THYROID HORMONE ANALOGS:

SYNTHESIS, THYROMIMETIC ACTIVITIES, MOLECULAR ORBITAL STUDIES, AND QUANTITATIVE STRUCTURE-ACTIVITY CORRELATIONS

by

Stephen Winters Dietrich

B.A., Pomona College, 1972

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

in the

GRADUATE DIVISION

(San Francisco)

of the

UNIVERSITY OF CALIFORNIA

ABSTRACT

THYROID HORMONE ANALOGS:

SYNTHESIS, THYROMIMETIC ACTIVITIES, MOLECULAR ORBITAL STUDIES,

AND QUANTITATIVE STRUCTURE-ACTIVITY CORRELATIONS

Stephen Winters Dietrich

Ph. D. Dissertation

Department of Pharmaceutical Chemistry

School of Pharmacy

University of California

San Francisco

Experimental and theoretical studies were utilized in examining the structure-activity relationships of the thyroid hormones and analogs in both *in vivo* and *in vitro* assays, with emphasis in particular on investigating the specific molecular interactions involved in binding to nuclear receptors and plasma proteins. Nine new thyroid hormone analogs (halogen-containing and halogen-free) were synthesized and tested for thyromimetic activity in the rat antigoiter bioassay and/or in binding to solubilized rat hepatic non-histone nuclear protein receptors. Molecular orbital calculations (CNDO/2 and *ab initio*) were used to: (1) conduct the first extensive quantum mechanical study of intramolecular hydrogen bonding in ortho-substituted phenols and thiophenols; (2) examine the intermolecular hydrogen bonding of ortho-substituted phenols and phenoxides as model systems for binding of the outer (phenolic) ring of the thyroid hormones and analogs to nuclear receptors and plasma proteins, respectively; (3) investigate the conformations of ortho-alkyl phenols; and (4) provide a preliminary con-

formational study of the alanine side chain of thyroid hormone analogs. Quantitative structure-activity relationship studies of the thyroid hormones and analogs were used to examine: (1) in vivo antigoiter bioassay activities; (2) in vitro binding to intact rat hepatic nuclei, to solubilized rat hepatic non-histone nuclear protein receptors, and to the plasma protein thyroxine binding globulin (TBG); and (3) correlations between in vivo activities and in vitro binding to nuclear receptors. The substituent parameters 3'SIZE>I and INTERACT (derived from MO calculations and experimental data) were utilized as estimates of 3'-substituent "size" greater than iodine and the intramolecular interactions of the 3' and 5' substituents with the 4'-OH, respectively.

The results of these experimental and theoretical studies of the thyroid hormones and analogs can be summarized as follows:

(1) In vivo antigoiter activity and in vitro binding to nuclear receptors are enhanced by bulky, lipophilic 3 and 5 substituents and by size-limited, lipophilic 3' substituents.

(2) Any 5' substituent bulk or lipophilicity decreases in vivo activity and in vitro binding to nuclear receptors by interfering with 4'-OH hydrogen bonding with the receptor and/or by direct steric interaction with the receptor.

(3) In vivo activity is enhanced by electron-donating 3' and 5' substituents, which discourage plasma protein binding and encourage analog movement into cells.

(4) Binding to nuclear receptors probably involves hydrogen bond donation by the 4'-OH to the 5' side of the receptor, with attractive and/or repulsive interactions between the 3' and 5' substituents and the 4'-OH affecting the latter's 3'/5' orientation.

(5) With adjustments made for *in vivo* metabolism, *in vivo* activity correlates well with *in vitro* binding to nuclear receptors, indicating that the latter is probably the first step in initiating subsequent hormonal responses.

(6) Except for 3' and 5' substituent influences on 4'-OH ionization, distribution, at least within the range of analog lipophilicities studied, does not play a major role in determining whole animal activity.

(7) Binding to TBG is proportional to the degree of 4'-OH ionization.

(8) The free energy of *in vitro* binding to nuclear receptors can be partitioned into substituent contributions, which can themselves be partitioned into contributions due to the physico-chemical properties of the substituents.

ACKNOWLEDGEMENTS

I wish to express my sincerest appreciation and thanks to the following:

To Dr. Eugene C. Jorgensen for your guidance, encouragement, understanding, and support.

To Dr. Peter A. Kollman for your direction, criticism, enthusiasm, and patience.

To Dr. Roger G. Ketchum for reading my dissertation and your excellent teaching.

To all the other professors and fellow graduate students (too numerous to mention individually) who enriched my studies both in classes and in informal discussions.

To the Department of Pharmaceutical Chemistry and the School of Pharmacy staffs for help in times of need and of red tape.

To the Graduate Division for personal financial support (Regent's Fellowship) and for research financial support (Chancellor's Patent Fund).

To Tariq Andrea and Mike Bolger for your generous advice and assistance in the experimental and theoretical studies.

To the 10th floor remote computer terminal--just for being where it is.

To my parents for your love and support.

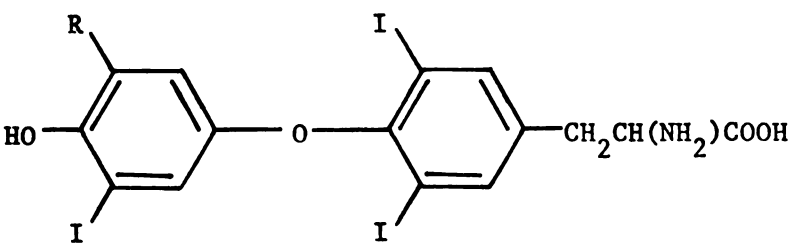
To Irak and Ramat for your devotion and love.

And especially to Sarah for your continuing support, understanding, and love.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

The two thyroid hormones, thyroxine (T_4; 1-1) and 3,5,3'-triiodothyronine (T_3; 1-2) elicit a multitude of biological responses and are

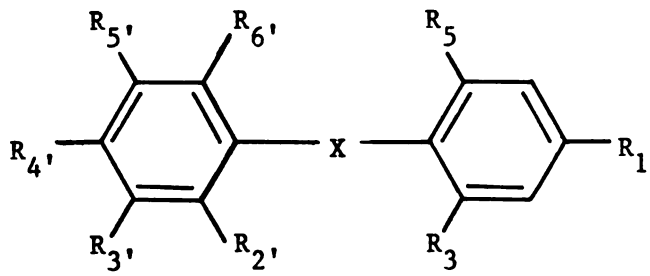


1-1, R = I

1-2, R = H

essential for normal growth and development.[1,2] A number of hypotheses have been proposed to relate the various structural features of the thyroid hormones and analogs (1-3) to the expression of their biological effects. These include: (a) the unique role of iodine excitation to

the long-lived, reactive triplet and its participation in energy transfer processes;[3-5] and (b) participation of the ether oxygen and phenolic hydroxyl in quinone-mediated electron transfer.[6,7]



1-3

Classical structure-activity studies[2,8,9] have ruled out these two hypotheses. In view of recent studes, it might be proposed that the association between hormone and receptor may induce in the latter a conformational change, which is responsible for initiating subsequent hormonal effects. This has refocused attention on the structural and stereochemical aspects of the hormone, as they relate to receptor interactions with emphasis in particular on: (a) the importance of the phenyl-X-phenyl conformation, as influenced by X and the 3, 5, 2', and 6' substituents; (b) specific hydrophobic, hydrogen bonding, and steric interactions of the 3', 4', and 5' substituents with receptors; and (c) the importance in transport, metabolism, and receptor binding of the 1 position side chain length, conformation, stereochemistry, and associated charges. A number of physico-chemical studies of the hormones and analogs utilizing X-ray crystallography,[10-14] NMR spectroscopy,[15,16] and theoretical MO calculations,[17] as well as analog structure-activity studies,[2,8,18-23] have consistently supported the structural and stereochemical dependence of thyromimetic activity.

An important recent advance in the study of the thyroid hormones and analogs has been the development of suitable *in vitro* assays for thyromimetic activity. These assays measure the binding affinities of thyroid hormones and analogs to isolated intact rat hepatic nuclei,[24,25] to solubilized rat hepatic nuclear non-histone proteins,[26-28] and to various purified plasma proteins.[29,30] The results of these studies have shown that: (a) There are binding sites with high affinity and low capacity for the thyroid hormones and analogs in the chromatin of cell nuclei,[31,32] T_3 (1-2) having an apparent binding constant $K = 6.1 \times 10^8\ M^{-1}$.[33] For rat hepatic cells these nuclear receptors have been solubilized[28] and characterized as acidic, non-histone proteins of approximately 60,000 molecular weight;[31,32] (b) for binding of analogs to both rat hepatic intact nuclei and solubilized nuclear protein, there is a quantitative 1:1 correlation between log K and log (*in vivo* rat antigoiter activity), once adjustments are made for well-established *in vivo* metabolism of certain analogs;[24-26] and (c) for thyroxine analogs there are many similarities, but also significant differences, between the structure-binding affinity relationships for interactions with nuclear receptors and with plasma proteins.

The apparent binding constants reflect the thermodynamics of binding rather than actual biological activity. Their *in vitro* measurement avoids the difficulties arising from distribution, metabolism, and the sequence of events between binding and biological response and provides a unique opportunity to examine the physical origins of the binding interactions without such complications. The *in vivo* binding of thyroxine and analogs to plasma proteins has primarily been viewed as a mechanism for transport and storage in the blood.[34] However,

the partial similarities of the structure-binding affinity relationships for analog binding to nuclear receptors and to certain plasma proteins, as well as the availability of the X-ray crystallographic structure of prealbumin,[35] encourages the further physico-chemical study and evaluation of analog binding to plasma proteins. It is hoped that elucidation of the specific physical interactions that occur upon binding to the plasma proteins may be applied, at least in part, to better understanding of those that occur upon binding to nuclear receptors. Thus, it appears that the *in vitro* "test systems" are among the best currently available for the physico-chemical study of thyromimetic activity.

In an attempt to better understand and explain the structure-activity relationships of these *in vivo* and *in vitro* assays, as well as to investigate the specific molecular interactions involved in the binding to nuclear receptors and plasma proteins, a number of experimental and theoretical studies were undertaken.

Nine new thyroid hormone analogs were synthesized in the course of these studies. The rationale for desiring to determine the thyromimetic activities of these analogs was based on a combination of molecular orbital studies, quantitative structure-activity relationship considerations, and the qualitative structure-activity relationships of the thyroid hormones and analogs. Chapter Two describes this rationale, as well as the synthetic procedures.[36]

A number of these analogs were tested for thyromimetic activity in the rat antigoiter assay and/or in binding to solubilized rat hepatic nuclear protein. Descriptions of these assays and the results obtained are presented in Chapter Three.[26,36] A description of the detailed

recalculation and standardization of analog activities in the rat antigoiter assay is also included.

Chapter Four describes a number of quantum mechanical studies on the thyroid hormones and analogs. As a preliminary study, the first extensive molecular orbital study of intramolecular hydrogen bonding in ortho-substituted phenols and thiophenols was conducted.[37] As a model system for the binding to nuclear receptors and plasma proteins of the outer (phenolic) ring of the thyroid hormones and analogs, subsequent calculations were conducted on the intermolecular hydrogen bonding of ortho-substituted phenols and phenoxides, as well as on the conformations of ortho-alkyl phenols.[38] The results of some preliminary calculations on the conformation of the alanine side chain are also described.[38]

The results of a large number of quantitative structure-activity relationship studies of the thyroid hormones and analogs are presented in Chapter Five.[25,26,39,40] Rat antigoiter activities, binding to rat hepatic intact nuclei and solubilized nuclear protein and to thyroxine binding globulin, and correlations between the in vivo and in vitro thyromimetic activities were examined.

Comments on future experimental and theoretical investigations, including utilization of analog structure-activity studies (especially in elucidating the specific molecular interactions involved in binding to macromolecules), are presented in Chapter Six.

Appendix I contains a detailed tabulation of the recalculated and standardized rat antigoiter bioassay activities.

Appendix II contains listings of computer programs which were written and used in the quantitative structure-activity relationship studies.

CHAPTER TWO: SYNTHESIS OF THYROID HORMONE ANALOGS

QUALITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS

A series of experimental and theoretical analog studies[2,16,17,38,41] have established that with certain (large) 3,5 substituents there are two local minima as a function of ϕ_1, ϕ_2 (the dihedral angles connecting the phenyl rings) in which the two aromatic rings are mutually perpendicular. These conformations ("distal" has the R_3, away from the inner ring; "proximal" has the R_3, toward the inner ring) are of approximately equal energy and are readily interconvertible at room temperature. In vivo rat antigoiter assay activities and in vitro binding to nuclear receptors directly reflect the dependence of thyromimetic activity on the ability of the 3,5 substituents to influence the diphenyl ether conformation. Adding a 2'-CH_3 substituent leaves only one local minimum in which the 2' group is distal to the inner ring. Bearing this in mind, the analog structure-activity relationships for the in vivo rat antigoiter assay[2,8] and in vitro binding to nuclear receptors[24-26,42,43] can be briefly summarized as follows for the 3', 4', and 5' substituents:

1. Maximal activity results from monosubstitution ortho to the 4'-OH by a moderately lipophilic alkyl or halogen 3'-substituent, which binds in a size-limited pocket (approximately the size of iodine) of the nuclear receptor.

2. It is the distal conformation (Figure 2-1) and not the proximal conformation (Figure 2-2) which is the active form of analogs monosubstituted ortho to the 4'-OH (as shown, in particular, by the activities of 2'-CH_3 conformationally "locked" analogs).

3. Disubstitution (3' and 5') ortho to the 4'-OH decreases activity (as compared to monosubstitution) in direct proportion to the size of the second ortho substituent.

4. A 4'-OH imparts maximal activity, with 4'-NH_2, 4'-OCH_3, and 4'-H groups decreasing activity by 10 to 100 fold. (In <u>in vivo</u> studies the phenolic OH may be replaced by functional groups which can be metabolically converted to it. Such compounds exhibit significant but not full thyromimetic activity as compared to the corresponding free phenolic compounds.)

Similarly, the structure-activity relationships for <u>in vitro</u> binding of analogs to the human plasma protein, thyroxine binding globulin (TBG),[29,30] can be briefly summarized as follows for the 3', 4', and 5' substituents:

1. Maximal binding results from disubstitution ortho to the 4'-OH by moderately lipophilic, electron-withdrawing 3' and 5' substituents.

2. Monosubstitution (relative to disubstitution) ortho to the 4'-OH and electron-donating 3',5' substituents significantly decrease binding.

