, the ability to cyclize to give a diketopiperazine may be reflected by an enhanced rate of racemization.

---

*By 11.5 × for plgg, 25.0 × for gplg.

Table 62. Proline Epimerization in Recovered Substrate, 148.5°, pH 6.8[b]

Peptide[c]	(D/L)	$(D/L)_{t=0}$	10^5 k (s^{-1})[d]
plgg[e]	0.09563	0.0081	1.6 ± 0.2
Z-plgg	0.01810	0.0131	0.09 ± 0.1
Z-Plgg-NHPh	0.02556	0.0081	0.32 ± 0.10
plgg-NHPh	0.1744	0.0094	3.1 ± 0.2
gplg	0.2520	0.0145	2.7 ± 0.2
Z-gplg	0.02840	0.0142	0.16 ± 0.12
Z-gplg-NHPh	0.02413	0.0161	0.88 ± 0.12
gplg-NHPh	0.3143	0.0182	3.42 ± 0.15

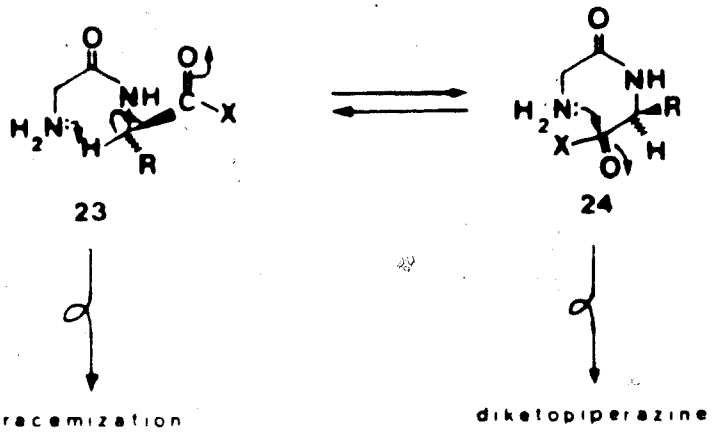
a) See footnote a), Table 44

b) 90 min for plgg and derivatives, 150 min for gplg and derivatives.

c) For abbreviations see footnote c), Table 61

d) Assume canonical expression (Table 61)

e) Only DD/LL measured



(2) The removal of the positive charge from nitrogen by substitution eliminates the inductive stabilization of the incipient carbanion.

(3) The presence of the hydrophobic CBZ group disrupts the solvation shell so that the approach of water, acting as a base to remove the chiral proton, is impeded. When the N-terminus is not blocked, racemization may be aided by more tightly packed solvent molecules.

Since the anilide derivatives are amines, at pH 6.8 they will be predominantly in the protonated form.*

---

*pKa for $H_2NCH_2CO_2CH_2CH_3$ = 7.5 ~ 10% free base at pH 6.8.[99]

Therefore, it is not surprising that the anilide derivative behaves in the same way as the parent tetrapeptide, giving both relatively rapid racemization and a large amount of diketopiperazine.

The derivatives can then be grouped according to their racemization behaviour. Those derivatives which racemize relatively quickly have positively charged nitrogen termini. The derivatives with blocked nitrogen termini racemize more slowly regardless of the charge on the C-terminus. The derivative with a protected nitrogen terminus (e.g. CBZ-plgg) will be mostly in the form of an anion.*

Since the anion on the C-terminus does not appear to increase the rate of racemization, the anion cannot be acting as a base. It is also noteworthy that the presumed increase in solvation of the anion (compared with the uncharged derivative) is also ineffective in promoting racemization. On the contrary, the leucine peptide unit of derivatives blocked only at the C-terminus racemize slower than those derivatives having both termini blocked (Table 61). This observation is true provided that the error limits are considered to be representing the worst

---

*pKa for $CH_3CONH-CH_2CO_2H$ = 3.67 ~ 99.8% anion at pH 6.8.[99]

case. The error limits were calculated estimating the error in the D/L ratios to be ±0.01. This may be too harsh an estimate. It appears that the presence of a carboxylate anion inhibits formation of an anion at the asymmetric centre. This is consistent with the slower rate of racemization observed for pro-leu compared with pro-leu-gly-gly.

The racemization of the peptide units in the total hydrolysate is given in Table 63 for leucine and in Table 64 for proline. The behaviour is the same as the substrates with one exception. The D/L ratio for the proline peptide unit in the total hydrolysate reaches a value of 1.97 for plgg and 1.94 for plggNHPh. This is an indication that plggNHPh is decomposing by internal aminolysis to give the diketopiperazine cpl. The diketopiperazine then epimerizes to the preferred diastereomer, cDpl. The D/L ratios for proline in the total hydrolysates of gplg and gplgNHPh (1.12 and 1.07 respectively) may be indicating only total racemization of proline and the apparent excess of D-proline may be an artifact. However, there was some indication of diketopiperazine in the chromatogram. This was not investigated further.

The observation of a reduced rate of racemization in the total hydrolysate of the CBZ derivatives is consistent with reduced diketopiperazine formation. This is borne

Table 63. Leucine Epimerization[a] in Total Hydrolysate of Model Peptides, 148.5°, pH 6.8[b]

Peptide[c]	(D/L)	$(D/L)_{t=0}$
plgg	0.1688	0.01049
Z-plgg	0.05292	0.01097
Z-plgg-NHPh	0.1002	0.01159
plgg-NHPh	0.1815	0.01191
gplg	0.3659	0.01759
Z-gplg	0.01803	0.01871
Z-gplg-NHPh	0.07412	0.02014
gplg-NHPh	0.4393	0.02218
glgg	0.4456	0.01939
Z-glgg	0.1975	0.01888
Z-glgg-NHPh	0.2929	0.01770
glgg-NHPh	0.4691	0.01563

a) See footnote a), Table 44

b) See footnote b), Table 62

c) For abbreviations, see footnote c), Table 61

Table 64. Proline Epimerization[a] in Total Hydrolysate of Model Peptides, 148.5°, pH 6.8[b]

Peptide[c]	(D/L)	$(D/L)_{t=0}$
plgg	1.9656	0.0081
Z-plgg	0.1641	0.0131
Z-plgg-NHPh	0.2749	0.0081
plgg-NHPh	1.9423	0.0094
gplg	0.2520	0.0185
Z-gplg	0.02186	0.0143
Z-gplg-NHPh	0.1100	0.0161
gplg-NHPh	1.0722	0.0185

a) See footnote a), Table 44

b) See footnote b), Table 62

c) For abbreviations see footnote c), Table 61

out by the failure to observe peaks in the chromatograms corresponding to the diketopiperazines.

The competitive rate experiments have indicated that proximity to a protonated nitrogen enhances racemization. Conversely, proximity to a carboxylate anion decreases the rate of racemization. This does not mean that the rate of racemization for the individual peptide units in a peptide or peptide chain will increase monotonically from the C-terminus to the N-terminus. Such a conclusion ignores possible rate enhancement/retardation by a nearby charge centre which is not a terminus (e.g. $^{+}NH_3$ of lys or CO_2^{-} of glu). Any rate enhancement or retardation also increases the chance of asymmetric induction. A uniform distribution of water along the peptide chain should also not be assumed. Hare observed little racemization in bones heated at 157° for 3 days under anhydrous conditions while there was essentially complete racemization in the presence of water.[100] By analogy, a lower rate of racemization is expected for peptide units in a region of peptide where water is excluded. Clearly, it is difficult to generalize protein racemization behaviour based on any model study because of the individual characteristics of each protein.

CHAPTER 5

CONCLUSION

The use of amino acid racemization as a tool for geological dating provided the motivation for this study. It is appropriate to include a discussion of the geochemical implications of the results.

The general goal of this work was to investigate some of the unique properties of peptides which can complicate the interpretation of peptide racemization measurements used for dating fossils. In the broadest sense, there are two structural features which make peptides unlike amino acids. First, the presence of more than one peptide unit gives rise to the particular structure of the peptide or protein. Here, structure means the primary, secondary, tertiary and quaternary structure of proteins. Second, the tendency of proteins to decompose by breaking the peptide bond gives rise to a complex mixture of species all having different rates of racemization. A consideration of these properties of peptides leads to a number of questions of concern to geochemists.

The first characteristic raises three questions.

(1) What is the effect of different neighbouring peptide units on the rate of racemization, i.e. how does the rate of racemization of Y differ between the sequence -X-Y- and -?-Y-?

It was found that the leucine peptide unit in glgg racemizes about three times faster than in plgg. The factor of three indicates a range of rates in peptides of similar structure. Also, these two tetrapeptides were considered to be realistic models of collagen which contains the repeating sequence -gly-X-Y- with about 10% -gly-X-pro-.[3] The result helps to explain the variation in D/L ratios in fossils of the same age but of different species. The range of D/L ratios in fossil shells has been reported to span at least a factor of two[101], an observation which compares favourably with the measurement for plgg and glgg.

(2) What is the variability of the racemization rate constant within a protein?

The results from studies on peptide derivatives give an indication of the racemization in the interior of proteins. The peptide units of the derivatives blocked at the nitrogen terminus were observed to racemize 6 to 25 times slower than the unblocked parent peptides. This agrees with previous model studies which reported that the interior peptide units racemize slower than those at the

termini. From the results on the peptide derivatives, racemization rate is dependent on the proximity to the nitrogen terminus as well.

The exceptionally rapid epimerization of the proline peptide unit in plgg (at least 100 times faster than in CBZ-gplg) is significant. Clearly, the D/L ratio measured for a protein hydrolysate cannot be representative of the original peptide if there is a range of racemization rates of this magnitude. This hazard can be avoided by analyzing only larger peptide fragments where there are on average essentially no terminal peptide units.

Results concerning the effect of the C-terminus are in discord with the previous model studies.[41,45,46,47] This work showed that the carboxylate anion at the C-terminus decreases the rate of racemization.

Racemization of the glycine peptide unit in the dipeptide pro-leu is about 18 times slower than in plgg while the proline peptide unit in the dipeptide racemizes at about one quarter the rate of the tetrapeptide. Blocking the N-terminus (producing a carboxylate anion at pH 6.8) gives species having the slowest epimerization rate.

(3) Is it possible to observe D/L ratios greater than unity?

Since there is more than one chiral centre in

peptides, there is the opportunity for observing asymmetric induction. The chances of this are especially acute if neighbouring peptide units racemize at significantly different rates. The computer simulations of systems containing two chiral centres showed that the conditions necessary to achieve D/L ratios greater than one are not that restrictive. The occurrence of this phenomenon is a major deviation from the pseudo first order kinetics found for amino acids.

The possibility of asymmetric induction (giving D/L > 1) has largely been ignored in protein studies. The observation of asymmetric induction in the proline peptide unit in plgg and pl in the total hydrolysate underscores the problem of using the total peptide mixture present in the fossil. The source of the asymmetric induction was found to be the diketopiperazine rather than the tetrapeptide. This does not exclude the occurrence of the phenomenon in peptides or proteins.

Anomalously high D/L ratios (but not greater than unity) for alanine, glutamic acid and aspartic acid have been reported in certain soils[102] and sediments.[103] These ratios were attributed to the addition of D-ala and D-glu from the peptidoglytan making up the cell walls of gram positive and gram negative bacteria.[104] Asymmetric induction provides another mechanism for high D/L ratios.

Protein hydrolysis raises three more questions relevant to amino acid dating.

(4) Are rate constants measured from the total hydrolysate near time zero representative of the original peptide or protein?

Although this is theoretically possible, the time studies show the practical limitations of the procedure. The rate constant extracted in this manner has a value between the rate constants for the substrate and the more quickly epimerizing species. The rate of leucine epimerization was larger in the total hydrolysate by a factor of 3.6 for plgg, 1.8 for glgg and 18 for pl. Structure reactivity relationships measured by the racemization of the total hydrolysate are to be taken only as approximations.

(5) Do peptides obey the canonical rate law?

Except for the cases where there is asymmetric induction, straight-line canonical plots are observed.

This is clearly demonstrated for the time study of the peptide glgg for which all of the isolated fractions exhibit linear canonical behaviour. It is important to emphasize that a linear canonical plot does not necessarily give a chemically or mechanistically meaningful rate constant. At the same time, this does not mean that these "rate constants" are not valuable. To say

this would do an injustice to the studies where empirical relationships between racemization and time have been used to advantage.[105-107] Recently, linear canonical behaviour was demonstrated in a report of the racemization of aspartic acid peptide units in the white matter of the human brain.[108] This study on presumably intact protein bodes well for the observation of similar behaviour in intact fossil protein.

(6) What are the species which result from decomposition of peptides and what are their characteristic rates of racemization?

For plgg and to a lesser extent for glgg, the predominant decomposition pathway is via the diketopiperazine. The presence of the diketopiperazines in the mixture is significant since they were also identified as rapidly epimerizing species. Since these species make an important contribution to the composition of the hydrolysate, the rate constant observed for the hydrolysate will be weighted towards the value for the diketopiperazines. This effect can be accentuated by the apparently slow racemization of peptide units in the interior of proteins.

In order to obtain well behaved kinetics it will be necessary to isolate the intact protein or large peptide fragments. The rate of protein hydrolysis becomes

important in determining the precision of the fossil ages which can be measured. The tetrapeptide plgg provides an extreme case. When the peptide has decomposed to about 10% of its original concentration (ca. 90 min, Figure 31), the leucine peptide unit has a D/L ratio of about 0.06 (Figure 37). This value must be considered unusable for accurate dating due to a measurement error of ±0.01. This situation could be ameliorated by using a more rapidly racemizing peptide unit. (Aspartic acid is used just for this reason. Its rate of racemization is about 7.5 times that of leucine at 142°.)[10]

The situation for glgg is more encouraging. When the peptide has decomposed to about 1/3 its original concentration (ca. 60 min, Table 59), the D/L ratio is about 0.10 (Table 60). Thus, the window for age measurements is determined on the low end of the time scale by the rate of racemization and on the high end by the rate of hydrolysis.

Amino acid dating does have the potential to be an important tool for geologists owing to its simplicity and speed of analysis. The uncertainty in the age determination is primarily due to an unknown temperature history but there is a certain amount of uncertainty due to chemistry as well.[109] If asymmetric induction is restricted primarily to diketopiperazines and provided

that rapid racemization likewise occurs in only diketopiperazines and near nitrogen termini, the chemical variability can be controlled by judicious selection of the higher molecular weight fraction of the protein hydrolysate. Partial separation of the protein hydrolysate may result in linear canonical plots which, while not generally mechanistically meaningful, can be empirically valuable.

REFERENCES

1. Abelson, P.H. Carnegie Inst. Wash. Year Book **1954**, 53, 97-101.
2. Hare, P.E.; Mitterer, R.M. Ibid. **1968**, 67, 205-208.
3. "Nuclear and Chemical Dating Techniques: Interpreting the Environmental Record"; Currie, L.A., Ed.; American Chemical Society: Washington D.C., 1982.
4. "Biogeochemistry of Amino Acids"; Hare, P.E., Hoering, T.C., King, K. Jr., Eds.; John Wiley and Sons: New York, 1980.
5. Rutter, M.W.; Crawford, R.J.; Hamilton, R.D. Geoscience Canada **1979**, 6, 122-128.
6. Hedges, R.E.M.; Wallace, C.J.A. Ref. 4, p 35-40.
7. Engel, M.H.; Zumberge, J.E.; Nagy, B.; Van Devender, T.R. Phytochemistry **1978**, 17, 1559-1562.
8. King, K., Jr.; Neville, C. Science (Washington D.C.) **1977**, 195, 1333-1335.
9. Bada, J.L.; Schroeder, R.A.; Protsch, R.; Berger, R. Proc. Nat. Acad. Sci. USA **1974**, 71, 914-917.
10. Bada, J.L.; Shou, M.-Y. Ref. 4, p 235-255.
11. Wehmiller, J.F. Geochim. Cosmochim. Acta **1981**, 45, 261-264.

12. Weiner, S.; Kustanovich, Z.; Gil-Av, E.; Traub, W. Nature **1980**, 287, 820-823.

13. Traub, W.; Piez, K.A. In "Advances in Protein Chemistry"; Anfinsen, C.B. Jr., Edsall, J.T., Richards, Eds.; Academic Press: New York, 1971; Vol. 25, p 243-352.

14. Jope, E.M. Ref. 4, p 23-33.

15. Tuoss, N.; Eyre, D.R.; Holtrop, M.E.; Glimcher, M.J.; Hare, P.E. Ref. 4, p 53-63.

16. Hauschen, P.V. Ref. 4, p 75-82.

17. Neuberger, A. In "Advances in Protein Chemistry"; Anson, M.L. Edsall, J.T., Eds.; Academic Press: New York, 1948; Vol. 4, p 298-383.

18. Smith, G.G.; Sivakua, T. J. Org. Chem. **1983**, 48, 628-634.

19. Stroud, E.D.; Fite, D.J.; Smith, G.G. Ibid. **1983**, 48, 5368-5369.

20. Matsuo, H.; Kawazoe, Y.; Sato, M.; Ohnishi, M.; Tatsuno, T. Chem. Pharm. Bull. **1967**, 15, 391-398.

21. Sato, M.; Tatsuno, T.; Matsuo, H. Ibid. **1970**, 18, 1794-1798.

22. Smith, G.G.; Williams, K.M.; Wonnacott, D.M. J. Org. Chem. **1978**, 43, 1-5.

23. Rada, P.L. J. Am. Chem. Soc. **1972**, 94, 1371-1373.

24. Bordwell, F.G.; Boyle, W.J., Jr. Ibid. **1975**, 97,

3447-3452.

25. Smith, G.G.; Evans, R.C. Ref. 4, p 257-282.

26. Cotrait, M.; Geoffre, S.; Hospital, M.; Preciğoux, G. Acta Cryst. **1979**, B35, 114-118.

27. Levene, P.A.; Bass, L.W. J. Biol. Chem. **1929**, 82, 171-190.

28. Levene, P.A.; Yang, P.S. Ibid. **1933**, 99, 405-416.

29. Levene, P.A.; Steiger, R.E.; Marker, R.E. Ibid. **1931**, 93, 605-621.

30. Degeilh, R.; Marsh, R.E. Acta Cryst. **1959**, 12, 1007-1014.

31. Corey, R.B. J. Am. Chem. Soc. **1938**, 60, 1598-1604.

32. Sletten, E. Ibid. **1970**, 92, 172-177.

33. Benedetti, E.; Corradini, P.; Pedone, C. J. Phys. Chem. **1969**, 73, 2891-2895.

34. Hare, P.E. Carnegie Inst. Wash. Year Book **1971**, 70, 256-258.

35. Hare, P.E. Ibid. **1973**, 72, 690-694.

36. Wehmiller, J.; Hare, P.E. Science (Washington D.C.) **1971**, 173, 907-911.

37. Levene, P.A.; Pfaltz, M.H. J. Biol. Chem. **1925**, 63, 661-668.

38. Levene, P.A.; Bass, L.W. Ibid. **1927**, 74, 727-737.

39. Levene, P.A.; Bass, L.W. Ibid. **1928**, 78, 145-157.

40. Darge, W.; Theimann, W. 1st Eur. Biophys. Congress

1971, 1, 133-137.

41. Kriausakul, N.; Mitterer, R.M. Science (Washington D.C.) **1978**, 201, 1011-1014.

42. Freidman, M.; Masters, P.M. J. Food Sci. **1982**, 47, 760-764.

43. Lajoie, K.R.; Wehmiller, J.F.; Kennedy, G.L. Ref. 4, p 305-340.

44. Wehmiller, J.F. Ref. 4, p 341-355.

45. Kriausakul, N.; Mitterer, R.M. Geochim. Cosmochim. Acta **1980**, 44, 753-758.

46. Kriausakul, N.; Mitterer, R.M. Ibid. **1983**, 47, 963-966.

47. Smith, G.G.; de Sol, B.S. Science (Washington D.C.) **1980**, 207, 765-767.

48. Dorman, D.E.; Bovey, F.A. J. Org. Chem. **1973**, 38, 2379-2383.

49. Wüthrich, K.; Tun-Kyi, A.; Schwyzer, R. F.E.B.S. Lett. **1972**, 25, 104-108.

50. Grathwohl, C.; Wüthrich, K. Biopolymers **1976**, 15, 2025-2041.

51. Grathwohl, C.; Wüthrich, K. Ibid **1976**, 15, 2043-2057.

52. Grathwohl, C.; Wüthrich, K. Ibid. **1981**, 20, 2623-2633.

53. Newman, M.S. "Steric Effects in Organic Chemistry";

John Wiley and Sons: New York, 1956; p 117-120.

54. Brewerton, A.A.; Long, D.A.; Truscott, T.G. Trans. Faraday Soc. **1970**, 66, 2297-2304.

55. Epand, R.M.; Epand, R.F. Can. J. Biochem. **1973**, 51, 140-147.

56. Pickering, B.T.; Li, C.H. Arch. Biochem. Biophys. **1964**, 104, 119-127.

57. Geschwind, I.I.; Li, C.H. Ibid. **1964**, 106, 200-206.

58. Lee, T.H.; Buettener-Janusch, V. J. Biol. Chem. **1963**, 238, 2012-2015.

59. Fridkin, M.; Wilchek, M. Biochem. Biophys. Res. Commun. **1970**, 38, 458-464.

60. Smith, G.G. Science (Washington D.C.) **1976**, 191, 102-103.

61. Wernstedt, F.L. "World Climatic Data"; Climatic Data Press: Lemont Pensylvania, 1972; p 385.

62. Hill, R.L. In "Advances in Protein Chemistry"; Anfinsen, C.B. Jr., Anson, M.L., Edsall, J.T., Richards, F.M., Eds.; Academic Press: New York, 1965; Vol. 20, p 37-107.