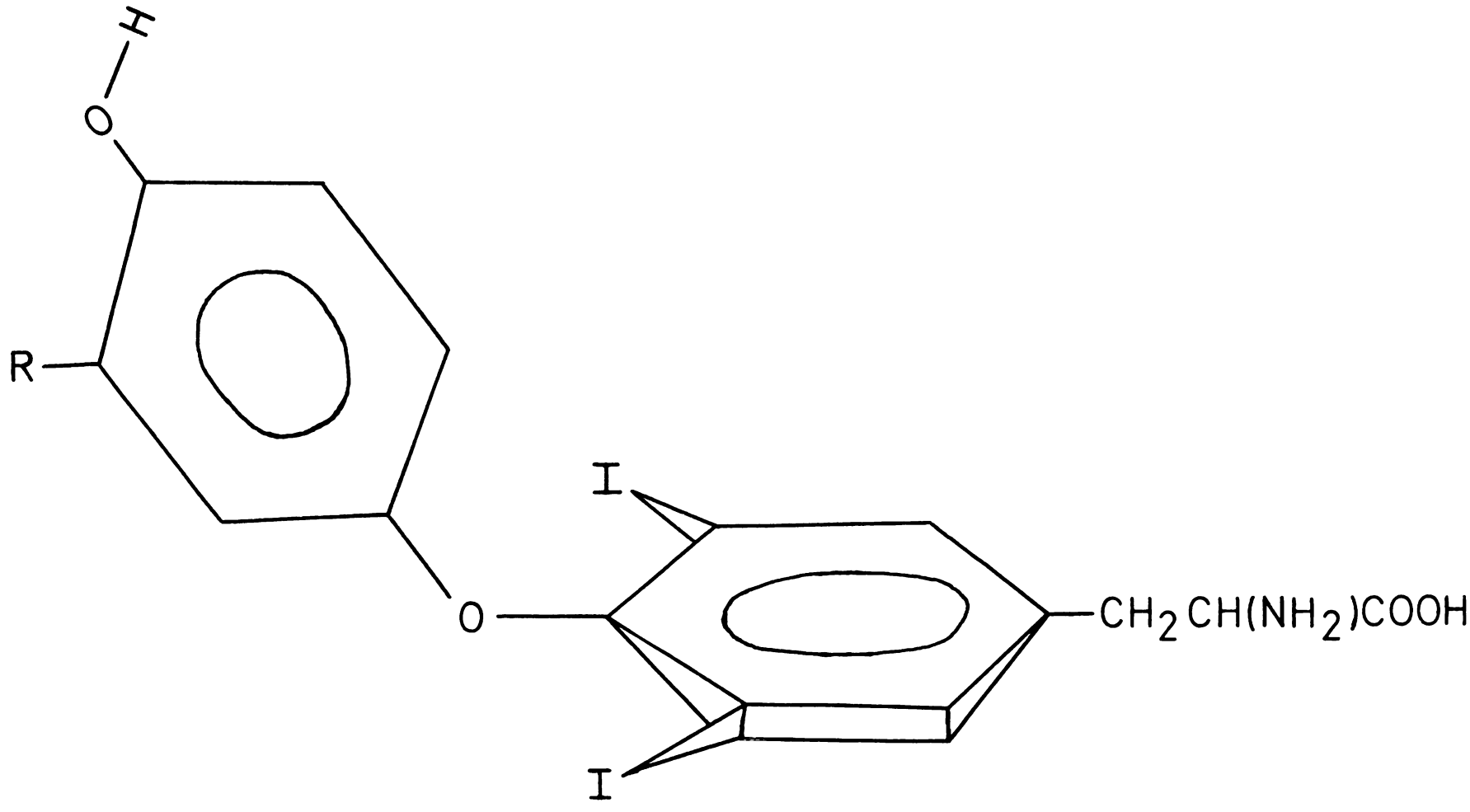


Figure 2-1. The distal conformation of thyroid hormones and analogs.

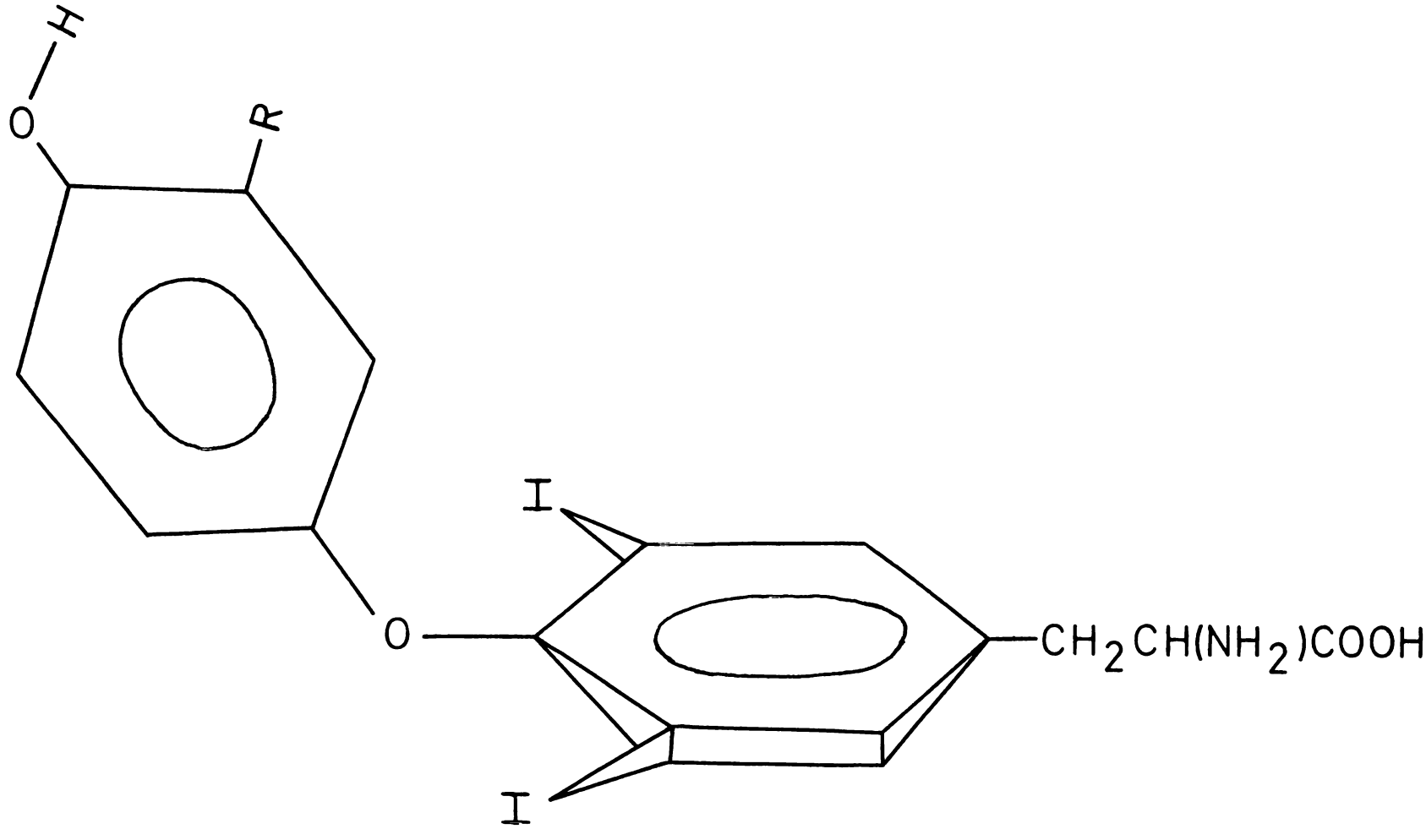


Figure 2-2. The proximal conformation of thyroid hormones and analogs.

3. For monosubstitution ortho to the 4'-OH, the same size limitations apparently exist for the 3' substituent as with the nuclear receptors.

4. 4'-OCH_3, 4'-NH_2, and 4'-H substituents result in a 30 to 50 fold decrease in binding, as compared to the 4'-OH.

On the basis of the *in vivo* rat antigoiter activities of analogs,[2,8] the binding of analogs to intact rat hepatic nuclei[24,25] and to solubilized nuclear protein,[26,43] and the pH dependence of *in vivo* binding of T_3 and T_4 to rat hepatic nuclear non-histone proteins,[44] it appears that the un-ionized 4'-phenolic hydroxyl is forming an intermolecular hydrogen bond with some appropriate functional group of the "receptor". In contrast, studies of the relative binding affinities of analogs to TBG[29,30] show that it is probably the 4'-phenoxide ion that binds to this plasma protein.

Prealbumin has also been found to strongly bind the thyroid hormones and analogs with a qualitative structure-activity picture similar to TBG: i.e.,binding affinity increases as the pK_a of the 4'-OH decreases.[45-48] In addition, X-ray crystallographic studies[35,49] of prealbumin show that the vicinity of the binding site where the 4'-O^- apparently binds contains no charged amino acid side chains, but rather a serine hydroxyl. On the basis of these X-ray studies and the similarity of TBG and prealbumin dependence of binding of thyroxine analogs on the 4'-OH pK_a, we have assumed that the 4'-O^- binds in both proteins *via* a hydrogen bond.

The measurement of *in vitro* equilibrium binding affinities of thyroxine analogs as apparent binding constants K permits the analysis of such binding affinities in terms of individual group contributions to the free energy of binding, according to the following approach.[26,42] The apparent free energy of binding (ΔG) of an analog to nuclear

receptors or plasma proteins can be estimated by:

$$\Delta G = -RT \ln K \qquad \text{(Eqn. 2-1)}$$

where: R = the ideal gas constant

$= 1.987 \times 10^{-3}$ kcal/(mole•deg)

T = the experimental temperature (°K)

The contribution of group(s) [$\Delta G(X)$] present at a certain position(s) on the molecule to the free energy of binding is determined relative to group(s) Y present at the same position(s) by:

$$\Delta G(X) = \Delta G(AX) - \Delta G(AY)$$

$$= -RT \ln [K_{AX}/K_{AY}] \qquad \text{(Eqn. 2-2)}$$

where: $\Delta G(AX)$ = the apparent free energy of binding of the analog containing group(s) X at a certain position(s)

$\Delta G(AY)$ = the apparent free energy of binding of the analog having the reference group(s) Y at the same position(s)

K_{AX} and K_{AY} = the corresponding analog binding constants

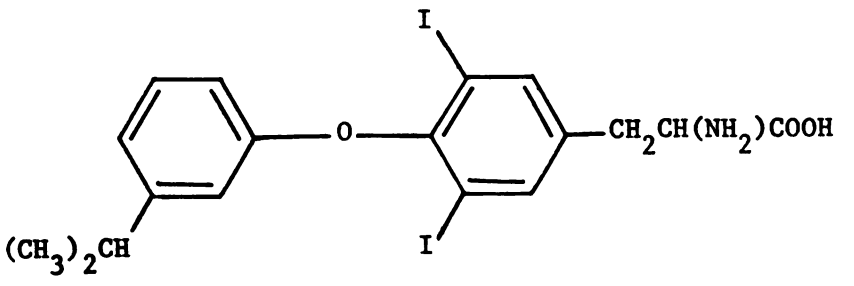
The validity of this "additivity" assumption is best verified if $\Delta G(X)$ values determined from more than one pair of structurally different compounds are similar. When two or more groups are far apart on the molecule, the additivity assumption has been found to be valid.[26,42] When the groups are close together (e.g., 3',4' disubstitution), interactions between them result in significant deviations from additivity.

SYNTHETIC RATIONALE

Preliminary molecular orbital calculations on the intramolecular hydrogen bonding in ortho-substituted phenols, as well as subsequent calculations on the interactive effects of ortho-substituents on the

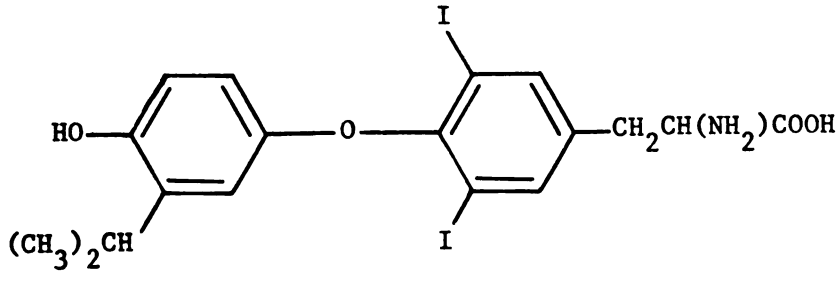
intermolecular hydrogen bonding of ortho-substituted phenols (see Chapter Four), suggested that 3' and 5' substituents can affect the hydrogen bonding of the 4'-hydroxyl with nuclear receptors. This assumption was verified by the finding that the contributions of 3' and 4' substituents to the free energy of binding to solubilized rat hepatic nuclear protein significantly deviate from simple additivity due to interactions between these substituents. In particular, it was qualitatively observed[26,38] and then semi-quantitatively shown (see Chapter Five) that the contribution of the 4'-OH to the free energy of binding is (a) increased by steric repulsion with and electron withdrawal by the 3'substituent, and (b) decreased by hydrogen bond formation with and electron donation by the 3'-substituent. These findings suggested that the 4'-OH is binding to the nuclear receptor by hydrogen bond donation toward the 5' side of the phenolic ring (i.e., away from the 3' substituent).

4'-Deoxy-3,5-diiodo-3'-isopropyl-L-thyronine (2-1) was specifically synthesized for testing in binding studies in order to enlarge the number of data points with which these 3'/4' substituent interactions could be studied. In addition, it would provide a direct estimate of the contribution of the 3'-isopropyl to the free energy of binding to the nuclear receptor. There is direct evidence for in vivo metabolic



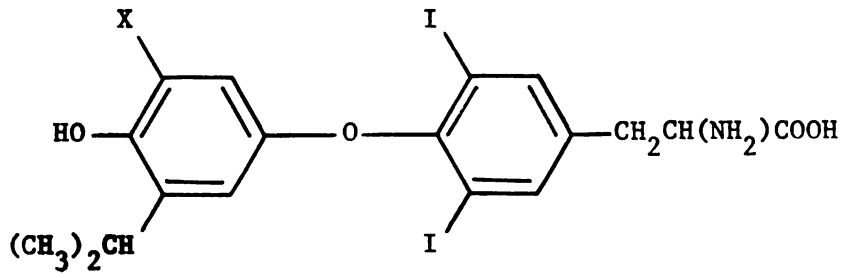
2-1

4'-hydroxylation of 4'-deoxy analogs.[50] This is supported by the observed enhancement of the activities of 4'-deoxy analogs in _in vivo_ assays,[23,51] as compared to their binding affinities to solubilized rat hepatic nuclear protein.[26] This enhancement due to 4'-hydroxylation _in vivo_ can be accounted for in correlations between _in vivo_ activities and _in vitro_ nuclear binding by use of an indicator variable.[26] 2-1 should be hydroxylated _in vivo_ to the most active thyroid hormone known, 2-2. Hence, of equal importance for the synthesis of 2-1 was to test the reliability of correlations between _in vivo_ antigoiter activities and _in vitro_ binding affinities to solubilized rat hepatic nuclear protein in predicting



2-2

the former based on the observed values for the latter (see Chapters Three and Five).



2-3, X = Cl

2-4, X = Br

2-5, X = I

Synthesis of the 3,5-diiodo-3'-halo-5'-isopropyl-L-thyronines (2-3, 2-4, and 2-5) was desired for several reasons:

1. Of the over 500 thyroxine analogs synthesized,[52] essentially all 3',5'-disubstituted analogs possess identical 3' and 5' substituents: e.g., dimethyl, dichloro, dibromo, diiodo, etc.

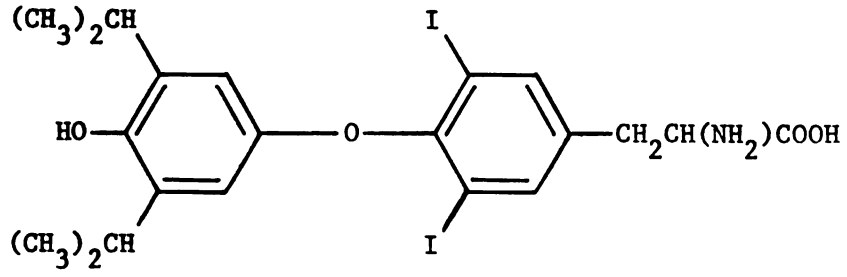
These three analogs provide a means for deviating from this trend and for the mixing of 3' and 5' alkyl and halo substituents.

2. This mixing of 3' and 5' alkyl and halo substituents provides analogs with total 3' and 5' electronic contributions (i.e., sum of sigma constants for 3' and 5' substituents) lying between those of the 3',5'-dialkyl analogs and those of the 3',5'-dihalo analogs. This increased randomness of 3',5' substituent electronic effects, as compared with 3',5' substituent hydrophobicities, helps to insure a greater lack of colinearity between these physico-chemical properties. Orthogonality of the associated substituent parameters enables more precise estimation of the relative importance of 3',5' substituent electronic and lipophilic properties.

3. According to the model of outer ring binding to receptors, the inherent lack of substitution symmetry of the phenolic ring of each of these analogs should be reflected in their nuclear receptor binding affinities by a combination of effects, including: (a) inverse correlation of thyromimetic activity with the proximal substituent size; (b) occupation of the distal position by a larger substituent (i.e., the isopropyl here); and (c) orientation of the 4'-OH away from the distal 3'-isopropyl and towards the proximal 5'-halogen: i.e., in the direction in which the 4'-OH is proposed to donate a hydrogen bond to the receptor.

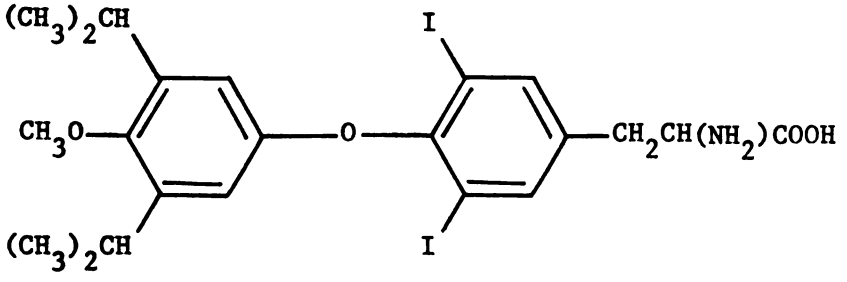
3,5-Diiodo-3',5'-diisopropyl-thyronine (2-6) had previously been synthesized as the DL analog[53] and demonstrated very low hypochlesteremic activity (ability to lower plasma cholesterol levels) (<0.1 relative

to L-T_3 = 100)[53] and fairly weak binding to intact rat hepatic nuclei (1.4% relative to L-T_3 = 100%).[24,25] It was never tested for rat



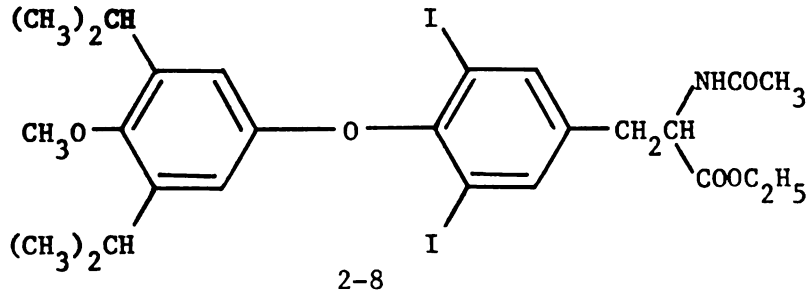
2-6

antigoiter activity. Thin layer chromatographic inspection of a sample of this compound indicated significant contamination with the 4'-O-methyl ether (2-7) due, apparently, to incomplete hydrolysis of the protected analog (2-8). In view of the importance of 5' substituent bulk in decreasing antigoiter activity and nuclear binding and in increasing TBG binding, the L-analog of 2-6 was synthesized in order to provide a



2-7

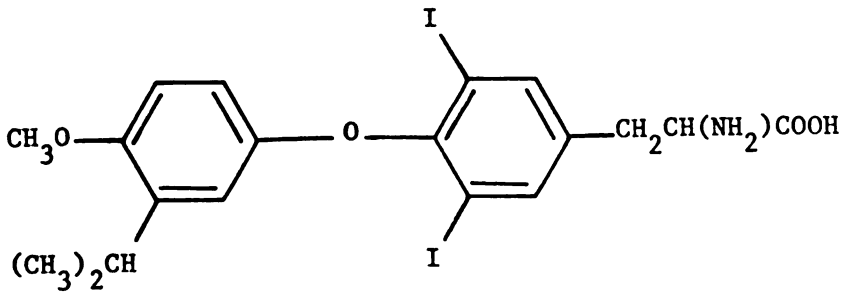
sample of this analog which would be free of 4'-O-methyl ether (2-7) contamination and whose thyromimetic activities could be compared



2-8

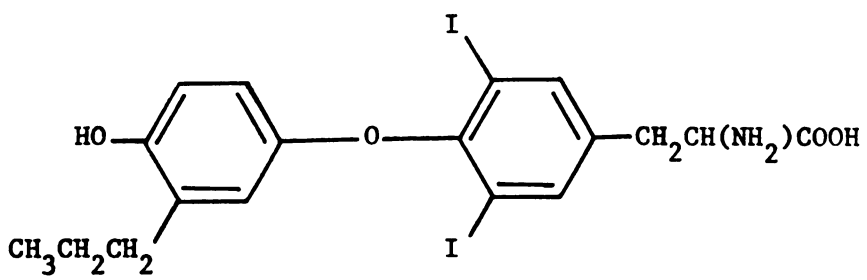
directly with other L-analogs without the complications of side chain stereochemistry differences. 4'-O-Methyl-3,5-diiodo-3'5'-diisopropyl-L-thyronine (2-7) was initially synthesized to insure that samples of 2-6 contained none of this compound as impurity.

4'-O-Methyl analogs are apparently O-demethylated *in vivo*.[8,54,55] The enhancement of the *in vivo* antigoiter activities of these 4'-O-methyl analogs, as compared to their much lower *in vitro* binding affinities to solubilized rat hepatic nuclear protein, can be accounted for in correlation of these *in vivo* activities with *in vitro* binding affinities by use of an indicator variable which takes into account their *in vivo* metabolic transformation to the more active 4'-OH compounds[26,39] (and see Chapter Five). 4'-O-Methyl-3,5-diiodo-3'-isopropyl-L-thyronine (2-9) was synthesized in order to increase the limited number of 4'-O-methyl analogs with which such correlations could be made. 2-7, although not originally synthesized with this purpose in mind, could also be used to extend such 4'-O-methyl analog studies.



2-9

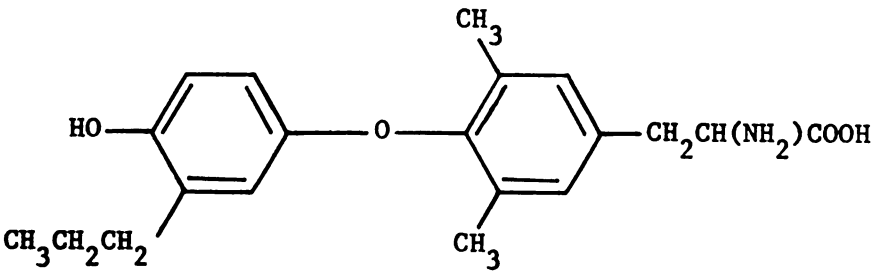
For rat antigoiter activities, binding to rat hepatic intact nuclei and solubilized nuclear protein, and binding to TBG, the 3' substituent apparently binds in a size-limited, hydrophobic pocket approximately the size of iodine.[25,30,38,39,56] That is, a negative steric effect is observed, especially utilizing quantitative structure-activity studies[25,39,56] (and see Chapter Five), for 3' substituents larger than iodine in direct proportion to the distance which such substituents extend out from the 3'-ring carbon further than a 3'-iodine. 3,5-diiodo-3'-n-propyl-L-thyronine (2-10) was synthesized with the specific intent of comparing its thyromimetic activities with those of the isomeric 3,5-diiodo-3'-isopropyl-L-thyronine (2-2). The lipophilicities of the



2-10

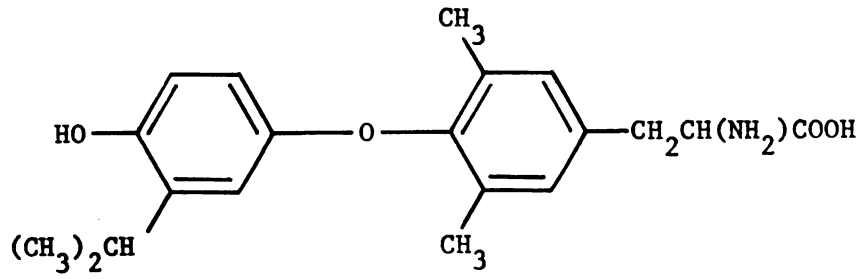
3'-n-propyl of 2-10 and of the 3'-isopropyl of 2-2 are approximately the same (π(n-propyl/benzene system) = 1.55; π(isopropyl/benzene system) = 1.53).[57] The "sizes" of these two substituents as they extend out from the 3'-ring carbon are, however, quite different. In addition, the "steric bulk" of these two substituents, as reflected in their steric interactions with the 4'-hydroxyl, are also quite different (see Chapter Four). Hence, the activity of the 3'-n-propyl analog (2-10) was expected to be somewhat less than that of the 3'-isopropyl analog (2-2). As it turned out (see Chapters Three and Five), the thyromimetic antigoiter and nuclear binding activities of the 3'-n-propyl analog (2-10) were found to be considerably less than those of the corresponding 3'-isopropyl analog (2-2), indicating just how crucial the exact interactions of the 3'-substituent with both the 4'-OH and with the nuclear receptor are.

3,5-Dimethyl-3'-n-propyl-L-thyronine (2-11) was synthesized for



2-11

testing for thyromimetic activities in order to: (a) increase the number of halogen-free, 3,5-dimethyl analogs for which thyromimetic activities are available for analysis; (b) compare its activities (just as with 2-10) with the corresponding 3'-isopropyl analog (2-12); and (c) further investigate the additivity (or lack of) of



2-12

3, 5, and 3' substituent contributions to thyromimetic activity.

SYNTHETIC SCHEMES

The synthetic pathways to the desired thyroid hormone analogs are presented in Figures 2-3 through 2-8. In most cases, well established general procedures for the synthesis of intermediates and the final analogs were utilized, although modifications were required for several reactions. It was found that several of the desired amino acid analogs could not be sufficiently purified by isoelectric reprecipitation/ recrystallization from acidic or basic ethanolic solutions. Final purification of these compounds (to give analytical purity suitable for physical measurements and bioassay) was accomplished by means of preparative thin layer chromatography. All of the amino acid side chains were of L-stereochemistry (S configuration).

Protection of the amino acid side chain of 3,5-diiodo-L-tyrosine (2-13) was accomplished by N-acetylation with acetic anhydride in base[58] and then carboxyl esterification with EtOH utilizing p-toluenesulfonic acid[58] to give the desired N-acetyl-3,5-diiodo-L-tyrosine ethyl ester (2-15). The unhindered 2-isopropylphenol (2-16) and 2-n-propylphenol (2-18) were O-methylated in Claisen's alkali[59] with Me_2SO_4 to give the

corresponding substituted anisoles[60] (2-19 and 2-21). O-Methylation of the very hindered 2,6-diisopropylphenol (2-17) was accomplished by treating sodium 2,6-diisopropylphenoxide in dioxane[61] with Me_2SO_4. Attempts to convert 2-isopropylanisole (2-19) and 2-n-propylanisole (2-21) to the corresponding di-(3-substituted-4-methoxyphenyl)-iodonium iodides (2-22) and 2-24) by general synthetic methods[53] utilizing iodine tris-(trifluoroacetate) resulted in poor yields. Drastic modification of the reaction workup conditions resulted in increased, acceptable reaction yields. The yield for conversion of 2,6-diisopropylanisole (2-20) to di-(3,5-diisopropyl-4-methoxyphenyl)-iodonium iodide (2-23) with iodine tris-(trifluoroacetate) was doubled (as compared to the literature preparation[53]) by allowing the reaction to proceed for 20 hours at room temperature (literature[53] reaction conditions: refrigerated overnight and then 3 hours at room temperature). Apparently the two ortho isopropyl groups of 2,6-diisopropylanisole (2-20) sterically force the OCH_3 out of coplanarity and, hence, conjugation with the aromatic ring. The resulting decrease in electron donation to the ring carbon para to the OCH_3 group reduces the reactivity of this position, necessitating the more "drastic" reaction conditions. In contrast, the essentially unhindered conjugation of the OCH_3 of 2-isopropylanisole (2-19) and 2-n-propylanisole (2-21) provides adequate electron density at the 4-position for rapid reaction of these compounds. The same conjugation of the OCH_3, however, also increases the instability of the resulting substituted dianisyl-iodonium iodides (2-22 and 2-24). The modified reaction workup for these two compounds apparently avoids much of the product decomposition. Condensation[53] of 2-15 with the substituted

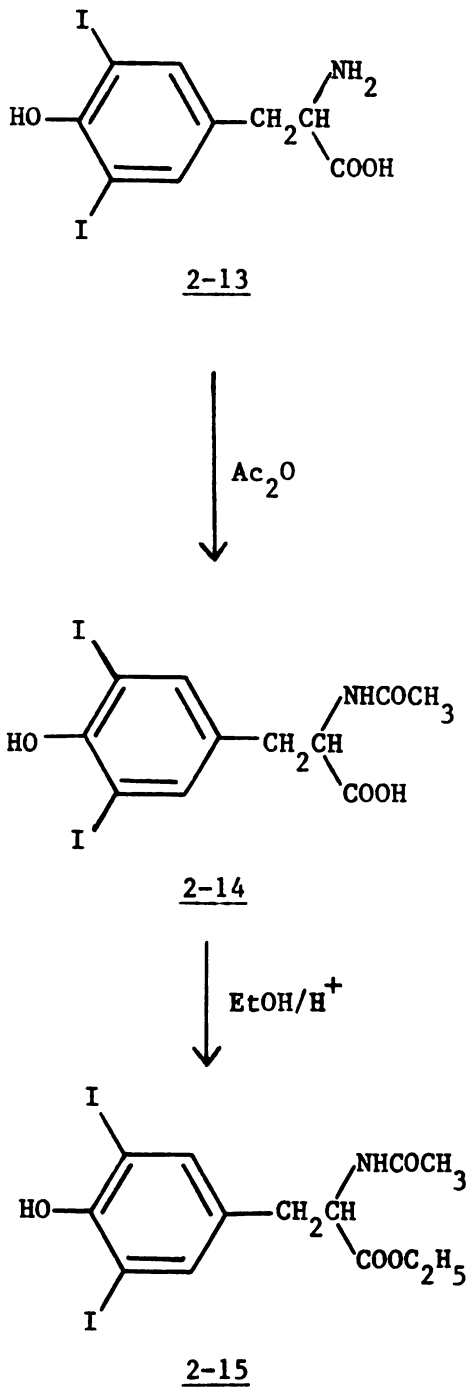
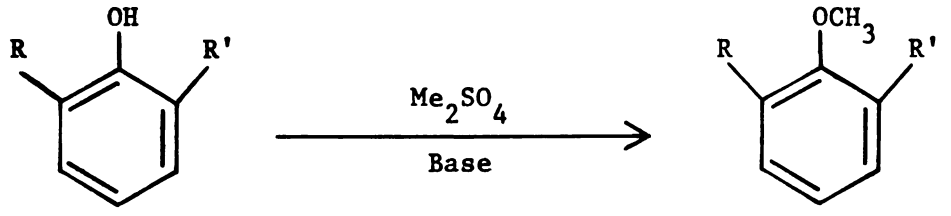


Figure 2-3. Synthetic pathway to N-acetyl-3,5-diiodo-L-tyrosine ethyl ester (2-15).