63. Sanger, F.; Thompson, E.O.P. Biochem. J. **1956**, 53, 353-366.

64. Sanger, F.; Thompson, E.O.P. Ibid. **1956**, 53, 366-374.

65. Hirohata, R.; Kanda, Y.; Nakamura, M.; Tzumiza, N.;

Nagamatsu, A.; Ono, T.; Fujui, S.; Kimitsuku, M. Z. Physiol. Chem. **1953**, 295, 368-377.

66. Hirohata, R.; Muramatu, M. Ibid. **1963**, 332, 271-275.

67. Bergmann, M.; Smith, E.L. J. Biol. Chem. **1944**, 153, 627-651.

68. Abderhalden, E.; Komm, E. Z. Physiol. Chem. **1923**, 132, 1-11.

69. Abderhalden, E.; Komm, E. Z. Physiol. Chem. **1924**, 134, 121-128.

70. Abderhalden, E.; Komm, E. Ibid. **1924**, 139, 147-204.

71. Bada, J.L.; Steinburg, S.M. J. Org. Chem. **1983**, 48, 2295-2298.

72. Bada, J.L.; Steinburg, S.M. Science (Washington D.C.) **1981**, 213, 544-545.

73. Long, D.A.; Lillycrop, J.E. Trans. Faraday Soc. **1963**, 59, 907-917.

74. Moore, S.; Stein, W.H. J. Biol. Chem. **1948**, 176, 367-388.

75. Long, D.A.; Truscott, T.G. Trans. Faraday Soc. **1963**, 59, 1833-1841.

76. Cronin, J.R.; Long, D.A.; Truscott, T.G. Ibid. **1971**, 67, 2096-2100.

77. Hammel, E.F., Jr.; Glasstone, S. J. Am. Chem. Soc. **1954**, 76, 3741-3745.

78. Lawrence, L.; Mooré, W.J. Ibid. **1951**, 73, 3973-3977.

79. Lee, R.G.; Long, D.A.; Truscott, T.G. Trans. Faraday Soc. **1969**, 65, 503-508.

80. Muramatu, M.; Hirohata, R.; Kanda, Y.; Shibuya, S.; Fujii, S.; Nagamatsu, A.; Ono, T. Z. Physiol. Chem. **1963**, 332, 256-262.

81. Muramatu, M.; Hirohata, R.; Kanda, Y.; Shibuya, S. Ibid. **1963**, 332, 263-270.

82. Bodanszky, M. "Peptide Synthesis"; Wiley Interscience: New York, 1976.

83. Bodanszky, M.; Bodanszky, A. "The Practice of Peptide Synthesis" In 'Structure Concepts in Organic Chemistry'; Hafner, K., Lehn, J-M., Rees, C.W., Hofmann, F.R.S., Scheyer, P.V.R., Trost, B.M., Zahradnik, R., Eds.; Springer-Verlag: New York, 1984.

84. Kopple, K.D.; Ghazarian, H.G. J. Org. Chem. **1968**, 33, 862-864.

85. Nitecki, D.E.; Halpern, B.; Westley, J.W. J. Org. Chem. **1968**, 33, 864-866.

86. Steinblatt, M. J. Am. Chem. Soc. **1966**, 88, 2845-2848.

87. Rabenstein, D.L. J. Am. Chem. Soc. **1973**, 95, 2797-2803.

88. Rabenstein, D.L.; Sayer, T.L. Anal. Chem. **1976**, 48, 1141-1146.

89. Rabenstein, D.L.; Greenberg, M.S.; Evans, C.A. Biochemistry **1977**, 16, 977-981.

90. Lucente, G.; Pinnen, F.; Zanotti, G.; Cerrini, S.; Mazza, F. J. Chem. Soc., Perkin Trans. I **1980**, 1499-1506.

91. Leucente, G.; Pinnen, F.; Romeo, A.; Zanotti, G. J. Chem. Soc., Perkin Trans. I **1983**, 1127-1130.

92. Ott, H.; Frey, A.J.; Hofmann, A. Tetrahedron **1963**, 19, 1675-1684.

93. Smith, S.; Timmins, G.M. J. Chem. Soc. **1937**, 396-401.

94. Stoll, A.; Hofmann, A.; Becker, B. Helv. Chim. Acta **1943**, 26, 1602-1613.

95. Stoll, A.; Hofmann, A. Ibid. **1950**, 33, 1705-1711.

96. Purdie, J.E.; Benoiton, N.L. J. Chem. Soc., Perkin Trans. 2 **1973**, 13, 1845-1852.

97. Lucente, G.; Pinnen, F.; Zanotti, G.; Cerrini, S.; Mazza, F.; Segre, A.L. J. Chem. Soc., Perkin Trans. 2 **1982**, 1169-1174.

98. Conti, F.; Lucente, G.; Romeo, A.; Zanotti, G. Int. J. Pept. Protein Res. **1973**, 5, 353-357.

99. "Handbock of Biochemistry"; Weast, R.C., Ed.; The Chemical Rubber Co.: Cleveland Ohio, 1970; 2nd Ed.,

p. J-59.

100. Hare, P.E. Carnegie Inst. Wash. Year Book **1974**, 73, 576-581.

101. King, K. Jr. Ref. 4, p. 377-391.

102. Pollock, G.E.; Cheng, C-N.; Cronin, S.E. Anal. Chem. **1977**, 49, 2-7.

103. Dungworth, G.; Vrenker, J.A.Th.; Schwartz, A.W. Comp. Biochem. Physiol. **1975**, 51B, 331-335.

104. Crout, D.H.G. "Peptide and Amino Acid Biosynthesis" In 'International Review of Science: Organic Chemistry Series Two'; Rydon, H.N., Ed.; Butterworth: London, 1976; Vol. 6, pp 281-234.

105. Brigham, J.K. Can. J. Earth Sci. **1983**, 20, 577-598.

106. Brigham, J.K. Ph.D. dissertation, University of Colorado, 1985.

107. Wehmiller, J.F. Quat. Sci. Rev. **1982**, 1, 83-120.

108. Mann, E.H.; Sandhouse, M.E.; Burg, J.; Fisher, G.H. Science (Washington D.C.) **1983**, 220, 1407-1408.

109. Williams, K.M.; Smith, G.G. Origins of Life **1977**, 8, 91-144.

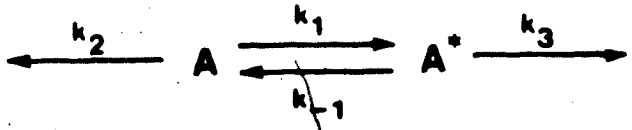
110. James, M.L.; Smith, G.M.; Wolford, J.C. "Applied Numerical Methods for Digital Computation with FORTRAN"; International Textbook Company: Scranton Pennsylvania, 1970; p 231-252.

111. Beckett, R.; Hunt, J. "Numerical Calculations and Algorithms"; McGraw-Hill: New York, 1967; p 30-32.

APPENDIX 1

ALGEBRAIC SOLUTION OF SCHEME 3

A mathematical description is presented here for racemization and irreversible hydrolysis of a peptide containing one epimerizing centre. This situation is depicted in Scheme 3.



SCHEME 3

The peptide substrate A is forming an equilibrium mixture with its epimer A*. The epimerizing peptide unit may have another chiral centre (as in ile or thr) so that $k_1 \neq k_{-1}$. The hydrolysis rates of the two diastereomers may be different as well, $k_2 \neq k_3$.

Scheme 3 is represented by the following pair of differential equations. Concentration brackets are left off for clarity.

$$\frac{dA}{dt} = -(k_1 + k_2)A + k_{-1}A^* \qquad (A1\text{-}1)$$

$$\frac{dA^*}{dt} = k_1A - (k_{-1} + k_3)A^* \qquad (A1\text{-}2)$$

The system of linear differential equations A1-1 and A1-2 has a solution:

$$A = a_1 e^{\lambda t} \qquad \text{(A1-3)}$$

$$A^* = a_2 e^{\lambda t} \qquad \text{(A1-4)}$$

where a_1, a_2 and λ are constants.

Substituting A1-3 and A1-4 in A1-1 and A1-2 yields the characteristic determinant. There is a non-trivial solution of A1-1 and A1-2 when:

$$\begin{vmatrix} -(k_1+k_2)-\lambda & k_{-1} \\ k_1 & -(k_{-1}+k_3)-\lambda \end{vmatrix} = 0$$

$$2\lambda = (-(k_1+k_2+k_{-1}+k_3) \pm ((K_1+k_2+k_{-1}+k_3)^2 -$$

$$4(k_1k_3+k_{-1}+k_2k_3))^{1/2}$$

Let

$$\Sigma k_i = -(k_1+k_2+k_{-1}+k_3)$$

and

$$I = ((k_1+k_2+k_{-1}+k_3)^2 - 4(K_1k_3+k_{-1}k_2+k_2k_3))^{1/2}$$

The two Eigenvalues λ_1 and λ_2 are then:

$$\lambda_1 = (\Sigma k_i - I)/2$$

$$\lambda_2 = (\Sigma k_i + I)/2$$

The general solution of system A1-1 and A1-2 is:

$$A = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} \qquad \text{(A1-5)}$$

$$A^* = c_3 e^{\lambda_1 t} + c_4 e^{\lambda_2 t} \qquad \text{(A1-6)}$$

where c_1 through c_4 are constants determined by the initial conditions

$$A_0 = c_1 + c_2 \qquad \text{(A1-7)}$$

$$A_0^* = c_3 + c_4 \qquad \text{(A1-8)}$$

From the derivative of A1-5, A1-1 and time zero conditions A1-7 and A1-8:

$$c_1 = (-(k_1+k_2+\lambda_2)A_0 + k_{-1}A_0^*)/(-I)$$

$$c_2 = A_0 - c_1$$

From the derivative of A1-6, A1-2 and time zero conditions A1-7 and A1-8:

$$c_3 = (k_1A_0 - (k_{-1}+k_3 + \lambda_2)A_0^*)/(-I)$$

$$c_4 = A_0^* - c_3$$

An interesting degenerate case develops when $k_1 \neq k_{-1}$ and $k_3 = k_2$. In this case:

$$\lambda_1 = -(k_1+k_{-1}+k_2)$$

$$\lambda_2 = -k_2$$

$$c_1 = \frac{k_1A_0 - k_{-1}A_0^*}{k_1+k_{-1}}$$

$$c_2 = \frac{k_{-1}}{k_1+k_{-1}}(A_0+A_0^*)$$

$$c_3 = -\left(\frac{k_1A_0 - k_{-1}A_0^*}{k_1+k_{-1}}\right) = -c_1$$

$$c_4 = \frac{k_1}{k_1+k_{-1}}(A_0+A_0^*)$$

Substituting into A1-5 and A1-6 yields:

$$A = \frac{k_1A_0 - k_{-1}A_0^*}{k_1+k_{-1}} e^{-(k_1+k_{-1}+k_2)t} + \frac{k_{-1}}{k_1+k_{-1}} (A_0+A_0^*)e^{-k_2t} \qquad \text{(A1-9)}$$

$$A^* = -\left(\frac{k_1A_0 - k_{-1}A_0^*}{k_1+k_{-1}}\right) e^{-(k_1+k_{-1}+k_2)t} + \frac{k_1}{k_1+k_{-1}} (A_0+A_0^*)e^{-k_2t} \qquad \text{(A1-10)}$$

Adding A1-9 to A1-10 gives:

$$A + A^* = (A_0+A_0^*)e^{-k_2t} \qquad \text{(A1-11)}$$

Let $K' = \frac{k_{-1}}{k_1}$

Then:

$$A - K'A^* = \frac{k_1^2A_0 - k_1k_{-1}A_0^* + k_1k_{-1}A_0 - k_{-1}^2A_0^*}{k_1(k_1+k_{-1})} e^{-(k_1+k_{-1}+k_2)t}$$

$$= (A_0 - K'A_0^*)e^{-(k_1=k_{-1}+k_2)t} \qquad \text{(A1-12)}$$

Dividing A1-12 by A1-11 gives

$$\frac{A-K'A^*}{A+A^*} = \left(\frac{A_0-K'A_0^*}{A_0+A_0^*}\right) e^{-(k_1+k_{-1})t} \qquad \text{(A1-13)}$$

Multiplying the left-hand side of Al-13 by 1/A/1/A and the right-hand side by $1/A_0/1/A_0$ and taking the natural logarithm of both sides yields the familiar canonical equation:

$$\ln\left(\frac{1 - K'\frac{A^*}{A}}{1 + \frac{A^*}{A}}\right) - \ln\left(\frac{1 - K'\frac{A^*}{A}}{1 + \frac{A^*}{A}}\right)_{t=0} = -(k_1+k_{-1})t \qquad \text{(Al-14)}$$

This result means that a peptide undergoing racemization and competitive irreversible hydrolysis will follow the above reversible first-order expression if the rates of hydrolysis of the diastereomers are equal.

The case where $k_3 \neq k_2$ and $k_1 \neq k_{-1}$ will now be examined. Now the constants c_1 through c_4 take on the values:

$$c_1 = (-(k_1+k_2+\lambda_2)A_0+k_{-1}A_0^*)/(-I)$$

$$c_2 = (-(k_1+k_2+\lambda_1)A_0+k_{-1}A_0^*)/I$$

$$c_3 = (-(k_{-1}+k_3+\lambda_2)A_0^*+k_1A_0)/(-I)$$

$$c_4 = (-(k_{-1}+k_3+\lambda_1)A_0^*+k_1A_0)/I$$

In order to produce an expression like the canonical equation, expressions like A1-11 and A1-12 must be found.

Hence find z such that $c_1 + zc_3 = 0$.

$$z = \frac{c_1}{-c_3} = \frac{-(k_1+k_2+\lambda_2)A_0+k_{-1}A_0^*}{(k_{-1}+k_3+\lambda_2)A_0^*-k_1A_0} \qquad \text{(A1-15)}$$

so that

$$A + zA^* = (c_2 + zc_4)e^{\lambda_2 t} \qquad \text{(A1-16)}$$

Also find x such that $c_2 - xc_4 = 0$

$$x = \frac{c_2}{c_4} = \frac{-(k_1+k_2+\lambda_1)A_0+k_{-1}A_0^*}{-(k_{-1}+k_3+\lambda_1)A_0^*+k_1A_0} \qquad \text{(A1-17)}$$

so that

$$A = xA^* = (c_1 + xc_3)e^{\lambda_1 t} \qquad \text{(A1-18)}$$

Dividing A1-18 by A1-16 gives:

$$\frac{A - xA^*}{A + zA^*} = \left(\frac{c_1 - xc_3}{c_2 + zc_4}\right) e^{-It} \qquad \text{(A1-19)}$$

$$\frac{c_1 - xc_3}{c_2 + zc_4} = \frac{c_1 - \frac{c_2}{c_4} \cdot c_3}{c_2 + \frac{c_1}{-c_3} \cdot c_4}$$

$$= \frac{\frac{c_1c_4 - c_2c_3}{c_4}}{\frac{-c_2c_3 + c_1c_4}{-c_3}}$$

$$= -\frac{c_3}{c_4}$$

Al-18 becomes:

$$\frac{A - xA^*}{A + zA^*} = ye^{-It} \qquad \text{(Al-20)}$$

where

$$y = \frac{-(k_{-1}+k_3+\lambda_2)A_0^* + k_1A_0}{-(k_{-1}+k_3+\lambda_1)A_0^* + k_1A_0} \qquad \text{(Al-21)}$$

Computer Program SGC

A computer program was written* to simulate the situation illustrated by Scheme 3. The user is asked to input the values of the rate constants, the initial

*The language used was Commodore BASIC V.2.

conditions of A and A*, the duration of the experiment and the sampling interval. The user must then choose the equation in which the calculated values of A and A* are to be substituted. A "Y" input to the prompt "NORMAL PLOT?" uses equation A1-14 with K' = 1. A "N" input to the same prompt followed by a "Y" input to the prompt "K' PLOT" uses equation A1-14 with $K' = k_{-1}/k_1$. Answering "N" to both prompts gives output to the printer using equation A1-20 (in natural log form). The program listing follows. Note that the printer is the output device. The values of A, A*, (A-A*)/(A+A*) and ln((A-A*)/(A+A*)) are printed out when "Y" is entered in response to the first prompt; A, A*, (A-K'A*)/(A+A*) and ln((A-A*)/(A+A*)) when "Y" is entered in response to the second prompt and A, A*, (A-xA*)/(A+zA*) and ln((A-xA*)/(A+zA*)) otherwise. The program is protected against calculating the logarithm of a negative number.

```
SGC

100 PRINT"RATE CONSTANTS:"
105 INPUT"K1";K1
110 INPUT"K-1";K0
120 INPUT"K2";K2
130 INPUT"K3";K3
140 PRINT"INITIAL COND."
150 INPUT"A  ";A0
160 INPUT"A$ ";S0
170 INPUT"DURATION";DU
180 INPUT"TIME INC.";IT
185 INPUT"NORMAL PLOT";SP$:IFSP$="Y"THENSP$="R":GOTO190
186 INPUT"K' PLOT";SP$:IFSP$="Y"THEN190
190 OPEN4,4:PRINT#4,CHR$(17)"RATE CONSTANTS:"CHR$(145)"  K1 ="K1" K-1 ="K0"
K2 = "K2" K3 ="K3
192 PRINT#4,CHR$(17)"INITIAL CONDITIONS:"CHR$(145)"  A ="A0" A$ ="S0
194 PRINT#4:PRINT#4
195 PRINT#4,CHR$(16)"02TIME"CHR$(16)"17A"CHR$(16)"33A$";
196 IFSP$="R"THENPRINT#4,CHR$(16)"44(A-A$)/(A+A$)";:GOTO199
197 IFSP$="Y"THENPRINT#4,CHR$(16)"43(A-K'A$)/(A+A$)";:GOTO199
198 PRINT#4,CHR$(16)"41(A-XA$)/(A+ZA$)";
199 PRINT#4,CHR$(16)"67LN":PRINT#4
200 SK=-(K0+K1+K2+K3)
210 I=SQR(SK^2-4*(K1*K3+K0*K2+K2*K3))
220 L1=(SK-I)/2:L2=(SK+I)/2
230 C1=(-(K1+K2+L2)*A0+K0*S0)/(-I):C2=A0-C1
240 C3=(-(K0+K3+L2)*S0+K1*A0)/(-I):C4=S0-C3
245 GOSUB600
250 FORT=0TODUSTEPIT
260 A=C1*EXP(L1*T)+C2*EXP(L2*T)
270 AS=C3*EXP(L1*T)+C4*EXP(L2*T)
275 IFSP$="R"THENCN=(A-AS)/(A+AS):GOTO290
276 IFSP$="N"THENCN=(A-XX*AS)/(A+ZZ*AS):GOTO290
280 CN=(A-(K0/K1)*AS)/(A+AS)
290 PRINT#4,TCHR$(16)"11"ACHR$(16)"27"AS;CHR$(16)"43"CN;
300 IFCN>0THENPRINT#4,CHR$(16)"59"LOG(CN):GOTO320
310 PRINT#4
320 NEXTT
500 CLOSE4:END
600 XX=(-(K1+K2+L1)*A0+K0*S0)/(-(K0+K3+L1)*S0+K1*A0)
610 ZZ=(-(K1+K2+L2)*A0+K0*S0)/((K0+K3+L2)*S0-K1*A0)
620 RETURN
1000 OPEN15,8,15:INPUT#15,A$,B$,C$,D$:PRINTA$B$C$D$:CLOSE15
```

APPENDIX 2

ALGEBRAIC SOLUTION OF SCHEME 5

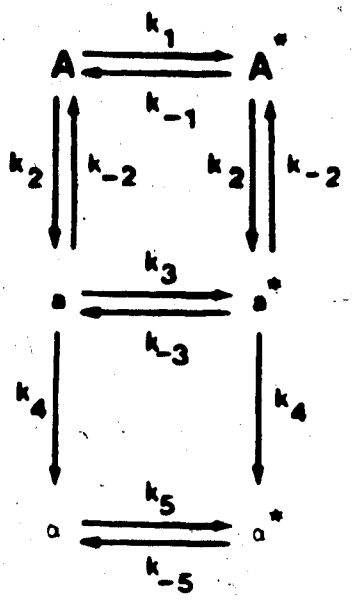
A mathematical description is presented here for racemization and hydrolysis of a peptide with only one chiral centre. This chiral centre is also the one which is racemizing. This situation is represented in Scheme 5.

Substrate peptide A can reversibly form intermediate a. Intermediate a can hydrolyze to the amino acid α. Simultaneously, A, a and α are forming a racemic mixture with enantiomers A*, a* and α*, respectively. There is mirror image hydrolytic pathway linking A*, a* and α*. Since there is only one chiral centre, $k_1 = k_{-1}$, $k_3 = k_{-3}$ and $k_5 = k_{-5}$.