2-16, R = iPr; R' = H

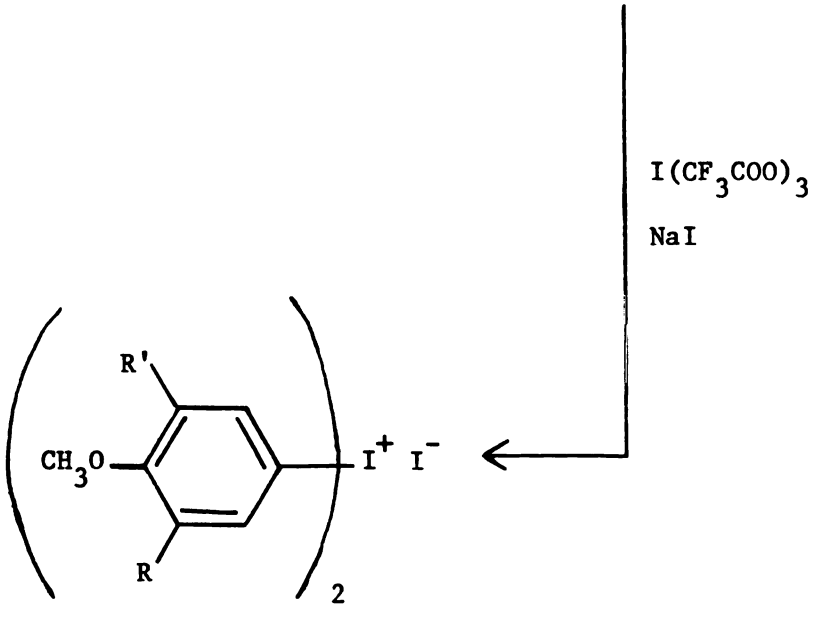
2-17, R = R' = iPr

2-18, R = nPr; R' = H

2-19, R = iPr; R' = H

2-20, R = R' = iPr

2-21, R = nPr; R' = H



2-22, R = iPr; R' = H

2-23, R = R' = iPr

2-24, R = nPr; R' = H

Figure 2-4. Synthetic pathway to the substituted di-(p-anisyl)-iodonium iodides.

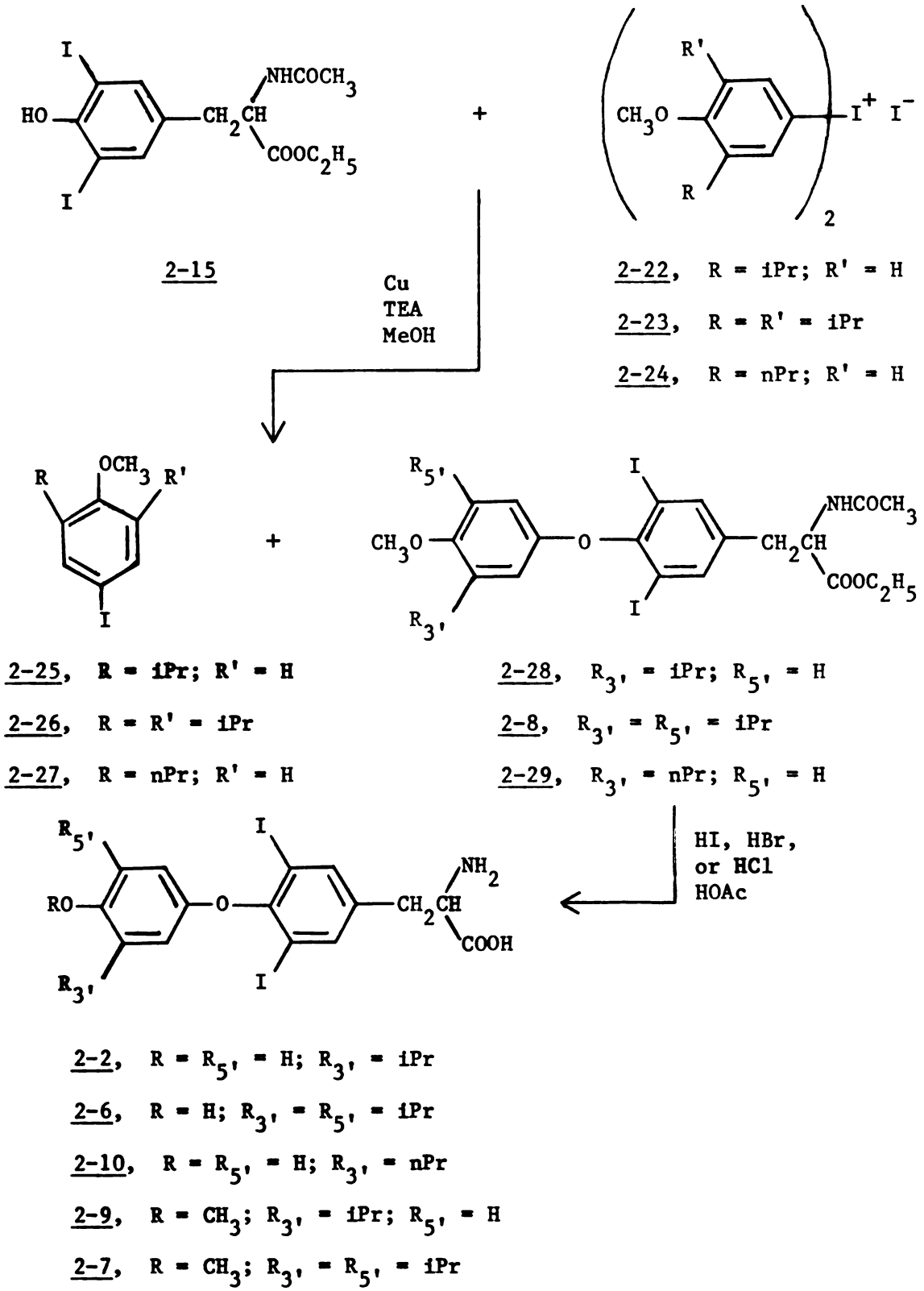
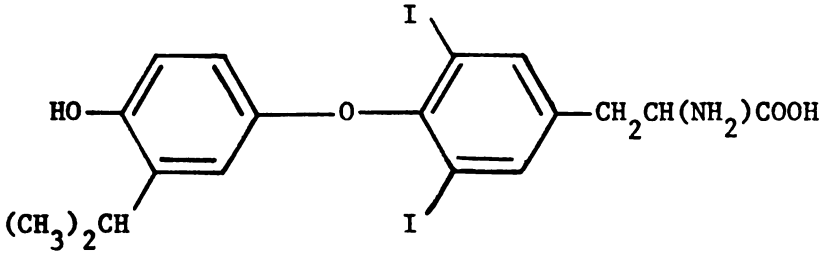
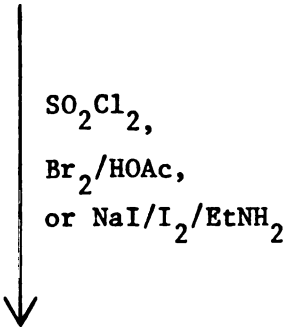
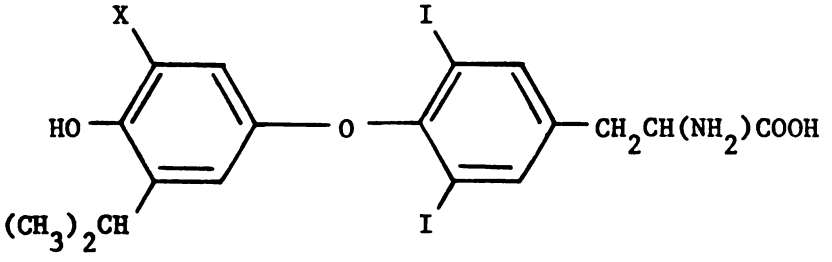


Figure 2-5. Synthetic pathways to the 3,5-diiodo-L-thyronine analogs.



2-2



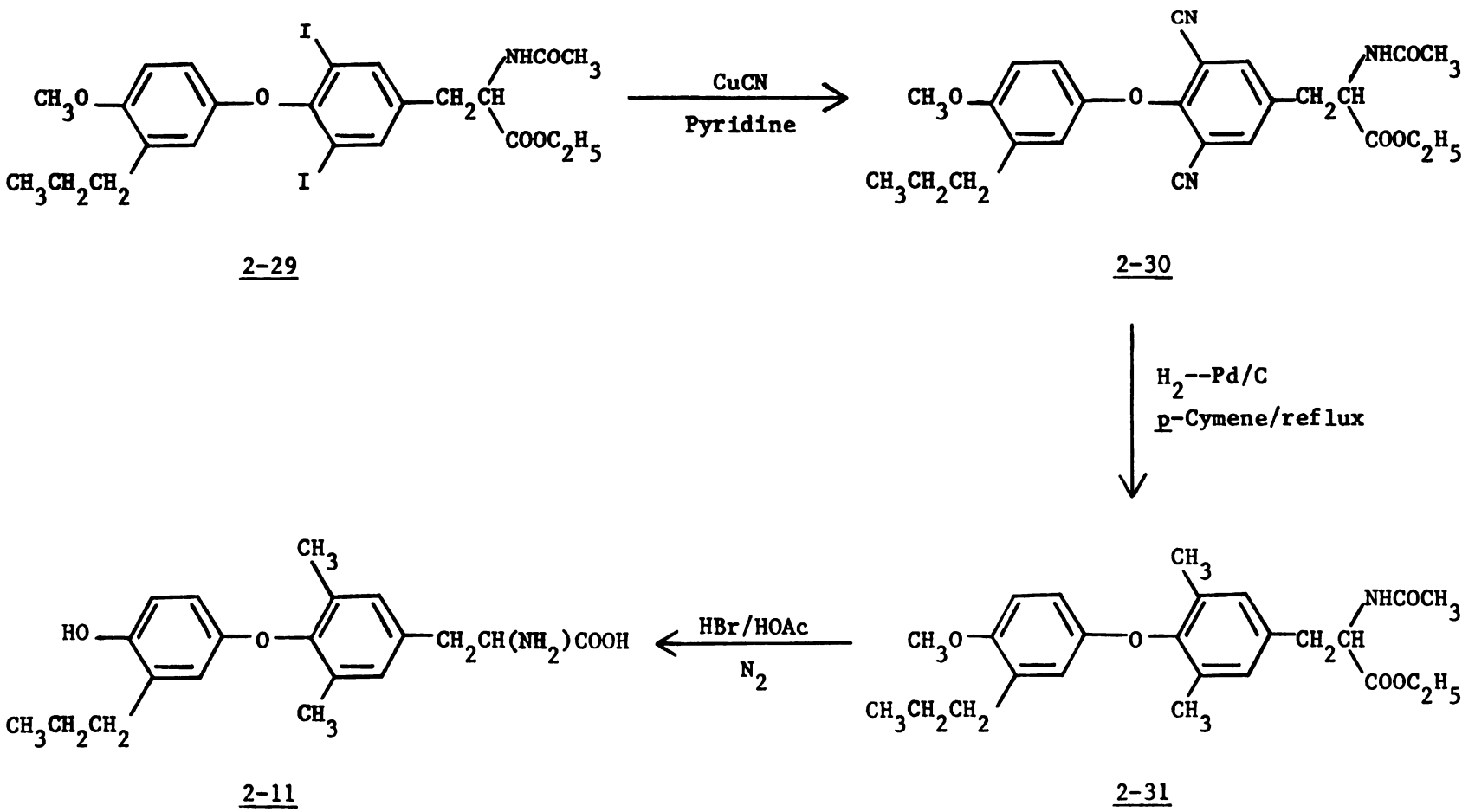


2-3, X = Cl

2-4, X = Br

2-5, X = I

Figure 2-6. Synthetic pathway to the 3,5-diiodo-3'-halo-5'-isopropyl-L-thyronines.



2-29 2-30

2-11 2-31

Figure 2-7. Synthetic pathway to 3,5-dimethyl-3'-n-propyl-L-thyronine (2-11).

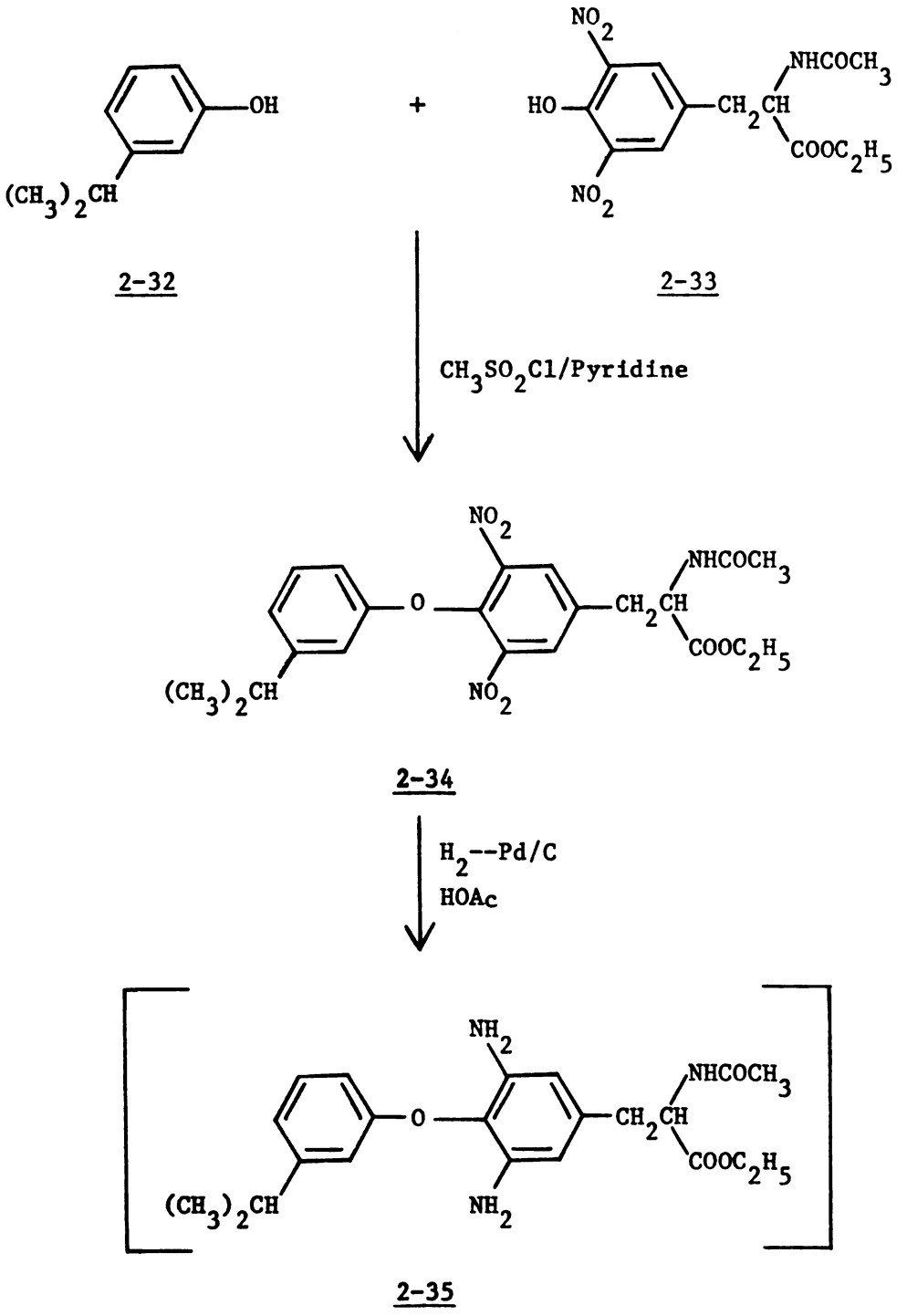


Figure 2-8. Synthetic pathway to 4'-deoxy-3,5-diiodo-3'-isopropyl-L-thyronine (2-1).

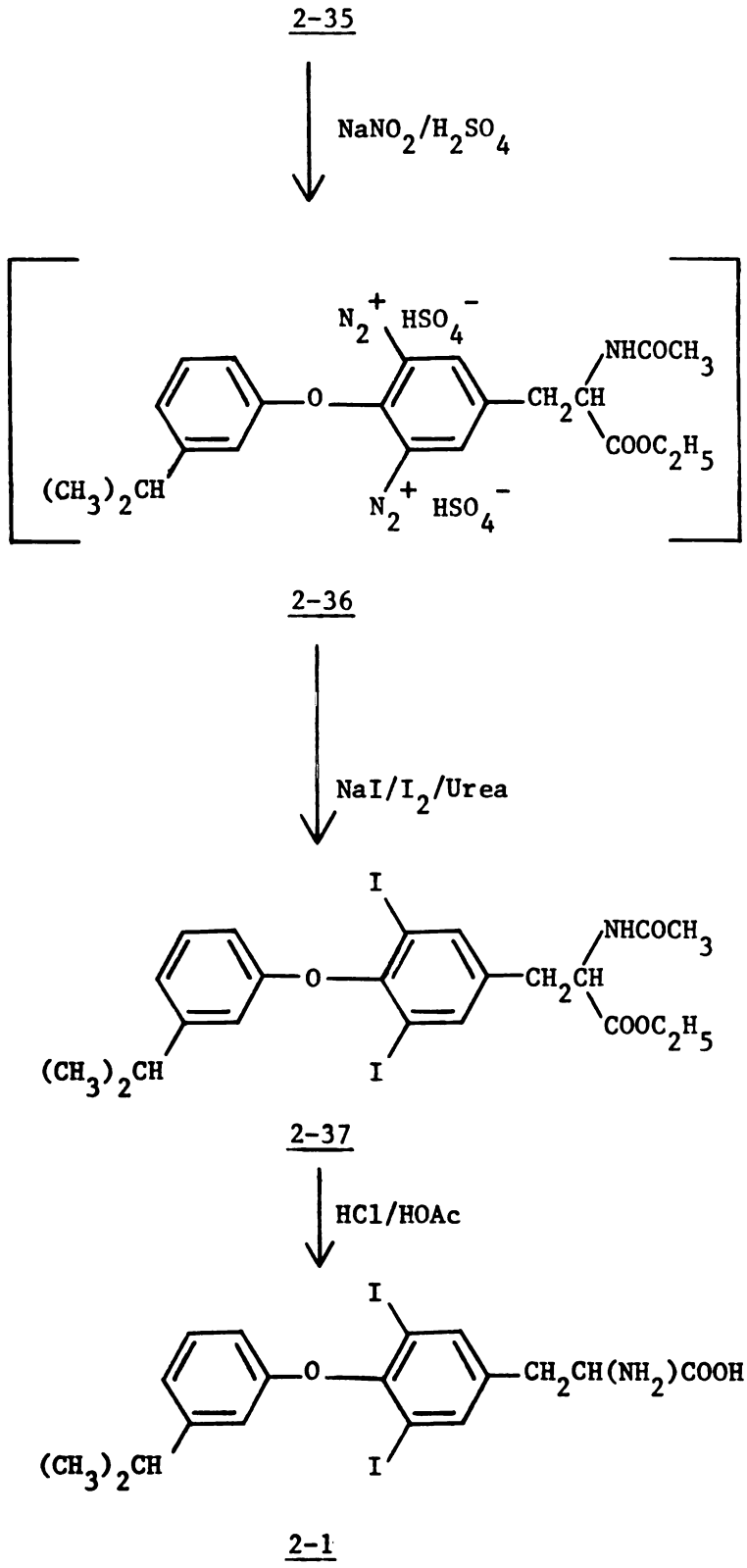


Figure 2-8. (Continued)

dianisyl-iodonium iodides (2-22, 2-23, and 2-24) in MeOH with copper powder and triethylamine yielded the protected thyroid hormone analogs (2-28, 2-8, and 2-29). As further verification of the structures of the substituted dianisyl-iodonium iodides (2-22, 2-23, and 2-24), the previously unreported substituted 4-iodo-anisoles (2-25, 2-26, and 2-27) were isolated and characterized as by-products of these condensation reactions. Hydrolysis[53] of 2-28, 2-8, and 2-29 with HI (or HBr) in HOAc yielded the desired 4'-OH thyroid hormone analogs (2-2, 2-6, and 210). Hydrolysis[53] of 2-28 and 2-8 with HCl in HOAc yielded the corresponding 4'-OCH_3 thyroid hormone analogs (2-9 and 2-7). Treatment of 3,5-diiodo-3'-isopropyl-L-thyronine (2-2) with SO_2Cl_2, Br_2 in HOAc, and I_2/NaI in 70% aqueous ethylamine yielded the desired 3,5-diiodo-3'-halo-5'isopropyl-L-thyronines (2-3, 2-4, and 2-5).

Treatment of N-acetyl-3,5-diiodo-4-(3'-n-propyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-29) with CuCN in refluxing pyridine[9,62] gave the corresponding 3,5-dicyano compound (2-30) in excellent yield. Hydrolytic reduction of 2-30 with H_2--Pd/C in refluxing, purified p-cymene[9,62] provided the corresponding 3,5-dimethyl compound (2-31). (Careful purification and drying of the p-cymene[62] and anhydrous reaction conditions were found to be extremely critical for this reaction. The slightest moisture or impurities in the p-cymene very effectively poisoned the catalyst.) Hydrolysis[9,26] of the protected analog (2-31) with HBr in HOAc under N_2 yielded the desired 3,5-dimethyl-3'-n-propyl-L-thyronine (2-11).

N-Acetyl-3,5-dinitro-L-tyrosine ethyl ester (2-33)[9,63] in pyridine was treated[9,53] with $MeSO_2Cl$ and then condensed with freshly distilled 3-isopropylphenol (2-32) to give N-acetyl-3,5-dinitro-4-(3'-isopropylphenoxy)-L-phenylalanine ethyl ester (2-34). Without isolation of intermediates,[53] the NO_2 groups of 2-34 were reduced (H_2--Pd/C in HOAc) to give 2-35, the NH_2 groups of 2-35 were diazotized (nitrosyl sulfuric acid) to give 2-36, and the diazonium groups of 2-36 were replaced (NaI/I_2/urea) to give the 3,5-diiodo compound (2-37). Hydrolysis[53] of 2-37 with HCl in HOAc gave the desired 4'-deoxy-3,5-diiodo-3'-isopropyl-L-thyronine (2-1).

EXPERIMENTAL SECTION

Melting points, determined with a Thomas-Hoover Uni-Melt stirred oil capillary tube melting point apparatus, are uncorrected. 60 MHz proton magnetic resonance (PMR) spectra were determined with a Varian Model A-60A PMR spectrometer. PMR chemical shift values are expressed in δ units (parts per million) relative to a TMS internal standard. For the presentation of the PMR spectra, the following abbreviations are used: s= singlet, d = doublet, t = triplet, q = quartet, m = multiplet, comp m = complex multiplet. Infrared (IR) spectra of liquid phenols and anisoles were recorded neat as thin films with a Perkin-Elmer Model 337 grating infrared spectrometer. Microanalyses were performed by the Microanalytical Laboratory, University of California, Berkeley, Calif. Optical rotations were measured with a Perkin-Elmer Model 141 Polarimeter (microcell: 10 cm path length, 1 ml cell volume).

Thin layer chromatography (TLC) was routinely used to check the purity of samples, to follow the progress of reactions, and to analyze

elutent fractions obtained with column chromatography of samples. For presentation of the TLC data, the following abbreviations are used for plate types: A = pre-coated silica gel sheets with flourescent indicator (100 μ coating on flexible plastic sheets; Eastman Kodak Company, #13181); B = pre-coated 4-channel silica gel plates with flourescent indicator (250 μ coating on glass with pre-absorbent loading zone; Quantum Industries, Inc., #5052); C = pre-coated alumina sheets with flourescent indicator (100 μ coating on flexible plastic sheets; Eastman Kodak Company, #6063).

The following general procedure was used for preparative TLC of several of the amino acids (50 to 120 mg per plate). 20 cm x 20 cm pre-coated silica gel plates with flourescent indicator (1000 μ coating on glass with pre-absorbent loading zone; Quantum Industries, Inc., #5080) were developed with $CHCl_3$-MeOH-conc. NH_4OH (20:10:1). After air drying 5 minutes and oven drying 10 minutes, the appropriate zones were manually removed with the aid of brief and careful UV visualization. The plate scrapings were extracted for 20 minutes with warm anhydrous EtOH and filtered through filter aid, washing with warm anhydrous EtOH. (Extraction of the scrapings with acidic or basic solutions was found to also extract the flourescent indicator.[64]) After addition of a small amount of water and several drops of concentrated HCl, the ethanolic solution was reduced to a minimal volume (rotary evaporation). This was heated on a steam bath for 1 minute, and then the pH was adjusted to 5.2 with hot H_2O and hot 2N sodium acetate. After allowing to cool to room temperature, the solution was refrigerated overnight. The precipitate was collected by centrifugation, washing with cold H_2O, and dried in vacuo to give the purified amino acid.

N-Acetyl-3,5-diiodo-L-tyrosine (2-14). 3,5-Diiodo-L-tyrosine (2-13) (50.00 g, 115.5 mmoles; Nutritional Biochemicals Corporation) was acetylated with acetic anhydride using the method of Barnes, et al.,[58] to give this compound (48.08 g, 87%; lit.[58] 87%). MP 113-119° (lit.[58] 112-118°).

N-Acetyl-3,5-diiodo-L-tyrosine Ethyl Ester (2-15). N-Acetyl-3,5-diiodo-L-tyrosine (2-14) (140.0 g, 294.7 mmoles) was esterified with EtOH using the method of Barnes, et al.,[58] to provide this compound (124.5 g, 84%; lit.[58] 88%). MP 154-155° (lit.[58] 154-155°). $[\alpha]_D^{31}$ = +13.1° (c, 2.0, EtOH) (lit.[58] $[\alpha]_D^{23}$ = + 15.4° (c, 2.0, EtOH)). TLC (UV) R_f (A: $CHCl_3$) 0.35, R_f (A: $CHCl_3$-EtOAc/9:1) 0.45, R_f (A: C_6H_6) 0.03.

2-Isopropylanisole (2-19). 2-Isopropylphenol (2-16) (312.00 g; 2.29 moles; Aldrich Chemical Company, Inc.) in Claisen's alkali[59] was methylated with Me_2SO_4 utilizing the general procedure of Dhami and Stothers[60] for the O-methylation of unhindered phenols. Distillation in vacuo of the crude reaction product yielded the desired anisole (237.83 g, 69%). BP 47-48°/1.2 mm Hg (lit.[65] 198-199°/751 mm Hg). PMR ($CDCl_3$) δ 1.20 (d, J = 7 Hz, 6H, iPr-CH_3), 3.36 (m, J = 7 Hz, 1H, iPr-CH), 3.79 (s, 3H, O-CH_3: present in 2-19, absent in 2-16), 5.50 (concentration dependent, s, 1H, O-H: present in 2-16, absent in 2-19), 6.7-7.4 (comp m, 4H, Ar-3,4,5,6 H).

2,6-Diisopropylanisole (2-20). This compound was prepared utilizing the method of Coffield, et al.,[61] for the preparation of sodium 2,6-diisopropylphenoxide in dioxane. A total of 16.55 g (720 mmoles) of sodium metal was dispersed in 1500 ml of dioxane at 101° with vigorous stirring. After cooling to 60°, a solution of 128.29 g (720 mmoles) of 2,6-diisopropylphenol (2-17) (Ethyl Corporation) in 150 ml of dioxane was added

dropwise over a 30 minute period. A slow evolution of hydrogen was observed. The still vigorously stirred mixture was heated to 101°. After 35 minutes no further hydrogen evolution occurred and the sodium metal dispersion had disappeared. The solution was cooled to 56° and 70.0 ml (93.31 g, 740 mmoles) of Me_2SO_4 was added dropwise over a 10 minute period, with a resultant slight warming. (All steps to this point were run under H_2SO_4-dried N_2.) After stirring at room temperature all night, the green-brown solution was filtered to remove the inorganic precipitate and the solvent was removed by rotary evaporation. The resulting red oil in 500 ml ether was washed with 20% NaOH (100 ml, 2x) and then with H_2O (100 ml, 2x). The organic phase was dried over Na_2SO_4 and filtered. After removal of solvent, the resulting red oil was distilled in vacuo to give the desired anisole (125.29 g, 90.5%). BP 48-48.5°/0.5 mm Hg (lit.[66] 74°/3 mm Hg). IR 3575 cm^{-1} (phenolic OH: present in 2-17, absent in 2-20). PMR ($CDCl_3$) δ 1.23 (d, J = 7 Hz, 12H, iPr-CH_3), 3.34 (m, J = 7 Hz, 2H, iPr-CH), 3.72 (s, 3H, O-CH_3: present in 2-20, absent in 2-17), 4.91 (s, 1H, O-H: present in 2-17, absent in 2-20), 7.10 (s, 3H, Ar-3,4,5 H).

An attempt to prepare 2-20 from 2-17 using the method of Zimmer[66] resulted in the formation of no 2-20 and recovery of only 2-17.

2-n-Propylanisole (2-21). 2-n-Propylphenol (2-18) (100.47 g, 738 mmoles: Aldrich Chemical Company, Inc.) in Claisen's alkali[59] was methylated with Me_2SO_4 utilizing the general procedure of Dhami and Stothers[60] for O-methylation of unhindered phenols. Distillation in vacuo of the crude reaction product yielded the desired anisole (81.32 g, 73%). BP 76-80°/1.8 mm Hg (lit.[65] 207-209°/757.7 mm Hg). IR 3500 cm^{-1} (phenolic OH: present in 2-18, absent in 2-21). PMR ($CDCl_3$) δ 0.95 (t, J = 7 Hz, 3H, nPr-CH_3), 1.65 (comp m, 2H, nPr-CH_2C-Ar),

2.64 (t, J = 8 Hz, 2H, nPr-CH_2-Ar), 3.80 (s, 3H, O-CH_3: present in 2-21, absent in 2-18), 4.94 (s, 1H, O-H: present in 2-18, absent in 2-21), 6.7-7.35 (comp m, 4H, Ar-3,4,5,6 H).

Di-(3-isopropyl-4-methoxyphenyl)-iodonium Iodide (2-22). A modification[67,68] of the synthetic procedure of Blank, et al.,[53] was used. To 38.3 ml (41.4 g, 406 mmoles) of acetic anhydride, cooled to +5°, there was added with stirring 13.7 ml of fuming HNO_3 (Sp. Gr. $\tilde{\sim}$ 1.5) with the temperature being allowed to rise to, but not beyond 15°. With the temperature kept below 20°, 12.69 g (50.0 mmoles) of finely powdered iodine and then 25.6 ml of CF_3COOH were added. The mixture was then stirred for 30 minutes, during which time the temperature rose to room temperature, the iodine completely dissolved, and nitrogen oxides were evolved. The clean light yellow-orange solution was concentrated under reduced pressure (H_2O aspirator and then vacuum pump) at 30° until all the colored fumes of oxides of nitrogen had disappeared. The resulting clear light yellow solution of iodine tris-(trifluoroacetate) was dissolved in 45 ml of acetic anhydride. This solution was kept between -15° and -30° while a solution containing 30.04 g (200 mmoles) of 2-isopropylanisole (2-19), 88 ml of acetic anhydride, and 12.8 ml of CF_3COOH was added in about 20 minutes. The dark green solution was stirred for 1 hour, the solution being allowed to rise to room temperature. After stirring for an additional 15 minutes, the dark green solution was poured onto 20.0 g solid $Na_2S_2O_5$ and 100 g solid NaI. With ice bath cooling, 1000 ml ice water was rapidly added to give a clear yellow aqueous phase and a dirty yellow-brown solid. Benzene (120 ml; just enough to disperse the solid) was added with very vigorous stirring to give a clear yellow aqueous phase and an

orange organic phase in which was suspended a yellow solid. Addition of 1350 ml heptane with very vigorous stirring yielded an orange organic phase and a pale yellow benzene-heptane-H_2O-precipitate emulsion. Filtration, washing with heptane (300 ml) gave a very pale yellow-white solid, which was dried in vacuo at room temperature and protected from light (44.64 g, 81%; lit.[53] prep 86%). MP 151-154° (decomp.) (lit.[53] 164-166°). (The purity of this compound is adequate for the subsequent condensation reaction.) TLC (B: $CHCl_3$) streak from origin (length of streak dependent on compound load on TLC plate).