The concentrations of A, a and α and the enantiomers A*, a* and α* respectively can be obtained by the solution of the following system of differential equations. Concentration brackets are left out for clarity.

$$\frac{dA}{dt} = -(k_1+k_2)A + k_1A^* + k_{-2}a \qquad (A2\text{-}1)$$

$$\frac{dA^*}{dt} = k_1A - (k_1+k_2)A^* + k_{-2}a^* \qquad (A2\text{-}2)$$



SCHEME 5

$$\frac{da}{dt} = k_2A - (k_3+k_{-2}+k_4)a + k_3a^* \qquad \text{(A2-3)}$$

$$\frac{da^*}{dt} d_2A^* + k_3a - (k_3+k_{-2}+k_4)a^* \qquad \text{(A2-4)}$$

$$\frac{d\alpha}{dt} = k_4a - k_5\alpha + k_5\alpha^* \qquad \text{(A2-5)}$$

$$\frac{d\alpha^*}{dt} k_4a^* + k_5\alpha - k_5\alpha^* \qquad \text{(A2-6)}$$

From A2-1 and A2-2:

$$\frac{d}{dt}(A+A^*) = -k_2(A+A^*) + k_{-2}(a+a^*) \qquad \text{(A2-7)}$$

From A2-3 and A2-4:

$$\frac{d}{dt}(a+a^*) = k_2(A+A^*) - (k_{-2}+k_4)(a+a^*) \qquad \text{(A2-8)}$$

The system of linear differential equations A2-7 and A2-8 is assumed to have the solution:

$$A+A^* = b_1e^{\lambda t} \qquad \text{(A2-9)}$$

$$a+a^* = b_2e^{\lambda t} \qquad \text{(A2-10)}$$

where b_1, b_2 and λ are constants.

Substituting A2-9 and A2-10 in A2-7 and A2-8, a non-

trivial solution occurs when:

$$(-k_2-\lambda)(-(k_{-2}+k_4)-\lambda) - k_2k_{-2} = 0$$

$$\lambda = (-(k_2+k_{-2}+k_4) \pm ((k_2+k_{-2}+k_4)^2 - 4k_2k_4)^{1/2})/2$$

Let

$$\lambda_1 = (-(k_2+k_{-2}+k_4) - ((k_2+k_{-2}+k_4)^2 - 4k_2k_4)^{1/2})/2$$

and

$$\lambda_2 = (-(k_2+k_{-2}+k_4) + ((k_2+k_{-2}+k_4)^2 - 4k_2k_4)^{1/2}/2$$

The general solution to system A2-7 and A2-8 is then:

$$A+A^* = c_1e^{\lambda_1 t} + c_2e^{\lambda_2 t} \qquad \text{(A2-11)}$$

$$a+a^* = c_3e^{\lambda_1 t} + c_4e^{\lambda_2 t} \qquad \text{(A2-12)}$$

where c_1 through c_4 are constants to be solved from the initial conditions:

$$(A+A^*)_0 = c_1+c_2 \qquad \text{(A2-13)}$$

$$(a+a^*)_0 = 0 \qquad \text{(A2-14)}$$

By differentiating A2-11, setting the result at time zero equal to A2-7 at time zero:

$$-k_2(A+A^*)_0 = \lambda_1 c_1 + \lambda_2 c_2 \qquad \text{(A2-15)}$$

Solving A2-15 and A2-13:

$$c_1 = \frac{-(A+A^*)_0}{\lambda_1 - \lambda_2} \cdot (k_2+\lambda_2)$$

$$c_2 = (A+A^*)_0 - c_1$$

By a similar route:

$$c_3 = \frac{k_2(A+A^*)_0}{\lambda_1 - \lambda_2}$$

$$c_4 = -c_3$$

From A2-5 and A2-6:

$$\frac{d}{dt}(\alpha+\alpha^*) = k_4(a+a^*)$$

$$\alpha+\alpha^* = k_4 c_3\left\{\frac{e^{\lambda_1 t}}{\lambda_1} - \frac{e^{\lambda_2 t}}{\lambda_2}\right\} + c_{10} \qquad \text{(A2-16)}$$

At t_∞, $\alpha+\alpha^* = (A+A^*)_0$ so that $c_{10} = (A+A^*)_0$.

The solution A2-16 will fail when $k_2 = k_4$ and $k_{-2} = 0$. Under these conditions, by inspection:

$$(A+A^*) = (A+A^*)_0 e^{-k_2 t} \qquad \text{(A2-17)}$$

From A2-8 and A2-17

$$\frac{d}{dt}(a+a^*) + k_2(a+a^*) = k_2(A+A^*)_0 e^{-k_2 t} \qquad \text{(A2-18)}$$

A particular solution to A2-18 is:

$$(a+a^*) = b_1 t e^{-k_2 t}$$

$$\frac{d}{dt}(a+a^*) = -k_2 b_1 t e^{-k_2 t} + b_1 e^{-k_2 t}$$

Substituting in A2-18 and equating coefficients:

$$b_1 = k_2(A+A^*)_0$$

The general solution of A2-18 is:

$$(a+a^*) = (k_2(A+A^*)_0 t + c_a)e^{-k_2 t}$$

From A2-14:

$$c_a = 0$$

Thus when $k_2 = k_4$ and $k_{-2} = 0$:

$$(a+a^*) = k_2(A+A^*)_0 te^{-k_2 t} \qquad \text{(A2-19)}$$

From A2-1 and A2-2:

$$\frac{d}{dt}(A-A^*) = -(2k_1+k_2)(A-A^*) + k_{-2}(a-a^*) \qquad \text{(A2-20)}$$

From A2-3 and A2-4:

$$\frac{d}{dt}(a-a^*) = k_2(A-A^*) - (2k_3+k_{-2}+k_4)(a-a^*) \qquad \text{(A2-21)}$$

The non-trivial solution to system A2-20 and A2-21 has Eigenvalues λ_3 and λ_4 when:

$$\lambda_3 = (\Sigma_2 k_i - I_2)$$

$$\lambda_4 = (\Sigma_2 k_i + I_2)$$

with

$$\Sigma_2 k_i = -(2k_1+k_2+k_{-2}+2k_3+k_4)$$

and

$$I_2 = ((\Sigma_2 k_i)^2 - 4(4k_1k_3+2k_1k_{-2}+2k_1k_4+2k_2k_3+k_2k_4))^{1/2}$$

The general solution of the system A2-20 and A2-21 is then:

$$A-A^* = c_5 e^{\lambda_3 t} + c_6 e^{\lambda_4 t} \qquad \text{(A2-22)}$$

$$a-a^* = c_7 e^{\lambda_3 t} + c_8 e^{\lambda_4 t} \qquad \text{(A2-23)}$$

where c_5 through c_8 are constants to be solved from the initial conditions:

$$(A-A^*)_0 = c_5 + c_6 \qquad \text{(A2-24)}$$

$$(\alpha-\alpha^*)_0 = 0 \qquad \text{(A2-25)}$$

From A2-20, the derivative of A2-22 and A2-24 and A2-25:

$$c_5 = \frac{-(A-A^*)_0}{\lambda_3-\lambda_4} \cdot (2k_1+k_2+\lambda_4)$$

$$c_6 = (A-A^*)_0 - c_5$$

From A2-21, the derivative of A2-23 and A2-24 and A2-25:

$$c_7 = \frac{k_2(A-A^*)_0}{\lambda_3-\lambda_4}$$

$$c_8 = -c_7$$

Hence

$$a-a^* = \frac{k_2(A-A^*)_0}{\lambda_3-\lambda_4}\{e^{\lambda_3 t} - e^{\lambda_4 t}\} \qquad \text{(A2-26)}$$

From A2-5 and A2-6:

$$\frac{d}{dt}(\alpha-\alpha^*) = -2k_5(\alpha-a^*) + k_4(a-a^*) \qquad \text{(A2-27)}$$

Substituting A2-26 in A2-27:

$$\frac{d}{dt}(\alpha-\alpha^*) + 2k_5(\alpha-\alpha^*) = k_4c_7e^{\lambda_3 t} - k_4c_7e^{\lambda_4 t} \qquad \text{(A2-28)}$$

A particular solution of A2-28 must be:

$$(\alpha-\alpha^*) = b_1e^{\lambda_3 t} + b_2e^{\lambda_4 t} \qquad \text{(A2-29)}$$

where b_1 and b_2 are constants.

Differentiating A2-29 gives:

$$\frac{d}{dt}(\alpha-\alpha^*) = \lambda_3 b_1 e^{\lambda_3 t} + \lambda_4 b_2 e^{\lambda_4 t} \qquad \text{(A2-30)}$$

Substituting A2-29 and A2-30 in A2-28 and equating terms in $e^{\lambda_3 t}$ and $e^{\lambda_4 t}$ gives:

$$b_1 = \frac{k_4 c_7}{\lambda_3 + 2k_5}$$

$$b_2 = \frac{-k_4 c_7}{\lambda_4 + 2k_5}$$

The general solution of A2-28 is:

$$\alpha-\alpha^* = c_9 e^{-2k_5 t} + \frac{k_4 c_7}{\lambda_3 + 2k_5} e^{\lambda_3 t} - \frac{k_4 c_7}{\lambda_4 - 2k_5} e^{\lambda_4 t} \qquad \text{(A2-31)}$$

From the time zero condition A2-21:

$$c_9 = k_4 c_7 \left\{ \frac{1}{\lambda_4 + 2k_5} - \frac{1}{\lambda_3 + 2k_5} \right\}$$

The solution A2-31 will fail when $\lambda_3 = -2k_5$ or when $\lambda_4 = -2k_5$.

When $\lambda_3 = -2k_5$ A2-28 becomes:

$$\frac{dy}{dt} + 2k_5 y = k_4 c_7 e^{-2k_5 t} - k_4 c_7 e^{\lambda_4 t} \qquad \text{(A2-32)}$$

where $y = \alpha - \alpha^*$. A particular solution of A2-32 is:

$$y = b_1 t e^{-2k_5 t} + b_2 e^{\lambda_4 t} \qquad \text{(A2-33)}$$

Substituting the derivative of A2-33 and A2-33 in A2-32 and equating coefficients of $e^{-2k_5 t}$ and $e^{\lambda_4 t}$:

$$b_1 = k_4 c_7$$

$$b_2 = \frac{-k_4 c_7}{2k_5 + \lambda_4}$$

The general solution of A2-32 is therefore:

$$y = (k_4 c_7 t + c_a) e^{-2k_5 t} - \frac{k_4 c_7}{2k_5 + \lambda_4} e^{\lambda_4 t}$$

At t_0, $\alpha - \alpha^* = 0$ so that

$$c_a = \frac{k_4 c_7}{2k_5 + \lambda_4}$$

Thus when $\lambda_3 = -2k_5$:

$$\alpha - \alpha^* = k_4 c_7 \left\{ \left(t + \frac{1}{2k_5 + \lambda_4}\right) e^{-k_5 t} - \frac{e^{\lambda_4 t}}{2k_5 + \lambda_4} \right\} \qquad \text{(A2-34)}$$

The solution of A2-24 when $\lambda_4 = -2k_5$ is arrived at in a similar manner. By analogy to A2-34, this solution can be written down:

$$\alpha-\alpha^* = k_4c_7\{\frac{e^{\lambda_3 t}}{2k_5+\lambda_3} - (t + \frac{1}{\lambda_3+2k_5})e^{-2k_5t}\} \quad (A2\text{-}35)$$

An algebraic solution of Scheme 5 is now complete.

A computer program was written* to simulate the kinetics of Scheme 5 given at the beginning of this appendix. The user of the program is asked to input values of the rate constants, the initial conditions of A and A*, the duration of the simulation and the sampling interval. The program generates the values of A, A*, a, a*, α, α^*, the argument of the canonical expressions (e.g. (A-A*)/(A+A*)) equivalent to (1-D/L)/(1+D/L) and the value of the canonical expressions. The program listing follows. Note that the printer is the output device.

*The language used was Commodore BASIC V.2.

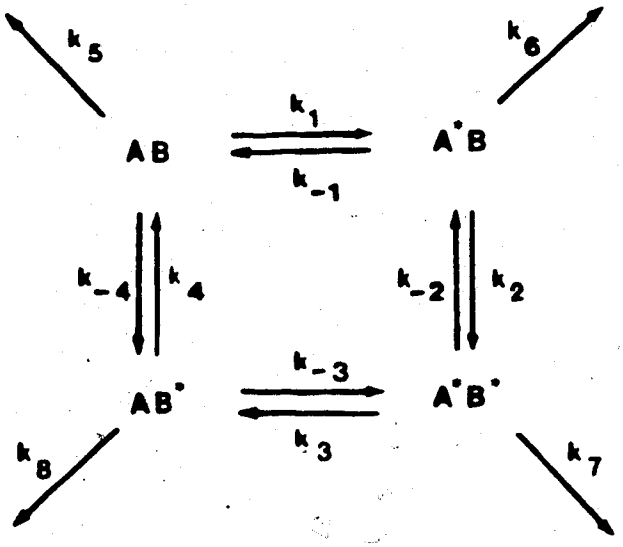
```
100 PRINT"RATE CONSTANTS:"
105 INPUT"K1";K1
110 INPUT"K2";K2
115 IFK2=0THENPRINT"TRIVIAL CASE":END
120 INPUT"K-2";K0
125 INPUT"K3";K3
130 INPUT"K4";K4
135 INPUT"K5";K5
140 PRINT"INITIAL COND."
145 INPUT"A  ";A0
150 INPUT"A$ ";B0
155 INPUT"DURATION";DU
160 INPUT"TIME INC.";IT
165 DIMA(INT(DU/IT))
170 OPEN4,4:PRINT#4,CHR$(17)"RATE CONSTANTS:"CHR$(145)"  K1 ="K1" K2 ="K2"
K-2 = "K0" K3 ="K3
175 PRINT#4,CHR$(16)"17";"K4 ="K4" K5 ="K5
180 PRINT#4,CHR$(17)"INITIAL CONDITIONS:"CHR$(145)"  A ="A0" A$ ="B0
185 PRINT#4:PRINT#4
190 PRINT#4,CHR$(16)"02TIME"CHR$(16)"17A"CHR$(16)"33A$";
195 PRINT#4,CHR$(16)"41(A-A$)/(A+A$)";
200 PRINT#4,CHR$(16)"67LN":PRINT#4
205 S2=-(2*K1+2*K3+K2+K0+K4)
210 S1=-(K0+K2+K4)
215 I2=SQR(S2^2-4*(4*K1*K3+2*K1*K0+2*K1*K4+2*K2*K3+K2*K4))
220 I1=SQR(S1^2-4*K2*K4)
225 L3=(S2-I2)/2:L4=(S2+I2)/2
230 L1=(S1-I1)/2:L2=(S1+I1)/2
235 C5=-(2*K1+K2+L4)*(A0-B0)/(L3-L4):C6=(A0-B0)-C5
240 C7=K2*(A0-B0)/(L3-L4):C8=-C7
245 IFK0=0ANDK2=K4THEN255
250 C1=-(A0+B0)*(K2+L2)/(L1-L2):C2=A0+B0-C1:C3=K2*(A0+B0)/(L1-L2):C4=-C3
255 GOSUB450
260 FORT=0TODUSTEPIT
265 AN=FNAN(T)
270 AP=FNAP(T)
275 CN=AN/AP
280 A=(AP+AN)/2:AS=(AP-AN)/2:A(INT(T/IT))=A
285 PRINT#4,TCHR$(16)"11"ACHR$(16)"27"AS;CHR$(16)"43"CN;
290 IFCN>0THENPRINT#4,CHR$(16)"59"LOG(CN);
295 PRINT#4:NEXTT:A$="A":B$="A$":GOSUB430
300 FORT=ITTODUSTEPIT
305 BN=FNBN(T)
310 BP=FNBP(T)
315 DN=BN/BP
320 B=(BP+BN)/2:BS=(BP-BN)/2:I=INT(T/IT):A(I)=A(I)+B
325 PRINT#4,TCHR$(16)"11"BCHR$(16)"27"BS;CHR$(16)"43"DN;
330 IFDN>0THENPRINT#4,CHR$(16)"59"LOG(DN);
335 PRINT#4:NEXTT
340 IFK4=0THEN395
345 A$="AA":B$="AA$":GOSUB430
350 C9=K4*C7*(1/(L4+2*K5)-1/(L3+2*K5))
355 FORT=ITTODUSTEPIT
360 BN=FNBN(T)
365 BP=FNBP(T)
370 HN=BN/BP
375 B=(BP+BN)/2:BS=(BP-BN)/2:I=INT(T/IT):A(I)=A(I)+B
380 PRINT#4,TCHR$(16)"11"BCHR$(16)"27"BS;CHR$(16)"43"HN;
385 IFHN>0THENPRINT#4,CHR$(16)"59"LOG(HN);
390 PRINT#4:NEXTT
```

```
395 A$="[illegible]":B$="[illegible]":GOSUB430:IC=A0+B0
400 FORT=0TOOUSTEP IT:A=A(INT(T/IT))
405 A$=IC-A:CN=(A-A$)/IC
410 PRINT#4,TCHR$(16)"11"ACHR$(16)"27"A$;CHR$(16)"43"CN;
415 IFCN>0THENPRINT#4,CHR$(16)"59"LOG(CN);
420 PRINT#4:NEXTT
425 PRINT#4:CLOSE4:END
430 PRINT#4
435 PRINT#4,CHR$(16)"02TIME"CHR$(17)CHR$(16)"12"A$CHR$(16)"33"B$;
440 PRINT#4,CHR$(17)CHR$(16)"41("A$"-"B$")/("A$"+"B$")"CHR$(145);
445 PRINT#4,CHR$(16)"67LN":PRINT#4:RETURN
450 DEFFNAN(T)=C5*EXP(L3*T)+C6*EXP(L4*T)
455 DEFFNAP(T)=C1*EXP(L1*T)+C2*EXP(L2*T)
460 DEFFN[illegible](T)=C7*EXP(L3*T)+C8*EXP(L4*T)
465 DEF[illegible]T)=C3*EXP(L1*T)+C4*EXP(L2*T)
470 DEF[illegible]T)=C9*EXP(-2*K5*T)+K4*C7*(EXP(L3*T)/(L3+2*K5)-EXP(L4*T)/(L4+2*K5))
475 DEF[illegible](T)=K4*C3*(EXP(L1*T)/L1-EXP(L2*T)/L2)+A0+S0
480 IFABS(L3+2*K5)<1E-4THENDEFFNBN(T)=K4*C7*((T+1/(2*K5+L4))*EXP(-2*K5*T)-EXP(L4
*T)/(2*K5+L4))
485 IFABS(L4+2*K5)<1E-4THENDEFFNBN(T)=K4*C7*(EXP(L3*T)/(2*K5+L3)-(T+1/(L3+2*K5))
*EXP(-2*K5*T))
490 IFK0>0ORK2<>K4THENRETURN
495 DEFFNAP(T)=(A0+B0)*EXP(-K2*T)
500 DEFFNBP(T)=K2*(A0+B0)*T*EXP(-K2*T)
505 DEFFNBP(T)=(A0+B0)*(1-(K2*T+1)*EXP(-K2*T))
510 RETURN
515 OPEN15,8,15:INPUT#15,A$,B$,C$,D$:PRINTA$B$C$D$:CLOSE15:END
```

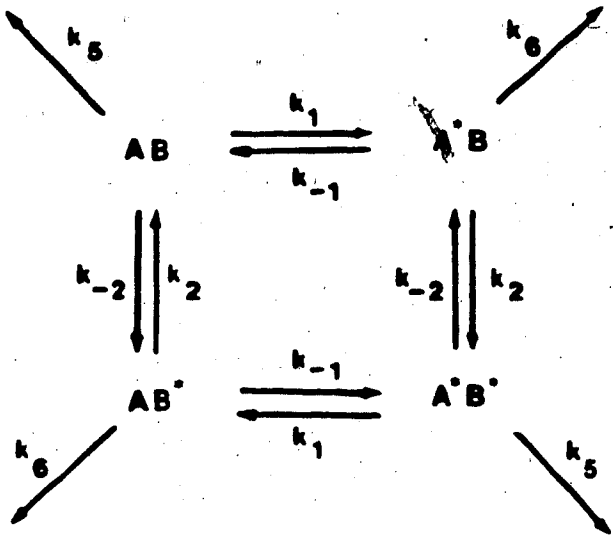
APPENDIX 3

ALGEBRAIC SOLUTION OF SCHEME 6

A mathematical description is presented here for racemization and irreversible hydrolysis of a peptide containing two epimerizing chiral centres. The peptide AB, where A and B denote the chiral centres, can undergo epimerization to diastereomers A*B, AB* and A*B*, where A* and B* are the enantiomers of A and B respectively. Assuming that epimerization occurs at one centre at a time and that hydrolysis and epimerization are independent processes, the process of epimerization and hydrolysis can be illustrated by the following scheme:



Since AB and A*B* are enantiomers and A*B and AB* are also enantiomers: $k_7 = k_5$, $k_8 = k_6$, $k_3 = k_1$, $k_{-3} = k_{-1}$, $k_4 = k_2$ and $k_{-4} = k_{-2}$. The kinetic scheme can be simplified (Scheme 6).