Di-(3,5-diisopropyl-4-methoxyphenyl)-iodonium Iodide (2-23). A modification[68] of the synthetic procedure of Blank, et al.,[53] was used. A solution of iodine tris-(trifluoroacetate) (prepared, as for 2-22, from 42.0 ml (45.4 g, 445 mmoles) of acetic anhydride, 16.2 ml of fuming nitric acid, 15.00 g (59.1 mmoles) of finely powdered iodine, and 30.3 ml of CF_3COOH) was dissolved in 45 ml of acetic anhydride. This solution was kept between -5° and -10° while a solution containing 45.46 g (236 mmoles) of 2,6-diisopropylanisole (2-20), 105 ml of acetic anhydride, and 15 ml of CF_3COOH was added in about 20 minutes. The clear yellow solution (protected from light) was then stirred at room temperature for 20 hours, during which time the solution turned red. The solvents were removed in vacuo (H_2O aspirator and then vacuum pump at 40°) and the resulting clear, yellow-orange oil was dissolved in 420 ml of absolute MeOH. The methanolic solution was cooled with an ice bath and then diluted with 150 ml of a 10% $Na_2S_2O_5$ solution. 120 g of KI in 750 ml H_2O was added dropwise with stirring over a 20 minute period, yielding a clear yellow solution and a bright yellow oil; the latter crystallized after overnight refrigeration of the mixture. The precipitate was collected by filtration, ground in mortar with hexane, and filtered again, yielding a light yellow-white solid, which was dried in vacuo at room temperature and

protected from light (48.51 g, 65%; lit.[53] prep 33%). MP 141-144° (decomp.) (lit.[53] 160-162°). (The purity of this material is adequate for the subsequent condensation reaction.) TLC (UV) (B: $CHCl_3$) streak from origin (length of streak dependent on compound load on TLC plate). PMR ($CDCl_3$) δ 1.17 (d, J = 7 Hz, 12H, iPr-CH_3), 3.30 (m, J = 7 Hz, 2H, iPr-CH), 3.74 (s, 3H, O-CH_3), 7.67 (s, 2H, Ar-2,6 H).

Di-(3-n-propyl-4-methoxyphenyl)-iodonium Iodide (2-24). A modification[67,68] of the general procedure of Blank, et al.,[53] was used. A solution of iodine tris-(trifluoroacetate) (prepared, as for 2-22, from 35.5 ml (38.4 g, 376 mmoles) of acetic anhydride, 13.7 ml of fuming nitric acid, 12.69 g (50.0 mmoles) of finely powdered iodine, and 25.6 ml of CF_3COOH) was dissolved in 45 ml of acetic anhydride. This solution was kept between -15° and -30° while a solution containing 30.04 g (200.0 mmoles) of 2-n-propylanisole (2-21), 88 ml of acetic anhydride, and 12.8 ml of CF_3COOH was added in about 20 minutes. The dark green-black solution was stirred for 1 hour, the temperature being allowed to rise to room temperature. The solution was then poured onto 20.0 g solid $Na_2S_2O_5$ and 100 g solid NaI. With ice bath cooling, 1000 ml of ice water was rapidly added to give a dark oil and a dirty red-brown aqueous phase. Benzene (200 ml; just enough to dissolve the oil) was added with very vigorous stirring to give a clear light yellow aqueous phase and a dark red organic phase. Addition of 1100 ml of heptane with very vigorous stirring for 30 minutes yielded an orange organic phase and an emulsion of H_2O, organic solvents, and a yellow-white solid. Filtration, washing first with H_2O (100 ml) and then with heptane (600 ml) yielded the product as a white solid (tinted slightly yellow), which was dried in vacuo at room temperature and protected from light

(22.05 g, 40%). MP 137-139° (decomp.). (The purity of 2-24 at this point is adequate for its use in the subsequent condensation reaction.) An analytical sample was prepared by recrystallization (three times) at room temperature as follows. Crude 2-24 (500 mg) was dissolved in about 10 ml of benzene. Heptane was added slowly with stirring to give a fine precipitate. Filtration yielded the recrystallized compound, which was dried in vacuo at room temperature and protected from light (213 mg of highly purified material after three recrystallizations). MP 145.5-147° (decomp.). TLC (UV) (B: $CHCl_3$) streak from origin (length of streak dependent on compound load on TLC plate). PMR ($CDCl_3$) δ 0.86 (t, J = 7 Hz, 3H, nPr-CH_3), 1.56 (comp m, 2H, nPr-CH_2-C-Ar), 2.54 (t, J = 7.5 Hz, 2H, nPr-CH_2-Ar), 3.78 (s, 3H, O-CH_3), 6.78 (d, J = 8.5 Hz, 1H, Ar-5 H), 7.70 (d, J = 2 Hz, 1H, Ar-2 H), 7.89 (q, J = 2 Hz, J = 8.5 Hz, 1H, Ar-6 H). Analysis: $C_{20}H_{26}I_2O_2$: Calculated C, 43.50; H, 4.75; I, 45.96; Found C, 43.31; H, 4.50; I, 46.04.

N-Acetyl-3,5-diiodo-4-(3'-isopropyl-4'-methoxyphenoxy)-L-phenylalanine Ethyl Ester (2-28). The general procedure of Blank, et al.,[53] was used. A mixture of 32.72 g (59.3 mmoles) of di-(3-isopropyl-4-methoxyphenyl)-iodonium iodide (2-22), 16.56 g (32.9 mmoles) of N-acetyl-3,5-diiodo-L-tyrosine ethyl ester (2-15), 5.0 ml of triethylamine, and 329 mg of copper powder in 395 ml of anhydrous MeOH was stirred vigorously at room temperature for 27.5 hours (the reaction flask being protected from light). The mixture was filtered and the filtrate was evaporated to a syrup, which was dissolved in 275 ml of benzene. The benzene solution was vigorously stirred for 30 minutes with 82 ml of 3% aqueous HCl. Precipitated triethylamine hydrochloride was removed by filtration, washing with an additional 150 ml of benzene. The organic phase was

washed with H_2O (175 ml, 2x), 10% NaOH (175 ml, 2x), and again with H_2O (175 ml, 2x), dried over Na_2SO_4, and filtered. Removal of the solvent yielded an orange, "wet" solid, which was ground up in a mortar with 50 ml of hexane. Filtration, washing with hexane, gave a clean white solid, which was air dried (8.87 g, 41%; lit.[53] prep. from the 3,5-diNO_2 analog of 2-28 gave 72%). MP 129-130° (lit.[53] 129-131°). $[\alpha]_D^{31}$ = + 49.6° (c, 1.0, $CHCl_3$) (lit.[53] $[\alpha]_D^{25}$ = + 41.6° (c, 1.0, $CHCl_3$)). TLC (UV) R_f (A: $CHCl_3$) 0.48.

N-Acetyl-3,5-diiodo-4-(3',5'-diisopropyl-4'-methoxyphenoxy)-L-phenylalanine Ethyl Ester (2-8). Conditions similar to those of Blank, et al.,[53] for preparation of the DL-analog were used. A mixture of 19.83 g (312 mmoles) of di-(3,5-diisopropyl-4-methoxyphenyl)-iodonium iodide (2-23), 8.71 g (173 mmoles) of N-acetyl-3,5-diiodo-L-tyrosine ethyl ester (2-15), 2.6 ml of triethylamine, and 180 mg of copper powder in 210 ml of anhydrous MeOH was stirred vigorously at room temperature for 24 hours (the reaction flask being protected from light). The mixture was filtered and the filtrate was evaporated to a syrup, which was dissolved in 70 ml of benzene. The benzene solution was vigorously stirred for 10 minutes with 45 ml of 3% aqueous HCl. Precipitated triethylamine hydrochloride was removed by filtration, washing with benzene. The organic phase was washed with H_2O (35 ml), 10% NaOH (35 ml), and again with H_2O (35 ml), dried over Na_2SO_4, and filtered. Removal of the solvent yielded an orange oil, which could not be triturated or crystallized. The oil, initially dissolved in 6 ml of benzene, was chromatographed on a column of 300 g of Silica gel (60-200 mesh, Grade 950; Matheson Coleman & Bell), eluting throught the elutropic solvent series $C_6H_6/CHCl_3$/EtOAc/EtOH (total elutent volume = 2.5 l).

TLC inspection enabled combination of the appropriate 100 ml elutent fractions #14-20 (EtOAc/EtOH solvent range approximately), complete removal of solvents from which yielded the desired compound in analytically pure form (3.72 g, 31%; lit.[53] 18% for the DL-analog). MP 68-70° (lit.[53] 147-148° for the DL-analog). $[\alpha]_D^{29}$ = + 41.0° (c, 2.0, $CHCl_3$). TLC (UV) R_f (A: C_6H_6) 0.03, R_f (A: $CHCl_3$) 0.47, R_f (A: EtOAc) 0.55, R_f (A: acetone) 0.73, R_f (A: MeOH) 0.76. PMR ($CDCl_3$) δ 1.18 (d, J = 7 Hz, 12H, iPr-CH_3), 1.30 (t, J = 7.5 Hz, 3H, Et-CH_3), 2.03 (s, 3H, Ac-CH_3), 3.10 (d, 2H, β-CH_2), 3.30 (m, J = 7 Hz, 2H, iPr-CH), 3.72 (s, 3H, O-CH_3), 4.28 (q, J = 7.5 Hz, 2H, Et-CH_2), 4.81 (m, 1H, α-CH), 6.48 (s, 2H, Ar-2',6' H), 6.72 (shift concentration dependent, d, J = 8 Hz, 1H, N-H), 7.72 (s, 2H, Ar-2,6 H). Analysis: $C_{26}H_{33}I_2N_1O_5$: Calculated C, 45.04; H, 4.80; I, 36.61; Found C, 45.21; H, 4.91; I, 36.46.

N-Acetyl-3,5-diiodo-4-(3'-n-propyl-4'-methoxyphenoxy)-L-phenylalanine Ethyl Ester (2-29). The general procedure of Blank, et al.,[53] was followed. A mixture of 9.50 g (17.2 mmoles) of di-(3-n-propyl-4-methoxyphenyl)-iodonium iodide (2-24), 4.81 g (9.56 mmoles) of N-acetyl-3,5-diiodo-L-tyrosine ethyl ester (2-15), 1.43 ml of triethylamine, and 96 mg of copper powder in 115 ml of anhydrous MeOH was stirred vigorously at room temperature for 24 hours, the reaction flask protected from light. The mixture was filtered and the filtrate was evaporated to a syrup, which was dissolved in 45 ml of benzene. The benzene solution was stirred vigorously for 30 minutes with 24 ml of 3% aqueous HCl. Precipitated triethylamine hydrochloride was removed by filtration, washing with an additional 150 ml of benzene. The organic phase was washed with H_2O (30 ml, 1x; 50 ml, 1x), 10% NaOH (50 ml), and again with H_2O (50 ml, 2x), dried over Na_2SO_4, and filtered. Removal of the solvent yielded a

dark brown-red oil. Repeated triturations (with ether and heptane) yielded a white solid, which was dried *in vacuo* (3.54 g, 57%). MP 136-137°. An analytical sample was recrystallized from hot ether/heptane: MP 137-137.5°. $[\alpha]_D^{32}$ = + 48.8° (*c*, 2.0, $CHCl_3$). TLC (UV) R_f (A: $CHCl_3$) 0.49. PMR ($CDCl_3$) δ 0.92 (t, J = 6.5 Hz, 3H, nPr-CH_3), 1.28 (t, J = 7 Hz, 3H, Et-CH_3), 1.57 (m, 2H, nPr-CH_2-C-Ar), 2.04 (s, 3H, Ac-CH_3), 2.55 (t, J = 7 Hz, 2H, nPr-CH_2-Ar), 3.03 (d, J = 6 Hz, 2H, β-CH_2), 3.75 (s, 3H, O-CH_3), 4.20 (q, J = 7 Hz, 2H, Et-CH_2), 4.80 (m, 1H, α-CH), 6.25-6.70 (m, 3H, Ar-2',5',6' H), 6.80 (d, 1H, N-H), 7.62 (s, 2H, Ar-2,6 H). Analysis: $C_{23}H_{27}I_2N_1O_5$: Calculated C, 42.41; H, 4.18; I, 38.97; Found C, 42.29; H, 4.10; I, 39.02.

N-Acetyl-3,5-dinitro-4-(3'-isopropylphenoxy)-L-phenylalanine Ethyl Ester (2-34). The general reaction conditions of Blank, *et al.*,[53] as modified by Jorgensen, *et al.*,[9] were used with slight modifications. Technical (60%) 3-isopropylphenol (2-32) (Aldrich Chemical Company, Inc.) was purified by distillation *in vacuo* (BP 66.5-68°/0.25 mm Hg; 86-87.5°/0.65 mm Hg; lit.[65] 228°/760 mm Hg). Purity of 2-32 was examined by: (1) Gas chromatography[69] with a Varian 2100 series Chromatographic Instrument (with flame ionization detector); > 99% purity but not separated from commercial *p*-isopropylphenol (Dow Chemical Company) on two different columns; retention time = 5.70 min., 3% OV-225, 6' x 2 mm glass column, flow = 18 ml N_2/min., 140° C; retention time = 3.94 min., Carbowax 20-M, 6' x 2 mm glass column, flow = 24 ml N_2/min., 200°C; (2) Comparison of thin film IR spectrum of 2-32 with Sadtler Standard grating IR spectrum #1073 of p-isopropylphenol; qualitatively indicated 2-32 to be essentially pure, but minor *p*-isopropylphenol contamination could not be rule out; (3) Comparison of PMR spectra of 2-32, *p*-isopropylphenol, and mixtures;

also indicated that 2-32 was essentially pure, but could contain minor (∿10%) p-isopropylphenol impurity.

To 44.09 g (129 mmoles) of N-acetyl-3,5-dinitro-L-tyrosine ethyl ester[9,63] (2-33) in 300 ml of dry pyridine there was added dropwise with stirring 11.0 ml (16.28 g, 142 mmoles) of $MeSO_2Cl$, and the mixture was heated under reflux for 2 minutes. After some cooling, 34.31 g (252 mmoles) of freshly distilled 3-isopropylphenol (2-32) was added with stirring and the mixture was heated under reflux for 20 minutes. Most of the pyridine was removed *in vacuo* and the residue was taken up in $CHCl_3$ (470 ml). The $CHCl_3$ solution was washed successively with 2N HCl (235 ml, 2x), 2N NaOH (235 ml, 1x; 120 ml, 3x) (to a light yellow aqueous solution), and saturated aqueous NaCl (235 ml) and was dried over Na_2SO_4. After filtration the $CHCl_3$ was removed under reduced pressure to give the product first as an oil and, with more complete solvent removal, finally as a glass (44.04 g, 75%). Attempts to triturate the oil or crystallize the oil or glass yielded only oils. The glass was judged to be analytically pure by TLC, PMR, and elemental analysis (see below). (As determined by weight changes upon drying the oil to the glass, the former contained almost imperceptible amounts of solvent.) MP 54-56°. $[\alpha]_D^{29}$ = + 42.8° (c, 2.0, $CHCl_3$). TLC (UV) R_f (A: C_6H_6) 0.06, R_f (B: $CHCl_3$) 0.21, R_f (A: $CHCl_3$) 0.49, R_f (B: EtOAc) 0.53, R_f (C: $CHCl_3$) 0.69. PMR ($CDCl_3$) δ 1.20 (d, J = 7 Hz, 6H, iPr-CH_3), 1.25 (t, J = 7 Hz, 3H, Et-CH_3), 1.98 (s, 3H, Ac-CH_3), 2.84 (m, 1H, iPr-H), 3.25 (d, J = 6 Hz, 2H, β-CH_2), 4.22 (q, J = 7 cps, 2H, Et-CH_2), 4.88 (m, 1H, α-CH), 6.6 (d, 1H, N-H), 6.4-7.4 (comp m, 4H, Ar-2',4',5',6' H), 7.96 (s, 2H, Ar-2,6 H). Analysis: $C_{22}H_{25}N_3O_8$: Calculated C, 57.51; H, 5.48; N, 9.15; Found C, 57.37; H, 5.42; N, 9.06.