SCHEME 6

To simplify the algebraic statements, let A be the sum of the concentrations of those species which contain centre A and not A*:

$$A = AB + AB^* \qquad (A3\text{-}1)$$

Similarly

$$A^* = A^*B + A^*B^* \qquad \text{(A3-2)}$$

$$B = A^*B + AB^* \qquad \text{(A3-3)}$$

$$B^* = AB^* + A^*B^* \qquad \text{(A3-4)}$$

Concentration brackets are left off for clarity. Scheme 6 is represented by the following set of differential equations:

$$\frac{d}{dt}AB = k_2AB^* + k_{-1}A^*B - (k_1+k_{-2}+k_5)AB \qquad \text{(A3-5)}$$

$$\frac{d}{dt}A^*B = k_1AB + k_{-2}A^*B^* = (k_{-1}+k_2+k_6)A^*B \qquad \text{(A3-6)}$$

$$\frac{d}{dt}AB^* = k_{-2}AB + k_1A^*B^* - (k_{-1}+k_2+k_6)AB^* \qquad \text{(A3-7)}$$

$$\frac{d}{dt}A^*B^* = k_{-1}AB^* + k_2A^*B - (k_1+k_{-2}+k_5)A^*B^* \qquad \text{(A3-8)}$$

This system of linear differential equations can be solved directly but the algebra can be simplified by combining the variables to produce two systems of two variables each. Hence:

$$\frac{d}{dt}(A-A^*) = -(2k_{-1}+k_6)(AB^*-A^*B) - (2k_1+k_5)(AB-A^*B^*) \qquad \text{(A3-9)}$$

From A3-1 through A3-4:

$$AB^*-A^*B = ((A-A^*) - (B-B^*))/2 \qquad \text{(A3-10)}$$

and

$$AB-A^*B^* = ((A-A^*) + (B-B^*))/2 \qquad \text{(A3-11)}$$

Substituting A3-10 and A3-11 in A3-9 gives:

$$\frac{d}{dt}(A-A^*) = -(k_1+k_{-1}+\frac{k_6+k_5}{2})(A-A^*) + (k_{-1}-k_1+\frac{k_6k_5}{2})(B-B^*) \qquad \text{(A3-12)}$$

Similarly

$$\frac{d}{dt}(B-B^*) = (2k_1+k_6)(AB^*-A^*B) = (2k_{-2}+k_5)(AB-A^*B^*)$$

$$= (k_2-k_{-2}+\frac{k_6-k_5}{2})(A-A^*) - (k_2+k_{-2}+\frac{k_6+k_5}{2})(B-B^*) \qquad \text{(A3-13)}$$

Let

$$W = k_1+k_{-1} + \frac{k_6+k_5}{2}$$

$$X = k_{-1}-k_1 + \frac{k_6-k_5}{2} \qquad \text{(A3-14)}$$

$$Y = k_2 - k_{-2} + \frac{k_6 - k_5}{2}$$

$$Z = k_2 + k_{-2} + \frac{k_6 + k_5}{2}$$

Suppose that

$$A - A^* = c_1 e^{\lambda t} \qquad \text{(A3-15)}$$

and

$$B - B^* = c_2 e^{\lambda t} \qquad \text{(A3-16)}$$

is a solution to A3-12 and A3-13.

Substitution of A3-15, A3-16 and A3-14 into A3-12 and A3-13 gives:

$$\lambda c_1 e^{\lambda t} = -W c_1 e^{\lambda t} + X c_2 e^{\lambda t}$$

$$\lambda c_2 e^{\lambda t} = Y c_1 e^{\lambda t} - Z c_2 e^{\lambda t}$$

A non-trivial solution exists only when

$$\begin{vmatrix} \lambda + W & -X \\ -Y & \lambda + Z \end{vmatrix} = 0$$

Therefore:

$$\lambda = (-(W+Z) \pm ((W+Z)^2 - 4(WZ-XY))^{1/2})/2$$

The general solution to system A3-12 and A3-13 is:

$$A-A^* = c_1 \exp\{\frac{-(W+Z+I)}{2}\} t + c_3 \exp\{\frac{-(W+Z-I)}{2}\} t \quad \text{(A3-17)}$$

$$B-B^* = c_2 \exp\{\frac{-(W+Z+I)}{2}\} t + c_4 \exp\{\frac{-(W+Z-I)}{2}\} t \quad \text{(A3-18)}$$

where $I = ((W+Z)^2 - 4(WZ-XY))^{1/2}$.

With a solution for A+A*, the mathematical solution of Scheme 6 can be expressed in a canonical form. For A+A*:

$$\frac{d}{dt}(A+A^*) = -k_6(AB^*+A^*B) - k_5(AB+A^*B^*)$$

$$= -(\frac{k_6+k_5}{2})(A+A^*) + \frac{k_6-k_5}{2}(AB+A^*B^* - AB^*-A^*B) \quad \text{(A3-19)}$$

$$\frac{d}{dt}(AB+A^*B^* - AB^*-A^*B) = (2k_{-1}+2k_2+k_6)(AB^*+A^*B) -$$

$$(2k_1+2k_{-2}+k_5)(AB+A^*B^*)$$

$$= (k_{-1}-k_1+k_2-k_{-2}+\frac{k_6-k_5}{2})(A+A^*) - (k_{-1}-k_1+k_2-k_{-2}+\frac{k_6+k_5}{2})$$

$$(AB+A^*B^* - AB^*+A^*B) \qquad \text{(A3-20)}$$

A3-19 and A3-20 form another system of linear differential equations which can be solved in the same manner as before.

Let

$$Q = (k_6+k_5)/2$$

$$R = (k_6-k_5)/2$$

$$S = k_{-1}-k_1+k_2-k_{-2} + R$$

$$T = k_1+k_{-1}+k_2+k_{-2} + Q$$

Then the solution of A3-19 and A3-20 is:

$$A+A^* = c_a\exp\{\frac{-(Q+T)+I'}{2} \cdot t\} + c_b\exp\{\frac{-(Q+T)-I'}{2} \cdot t\} \qquad \text{(A3-21)}$$

$$AB+A^*B^* - AB^*-A^*B =$$

$$c_c \exp\left\{\frac{-(Q+T)+I'}{2} \cdot t\right\} + c_d \exp\left\{\frac{-(Q+T)-I'}{2} \cdot t\right\} \qquad \text{(A3-22)}$$

where $I' = ((Q+T)^2 - 4(QT-SR))^{1/2}$.

Constants c_1 through c_4 and c_a through c_d can be determined from the initial conditions: A_0, the initial D/L ratio for the A centre; B_0, the initial D/L ratio for the B centre; R_0, the initial ratio of diastereomers $(AB+A^*B^*)/(AB^*+A^*B)$; I_0, the initial total concentration of peptide.

From these definitions the initial concentrations of each of the diastereomers can be calculated.

$$A^*B_0 = I_0(B_0/(B_0+1) - R_0$$

$$A^*B_0^* = I_0/(A_0+1) - A^*B_0$$

$$AB_0^* = I_0/(1-R_0) - A^*B_0 - A^*B_0^* \qquad \text{(A3-23)}$$

$$AB_0 = I_0 - A^*B_0 - A^*B_0^* - AB_0^*$$

By employing the definitions in A3-1 through A3-4 and substituting A3-23, values of $(A-A^*)_0$, $(B-B^*)_0$, $(A+A^*)_0$ and $(AB+A^*B^*-A^*B-AB^*)_0$ are obtained.

From A3-12 and the derivative of A3-17 at time zero:

$$c_3 = (-W(A-A^*)_0 + X(B-B^*)_0 - (-W-Z+I)(A-A^*)_0/2)/(-I) \quad \text{(A3-24)}$$

$$c_1 = (A-A^*)_0 - c_3 \quad \text{(A3-25)}$$

From A3-13 and the derivative of A3-18 at time zero:

$$c_4 = (Y(A-A^*)_0 + Z(B-B^*)_0 - (-W-Z+I)(B-B^*)_0/2)/(-I) \quad \text{(A3-26)}$$

$$c_2 = (B-B^*)_0 - c_4 \quad \text{(A3-27)}$$

Similarly from A3-19, A3-20 and the derivatives of A3-21 and A3-22 at time zero:

$$c_b = (-Q(A+A^*)_0 + R(AB_0+A^*B_0^*-A^*B+A^*B^*)_0 - (-Q-T+I')(A+A^*)_0/2))/(-I') \quad \text{(A3-28)}$$

$$c_a = (A+A^*)_0 - c_b \quad \text{(A3-29)}$$

$$c_d = (S(A+A^*)_0 - T(AB+A^*B^*-A^*B-AB^*)_0 -$$

$$(Q-TX+I')(AB+A^*B^*-A^*B-AB^*)_0/2)/(-I') \qquad \text{(A3-30)}$$

$$c_c = (AB+A^*B^*-A^*B-AB^*)_0 - c_d \qquad \text{(A3-31)}$$

Conditions for Asymmetric Induction

Asymmetric induction occurs, for example, when starting from excess A or B, the course of the reaction would lead to an excess of A* or B* at some time. Suppose that excess A* is produced. The necessary condition is that A-A* = 0 for some positive time tz. From equation A3-17 this occurs when:

$$tz = \frac{1}{I} \ln\left(\frac{c_1-f}{c_1}\right) \qquad \text{(A3-32)}$$

where $f = (AB+AB^*-A^*B-A^*B^*)_0$.

For positive real tz

$$\frac{1}{I} \ln\left(\frac{c_1-f}{c_1}\right) > 0$$

By definition I > 0 so that:

$$\frac{c_1 - f}{c_1} > 1$$

Since $f > 0$ when starting with excess A:

$$c_1 < 0 \qquad \text{(A3-33)}$$

The algebra is simplified considerably if it is assumed that the experiment begins with pure AB and that AB = 1. Presumably, the conditions for asymmetric induction will not be affected when AB $\neq$ 1.

From equation A3-24

$$c_1 = (-(2k_1+k_5) + (W+Z+I)/2)/(I)$$

From A3-33

$$-(2k_1+k_5) + \frac{W+Z+I}{2} < 0$$

$$I^2 < (2(2k_1+k_5) - (W+Z))^2$$

$$XY-WZ < (2k_1+k_5)^2 - (2k_1+k_5)(W+Z) \qquad \text{(A3-34)}$$

Expanding XY-WZ and simplifying with the aid of the microscopic reversibility condition $k_1k_2 = k_{-1}k_{-2}$ gives:

$$XY-WZ = -4k_{-1}k_{-2} - k_{-1}k_5 - k_1k_6 - k_2k_5 - k_{-2}k_6 - k_5k_6 \qquad \text{(A3-35)}$$

Expanding the right-hand side of A3-34 gives

$$(2k_1+k_5)^2 - (2k_1+k_5)(W+Z) =$$

$$2k_1{}^2+k_1k_5 - 2k_1k_{-1} - 2k_{-1}k_{-2} - 2k_1k_{-2} - 2k_1k_6 - k_{-1}k_5$$

$$+ k_2k_5 - k_{-2}k_5 - k_5k_6 \qquad \text{(A3-36)}$$

Substituting A3-36 and A3-35 in A3-34 gives

$$0 < 2k_{-1}k_{-2}+2k_1^2-2k_{-1}k_1-2k_1k_{-2}-k_1k_6+k_{-2}k_6+k_1k_5-k_{-2}k_5$$

$$\therefore \quad 2(k_1-k_{-2})(k_{-1}-k_1) < k_1-k_{-2})(k_5-k_6)$$

Which yields the solutions:

$$k_1 > k_{-2} \text{ and } k_5-k_6 > 2(k_{-1}-k_1) \qquad \text{(A3-37)}$$

or

$$k_1 < k_{-2} \text{ and } k_5-k_6 < 2(k_{-1}-k_1) \qquad \text{(A3-38)}$$

It can be shown which of the inequalities A3-37 and A3-38 is true. From A3-33:

$$-2(2k_1+k_5) + W+Z+I < 0 \qquad \text{(A3-39)}$$

Since $I > 0$ and after substituting for W and Z in A3-39:

$$3k_1-k_{-1}-k_2-k_{-2} + k_5-k_6 > 0 \qquad \text{(A3-40)}$$

Inequality A3-37 can be written as

$$k_1-k_{-1} + k_1-k_{-1} + k_5-k_6 > 0,\ k_1 > k_{-2} \qquad \text{(A3-41)}$$

From microscopic reversibility:

$$k_1 = \frac{k_{-1}k_{-2}}{k_2}$$

Substituting into $k_1 > k_{-2}$ gives:

$$k_{-1} > k_2$$

$$-k_2 > -k_{-1}$$

$$k_1-k_2 > k_1-k_{-1}$$

Substituting into A3-41 gives:

$$(k_1-k_{-1}) + (k_1-k_2) + (k_5-k_6) > 0$$

Since

$k_1 > k_{-2}$, $k_1-k_{-2} > 0$ and $(k_1-k_{-1}) + (k_1-k_2) + (k_1-k_{-2}) +$

$(k_5-k_6) > 0$

$$\therefore \quad 3k_1 = k_{-1}-k_{-2}-k_2 + (k_5-k_6) > 0 \qquad \text{(A3-42)}$$

which is the original condition for $c_1 < 0$, A3-40.

Rearranging inequality A3-38 gives:

$$k_1-k_{-1} + k_1-k_{-1} + k_5-k_6 < 0, \; k_1 < k_{-2} \qquad \text{(A3-43)}$$

Since $k_1 < k_{-2}$, microscopic reversibility gives:

$$k_{-1} < k_2$$

and

$$k_{-1}-k_2 < 0$$

Substituting this and $k_1-k_{-2} < 0$ in A3-43 gives:

$$3k_1-k_{-1}-k_{-2}-k_2 + k_5-k_6 < 0$$

contrary to inequality A3-40.

Thus to get asymmetric induction in A (zero crossing in A-A*, or A*/A > 1) the following inequality must be satisfied:

$$k_5-k_6 > 2(k_{-1}-k_1) \text{ and } k_1 > k_{-2} \qquad \text{(A3-44)}$$

The condition for zero crossing in B-B* (asymmetric induction at the B centre) is a mirror image case of that for the A centre.

Thus

$$tz = \frac{1}{I} \ln\left(\frac{c_2-f}{c_2}\right) \qquad \text{(A3-45)}$$

$c_2 < 0$ when $AB_0 = 1$, $AB_0^* = A^*B_0 = A^*B_0^* = 0$

Going through the same algebra leads to:

$$k_5-k_6 > 2(k_2-k_{-2}) \text{ and } k_{-2} > k_1 \qquad \text{(A3-46)}$$

or

$$k_5-k_6 < 2(k_2-k_{-2}) \text{ and } k_1 > k_{-2} \qquad \text{(A3-47)}$$

Again, it is possible to find out which of A3-45 and A3-46 is true. From the condition for asymmetric induction of the B centre:

$$c_2 < 0$$

Substituting for c_2:

$$-(2k_{-2}+k_5) + \frac{W+Z+I}{2} < 0$$

$$3k_{-2}-k_1-k_{-1}-k_2 + k_5-k_6 > 0 \qquad \text{(A3-48)}$$

A3-45 leads to (after using microscopic reversibility):

$$3k_{-2}-k_1-k_{-1}-k_2 + k_5-k_6 > 0$$

whereas A3-46 leads to:

$$3k_{-2}-k_1-k_{-1}-k_2 + k_5-k_6 < 0$$

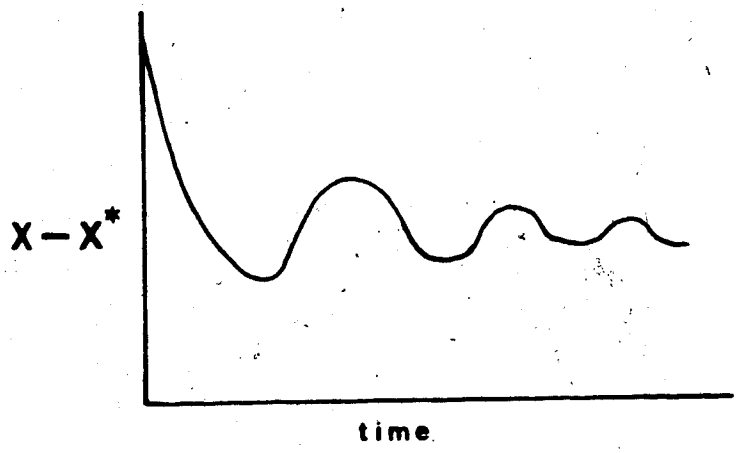
Therefore A3-46 is contrary to the original condition for asymmetric induction. Hence for asymmetric induction at the B centre the following inequality must be obeyed:

$$k_5-k_6 > 2(k_2-k_{-2}),\ k_{-2} > k_1 \qquad \text{(A3-49)}$$

It is noteworthy that A3-44 and A3-49 are contradictory statements. This means that it is not possible to observe asymmetric induction at both centres in the same experiment (as intuition would predict).

Potential for Oscillating Solutions

Intuitively, it seems clear that there would be no solutions of A-A* or B-B* whose graph would appear as follows:



This can be proved. Such solutions would occur when $I^2 < 0$. For this to be the case, from A3-18:

$$(W+Z)^2 - 4(WZ-XY) < 0 \qquad \text{(A3-50)}$$

After expanding and simplifying again with the aid of the microscopic reversibility condition $k_1k_2 = k_{-1}k_{-2}$ A3-50 becomes

$$(k_1-k_6+k_5-k_2-k_{-1}+k_{-2})^2 + 4(k_1-k_{-2})(k_{-1}-k_2) < 0$$

Since

$$(k_1-k_6+k_5-k_2-k_{-1}+k_{-2})^2 > 0$$

then

$$(k_1-k_{-2})(k_{-1}-k_2) < 0$$

so that

$$k_1-k_{-2} > 0 \text{ and } k_{-1}-k_2 < 0 \qquad \text{(A3-51)}$$

or

$$k_1-k_{-2} < 0 \text{ and } k_{-1}-k_2 > 0 \qquad \text{(A3-52)}$$

From A3-51 suppose that:

$$k_1-k_{-2} > 0$$

From microscopic reversibility substituting for k_1,

$$\frac{k_{-1}k_{-2}}{k_2} > k_{-2}$$

$$k_{-1} > k_2$$

contrary to the second part of A3-51.

From A3-52 suppose that:

$$k_1 - k_{-2} < 0$$

From microscopic reversibility stability for k_1,

$$\frac{k_{-1}k_{-2}}{k_2} < k_{-2}$$

$$k_{-1} < k_2$$

contrary to the second part of A3-52.

Therefore (and as a result of microscopic reversibility) there can be no oscillating solutions to A-A* or B-B*.

Computer Program RACSIM

A computer program was written* to simulate the

*The language used was Commodore BASIC V.2.

situation illustrated by Scheme 6. The user is asked to input the number of decimal places to be displayed (for a more readable screen output), the values of the rate constants (k_{-2} is calculated by the program as $k_{-2} = k_1k_2/k_{-1}$), the initial values of A*/A and B*/B, the initial ratio of diastereomers, (AB+AB*)/(A*B+A*B*) (the program calculates the range of possible values), the duration of the simulation and the sampling interval. The program determines if there is zero crossing of A-A* or B-B* (asymmetric induction) from the conditions for asymmetric induction (A3-44 and A3-49) and determines where it is (A3-32 and A3-45). The minimum in A-A* or B-B* is also calculated (from D_t(A-A*) = 0 or D_t(B-B*) = 0). The values output to the screen are (A-A*)/(A+A*), (B-B*)/(B+B*), their natural logarithms (if defined) and the values of AB, AB*, A*B and A*B*. The program listing follows.

RACSIM

```
5 PRINTCHR$(147);:FORI=1TO4:PRINTCHR$(29);:NEXT:PRINT"SIMULATION OF"
:FORI=1TO6:PRINTCHR$(29);:NEXT
6 PRINT"TWO CENTER":FORI=1TO5:PRINTCHR$(29);:NEXT:PRINT"RACEMIZATION"
7 FORI=1TO3:PRINTCHR$(29);:NEXT:PRINT"AND  HYDROLYSIS"
9 PRINT:PRINT"DECIMAL PLACES":INPUTPL:PRINTCHR$(147)
10 GOSUB1000
22 GOSUB200:GOTO30
30 OP$="":INPUT"ANOTHER RUN (Y/N)";OP$
40 IFOP$="Y"THENRUN
100 END
200 FORT=0TODSTEPIT:PRINTCHR$(147);:REM OUTPUT TO SCREEN
210 GOSUB900
220 NU=AM:GOSUB9000:PRINT"A-A$= "N$
230 NU=AP:GOSUB9000:PRINT"A+A$= "N$:PRINT""
232 NU=AM/AP:GOSUB9000:PRINT"(A-A$)/(A+A$)=":PRINTN$
235 IFAM<=0THENPRINT:GOTO252
240 PRINT"LN((A-A$)/A+A$))="
250 NU=LOG(AM/AP):GOSUB9000:PRINTN$:PRINT
252 NU=BM:GOSUB9000:PRINT"B-B$= "N$
254 NU=BM/AP:GOSUB9000:PRINT"(B-B$)/(B+B$)=":PRINTN$
255 IFBM<=0THENPRINT:GOTO300
256 PRINT"LN((B-B$)/(B+B$)="
258 NU=LOG(BM/AP):GOSUB9000:PRINTN$:PRINT
300 D$(1)="[AB]=    ":D$(2)="[AB$]=   ":D$(3)="[A$B]=   ":D$(4)="[A$B$]= "
305 FORJ=1TO4:PRINTD$(J);:NU=D(J):GOSUB9000:PRINTN$:NEXTJ
310 PRINT:PRINT"TIME ="T
318 PRINT"PRESS ANY KEY"
320 GETA$:IFA$=""THEN320
330 NEXT
350 RETURN
900 AM=C1*EXP((-W-Z+I)/2*T)+C3*EXP((-W-Z-I)/2*T)
905 BM=C2*EXP((-W-Z+I)/2*T)+C4*EXP((-W-Z-I)/2*T)
910 AP=CA*EXP((-Q-TX+IP)/2*T)+CB*EXP((-Q-TX-IP)/2*T)
912 DM=CC*EXP((-Q-TX+IP)/2*T)+CD*EXP((-Q-TX-IP)/2*T)
914 D(1)=(AM+BM+AP+DM)/4:D(2)=(AM+AP-BM-DM)/4:D(4)=(AP+DM-AM-BM)/4
:D(3)=(-AM+AP+BM-DM)/4
915 FORJ=1TO4:IFD(J)<1E-5THEND(J)=0:NEXT
920 RETURN
1000 PRINT"K1= ":INPUTP1
1015 PRINT"K-1= ":INPUTM1
1020 PRINT"K2= ":INPUTP2
1025 M2=P1*P2/M1
1030 PRINT"K5= ":INPUTK5
1035 PRINT"K6= ":INPUTK6
1040 PRINT"TIME INCREMENT"
1041 INPUTIT
1045 PRINT"DURATION OF EXPERIMENT":INPUTD
1050 Q=(K6+K5)/2
1070 R=(K6-K5)/2
1080 S=M1-P1+P2-M2+R
1090 TX=M1+P1+P2+M2+Q
1100 IP=SQR((Q+TX)^2-4*(Q*TX-S*R))
1120 W=P1+M1+Q
```

```
1130 X=M1-P1+R
1140 Y=P2-M2+R
1150 Z=P2+M2+Q
1160 I=SQR((W+Z)^2-4*(W*Z-X*Y)):IFI=0THENPRINT"TRIVIAL SOLUTION":END
1200 REM CALCULATION OF PRE-EXPONENTIAL CONSTANTS
1210 GOSUB5000
1255 IFD2<0ORD3<0ORD4<0THENPRINT"BAD INITIAL VALUES":GOTO1210
1260 Z1=D1+D2-D3-D4:Z2=D1+D3-D2-D4:Z3=D1+D2+D3+D4:Z4=D1+D4-D2-D3
1270 Z5=-W*Z1+X*Z2:Z6=Y*Z1-Z*Z2:Z7=(-W-Z+I)/2
1275 Z8=S*Z3-TX*Z4:Z9=(-Q-TX+IP)/2:Z0=-Q*Z3+R*Z4
1280 C3=(Z1*Z7-Z5)/I:C1=Z1-C3:C4=(Z2*Z7-Z6)/I:C2=Z2-C4
1290 CD=(Z4*Z9-Z8)/IP:CC=Z4-CD:CB=(Z3*Z9-Z0)/IP:CA=Z3-CB
1298 PRINT""
1300 IF(P1>M2)AND((K5-K6)>(2*(M1-P1)))THENGOSUB2000:GOTO1320
1310 IF(M2>P1)AND((K5-K6)>(2*(P2-M2)))THENGOSUB3000:GOTO1320
1315 PRINT"NO ZERO CROSSING"
1320 PRINT"PRESS ANY KEY"
1322 B$="":GETB$:IFB$=""THEN1322
1330 RETURN
2000 PRINT"ZERO CROSSING IN A-A$":F=0
2010 PRINT"TZ =";:NU=LOG((C1-Z1)/C1)/I:GOSUB9000:PRINTN$:GOSUB4000
2020 RETURN
3000 PRINT"ZERO CROSSING IN B-B$":F=1
3010 PRINT"TZ =";:NU=LOG((C2-Z2)/C2)/I:GOSUB9000:PRINTN$:GOSUB4000
3020 RETURN
4000 C5=C3:C6=C1:IFF=1THENC5=C4:C6=C2:TM=LOG(C5*(W+Z+I)/(C6*(-W-Z+I)))/I
4005 TM=LOG(C5*(W+Z+I)/(C6*(-W-Z+I)))/I
4010 SD=(C1*(-W-Z+I)^2*EXP((-W-Z+I)/2*TM)+C3*(-W-Z-I)^2*EXP((-W-Z-I)/2*TM))/4
4020 A$="MIN.":IFSD<0THENA$="MAX."
4030 PRINT""A$" AT TIME":NU=TM:GOSUB9000:PRINTN$
4050 RETURN
5000 INPUT"INITIAL D/L FOR A";AI:IFAI<0THEN5000
5001 INPUT"INITIAL D/L FOR B";BI:IFBI<0THEN5001
5002 AA=(AI+1)*(BI+1):BB=AI+BI:IFAI=0ANDBI=0THEN5014
5003 PRINT"RATIO OF DIASTEREOMERSAT TIME ZERO MUST BE GREATER THAN ";PRINT
5004 CC=AA/(BB+2):IF(AA/(BB+2*AI*BI))>CCTHENCC=AA/(BB+2*AI*BI)
5006 PRINTCC:PRINT
5008 IFAI=BITHEN5012
5010 PRINT"AND LESS THAN":PRINTABS(((AI+1)*(BI+1))/(AI-BI)):PRINT
5012 INPUT"RATIO";RD:PRINT
5014 INPUT"INITIAL CONC.";IC
5016 IFAI=0ANDBI=0THEND1=IC:D2=0:D3=0:D4=0:GOTO5026
5018 D1=IC*(1/(AI+1)+1/(BI+1)-(1/RD))/2
5020 D2=-IC*(1/(BI+1)-(1/RD))+D1
5022 D4=IC/RD-D2
5024 D3=IC-D1-D2-D4
5026 PRINT:PRINT"AB = "D1
5028 PRINT"AB* = "D2
5030 PRINT"A*B* = "D3
5032 PRINT"A*B = "D4
5034 PRINT"TOTAL CONC. = "D1+D2+D3+D4
5036 INPUT"OK Y/N";OK$
5038 IFOK$<>"Y"THEN5000
5040 RETURN
9000 N$=STR$(NU):S$=LEFT$(N$,1):N$=MID$(N$,2,LEN(N$)-1):L=LEN(N$):EX$="
9030 IFPL=0THENPL=4
9040 IFPM=PLTHEN9050
```

```
9045 PL=PL+2:PM=PL
9050 Z$="":FORC=1TOPL-2:Z$=Z$+"0":NEXT
9110 IFL>4THENR$=RIGHT$(N$,4):IFLEFT$(R$,1)="E"THENEX$=R$:L=L-4:N$=LEFT$(N$,L)
9112 IFL=1THENN$=N$+"."+Z$:GOTO9180
9113 IFEX$<>"      "THEN9150
9115 DP=L+1
9120 FORC=1TOL-1
9122 IFMID$(N$,C,1)="."THENDP=C:C=L-1
9124 NEXT
9126 IFDP=2THEN9150
9128 IFDP=1THENEX=0:GOSUB9650:GOTO9150
9130 IFDP<LTHENEX=DP-2:GOSUB9500:GOTO9150
9140 N$=N$+".0":L=LEN(N$):DP=L-1:EX=DP-2:GOSUB9500
9150 IFL>=PLTHEN9170
9160 BL$="":FORC=L+1TOPL:BL$=BL$+"0":NEXT:N$=N$+BL$
9170 IFL>PLTHENN$=LEFT$(N$,PL)
9180 N$=BB$+N$+" "+EX$
9200 RETURN
9500 EX$=STR$(EX):EX$=MID$(EX$,2,LEN(EX$)-1)
9510 IFLEN(EX$)=2THENEX$="E+"EX$:GOTO9530
9520 EX$="E+0"+EX$
9530 N$=LEFT$(N$,1)+"."+MID$(N$,2,DP-2)+MID$(N$,DP+1,L-DP):L=LEN(N$)
9540 RETURN
9650 FORC=1TOL-1
9652 IFMID$(N$,C+1,1)<>"0"THENEX=C:C=L-1
9654 NEXT
9656 EX$=STR$(EX):EX$=MID$(EX$,2,LEN(EX$)-1)
9658 IFLEN(EX$)=2THENEX$="E-"+EX$:GOTO9660
9659 EX$="E-0"+EX$
9660 IFEX=L-1THENN$=RIGHT$(N$,1)+"."+Z$:L=LEN(N$):RETURN
9670 N$=MID$(N$,EX+1,1)+"."+MID$(N$,EX+2,L-EX-1):L=LEN(N$)
9680 RETURN
10000 OPEN15,8,15:INPUT#15,A$,B$,C$,D$:PRINTA$,B$,C$,D$:CLOSE15
```

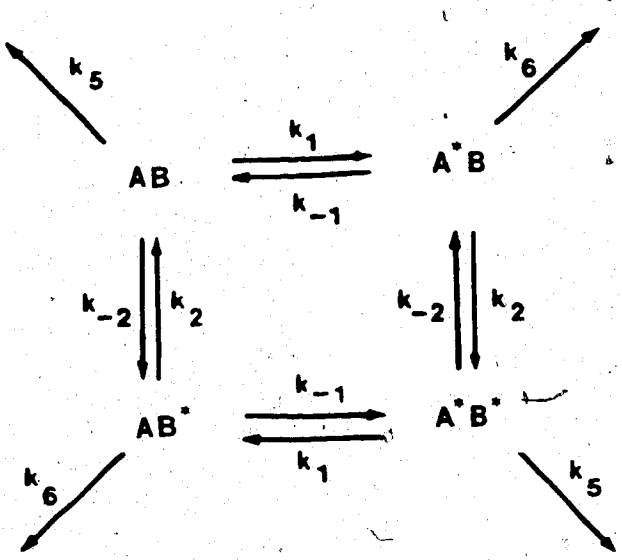
APPENDIX 4

NUMERICAL SOLUTION OF SCHEMES 6, 7 AND 8

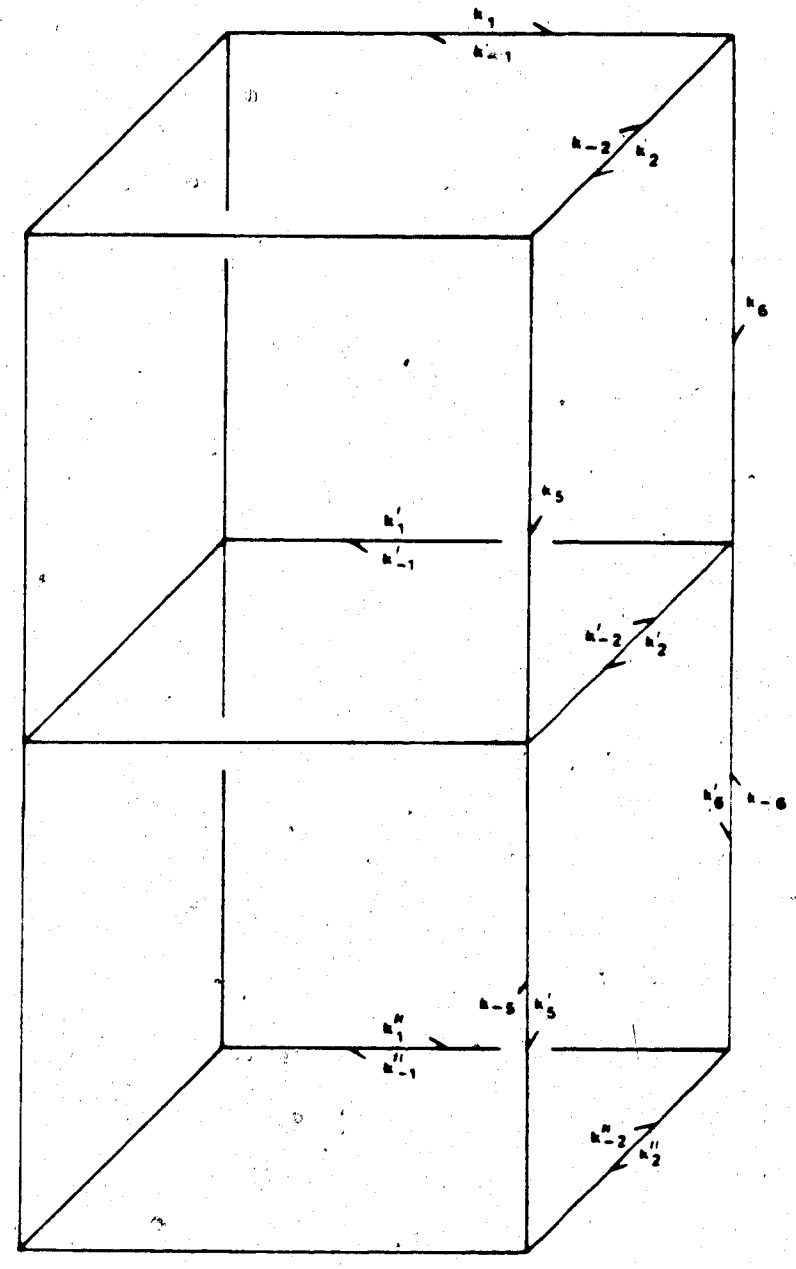
Scheme 7 shows the more general situation of a peptide containing two epimerizing centres undergoing competitive hydrolysis. There is one species at each vertex of Scheme 7. The top plane contains the substrate peptides diastereomers AB, AB*, A*B and A*B* interconverting as in Scheme 6. The substrate can hydrolyze irreversibly to a second set of interconverting diastereomers denoted by AB', AB*', A*B' and A*B*'.

The primed set of diastereomers in the middle plane of Scheme 7 hydrolyze irreversibly to a third set of interconverting diastereomers AB", AB*", A*B" and A*B*" in the bottom plane of Scheme 7. This hypothetical situation was designed to model a peptide with two epimerizing centres which hydrolyzes irreversibly to the diketopiperazine containing two chiral centres. The diketopiperazine is in equilibrium with the corresponding dipeptide.

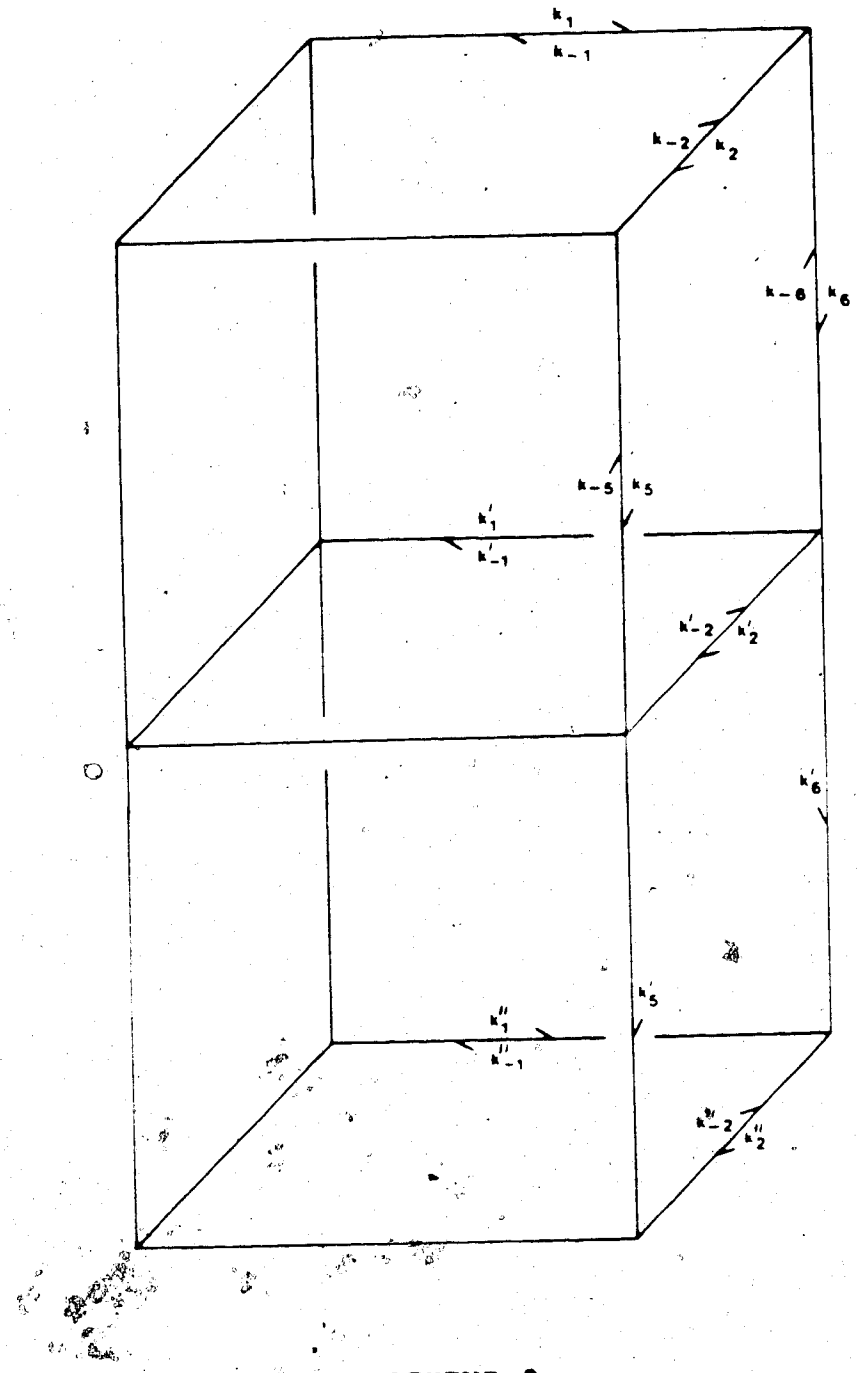
A similar situation is depicted in Scheme 8. Here, the species in the topmost plane are interconverting with the corresponding species in the middle plane. The diastereomers in the middle plane hydrolyze irreversibly



SCHEME 6



SCHEME 7



SCHEME 8

to the species in the bottom plane. The situation is designed to model the peptide with two epimerizing chiral centres which can form some intermediate (the species in the middle plane) by internal aminolysis. This is followed by cleavage to the diketopiperazine containing the chiral centres (Figure 4).

The algebraic solution of the set of 12 differential equations corresponding to the above 12 species was avoided by persuing a numerical solution. Figure 67 gives the matrix representation of the situation depicted in Scheme 7. Figure 68 shows the equations for the situation given by Scheme 8. As in Appendix 3, it is assumed that epimerization can occur at only one centre at a time and that epimerization and hydrolysis are independent processes.