N-Acetyl-3,5-diiodo-4-(3'-isopropylphenoxy)-L-phenylalanine Ethyl Ester (2-37). The general reaction procedure of Blank, et al.,[53] was used. A solution of N-acetyl-3,5-dinitro-4-(3'-isopropylphenoxy)-L-phenylalanine ethyl ester (2-34) (7.95 g, 17.3 mmoles; as a very viscous oil) in 173 ml of glacial acetic acid was reduced in a Parr apparatus in the presence of 1.73 g of 10% Pd/C under an initial pressure of 33.5 p/i^2 (2.28 atm) of hydrogen. When no further pressure drop had been observed for 25 minutes (50 minutes total; 85% of the theoretical H_2 pressure decrease), the catalyst was removed by filtration through filter aid and the filtrate was added to a stirred, cooled nitrosyl sulfuric acid solution (prepared by slowly adding 6.81 g of sodium nitrite to a mixture of 144 ml of sulfuric acid and 58 ml of glacial acetic acid at 60-70°) at such a rate that the temperature was maintained at 0-5°. After all the amine had been added, the tetrazonium solution was stirred and cooled an additional hour. It was then added rapidly to a mixture of 15.23 g of sodium iodide, 19.03 g of iodine, and 3.46 g of urea in 317 ml of H_2O and 317 ml of $CHCl_3$. Stirring was continued for 1.5 hours at room temperature. The aqueous layer was extracted with $CHCl_3$ (100 ml, 3x) and the $CHCl_3$ phase and extracts were washed in turn with H_2O (200 ml, 2x), 5% sodium carbonate (200 ml, 2x), and H_2O (200 ml, 2x). After drying over calcium chloride and then filtration, the solvent was removed to give a dark red oil, which gave only an oil upon repeated attempts to triturate with various solvents. After removal of the solvents to give an oil again, attempted crystallization from aqueous EtOH yielded only an oil. Complete removal of the aqueous EtOH in vacuo yielded a red glass, which becomes a red-brown powder upon being pulverized (8.63 g, 80% crude yield). MP 64-67°. TLC (UV) R_f (B: $CHCl_3$)

0.29, R_f (B: EtOAc) 0.55, R_f (C: $CHCl_3$) 0.75. Analysis: $C_{22}H_{25}I_2N_1O_4 \cdot C_2H_5OH$: Calculated C, 43.19; H, 4.68; I, 38.03; N, 2.10; Found C, 43.42; H, 4.43; I, 37.92; N, 2.41.

Elution of the glass from a column packed with acid alumina (Brockmann Activity Grade 1; J. T. Baker Chemical Co.) with $CHCl_3$, subsequent combination of the appropriate elutent fractions (as determined by TLC: see above), removal of solvent, failure to crystallize from aq. EtOH, and final complete (as possible) removal of solvent in vacuo yielded a red glass, which gave a yellow-brown powder upon pulverization. MP 64-68°. $[\alpha]_D^{29}$ = + 46.6° (c, 2.0, $CHCl_3$). PMR ($CDCl_3$) δ 1.22 (d, J = 7 Hz, 6H, iPr-CH_3), 1.28 (t, J = 7 Hz, 3H, Et-CH_3), 2.02 (s, 3H, Ac-CH_3), 2.87 (m, 1H, iPr-CH), 3.03 (d, J = 6 Hz, 2H, β-CH_2), 4.20 (q, J = 7 Hz, 2H, Et-CH_2), 4.82 (m, 1H, α-CH), 6.38 (d, J = 8 Hz, 1H, N-H), 6.5-7.4 (comp m, 4H, Ar-2',4',5',6' H), 7.63 (s, 2H, Ar-2, 6 H). Analysis: $C_{22}H_{25}I_2N_1O_4 \cdot 1/2EtOH$: Calculated C, 42.87; H, 4.38; I, 39.40; N, 2.17; Found C, 43.26; H, 4.17; I, 39.44; N, 2.34.

N-Acetyl-3,5-dicyano-4-(3'-n-propyl-4'-methoxyphenoxy)-L-phenylalanine Ethyl Ester (2-30). The general procedure of Barnes, et al.,[70] as used by Jorgensen, et al.,[9] was utilized. A solution of N-acetyl-3, 5-diiodo-4-(3'-n-propyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-10) (3.938 g, 6.05 mmoles) in dry pyridine containing cuprous cyanide (2.65 g, 29.6 mmoles) was heated under reflux for 6 hours. After cooling to room temperature, the reaction mixture was poured into 120 ml of ice water. After stirring for 10 minutes, the resulting yellow-green solid was collected by filtration, washed with cold H_2O (400 ml), and then stirred for 30 minutes in a mixture of 125 ml of 2N NH_4OH and 90 ml of $CHCl_3$. After filtration through filter aid, the $CHCl_3$ layer was separated,

gently (to avoid emulsions) washed successively with 2N NH_4OH (50 ml), H_2O (50 ml), 2N HCl (50 ml), and H_2O (50 ml), and finally dried over Na_2SO_4. After filtration the $CHCl_3$ was removed under reduced pressure to give the crude product (2.63 g, 97%). Two recrystallizations from hot anhydrous EtOH with decolorizing carbon yielded (after drying in vacuo) a total of 2.14 g (79%) of the purified, fluffy white crystalline product. MP 154-155°. $[\alpha]_D^{29}$ = +54.6° (c, 2.0, $CHCl_3$). TLC (UV) R_f (B: $CHCl_3$) 0.09, R_f (A: $CHCl_3$) 0.36. PMR δ 0.92 (t, J = 7.5 Hz, 3H, nPr-CH_3), 1.27 (t, J = 7 Hz, 3H, Et-CH_3),1.60 (m, 2H, nPr-CH_2-C-Ar), 1.99 (s, 3H, Ac-CH_3), 2.58 (t, J = 8 Hz, 2H, nPr-CH_2-Ar), 3.13 (d, J = 7 Hz, 2H, β-CH_2), 3.78 (s, 3H, O-CH_3), 4.21 (q, J = 7 Hz, 2H, Et-CH_2), 4.78 (m, 1H, α-CH), 6.38 (d, J = 7 Hz, 1H, N-H), 6.75 (s, 3H, Ar-2',5',6' H), 7.63 (s, 2H, Ar-2,6 H). Analysis: $C_{25}H_{27}N_3O_5$: Calculated C, 66.80; H, 6.06; N, 9.35; Found C, 66.99; H, 6.07; N, 9.42.

N-Acetyl-3,5-dimethyl-4-(3'-n-propyl-4'-methoxyphenoxy)-L-phenylalanine Ethyl Ester (2-31). The hydrogenation was carried out under the conditions described by Block and Coy[62] and used by Jorgensen, et al.[9] p-Cymene (99+%, Aldrich Chemical Company, Inc.) was purified exactly as described by Block and Coy[62] and was stored under N_2. A three-necked flask (100 ml) was fitted with a thermometer, gas dispersion tube, and reflux condenser, the tip of which led to a second dispersion tube, dipping beneath the surface of H_2O (100 ml) containing Methyl-Red indicator. N-Acetyl-3,5-dicyano-4-(3'-n-propyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-30) (1.297 g, 2.89 mmoles) was dissolved in freshly purified p-cymene (50 ml) containing 10% Pd/C (0.70 g), the reaction system having been flushed with N_2. Hydrogen was bubbled through while the temperature was maintained at 168-171°. The ammonia evolved was

absorbed and titrated against 0.1N HCl. The reaction became slower near its end and virtually ceased when 98% of the theoretical amount of acid had been neutralized (1.75 hours). After allowing to cool to 60°, the catalyst was removed by filtration while hot through filter aid, washing copiously with acetone. The solvents were removed *in vacuo* (rotary evaporation: H_2O aspirator at 50° and then to 95°) to give a yellow oil. 20 ml heptane was added and the mixture was refrigerated. The oil completely crystallized over a period of 3 days. Pulverization and then filtration, washing with ice cold heptane, gave the white crystalline product, which was dried *in vacuo* (1.102 g, 89%). MP 86-88°. $[\alpha]_D^{30}$ = + 23.3° (*c*, 1.0, EtOH). TLC (UV) R_f (A: $CHCl_3$) 0.46. PMR ($CDCl_3$) 0.91 (t, J = 7 Hz, nPr-CH_3), 1.23 (t, J = 7 Hz, 3H, Et-CH_3), 1.58 (m, 2H, nPr-CH_2-C-Ar), 1.98 (s, 6H, Ac-CH_3), 2.08 (s, 3H, Ar-CH_3), 2.47 (t, J = 7.5 Hz, 2H, nPr-CH_2-Ar), 3.07 (d, J = 6 Hz, 2H, β-CH_2), 3.73 (s, 3H, O-CH_3), 4.13 (q, J = 7 Hz, 2H, Et-CH_2), 4.80 (m, 1H, α-CH), 6.30 (d, J = 7 Hz, 1H, N-H), 6.4-7.0 (s, 5H, Ar-2, 6, 2', 5',6' H). Analysis $C_{25}H_{33}N_1O_5$: Calculated C, 70.23; H, 7.78; N, 3.28; Found C, 70.18; H, 7.68; N, 3.37.

3,5-Diiodo-3'-isopropyl-L-thyronine (2-2). Utilizing the procedure of Blank, *et al.*,[53] N-acetyl-3,5-diiodo-4-(3'-isopropyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-28) (10.00 g, 15.35 mmoles) was hydrolyzed in a solution made of hydriodic and glacial acetic acids to yield the desired product (8.37 g, 96%; lit.[53] 84%). MP 223-225° (decomp.) (lit.[53] 225-226°). $[\alpha]_D^{31}$ = + 32.0° (*c*, 0.8, EtOH-1N HCl/3:1 by volume) (lit.[53] $[\alpha]_D^{25}$ = + 23.2° (*c*, 0.8, EtOH-1N HCl by volume)). TLC (UV, ninhydrin) R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/20:10:1) 0.32, R_f (B: iPrOH-conc. NH_4OH/4:1) 0.44, R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/10:5:1) 0.58, R_f (B: $CHCl_3$-MEOH-conc. NH_4OH/10:20:1) 0.78.

3,5-Diiodo-3',5'-diisopropyl-L-thyronine (2-6). A mixture of 1.01 g (1.46 mmoles) of N-acetyl-3,5-diiodo-4-(3',5'-diisopropyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-8) in 20 ml of a solution made of equal volumes of constant boiling hydrobromic (or hydriodic) and glacial acetic acids was heated under reflux for 5 hours, cooled, and poured into 85 ml of ice water. After adjustment of the pH to 5.1 with concentrated NH_4OH, the precipitate was collected by filtration, washed with H_2O, and dried in vacuo to give a very light brown solid (779 mg, 88% crude yield). MP 221-223° (decomp.). Repeated recrystallizations from acidic aqueous EtOH by adjustment of the pH to 5.2 with hot H_2O and hot 2N sodium acetate gave material that gave erratically erroneous elemental analyses. Final purification was accomplished with preparative TLC: 108 mg/plate; developed 15.7 cm; R_f 0.26 to 0.52 removed from plate. This yielded 40.1 mg of a pure white solid. MP 232-234° (decomp.) (lit.[53] 235-236° for the DL-analog). $[\alpha]_D^{30}$ = + 22.6° (c, 0.5, EtOH-1N HCl/3:1 by volume). TLC (UV, ninhydrin) R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/20:10:1) 0.37 (separated from 2-7, R_f 0.41). PMR (CF_3COOH) δ 1.28 (d, J = 7 Hz, 12H, iPr-CH_3), 3.47 (m, 2H, iPr-H), 3.67 (d, J = 7.5 Hz, 2H, β-CH_2), 5.00 (m, 1H, α-CH), 6.99 (s, 2H, Ar-2',6' H), 7.8-8.1 (broad peak, 1.8H by integration, NH_3^+), 8.35 (s, 2H, Ar-2,6 H); in particular, spectrum completely lacking any O-CH_3 peak. Analysis: $C_{21}H_{25}I_2N_1O_4 \cdot 1/4C_2H_5OH$ (the amino acid associates very strongly with EtOH): Calculated C, 41.60; H, 4.30; I, 40.89; N, 2.26; Found C, 42.02; H, 4.31; I, 40.82; N, 2.30.

3,5-Diiodo-3'-n-propyl-L-thyronine (2-10). A mixture of 910.4 mg (1.398 mmoles) of N-acetyl-3,5-diiodo-4-(3'-n-propyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-29) in 18.2 ml of a solution made of equal

volumes of glacial acetic and hydriodic (47-51%) acids was heated under reflux for 4 hours, cooled, and poured into 73 ml of ice water. After adjustment of the pH to 5.0 with concentrated NH_4OH, the mixture was cooled and filtered, washing with H_2O. The precipitate was dried *in vacuo* (726.1 mg, 91.6%). An analytical sample was recrystallized from hot aqueous EtOH containing several drops of concentrated HCl by the addition of hot H_2O and hot 2N sodium acetate to pH 5.1. The resulting precipitate was collected by centrifugation, washing with H_2O, and was dried *in vacuo*. MP 209-212° (decomp.). $[\alpha]_D^{31}$ = + 25.3° (*c*, 1.0, EtOH- 1N HCl /3:1 by volume). TLC (UV, ninhydrin) R_f (A: $CHCl_3$-MeOH-conc. NH_4OH/20:10:1) 0.31. Analysis: $C_{18}H_{19}I_2N_1O_4 \cdot 3/4H_2O$: Calculated C, 37.23; H, 3.56; I, 43.71; Found C, 37.53; H, 3.61; I, 43.60.

3,5-Diiodo-3'-chloro-5'-isopropyl-L-thyronine (2-3). Sulfuryl chloride (0.25 ml, 420 mg, 3.11 mmoles) was added in one portion to a suspension of 3,5-diiodo-3'-isopropyl-L-thyronine (2-2) (1.001 g, 1.77 mmoles) in 60 ml glacial acetic acid at room temperature. The mixture was stirred at room temperature for 1 hour, the solid dissolved, and the reaction solution turned clear light yellow. The reaction mixture was then poured into 150 ml 2N HCl. After adjustment of the pH to 5.2 with concentrated NH_4OH, the mixture was filtered and the solid was washed with H_2O to give the clean white product, which was dried *in vacuo* (971 mg, 91%). MP 228-229° (decomp.). $[\alpha]_D^{31}$ = + 26.8° (*c*, 1.0, EtOH-1N HCl/3:1 by volume). TLC (UV, ninhydrin) R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/20:10:1) 0.37 (separated from 2-2, R_f 0.32), R_f (B: iPrOH-conc. NH_4OH/4:1) 0.44 (not separated from 2-2, R_f 0.44), R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/10:5:1) 0.61 (barely separated from 2-2, R_f 0.58), R_f (B: $CHCl_3$-

MeOH-conc. NH_4OH/10:20:1) 0.78 (not separated from 2-2, R_f 0.78). Analysis: $C_{18}H_{18}Cl_1I_2N_1O_4$: Calculated C, 35.93; H, 3.02; Cl, 5.89; I, 42.19; N, 2.33; Found, C, 35.66; H, 3.02; Cl, 6.10; I, 41.94; N, 2.34.