Three other kinetic schemes can be solved by the method given in this appendix. Scheme 11 shows two planes connected by hydrolytic pathways. As shown here, the hydrolysis is reversible. This scheme represents the case where a dipeptide is interconverting with the corresponding diketopiperazine. The process of ring closure may be irreversible. This is depicted in Scheme 12. The simplest case shown in Scheme 13 is identical to the situation described algebraically in Appendix 3 with only one set of diastereomers undergoing irreversible

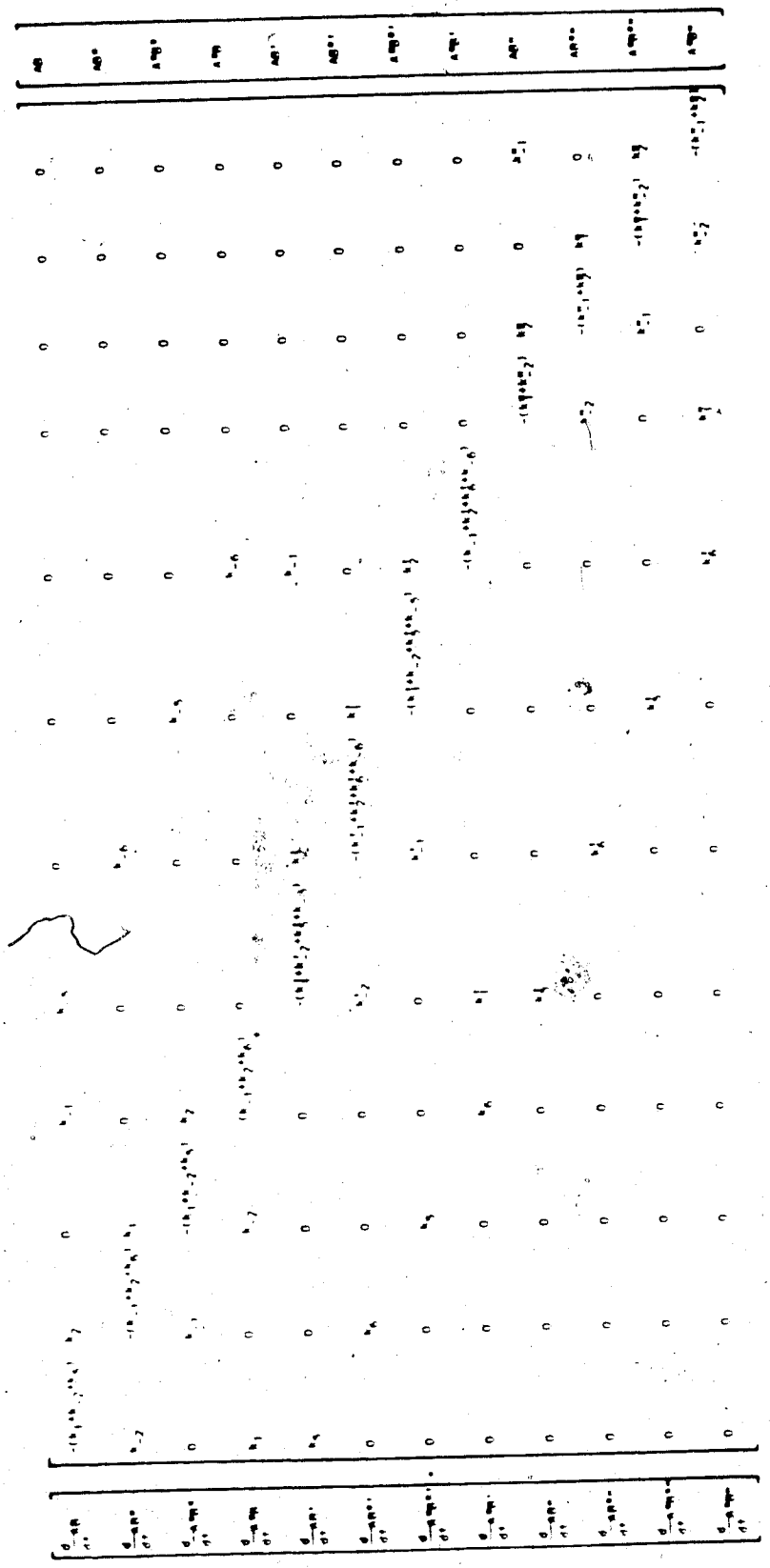


Figure 67. Matrix representation of the twelve differential equations describing Scheme 7.

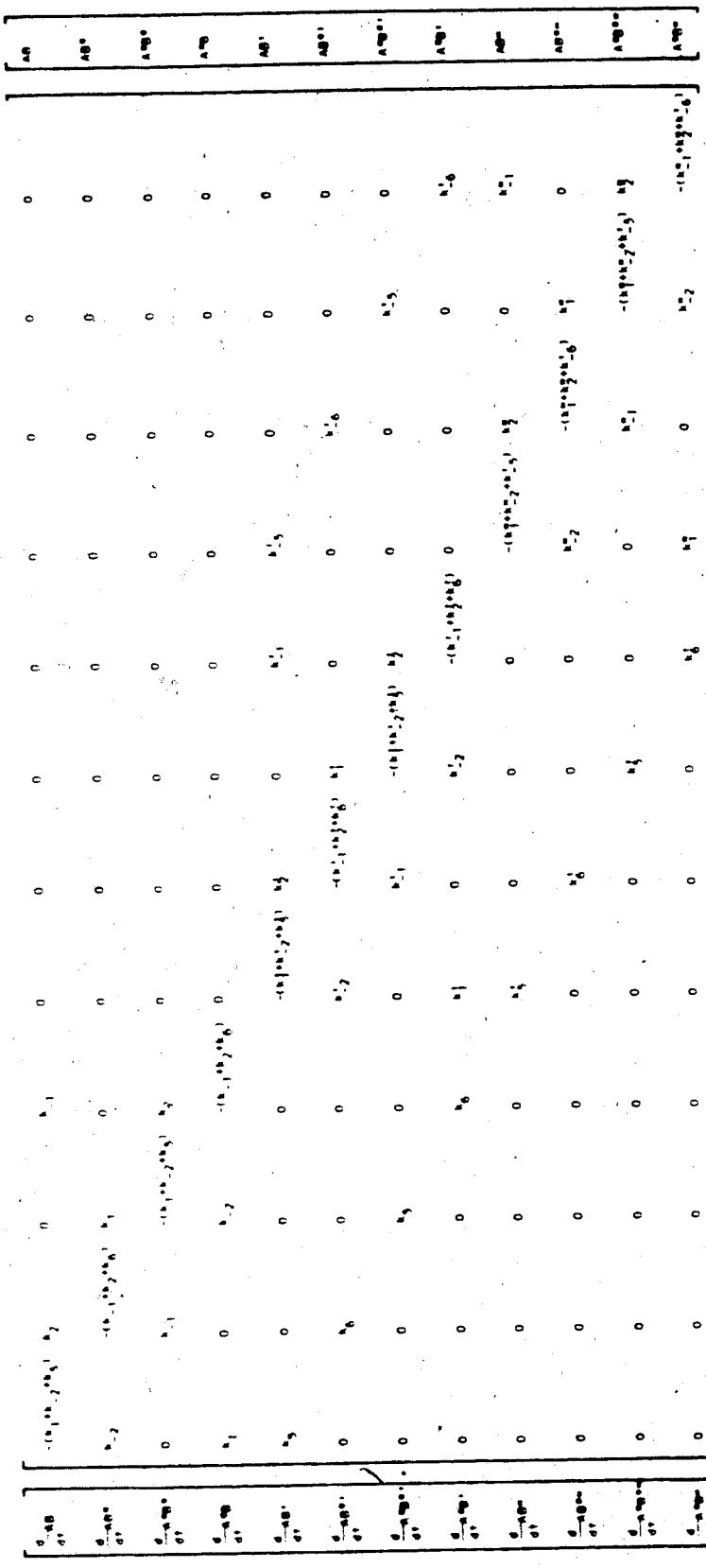
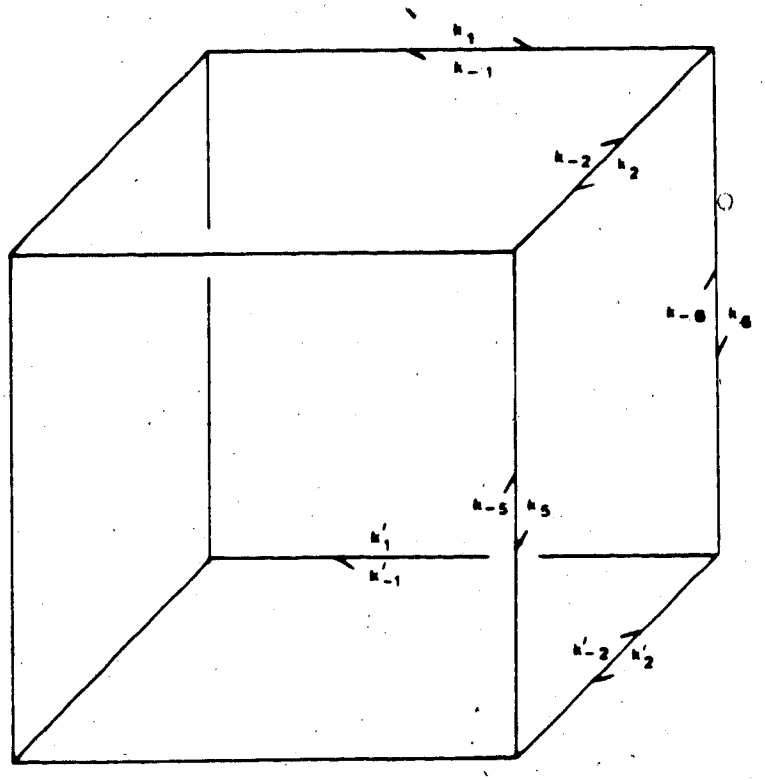
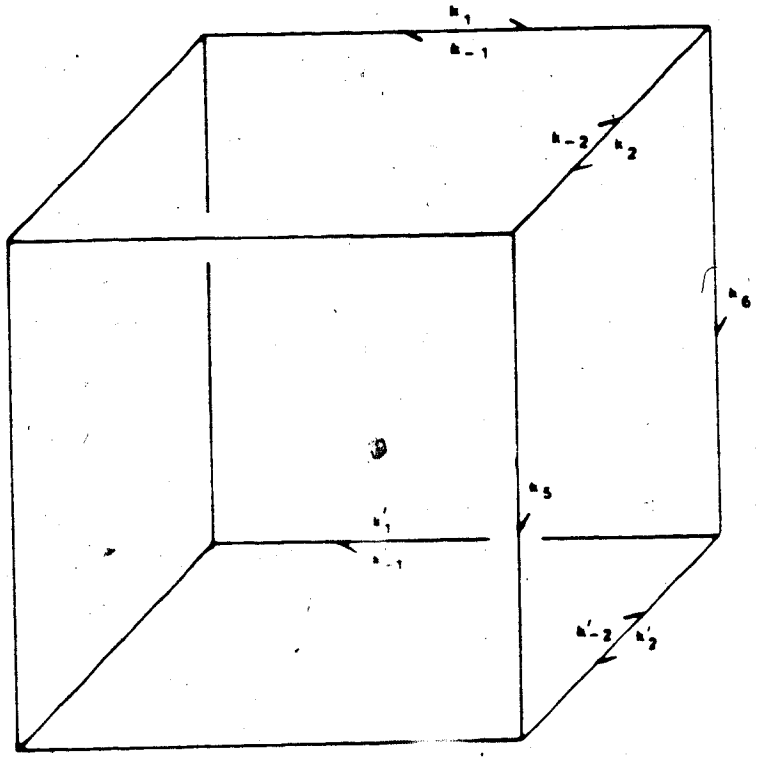


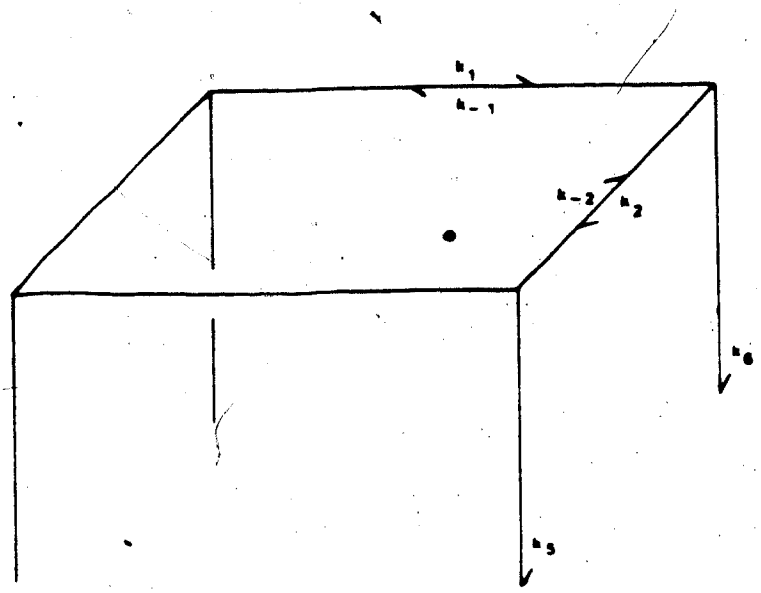
Figure 68. Matrix representation of the twelve differential equations describing Scheme 8.



SCHEME 11



SCHEME 12



SCHEME 13

hydrolysis.

The set of differential equations corresponding to any of the above schemes can be solved in the same way. The Eigenvalues of a matrix **A** (e.g. Figure 67) can be found by finding the characteristic polynomial and calculating the roots. The coefficients of the characteristic polynomial are determined by the Fadeev-Leverrier[110] method in which the n* coefficients p_i of the characteristic polynomial

$$(-1)^n(\lambda^n - p_1\lambda^{n-1} - \ldots - p_n\lambda^0) = 0$$

are determined by:

$$\mathbf{B}_1 = \mathbf{A};\ p_1 = \frac{\mathrm{tr}\mathbf{B}_n}{n}$$

$$\mathbf{B}_2 = \mathbf{A}(\mathbf{B}_1 - p_1\mathbf{I});\qquad p_2 = \frac{\mathrm{tr}\mathbf{B}_2}{2}$$

$$\vdots$$

$$\mathbf{B}_n = \mathbf{A}(\mathbf{B}_{n-1} - p_{n-1}\,\mathbf{I});\ p_n = \frac{\mathrm{tr}\mathbf{B}_n}{n}$$

where **A** is the matrix of coefficients for the n differential equations (Figure 67), tr**B** is the trace of B

---

*n is the number of species in the model, 4, 8 or 12.

and I is the identity matrix.

The Eigenvalues are determined by finding the roots of the characteristic polynomial by the Newton-Raphson method. Once the Eigenvalue is determined, the characteristic polynomial is factored to a smaller polynomial by synthetic division.[111]

The Eigenvector **X** corresponding to Eigenvalue λ is found by substituting λ into the characteristic equation:

$$(\mathbf{A} - \lambda I)\mathbf{X} = 0$$

The coefficients of the Eigenvector **X** are obtained by reducing the resulting matrix by the Gauss-Jordan elimination procedure.[110]

The Eigenvectors are assembled in a matrix such that the components of each column are the elements of one Eigenvector. The value of a species X_i can be calculated at any time t by the use of equation A4-1:

$$X_i = c_1 e_{i1} e^{\lambda_1 t} + c_2 e_{i2} e^{\lambda_2 t} + \ldots + c_n e_{in} e^{\lambda_n t} \quad \text{(A4-1)}$$

where species X_i uses the i^{th} row of a matrix B consisting of elements like:

$$B_{ij} = c_j e_{ij} e^{\lambda_j t} \quad \text{(A4-2)}$$

where $c_1 \ldots c_n$ are the coefficients to be determined and $e_{i1} \ldots e_{in}$ are the elements of one row of the matrix of Eigenvectors.

The n unknown coefficients can be found according to:

$$X_0 = c_1 e_{i1} + c_2 e_{i2} + \ldots + c_n e_{cn} \qquad (A4\text{-}3)$$

Where X_0 is the concentration of species X at time zero. Usually the n equations like A4-3 are sufficient to determine c_1 through c_8. Depending on the values of the rate constants, some of the n equations may not be independent. n more equations can be obtained by differentiating A4-1 and setting the result equal to a row of matrix **A** in Figure 67 at time zero.

$$\left(\frac{dX}{dt}\right)_0 = \lambda_1 c_1 e_{i1} + \lambda_2 c_2 e_{i2} + \ldots + \lambda_n c_n e_{in} \qquad (A4\text{-}4)$$

where λ_1 through λ_2 are the Eigenvalues determined previously. Once n independent equations are chosen, coefficients c_1 to c_8 can be calculated by reducing the matrix formed by n equations like A4-1 and A4-2 using Gaussian elimination.

Under certain circumstances, this method of solving a system of differential equations will fail. The source of

the problem is with the number of significant figures which are being used for calculation purposes. Since most computers are capable of handling 9 or 10 significant figures, it would seem that this should not present any difficulty. However, two considerations are important. First, the time needed to calculate a figure to one part in 10^9 or 10^{10} may be prohibitively long. Second, this kind of precision may not be possible to achieve because of round-off error.

These problems were encountered in the computer simulations using the method outlined above. The Newton-Raphson method for determining the roots of a polynomial in general converges quite rapidly for p[illegible]omials as high as eighth order, but is remarkably slow [illegible] twelfth order polynomials. The twelfth order polynomial used for the simulations typically has its roots within a very small range of λ. Thus, the derivative of the polynomial has a value of nearly zero in this range resulting in poor convergence. Over 1000 iterations of the Newton-Raphson method was typical for accuracy of one part in 10^5. The time taken to arrive at a solution can be minimized by reducing the desired precision or by choosing a better first guess. Since it is difficult to know a priori where the roots will be, the former alternative is preferred.

However, while good approximations of the roots may

be found, the inaccuracy of the Eigenvalues is compounded in the method for determining the Eigenvectors. In this method it is necessary that the value of some of the elements in the array used to calculate the Eigenvectors be exactly zero. Because of round-off error, elements whose absolute value is smaller than a preset small number ε are set to zero. If ε is too small, array elements which should be zero will not be set to zero. If ε is too large, elements which should not be zero will be set to zero. In either case, the Eigenvector cannot be determined. If a value of ε cannot be found so that the Eigenvector can be determined, a more accurate Eigenvalue must be supplied. It appears that the round-off error intrinsic to the solution of the 12 × 12 matrix used to find the twelve Eigenvectors demands a high degree of precision in the roots of the characteristic polynomial.

Computer Programs TLSA and TLSB

Program TLSA was written* to solve the kinetic situations given in Schemes 8 through 11.

Throughout the program, the user is asked to input certain values. The first prompt "# VARIABLES (4/8/12)"

*Language used was Commodore BASIC V.2.

requires the user to select the scheme to be used. Input 4, 8 or 12 for Scheme 13, 11 or 12, 7 or 8 respectively. The next prompt "REVERSIBLE or IRREVERSIBLE HYDROLYSIS" occurs if 8 has been entered for the first prompt and selects Scheme 11 if "R" is input or Scheme 12 if "I" is input.

The program will prompt the user for the values of the rate constants.* k_{-2} is calculated (likewise for k'_{-2} and k''_{-2}) from the relationship:

$$k_{-2} = \frac{k_1 k_2}{k_{-1}}$$

If Scheme 8 or 9 is selected, k'_{-6} is calculated by:

$$k'_{-6} = \frac{k'_6 k''_2 k'_{-5} k'_{-2}}{k'_2 k''_{-2} k'_5}$$

Note that if Scheme 11 is chosen, the matrix of coefficients to be solved is made up of the elements forming the 8 × 8 matrix in the bottom right-hand corner of Figure 67. Similarly, the 4 × 4 matrix in the top

---

*In simulations involving more than 4 variables, as a rule of thumb the user should restrict the values of the rate constants to the range $10^{-3} < k < 1$. The exponents of the coefficients of the characteristic polynomial can get quite large. The possibility of overflow error or serious round-off error is a real danger.

left-hand corner is used for Scheme 13 and the 8 × 8 matrix in the top left-hand corner is used for Scheme 12. The program adjusts the position of the elements to agree with the position of the species (e.g. AB, AB*, etc.) in Figure 67. The program displays the coefficients of the matrix **A** in the characteristic equation. The coefficients of the characteristic polynomial are calculated and displayed. The response to the prompt "ACCURACY OF ROOTS" should be 0.000001 or 0.0000001 for the systems involving 8 or 12 variables in order to minimize the problem of round-off error. The program may fail here, as convergence is not guaranteed. The roots (Eigenvalues) are displayed on the screen and are followed by the prompt "MINIMUM DIFFERENCE BETWEEN EIGENVALUES". If two or more of the calculated Eigenvalues have similar magnitudes (say within 0.0001), then they will yield the same set (2 or more) of linearly independent Eigenvectors. One of the Eigenvalues must be removed. Choosing a value of appropriate magnitude (in this case 0.0001) will fix the criterion for double roots.

The prompt "ZERO LIMIT" requires a response of between 0.001 and 0.1. Array elements less than this limit will be set to zero as required by the Gauss-Jordan elimination method. Larger arrays will generally require a larger "zero limit" to achieve a solution. The program

may fail here as a result of round-off error.

The response to "MINIMUM NON-ZERO EIGENVECTOR COMPONENT" determines the minimum non-zero value in the array of Eigenvectors.

The coefficients of the equations A4-1 are determined from the array of Eigenvectors and from the initial conditions. It is assumed that the only species present at time zero are those in the topmost plane of any of the schemes. The user supplies the initial D/L ratio, the initial ratio of diastereomers (within the calculated limits) and the total concentration.

The user must then choose the independent equations to be used to calculate the concentration of the species in solution. Generally, equations 1 through n should be chosen, where n is the number of variables. One or more of equations N+1 through 2n (from the derivatives of equations 1 through n) should be used when a solution cannot be achieved from equations 1 through n.

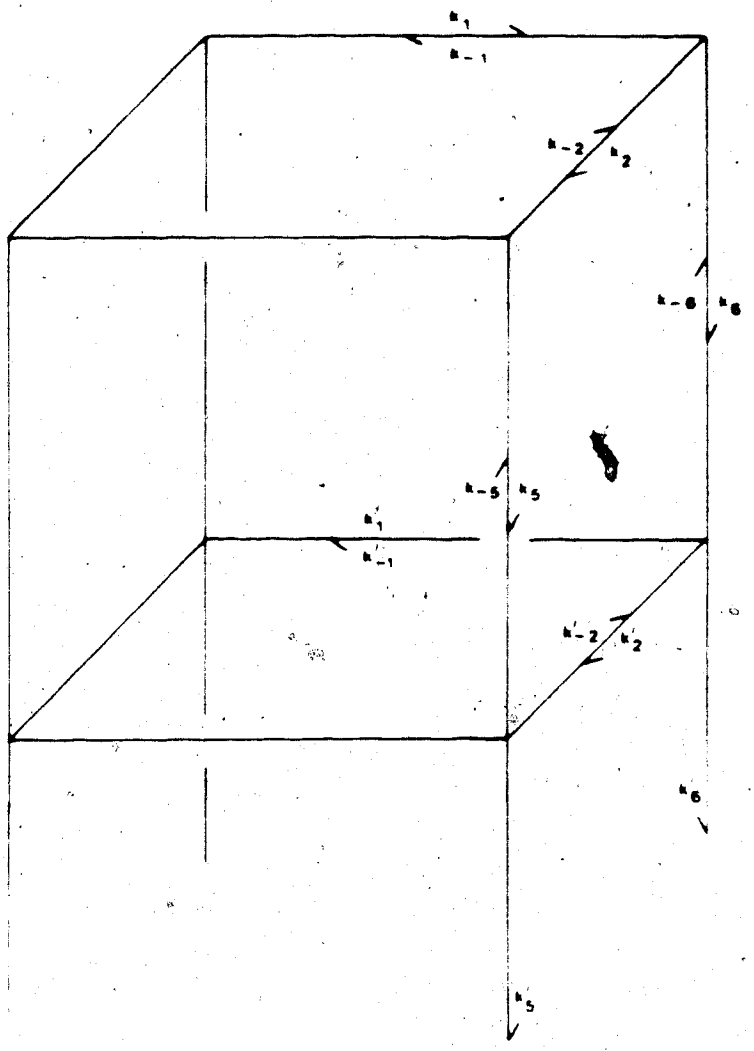
The user must then supply the sampling interval ("TIME INCREMENT") and the duration of the experiment. The range of variables which can be calculated are displayed. The user must be careful to choose from only those variables which are relevant to the kinetic scheme which is being simulated. Otherwise the program will terminate, potentially with the loss of all of the

calculated coefficients.

Program TLSB is to be used to simulate Schemes 7 and 14. The two programs, TLSA and TLSB, are essentially the same. The matrix of coefficients from the n differential equations is assembled from the matrix corresponding to the situation in Scheme 8 (Figure 68) and then adjusted appropriately. The prompt "INPUT 0 IF 3RD STAGE NOT DESIRED" is to be answered with "0" if Scheme 14 is to be simulated.

Program Documentation: TLSA

Line(s)	Description
100-199	Input rate constants and calculate k_{-2} and k''_{-6} from microscopic reversibility (lines 135, 152, 190, 195).
200-364	Create matrix of coefficients (**A**).
370-378	Display **A**.
380-795	Calculate coefficients of characteristic polynomial. Matrix **B** is found from $\mathbf{B}_n = A(\mathbf{B}_{n-1} - p_{n-1}I)$. Array P(N) contains the coefficients of the characteristic polynomial.
800-990	Calculate roots of characteristic polynomial.
1000-1040	Synthetic division.
1104-1117	Correct for double roots.



SCHEME 14

1119-1490 Calculate Eigenvectors and place as columns of **C**.

1500-1514 Set minimum value of C(I,J) and place in **AA**(I,J) and C(I,J).

1518-1546 Input initial conditions.

1548-1567 Calculate values of derivatives (e.g. (d/dt)[AB]) and concentrations (e.g. [AB]) at time zero. Place in **CM**.

1570-1590 Choose n equations. If derivative is used, adjust terms by multiplying by appropriate Eigenvalue. Copy **C** into **B**. Add **CM** to **B**.

1595-1800 Find n unknown coefficients by Gaussian elimination on **B**. Place in **P**. Matrix **AB**, the completed solution, is made by **AA** × P.

1804-1874 Select quantities to be calculated.

1875-1879 Select screen or printer output.

1900-1954 Calculate AB, AB*, etc.

2513-2560 Calculate chosen quantities.

2570-2578 Print headings on printer.

2600-2608 Print a line. Check for end of page.

2700-2740 Screen output.

2800-2815 Check if more quantities are desired using same rate constants.

2816-2819 Check if calculations to be done again using a different set of equations or different time

scale.

2830-5000 Check if program to be re-run.

6000-6050 Print rate constants, initial conditions, equations used for calculations.

7000-7010 Check to see if calculated values for AB, AB*, etc. at time zero differ from the actual values. If so, the selected equations are likely not independent. An alternate choice must be made.

10000 Read disk drive error channel.

Program Listings

The full listing for TLSA follows. After this is a listing of the program lines where TLSB differs from TLSA. Lines which have been omitted in TLSB are marked (*) in TLSA.

TLBA

```
 10 GOTO100
 100 PRINTCHR$(147)"8 VARIABLES (4/B/12)":INPUTN
 104 A$=CHR$(17):B$=CHR$(145):C$=CHR$(157)
 105 DIMRK(3,6),A(12,12)
*107 IFN=8THENPRINT"REVERSIBLE OR IRREVERSIBLE HYDROLYSIS":INPUTRV$
*108 IFRV$="R"THEN140
 109 FORI=0TO3:FORJ=0TO6:RK(I,J)=0:NEXT:NEXT
 110 PRINT"K"A$"1"B$"=";:INPUTRK(1,1):PRINT
 115 PRINT"K"A$"2"B$"=";:INPUTRK(1,2):PRINT
 120 PRINT"K"A$"-1"B$"=";:INPUTRK(1,3):PRINT
 125 PRINT"K"A$"5"B$"=";:INPUTRK(1,5):PRINT
 130 PRINT"K"A$"6"B$"=";:INPUTRK(1,6):PRINT
 135 RK(1,4)=RK(1,1)*RK(1,2)/RK(1,3)
*136 IFN=4THEN200
 140 PRINT"K'"C$;A$"1"B$"=";:INPUTRK(2,1):PRINT
 145 PRINT"K'"C$;A$"2"B$"=";:INPUTRK(2,2):PRINT
 150 PRINT"K'"C$;A$"-1"B$"=";:INPUTRK(2,3):PRINT
 152 RK(2,4)=RK(2,1)*RK(2,2)/RK(2,3)
*154 IFN=BANDRV$="I"THEN200
 155 PRINT"K'"C$;A$"5"B$"=";:INPUTRK(2,5):PRINT
 160 PRINT"K'"C$;A$"6"B$"=";:INPUTRK(2,6):PRINT
 170 PRINT"K'"C$;A$"1"B$"'=";:INPUTRK(3,1):PRINT
 175 PRINT"K'"C$;A$"2"B$"'=";:INPUTRK(3,2):PRINT
 180 PRINT"K'"C$;A$"-1"B$;C$"'=";:INPUTRK(3,3):PRINT
 185 PRINT"K'"C$;A$"-5"B$"=";:INPUTRK(3,5):PRINT
 190 RK(3,4)=RK(3,1)*RK(3,2)/RK(3,3)
 195 RK(3,6)=RK(2,6)*RK(3,2)*RK(3,5)*RK(2,4)/(RK(2,2)*RK(3,4)*RK(2,5))
 200 FORI=0TON:FORJ=0TON:A(I,J)=0:NEXT:NEXT
 230 FORI=1TO9STEP4
 235 A(I,I)=-(RK((I+3)/4,1)+RK((I+3)/4,4)+RK((I+3)/4,5))
 240 A(I+1,I+1)=-(RK((I+3)/4,2)+RK((I+3)/4,3)+RK((I+3)/4,6))
 245 A(I+2,I+2)=A(I,I)
 250 A(I+3,I+3)=A(I+1,I+1)
 255 NEXT
 258 FORI=0TO8STEP4
 260 A(I+1,I+2)=RK((I+4)/4,2)
 265 A(I+1,I+4)=RK((I+4)/4,3)
 270 A(I+2,I+1)=RK((I+4)/4,4)
 275 A(I+2,I+3)=RK((I+4)/4,1)
 280 A(I+3,I+2)=RK((I+4)/4,3)
 285 A(I+3,I+4)=RK((I+4)/4,2)
 290 A(I+4,I+1)=RK((I+4)/4,1)
 295 A(I+4,I+3)=RK((I+4)/4,4)
 300 NEXT
 310 A(5,1)=RK(1,5):A(7,3)=A(5,1)
 315 A(6,2)=RK(1,6):A(8,4)=A(6,2)
 320 A(9,5)=RK(2,5):A(11,7)=A(9,5)
 325 A(10,6)=RK(2,6):A(12,8)=A(10,6)
*330 A(5,9)=RK(3,5):A(7,11)=A(5,9)
*335 A(6,10)=RK(3,6):A(8,12)=A(6,10)
 350 IF(N=BANDRV$="I")ORN=12THEN370
 352 IFN=4THENFORI=5TO12:FORJ=1TO12:A(I,J)=0:NEXT:NEXT:GOTO370
 360 FORI=1TO8:FORJ=1TO8:A(I,J)=A(I+4,J+4):NEXT:NEXT
 362 FORI=9TO12:FORJ=1TO12:A(I,J)=0:NEXT:NEXT
 364 FORI=5TO8:FORJ=9TO12:A(I,J)=0:NEXT:NEXT
 370 FORI=1TON
 371 PRINTCHR$(147)"    CHARACTERISTIC":PRINT"      DETERMINANT"
 372 PRINTSPC(7)"ROW"I
```

```
373 FORJ=1TON:BB$="":IFJ<10THENBB$=" "
374 PRINTJ;BB$;"A(I,J):NEXT
375 PRINT:PRINT"PRESS ANY KEY"
376 BETA$:IFA$=""THEN376
378 NEXT
380 PRINTCHR$(147)"      CALCULATING":PRINT"      POLYNOMIAL"
500 REM FIND POLYNOMIAL COEFFICIENTS
510 DIMB(N+1,N+1),P(N),CB(N)
520 FORJ=1TON
530 FORI=1TON
540 B(I,J)=A(I,J)
550 NEXTI:NEXTJ
560 N=N-1
570 FORK=1TON
580 TR=0
590 FORI=1TON
600 TR=TR+B(I,I)
610 NEXTI
620 AK=K
630 P(K)=TR/AK
640 FORI=1TON
650 B(I,I)=B(I,I)-P(K)
660 NEXTI
670 FORJ=1TON
680 FORI=1TON
690 CB(I)=B(I,J)
700 NEXTI
710 FORI=1TON
720 B(I,J)=0
730 FORL=1TON
740 B(I,J)=B(I,J)+A(I,L)*CB(L)
750 NEXTL:NEXTI:NEXTJ:NEXTK
760 P(N)=B(1,1)
770 PRINT:PRINT"COEFFICIENTS":PRINT:P(0)=1
780 FORK=0TON
790 PRINTP(K)
795 NEXT
800 DIMRR(N),RT(N)
802 PRINT:INPUT"ACCURACY OF ROOTS";ER:PRINT
805 PRINT"      CALCULATING":PRINT"          ROOTS":PRINT
808 FORI=1TON:RR(I)=0:NEXT
810 FORI=1TON:P(I)=-P(I):NEXTI:P(0)=1
835 K=N
840 IF(RK(1,5)=0ANDRK(1,6)=0)ORN>4THENK=K-1
850 IFK>1THEN900
870 RR(1)=-P(1)/P(0)
880 PRINTCHR$(147)
890 FORI=1TON
891 PRINT"ROOT"I";"RR(I):PRINT
892 IFI<>4ANDI<>8ANDI<>12THENNEXT
893 PRINT"PRESS ANY KEY"
894 A$="":BETA$:IFA$=""THEN894
895 NEXT
896 GOTO1100
900 NI=0:KN=ER
910 FT=0:DV=0
915 FORJ=0TOK
920 FT=FT+P(K-J)*KN^J
950 DV=DV+J*P(K-J)*KN^(J-1)
960 NEXT
```

```
975 NX=KN-(FT/DV)
980 IFABS(KN-NX)<ERTHENRR(K)=KN:GOSUB1000:GOTO850
990 KN=NX:NI=NI+1:GOTO910
1000 REM SYNTHETIC DIVISION
1010 FORI=1TOK-1
1020 P(I)=P(I)+KN*P(I-1)
1030 NEXT
1040 K=K-1:RETURN
1100 REM EIGEN VECTORS
1101 PRINT:PRINTCHR$(34)"COL #"CHR$(34)" REFERS TO COL # OF LIN. IND. EIGEN
VEC TOR ASSOCIATED";
1102 PRINT" WITH A PARTICULAR EIGENVALUE"
1103 DIMAA(N+1,N),X(N,3),C(N+1,N+1)
1104 PRINT:PRINT"MINIMUM DIFFERENCE BETWEEN EIGENVALUES":INPUTER
1105 FORI=1TON+1:FORJ=1TON+1:C(I,J)=0:NEXT:NEXT:FORI=1TON:RT(I)=RR(I):NEXT
1106 NV=N:REM CORRECT FOR MULTIPLE ROOTS
1107 I=1
1108 J=I+1
1109 IFABS(ABS(RT(I))-ABS(RT(J)))>ERTHEN1114
1112 NV=NV-1:IFNV=1THEN1116
1113 FORK=JTONV-1:RR(K)=RR(K+1):NEXT:GOTO1109
1114 J=J+1:IFJ<NVTHEN1109
1115 I=I+1:IFI<NVTHEN1108
1116 IFNV=1AND(ABS(ABS(RT(1))-ABS(RT(2)))>ER)THENNV=2
1117 PRINTCHR$(17)"EIGENVALUES ARE":FORI=1TONV:PRINTRT(I);:NEXT:PRINT
1118 INPUTCHR$(17)"ZERO LIMIT";ER:PRINT
1119 FORKK=1TONV
1120 FORJ=0TON:FORI=0TON:AA(I,J)=A(I,J):B(I,J)=0:NEXT:NEXT
1122 REM RE-USING ARRAYS B(N,N) AND CB(N)
1125 FORI=0TON:CB(I)=0:NEXT
1140 FORJ=1TON:AA(J,J)=AA(J,J)-RR(KK):NEXT
1150 FORI=1TON:CB(I)=I:NEXT
1152 REM ARRAY CB(N) KEEPS TRACK OF COLUMN SWITCHES
1160 K=N
1200 M=N-K+1
1210 II=1:JJ=1:BG=ABS(AA(1,1))
1220 FORJ=1TOK:FORI=1TOK:DF=BG-ABS(AA(I,J))
1230 IFDF>=0THEN1250
1240 JJ=J:II=I:BG=ABS(AA(I,J))
1250 NEXT:NEXT
1260 IF(II-1)<=0THEN1280
1270 FORJ=1TOK:TP=AA(II,J):AA(II,J)=AA(1,J):AA(1,J)=TP:NEXT
1280 IF(JJ-1)<=0THEN1310
1290 TP=CB(M):L=JJ+N-K:CB(M)=CB(L):CB(L)=TP
1300 FORI=1TON:TP=AA(I,JJ):AA(I,JJ)=AA(I,1):AA(I,1)=TP:NEXT
1310 IFAA(1,1)=0THEN1400
1330 FORJ=2TOK:FORI=2TON
1340 B(I-1,J-1)=AA(I,J)-AA(1,J)*AA(I,1)/AA(1,1)
1350 Y=ABS(B(I-1,J-1))-ER*ABS(AA(I,J))
1360 IFY>=0THEN1380
1370 B(I-1,J-1)=0
1380 NEXT:NEXT
1382 FORJ=2TOK:B(N,J-1)=AA(1,J)/AA(1,1):NEXT
1384 K=K-1
1385 IFK>0THEN1395
1386 PRINT"RANK NOT LESS THAN":PRINT"ORDER. MORE ACCURATE":PRINT"EIGEN VALUE"K
1387 PRINT"IS NEEDED OR ADJUST":PRINT"ZERO LIMIT.":GOTO1104
1395 FORJ=1TOK:FORI=1TON:AA(I,J)=B(I,J):NEXT:NEXT:GOTO1200
```

```
1400 FORJ=1TOK:AA(J,J)=-1:NEXT
1420 FORMM=1TOK:FORJ=1TOK
1430 AA(N+1,J)=AA(1,J)
1440 FORI=1TON
1450 AA(I,J)=AA(I+1,J)
1455 NEXT:NEXT:NEXT
1460 FORJ=1TOK:FORI=1TON:FORNM=1TON
1470 IF(CB(NM)-I)=0THENX(I,J)=AA(NM,J)
1475 NEXT:NEXT:NEXT
1480 FORJ=1TOK
1483 PRINT"EV"KK"COL"J";";
1484 FORI=1TON:C(I,KK)=X(I,J):PRINTC(I,KK)TAB(5);:NEXT
1485 PRINT:NEXT
1490 NEXTKK
1500 REM ASSEMBLE COEFFICIENT MATRIX
1510 PRINT"MINIMUM NON-ZERO EIGENVALUE COMPONENT":INPUTER
1512 FORI=1TON:FORJ=1TON
1513 IFABS(C(I,J))<ERTHENC(I,J)=0
1514 AA(I,J)=C(I,J):NEXT:NEXT
1515 FORI=1TON:P(I)=0:NEXT:DIMPC(N)
1518 PRINTCHR$(147)"SUPPLY INITIAL":PRINT"CONDITIONS":PRINT
1520 INPUT"A$/A";AI:PRINT
1522 INPUT"B$/B";BI:PRINT
1523 AA=(AI+1)$(BI+1):BB=AI+BI:RD$="  ":IFAI=0ANDBI=0THEN1530
1524 PRINT"RATIO OF DIASTEREOMERS AT TIME ZERO MUST BE  GREATER THAN":PRINT
1525 CC=AA/(BB+2):IF(AA/(BB+2$AI$BI))>CCTHENCC=AA/(BB+2$AI$BI)
1526 PRINTCC:PRINT
1527 IFAI=BITHEN1529
1528 PRINT"AND LESS THAN":PRINT:PRINTABS(((AI+1)$(BI+1))/(AI-BI)):PRINT
1529 INPUT"RATIO";RD:PRINT:RD$=STR$(RD)
1530 INPUT"INITIAL CONC.";IC:PRINT
1531 IFAI=0ANDBI=0THENP(1)=IC:P(2)=0:P(3)=0:P(4)=0:GOTO1543
1536 P(1)=IC$(1/(AI+1)+1/(BI+1)-(1/RD))/2
1538 P(2)=-IC$(1/(BI+1)-(1/RD))+P(1)
1540 P(4)=IC/RD-P(2)
1542 P(3)=IC-P(1)-P(2)-P(4)
1543 FORI=1TON:PC(I)=0:NEXT:FORI=1TO4:PC(I)=P(I):NEXT
1544 FORI=1TO4:PRINT"D("MID$(STR$(I),2,1)")="P(I):NEXT:PRINT
1545 SS=0:FORI=1TO4:SS=SS+P(I):NEXT:PRINT"TOTAL CONCENTRATION ="SS:PRINT
1546 INPUT"OK Y/N";OK$:IFOK$<>"Y"THEN1518
1548 PRINT
1550 DIMCM(2$N),AB(N,N+1),VS$(50),EQ%(N-1)
1552 FORI=1TO2$N:CM(I)=0:NEXT
1554 FORI=1TO4:CM(I)=P(I):NEXT
1556 K=1:IFN=0ANDRV$="R"THENK=2
1558 CM(N+1)=-(RK(K,1)+RK(K,4)+RK(K,5))$P(1)+RK(K,3)$P(2)+RK(K,2)$P(4)
1560 CM(N+2)=RK(K,1)$P(1)-(RK(K,3)+RK(K,2)+RK(K,6))$P(2)+RK(K,3)$P(3)
1562 CM(N+3)=RK(K,2)$P(2)-(RK(K,1)+RK(K,4)+RK(K,5))$P(3)+RK(K,3)$P(4)
1564 CM(N+4)=RK(K,4)$P(1)+RK(K,1)$P(3)-(RK(K,2)+RK(K,3)+RK(K,6))$P(4)
1566 IFN=4THEN1570
1567 CM(N+5)=RK(K,5)$P(1):CM(N+6)=RK(K,6)$P(2):CM(N+7)=RK(K,5)$P(3)
:CM(N+8)=RK(K,6)$P(4)
1570 PRINT"CHOOSE"N" EQUATIONS TOSOLVE FOR CONSTANTS OFINTEGRATION. THE TOP
1571 FORI=0TON:FORJ=0TON+1:B(I,J)=C(I,J):NEXTJ,I
1572 PRINTN" EQUATIONS INVOLVE CONC. AT TIME ZERO.":PRINT"THE BOTTOM"N" EQNS
INVOLVE ";
```

```
1574 PRINT"THE CHANGE IN CONC. AT TIME ZERO.  THEY ARE ARRANGED TOP TO
BOTTOM ":
1576 PRINT"AB, AB$,   A$B$, A$B. BE SURE TO CHOOSE THE EQUATIONS  IN
NUMERICAL ORDER.":PRINT
1578 FORI=1TON:PRINT"EQUATION"I;
1579 EQ=0:INPUTEQ:IFEQ=0THEN1579
1580 B(I,N+1)=CM(EQ):EQ%(I-1)=EQ
1582 IFEQ<=NTHEN1590
1584 FORJ=1TON:B(I,J)=B(I,J)$RR(J):NEXT
1590 NEXT
1595 FORI=1TON:P(I)=0:NEXT
1600 REM GAUSSIAN ELIMINATION
1610 M=N+1:L=N-1
1620 FORK=1TOL
1630 JJ=K:BG=ABS(B(K,K)):KP=K+1
1640 FORI=KPTON
1650 AB=ABS(B(I,K))
1660 IF(BG-AB)<0THENBG=AB:JJ=I
1670 NEXTI
1680 IF(JJ-K)=0THEN1690
1685 FORJ=KTOM:TP=B(JJ,J):B(JJ,J)=B(K,J):B(K,J)=TP:NEXT
1690 FORI=KPTON:QT=B(I,K)/B(K,K)
1695 FORJ=KPTOM:B(I,J)=B(I,J)-QT$B(K,J)
1700 NEXT:NEXT
1710 FORI=KPTON:B(I,K)=0:NEXT
1720 NEXTK
1730 P(N)=B(N,M)/B(N,N)
1740 FORNN=1TOL
1750 SM=0:I=N-NN:IP=I+1
1760 FORJ=IPTON:SM=SM+B(I,J)$P(J):NEXT
1770 P(I)=(B(I,M)-SM)/B(I,I)
1780 NEXTNN
1790 PRINT:PRINT"CONSTANTS DETERMINED  BY INITIAL CONDITIONS":PRINT
1795 FORI=1TON:PRINT"B("MID$(STR$(I),2,1)")="P(I):NEXT:PRINT
1800 FORJ=1TON:PRINT"COL"J":";:FORI=1TON:AB(I,J)=AA(I,J)$P(J):PRINTAB(I,J);
:NEXT :PRINT:NEXTJ
1804 LE=66:KF=0
1805 PRINT:PRINT"TIME INCREMENT":INPUTIT
1807 PRINT:PRINT"DURATION OF EXPERIMENT":INPUTDU
1808 GOSUB1810:GOTO1814
1810 PRINTCHR$(147)"CHOOSE VARIABLES TO BE PRINTED":PRINT:RETURN
1814 FORI=0TO48:READVS$(I):NEXT
1816 DATAAB,AB$,A$B$,A$B
1818 DATAAB',AB$',A$B$',A$B'
1820 DATAAB'',AB$'',A$B$'',A$B''
1822 DATAA-A$,B-B$,A+A$
1824 DATAA-A$',B-B$',A+A$'
1826 DATAA-A$'',B-B$'',A+A$''
1828 DATAA-A$T,B-B$T
1830 DATA(A-A$)/(A+A$)
1831 DATA(B-B$)/(B+B$)
1832 DATA(A-A$)'/(A+A$)'
1833 DATA(B-B$)'/(B+B$)'
1834 DATA(A-A$)''/(A+A$)''
1835 DATA(B-B$)''/(B+B$)''
```

```
1836 DATA(A-A$)T/(A+A$)T
1837 DATA(B-B$)T/(B+B$)T
1838 DATAAB+A$B$,A$B+AB$
1840 DATAAB'+A$B$',A$B'+AB$'
1842 DATAAB''+A$B$'',A$B''+AB$''
1843 DATAAB$B$/AB,A$B/AB$,AB$B$'/AB',A$B'/AB$',AB$B$''/AB'',A$B''/AB$''
1844 DATA(AB+AB$)/(A$B+A$B$),(A$B$+AB$)/(A$B+AB)
1845 DATA(AB'+AB$')/(A$B'+A$B$'),(A$B$'+AB$')/(A$B'+AB')
1846 DATA(AB''+AB$'')/(A$B''+A$B$''),(A$B$''+AB$'')/(A$B''+AB'')
1849 GOSUB1850:GOTO1870
1850 FORI=0TO215STEP2:PRINTI+1;TAB(4)VB$(I)TAB(11)I+2;TAB(15)VB$(I+1):NEXT
:PRINTI+1TAB(4)VB$(I)
1852 PRINT:PRINT"PRESS"CHR$(34)"C"CHR$(34):PRINT"TO CHOOSE VARIABLES"
1854 PRINT:PRINT"PRESS"CHR$(34)"P"CHR$(34):PRINT"TO SEE VARIABLES";
1856 OK$="":GETOK$:IFOK$=""THEN1856
1858 IFOK$="C"THEN1869
1860 IFOK$<>"P"THEN1856
1861 GOSUB1810:IFI=49THEN1850
1862 IFI=37THEN1865
1863 IFI=47THEN1866
1864 GOSUB1810:FORI=23TO36:PRINTI+1VS$(I):NEXT:GOTO1852
1865 GOSUB1810:FORI=37TO46:PRINTI+1VS$(I):NEXT:GOTO1852
1866 GOSUB1810:FORI=47TO48:PRINTI+1VS$(I):NEXT:GOTO1852
1869 RETURN
1870 PRINT:INPUT"VARIABLE 1";V%(1)
1871 INPUT"VARIABLE 2";V%(2)
1872 INPUT"VARIABLE 3";V%(3)
1874 INPUT"VARIABLE 4";V%(4)
1875 FORI=1TO4:IFV%(I)=0THEN1870
1876 NEXT:PRINT:PRINT"PRINTER OR SCREEN";PRINT"(P/S)"
1877 PS$="":GETPS$:IFPS$<>"P"ANDPS$<>"S"THEN1877
1878 DF=2:IFPS$="P"THENDF=1
1890 RT=0:FORI=1TO4:IFV%(I)>37THENRT=IT
1891 NEXT
1900 FORT=RTTODUSTEPIT
1902 X=T:Y$="":K=0:X$=STR$(X):IFLEN(X$)>5THENX$=LEFT$(X$,5)
1904 ONDFGOSUB2600,2700
1910 FORI=1TON
1912 P(I)=0
1914 FORJ=1TON
1916 P(I)=P(I)+AB(I,J)*EXP(RR(J)*T)
1917 NEXTJ,I
1918 IFT=0THENGOSUB7000:IFPF=1THENRESTORE:GOTO1570
1919 IFPS$="S"THEN1922
1920 IFKF=0THENGOSUB6000
1921 IFT=RTTHENGOSUB2570
1922 FORK=1TO4:X=0
1923 IFV%(K)<13THENX=P(V%(K)):GOTO1950
1924 IFV%(K)=22ORV%(K)=23THENGOSUB2550:GOTO1950
1926 V%=V%(K)-12
1928 ONV%GOSUB2513,2514,2515,2516,2517,2518,2519,2520,2521:IFV%<10THEN1950
1930 V%=V%(K)-23
1932 ONV%GOSUB2524,2525,2526,2527,2528,2529,2530,2531,2532,2533,2534,2535
1933 IFV%<13THEN1950
1934 V%=V%(K)-35:ONV%GOSUB2536,2537,2538,2539,2540,2541,2542,2543,2544,2545,
2546
1935 IFV%<12THEN1950
1936 V%=V%(K)-46:ONV%GOSUB2547,2548,2549
```

```
1950 X$=STR$(X):ONDFBOSUB2600,2700
1952 NEXTK
1954 NEXTT:BOTO2800
2513 X=X+P(1)+P(2)-P(3)-P(4):RETURN
2514 X=X+P(1)+P(4)-P(2)-P(3):RETURN
2515 X=X+P(1)+P(2)+P(3)+P(4):RETURN
2516 X=X+P(5)+P(6)-P(7)-P(8):RETURN
2517 X=X+P(5)+P(8)-P(6)-P(7):RETURN
2518 X=X+P(5)+P(6)+P(7)+P(8):RETURN
2519 X=X+P(9)+P(10)-P(11)-P(12):RETURN
2520 X=X+P(9)+P(12)-P(10)-P(11):RETURN
2521 X=X+P(9)+P(10)+P(11)+P(12):RETURN
2524 BOSUB2515:Y=X:X=0:BOSUB2513:X=X/Y:RETURN
2525 BOSUB2515:Y=X:X=0:BOSUB2514:X=X/Y:RETURN
2526 BOSUB2518:Y=X:X=0:BOSUB2516:X=X/Y:RETURN
2527 BOSUB2518:Y=X:X=0:BOSUB2517:X=X/Y:RETURN
2528 BOSUB2521:Y=X:X=0:BOSUB2517:X=X/Y:RETURN
2529 BOSUB2521:Y=X:X=0:BOSUB2518:X=X/Y:RETURN
2530 BOSUB2550:X=X/IC:RETURN
2531 BOSUB2550:X=X/IC:RETURN
2532 X=P(1)+P(3):RETURN
2533 X=P(2)+P(4):RETURN
2534 X=P(5)+P(7):RETURN
2535 X=P(6)+P(8):RETURN
2536 X=P(9)+P(11):RETURN
2537 X=P(10)+P(12):RETURN
2538 X=P(3)/P(1):RETURN
2539 X=P(4)/P(2):RETURN
2540 X=P(7)/P(5):RETURN
2541 X=P(8)/P(6):RETURN
2542 X=P(11)/P(9):RETURN
2543 X=P(12)/P(10):RETURN
2544 X=(P(1)+P(2))/(P(3)+P(4)):RETURN
2545 X=(P(3)+P(2))/(P(1)+P(4)):RETURN
2546 X=(P(5)+P(6))/(P(7)+P(8)):RETURN
2547 X=(P(7)+P(6))/(P(5)+P(8)):RETURN
2548 X=(P(9)+P(10))/(P(11)+P(12))
2549 X=(P(11)+P(10))/(P(9)+P(12)):RETURN
2550 AD=1:IFV%(K)=23ORV%(K)=31THENAD=2
2552 ONADGOSUB2513,2514
2554 IFN=4THEN2560
2556 ONADGOSUB2516,2517
2557 IFN=8THEN2560
2558 ONADGOSUB2519,2520
2560 RETURN
2570 PRINT#4:PRINT#4,"TIME  ";
2573 FORZ=1TO4:L%=INT(LEN(VS$(V%(Z)-1)))/2
2574 TB%=15+18*(Z-1)-L%:TB$=MID$(STR$(TB%),2,2):IFLEN(TB$)=1THENTB$="0"+TB$
2575 PRINT#4,CHR$(16)LEFT$(TB$,1)RIGHT$(TB$,1)VS$(V%(Z)-1);
2577 NEXT
2578 PRINT#4:PRINT#4:LE=LE-3:RETURN
2600 SS=18
2602 IFK=0THENSS=6
2604 Z$="":FORZ=1TOSS-LEN(X$):Z$=Z$+" ":NEXT:Y$=Y$+X$+Z$
2606 IFK=4THENPRINT#4,Y$;:PRINT#4:LE=LE-1:IFLE=10THENGOSUB5500
2608 RETURN
```

```
2700 IFK=0THENPRINTCHR$(147)"TIME ="T:PRINT:GOTO2740
2710 PRINTVS$(VZ%(K)-1)"=":PRINTX:PRINT
2720 IFK<4THEN2740
2730 PRINT:PRINT"PRESS ANY KEY"
2732 OK$="":GETOK$:IFOK$=""THEN2732
2740 RETURN
2800 PRINT:PRINT"PRINT MORE VARIABLES"
2812 INPUTOK$
2814 IFOK$="Y"THENRESTORE:IFLE<15ANDPS$="P"THENGOSUB5500
2815 IFOK$="Y"THEN1808
2816 PRINT:PRINT"USE A DIFFERENT SET OF EQUATIONS"
2817 INPUTOK$:IFOK$<>"Y"THEN2830
2818 IFPS$="P"THENGOSUB5500:PRINT#4
2819 RESTORE:GOTO1570
2830 PRINT:INPUT"ANOTHER RUN";OK$
2832 IFOK$="Y"ANDPS$="S"THENRUN
2833 IFOK$="Y"THENGOSUB5500:PRINT#4:CLOSE4:RUN
2834 IFPS$="P"THENPRINT#4:CLOSE4
5000 END
5500 FORI=1TOLE:PRINT#4:NEXT:LE=66:RETURN
6000 KF=1:CLOSE4:OPEN4,4
6010 PRINT#4,CHR$(17)"RATE CONSTANTS"CHR$(145)CHR$(10):LE=LE-2
6015 IFN=8ANDRV$="R"THEN6025
6020 PRINT#4,"K1  ="RK(1,1)" K2  ="RK(1,2)" K-1  ="RK(1,3)" K-2  ="RK(1,4)
6021 PRINT#4,"K5  ="RK(1,5)" K6  ="RK(1,6)
6022 LE=LE-2:IFN=4THEN6040
6025 PRINT#4,"K1' ="RK(2,1)" K2' ="RK(2,2)" K-1' ="RK(2,3)" K-2' ="RK(2,4)
6026 PRINT#4,"K5' ="RK(2,5)" K6' ="RK(2,6)
6027 LE=LE-2:IFN=8ANDRV$<>"R"THEN6040
6030 PRINT#4,"K1''="RK(3,1)" K2''="RK(3,2)" K-1''="RK(3,3)" K-2''="RK(3,4)
6031 PRINT#4,"K5''="RK(3,5)" K6''="RK(3,6)
6032 LE=LE-2
6040 PRINT#4:LE=LE-1
6042 PRINT#4,CHR$(17)"INITIAL CONDITIONS:"CHR$(145):PRINT#4
6044 PRINT#4,"AI="AI" BI="BI" RD="RD$" IC="IC
6046 PRINT#4:LE=LE-4
6048 PRINT#4,CHR$(17)"EQUATIONS USED: ";:FOREQ=0TON-1:PRINT#4,EQ%(EQ);:NEXT
:PRINT#4,CHR$(146):LE=LE-2
6050 RETURN
7000 PF=0:FORPC=1TON:IFABS(P(PC)-PC(PC))>1E-4THENPF=1
7004 NEXT:IFPF=0THEN7010
7005 PRINT:PRINT"CHOICE OF EQUATIONS ISNO GOOD":PRINT
7010 RETURN
10000 OPEN15,8,15:INPUT#15,A$,B$,C$,D$:PRINT""A$B$C$D$:CLOSE15
```

TLBB

```
10 GOTO100
50 TP=RK(I,1):RK(I,1)=RK(I,4):RK(I,4)=TP
52 TP=RK(I,2):RK(I,2)=RK(I,3):RK(I,3)=TP:RETURN
100 PRINTCHR$(147):N=12
104 A$=CHR$(17):B$=CHR$(145):C$=CHR$(157)
105 DIMRK(3,6),A(12,12)
109 FORI=0TO3:FORJ=0TO6:RK(I,J)=0:NEXT:NEXT
110 PRINT"K"A$"1"B$"=";:INPUTRK(1,1):PRINT
115 PRINT"K"A$"2"B$"=";:INPUTRK(1,2):PRINT
120 PRINT"K"A$"-1"B$"=";:INPUTRK(1,3):PRINT
125 PRINT"K"A$"5"B$"=";:INPUTRK(1,5):PRINT
130 PRINT"K"A$"6"B$"=";:INPUTRK(1,6):PRINT
132 PRINT"K"A$"-5"B$"=";:INPUTRK(3,5):PRINT
135 RK(1,4)=RK(1,1)*RK(1,2)/RK(1,3):I=1:GOSUB50
140 PRINT"K'"C$;A$"1"B$"=";:INPUTRK(2,1):PRINT
145 PRINT"K'"C$;A$"2"B$"=";:INPUTRK(2,2):PRINT
150 PRINT"K'"C$;A$"-1"B$"=";:INPUTRK(2,3):PRINT
152 RK(2,4)=RK(2,1)*RK(2,2)/RK(2,3):I=2:GOSUB50
154 RK(3,6)=RK(1,6)*RK(2,2)*RK(3,5)*RK(1,4)/(RK(1,2)*RK(1,5)*RK(2,4))
155 PRINT"K'"C$;A$"5"B$"=";:INPUTRK(2,5):PRINT
160 PRINT"K'"C$;A$"6"B$"=";:INPUTRK(2,6):PRINT
170 PRINT"K'"C$;A$"1"B$"'=";:INPUTRK(3,1):PRINT
175 PRINT"K'"C$;A$"2"B$"'=";:INPUTRK(3,2):PRINT
180 PRINT"K'"C$;A$"-1"B$;C$"'=";:INPUTRK(3,3):PRINT
190 RK(3,4)=RK(3,1)*RK(3,2)/RK(3,3)
199 I=3:GOSUB50
200 FORI=0TON:FORJ=0TON:A(I,J)=0:NEXT:NEXT
230 FORI=1TO9STEP4
235 A(I,I)=-(RK((I+3)/4,1)+RK((I+3)/4,4)+RK((I+3)/4,5))
240 A(I+1,I+1)=-(RK((I+3)/4,2)+RK((I+3)/4,3)+RK((I+3)/4,6))
245 A(I+2,I+2)=A(I,I)
250 A(I+3,I+3)=A(I+1,I+1)
255 NEXT
258 FORI=0TO8STEP4
260 A(I+1,I+2)=RK((I+4)/4,3)
265 A(I+1,I+4)=RK((I+4)/4,2)
270 A(I+2,I+1)=RK((I+4)/4,1)
275 A(I+2,I+3)=RK((I+4)/4,4)
280 A(I+3,I+2)=RK((I+4)/4,2)
285 A(I+3,I+4)=RK((I+4)/4,3)
290 A(I+4,I+1)=RK((I+4)/4,4)
295 A(I+4,I+3)=RK((I+4)/4,1)
300 NEXT
310 A(5,1)=RK(1,5):A(7,3)=A(5,1)
315 A(6,2)=RK(1,6):A(8,4)=A(6,2)
320 A(9,5)=RK(2,5):A(11,7)=A(9,5)
325 A(10,6)=RK(2,6):A(12,8)=A(10,6)
330 A(5,5)=A(5,5)-RK(3,5):A(7,7)=A(5,5):A(6,6)=A(6,6)-RK(3,6):A(8,8)=A(6,6)
331 A(9,9)=A(9,9)+RK(3,5):A(11,11)=A(9,9):A(10,10)=A(10,10)+RK(3,6):
A(12,12)=A(10,10)
332 A(1,5)=RK(3,5):A(3,7)=A(1,5):A(2,6)=RK(3,6):A(4,8)=A(2,6)
```

APPENDIX 5

DETERMINATION OF ERROR IN TEMPERATURE MEASUREMENT DUE TO TEMPERATURE OSCILLATIONS

The following program calculates the difference between the true average temperature and the rate average temperature, the temperature determined by the progress of a chemical reaction. In this model, the temperature oscillates about the true average temperature in a sinusoidal fashion. The user must supply the values of the true average temperature and the activation energy for the chemical reaction.

```
90 PRINTCHR$(147):OPEN4,4
100 N=100:REM NUMBER OF SLICES FOR SIMPSON'S APPROX.
110 INPUT"LOWER TEMP. LIMIT";LL
122 INPUT"ACTIVATION ENERGY";EA
125 GOSUB1000:GOTO500
140 S4=0:S2=0:TS=0
150 FORT=1TON-1STEP2
160 GOSUB4000
170 S4=S4+4*E
180 NEXT
190 FORT=2TON-2STEP2
200 GOSUB4000
210 S2=S2+2*E
220 NEXT
230 T=0:GOSUB4000
240 TS=E+S2+S4
250 T=N:GOSUB4000
260 TS=TS+E
270 AR=TS/(3*N)
280 RETURN
500 FORTA=LLTOLL+25.005TEP5.00:PRINT#4
502 FORII=1TO6
507 READDT
510 DATA1,2,5,10,15,20
520 GOSUB140
530 TE=((-LOG(AR)*1.9872)/EA+TA^-1)^-1
540 GOSUB2000
550 NEXT:RESTORE:NEXT:CLOSE4:END
1000 PRINT#4,"TRUE AVERAGE   OSCILLATION    ACTIVATION          EFFECTIVE
ERROR"
1010 PRINT#4,"TEMPERATURE      AMPLITUDE        ENERGY          TEMPERATURE
(TE-TA)"
1020 PRINT#4
1030 RETURN
2000 PRINT#4,TA;CHR$(16)"16"DT;CHR$(16)"31"EA;CHR$(16)"46"TE;CHR$(16)"61"TE-T
2010 RETURN
4000 T1=T/N
4010 S=DT*SIN(2**T1)
4020 Q=(S*EA)/(1.9872*TA*(TA+S))
4030 E=EXP(Q)
4040 RETURN
```

APPENDIX 6

STANDARD D/L PROLINE CALIBRATION MIXTURES

D/L (calc)[a]	D/L (obs)[b]	D/L (calc)[a]	D/L (obs)[b]
0.01250	0.00837	0.8000	0.7501
0.02000	0.01605	1.000	0.8982
0.02500	0.02355	1.250	1.154
0.03333	0.02930	1.667	1.450
0.05000	0.04361	2.000	1.797
0.06667	0.05927	2.500	2.174
0.07500	0.07189	3.000	2.743
0.1000	0.09181	4.000	3.200
0.1500	0.1496	5.000	4.404
0.2000	0.1914	6.667	5.071
0.2500	0.2313	10.00	7.476
0.3333	0.3029	13.33	8.871
0.4000	0.3751	15.00	10.70
0.5000	0.4530	20.00	14.36
0.6000	0.5652	30.00	23.19

a) Standards were prepared by mixing appropriate volumes of solutions of D-proline and L-proline followed by preparation of the N-pentafluoropropionyl isopropyl esters

b) Measurements of the D/L ratio for proline in samples from the heating experiments were corrected as follows. For an observed D/L ratio between 0 and 0.1 the correction applied was 1.047 D/L (obs) + .0028; between 0.1 and 1, 1.078 D/L (obs) + .0074; between 1 and 5, 1.179 D/L (obs) - .0771; between 5 and 30, 1.314 D/L (obs) + .385