3,5-Diiodo-3'-bromo-5'-isopropyl-L-thyronine (2-4). To 3,5-diiodo-3'-isopropyl-L-thyronine (2-2) (1.002 g, 1.77 mmoles) dissolved in glacial acetic acid (53 ml) and concentrated hydrochloric acid (12 drops) was added dropwise (by injection through a rubber septum into the reaction vessel), at 50-60°, a glacial acetic acid solution (87 ml) of bromine (327 mg, 2.05 mmoles). After stirring an additional 20 minutes at 55-60°, the solution was allowed to cool to 40°, decolorized with sodium metabisulfite, diluted with 150 ml H_2O, and adjusted to pH 5.0 with concentrated NH_4OH and 2N sodium acetate. Filtration, washing with 200 ml H_2O, yielded a yellow filtrate and a clean white solid, which was recrystallized from hot aqueous EtOH containing a few drops concentrated HCl by adjustment of the pH to 5.0 with hot H_2O and hot 2N sodium acetate. After allowing to cool to room temperature, the solution was cooled on an ice bath. The precipitate was collected by filtration and washed with H_2O to give the product, which was dried in vacuo (919 mg, 81%). MP 225-227° (decomp.). (After refrigeration for 2 weeks, the yellow filtrate yielded an additional 118 mg (10%) of product upon filtration.) $[\alpha]_D^{32}$ = + 27.4° (c, 1.0, EtOH-1N HCl/3:1 by volume). TLC (UV, ninhydrin) R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/20:10:1) 0.36 (separated from 2-2, R_f 0.32), R_f (B: iPrOH-CONC. NH_4OH/4:1) 0.44 (not separated from 2-2, R_f 0.44). Analysis: $C_{18}H_{18}Br_1I_2N_1O_4$: Calculated C, 33.46; H, 2.81; Br, 12.37; I, 39.29; N, 2.17; Found C, 33.46; H, 2.88; Br, 12.21; I, 39.54; N, 2.23.

3,5,3'-Triiodo-5'-isopropyl-L-thyronine (2-5). To a stirred, ice bath cooled solution of 3,5-diiodo-3'-isopropyl-L-thyronine (2-2) (1.01 g, 1.78 mmoles) in 70% aqueous ethylamine (33 ml) a solution of I_2 (568.2 mg, 2.24 mmoles) in 40 ml of 1M aqueous KI was added dropwise in 5 minutes. The mixture was stirred for 15 minutes with and then 30 minutes without ice bath cooling. The excess iodine was reduced with an excess (10 ml of a 1M solution) of aqueous $NaHSO_3$. After stirring an additional 5 minutes at room temperature, the pH of the clear yellow-orange solution was adjusted to 5.0 with glacial acetic acid, cooling with an ice bath. After dilution with 200 ml H_2O, the precipitate was collected by filtration, washing copiously with H_2O, to give a clean white solid, which was dried in vacuo (1.02 g, 83%). MP 206-207° (decomp.). $[\alpha]_D^{32}$ = + 26.4° (c, 1.0, EtOH-1N HCl/3:1 by volume). TLC (UV, ninhydrin) R_f (B: $CHCl_3$-MeOH-conc.NH_4OH/20:10:1) 0.36 (separated from 2-2, R_f 0.32). Analysis: $C_{18}H_{18}I_3N_1O_4$: Calculated C, 31.19; H, 2.62; I, 54.94; N, 2.02; Found, C, 31.46; H, 2.73; I, 54.62; N, 2.15.

3,5-Dimethyl-3'-n-propyl-L-thyronine (2-11). To 504 mg (1.18 mmoles) of N-acetyl-3,5-dimethyl-4-(3'-n-propyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-31) dissolved in 10.0 ml glacial acetic acid was added 10.0 ml constant boiling hydrobromic acid. The mixture was refluxed for 5 hours under a positive N_2 atmosphere, cooled to room temperature, and poured into 80 ml ice water. With ice bath cooling, the pH was adjusted to 5.2 with concentrated NH_4OH. The solution was allowed to rise to room temperature while stirring vigorously. The precipitate was collected by filtration, washing with H_2O, and was dried in vacuo to give a light brown solid (340 mg, 84% crude yield).

Final purification was accomplished by preparative TLC: 109.4 mg on two plates; developed 16.0 cm; R_f 0.15 to 0.45 removed. This yielded a clean white solid (23.7 mg). (After removal of the EtOH from the preparative TLC filtrate, a second crop (16.4 mg) was obtained.) MP 207-209° (decomp.). $[\alpha]_D^{30}$ = + 21.4° (c, 0.4, EtOH-1N HCl/9:1 by volume). TLC (UV, ninhydrin) R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/20:10:1) 0.31. Analysis: $C_{20}H_{25}N_1O_4 \cdot 2/3\ H_2O$: Calculated C, 67.58; H, 7.47; N, 3.92; Found C, 67.50; H, 7.16; N, 3.81.

3,5-Diiodo-4-(3'-isopropylphenoxy)-L-phenylalanine (2-1). The general reaction conditions of Blank, et al.,[53] were used. A mixture of N-acetyl-3,5-diiodo-4-(3'-isopropylphenoxy)-L-phenylalanine ethyl ester (2-37) (1.00 g, 1.61 mmoles) in 20 ml of a solution made of equal volumes of concentrated hydrochloric and glacial acetic acids was heated under reflux for 4 hours (an additional 10 ml of concentrated hydrochloric acid being added after 2 hours), cooled, and poured into 120 ml ice water. Concentrated NH_4OH was added, with ice bath cooling, to pH 5.2. After further cooling, the precipitated solid was filtered, washed, and dried in vacuo to give a light brown solid (772 mg, 87% crude yield). This was recrystallized from hot aqueous EtOH containing several drops concentrated hydrochloric acid by addition of hot H_2O and hot 2N sodium acetate to pH 5.2. The solution was allowed to cool to room temperature and was filtered. The precipitate was washed with H_2O and dried in vacuo to give a light brown solid (686 mg, 77%). MP 216-218° (decomp.). Preparative TLC (see below) later showed this sample to contain minor impurities. Final purification was accomplished by preparative TLC: 113.2 mg on two plates; developed 14.8 cm; removed R_f 0.37 to 0.59. This yielded a clean buff-colored solid (84.8 mg). MP 220.5-222°

(decomp.). TLC (UV, ninhydrin) R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/20:10:1) 0.41. $[\alpha]_D^{30}$ = + 20.5° (c, 0.4, EtOH-1N HCl/9:1 by volume). Analysis: $C_{18}H_{19}I_2N_1O_3$: Calculated C, 39.22; H, 3.47; I, 45.82; N, 2.54; Found C, 39.76; H, 3.72; I, 45.82; N, 2.45.

3,5-Diiodo-4-(3'-isopropyl-4'-methoxyphenoxy)-L-phenylalanine (2-9). A mixture of 451.4 mg (0.693 mmole) of N-acetyl-3,5-diiodo-4-(3'-isopropyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-2) in 13.5 ml of a solution made of equal volumes of concentrated hydrochloric and glacial acetic acids was heated under reflux for 4 hours, cooled to room temperature, and poured into 55 ml of ice water. The pH of the solution was adjusted to 5.0 with concentrated and 2N NH_4OH. The white precipitate was collected by filtration, washed with H_2O, and dried in vacuo to give the clean white product (379 mg, 99%). An analytical sample was recrystallized from acidified aqueous EtOH by adjustment of the pH to 5.2 with 2N NaOH. The precipitate was collected by centrifugation, washing with H_2O, and was dried in vacuo. MP 216-217° (decomp.). $[\alpha]_D^{30}$ = +24.2° (c, 1.0, EtOH-1N HCl/3:1 by volume). TLC (UV, ninhydrin) R_f (A: iPrOH-conc. NH_4OH/4:1) 0.52. Analysis: $C_{22}H_{27}I_2N_1O_4$: Calculated C, 42.39; H, 4.37; I, 40.72; Found C, 42.49; H, 4.38; I, 40.60.

3,5-Diiodo-4-(3',5'-diisopropyl-4'-methoxyphenoxy)-L-phenylalanine (2-7). A mixture of 1.01 g (1.45 mmoles) of N-acetyl-3,5-diiodo-4-(3',5'-diisopropyl-4'-methoxyphenoxy)-L-phenylalanine ethyl ester (2-8) in 20.0 ml of a solution made of equal volumes of concentrated hydrochloric and glacial acetic acdis was heated under reflux for 4 hours, cooled, and poured into 80 ml of ice water. With ice bath cooling, the pH was adjusted to 5.2 with concentrated NH_4OH. The precipitate was collected by filtration, washed with H_2O, and dried in vacuo to give a tan solid.

This was dissolved in 190 ml EtOH plus several drops 2N NaOH, filtered to remove a small amount of undissolved material, and then diluted, first with 190 ml 2N NaOH, and then with 500 ml H_2O. With ice bath cooling, the pH was adjusted to 5.0 with hydrochloric acid. The solution was filtered, washing with 500 ml H_2O, to yield a light tan solid that was dried in vacuo (311 mg, 68%). MP 216-218° (decomp.). $[\alpha]_D^{30}$ = + 25.4° (c, 1.0, EtOH-1N HCl/3:1 by volume). TLC (UV, ninhydrin) R_f (A: iPrOH-conc. NH_4OH/4:1) 0.51, R_f (B: $CHCl_3$-MeOH-conc. NH_4OH/20:10:1) 0.41. Analysis: $C_{22}H_{27}I_2N_1O_4$: Calculated C, 42.39; H, 4.37; I, 40.72; Found C, 42.49; H, 4.38; I, 40.60.

4-Iodo-2-isopropylanisole (2-25). Complete removal of solvent (H_2O aspirator/rotary evaporation/50°) from the hexane filtrate obtained with synthesis of 2-28 yielded the crude product as a clear orange-red viscous liquid (23.52 g, 81.2% based on 41% yield for the condensation reaction). PMR ($CDCl_3$) δ 1.16 (d, J = 7 Hz, 6H, iPr-CH_3), 3.23 (m, J = 7 Hz, 1H iPr-CH), 3.74 (s, 3H, O-CH_3), 6.53 (q, J = 9 Hz, J = 2.5 Hz, 1H, Ar-6 H), 7.25-7.5 (comp m, 2H, Ar-3,5 H). An analytical sample was distilled in vacuo. BP 124.5-125.5°/1.0 mm Hg. Analysis: $C_{10}H_{13}I_1O_1$: Calculated C, 43.50; H, 4.74; I, 45.96; Found C, 43.57; H, 4.77; I, 45.88.

4-Iodo-2,6-diisopropylanisole (2-26). On the basis of TLC inspection, 100 ml elutent fractions #3-7 ($CHCl_3$ solvent range approximately) were combined from column chromatographic purification involved with synthesis of 2-8. Complete removal of solvent (H_2O aspirator/rotary evaporation/34°) yielded the crude product as a clear light yellow very viscous liquid (18.02 g, 99.4% based on 31% yield for the condensation reaction). TLC (UV) R_f (A: $CHCl_3$) 0.54. PMR ($CDCl_3$) δ 1.18 (d, J = 7 Hz, 12H, iPr-CH_3), 3.27 (m, J = 7 Hz, 2H, iPr-CH), 3.69 (s, 3H, O-CH_3), 7.40

(s, 2H, Ar-3,5 H). An analytical sample was distilled in vacuo. BP 103-104.5°/0.20 mm Hg. Analysis: $C_{13}H_{19}I_1O_1$: Calculated C, 49.07; H, 6.02; I, 39.88; Found C, 49.00; H, 5.94; I, 39.90.

4-Iodo-2-n-propylanisole (2-27). Removal of the solvents from the trituration mother liquor from synthesis of 2-29 (H_2O aspirator/ rotary evaporator/35°) yielded the crude product as a viscous red oil (7.10 g, 88.8% based on 57% yield for the condensation reaction). PMR ($CDCl_3$) δ 0.92 (t, J = 6.5 Hz, 3H, nPr-CH_3), 1.58 (m, 2H, nPr-CH_2-C-Ar), 2.53 (t, J = 7 Hz, 2H, nPr-CH_2-Ar), 3.73 (s, 3H, O-CH_3), 6.51 (d, J = 8.5 Hz, 1H, Ar-6 H), 7.25-7.55 (comp m, 2H, Ar-3,5 H). An analytical sample was distilled in vacuo. BP 92-93°/0.075 mmHg. Analysis: $C_{10}H_{13}I_1O_1$: Calculated C, 43.50; H, 4.74; I, 45.96; Found, C, 43.69; H, 4.90; I, 45.86.

CHAPTER THREE: THYROMIMETIC ACTIVITIES OF SOME THYROID HORMONE ANALOGS

In vivo biological test systems have been developed over the years for a large number of the various biological responses that the thyroid hormones and analogs elicit. These assays include: (1) induction of amphibian metamorphosis;[71-76] (2) elevation of various basal metabolic rates;[77,78] (3) goiter prevention;[9,79-82] (4) serum cholesterol lowering;[83,84] and (5) reversal of fetal hypothyroidism.[85] More recently, several *in vitro* assay procedures have been developed for measuring relative binding affinities of the thyroid hormones and analogs to: (a) various purified plasma proteins;[29,30,45-48] (b) intact cell nuclei;[24,25] and (c) solubilized nuclear proteins.[26-28] The most reliable and largest set of *in vivo* activities has been obtained utilizing the rat antigoiter bioassay.' The various *in vitro* binding affinities have shown good reproducibility, and the number of analogs for which various binding affinities have been measured is even now increasing rapidly.

The rat antigoiter activities and/or binding affinities to solubilized rat hepatic nuclear protein were determined for eight of the analogs newly synthesized for this study. This investigator participated in the planning and conducting of these antigoiter assays, while the binding assays were conducted *in toto* by another member of this research group.[86] Hence, a detailed description of the antigoiter assay procedure and only a brief description of the binding assay procedure are included in this chapter.

RAT ANTIGOITER BIOASSAYS

Five of the thyroid hormone analogs newly synthesized for this study and a number of other analogs were tested for their thyromimetic antigoitrogenic activities in three rat antigoiter bioassays. The results of these *in vivo* studies, as well as a detailed assay description, are presented below. A discussion of the recalculation and standardization of analog activities in the rat antigoiter bioassay is also presented in this section.

Assay Description.

The rat antigoiter bioassay is based on the "short loop" feedback control of pituitary thyrotropin (TSH; thyroid stimulating hormone) secretion by circulating thyroid hormones and analogs. An "antithyroid" drug such as thiouracil or propylthiouracil in the diet blocks the biosynthesis and release of T_3 and T_4 from the thyroid gland. The lowered levels of circulating thyroid hormones stimulate an increased thyrotropin secretion from the pituitary, leading to increased thyroid tissue development and circulation and eventually to an enlarged thyroid gland called a goiter. Such goiter formation is well achieved within

the ten day length of the bioassay. Graded doses of either the reference compound or the analog to be tested are administered daily by subcutaneous injection. The relative activity of an analog is estimated (based on a standard log dose vs. response curve) by the molar dose of the analog, relative to that of the standard (T_3 or T_4), required to cause 50% reversal of the drug-induced goiter.

The detailed bioassay procedure is as follows. Solutions of the analogs were made up a day or two before the injections were to begin and were prepared so that they could be compared to the reference compound on a molar basis. An arbitrary molar ratio value of 1.00 was assigned to the dose of L-T_4 containing 1 μg of L-T_4 per 100 g of rat body weight or to the dose of L-T_3 containing 0.25 μg of L-T_3 per 100 g of rat body weight, depending on which was used as the reference compound. Solutions were prepared so that the dose administered to a 100 g rat was contained in 0.125 ml of solution (the calibrated volume of a tuberculin syringe). Samples were weighed on a Cahn Electrobalance. Stock solutions were prepared by dissolving in normal saline (0.9% aqueous NaCl) 0.01 N in NaOH to a total volume of 10.0 ml or 25.0 ml (depending on compound solubility). Appropriate aliquots were diluted to 25.0 ml with normal saline 0.01 N in NaOH. Compounds which were found to have limited solubility in this saline solution or whose stability in basic solution was questionable (based on previous experience[9] or on color development upon solution) were dissolved in absolute EtOH to a total volume of 10.0 ml. Appropriate aliquots were diluted to 25.0 ml with normal saline. These solutions were decanted into 50 ml multiple dose vials, which were fitted with a septum, capped, and stored in the refrigerator when not in use.

Male Long-Evans (assays #1 and #2) or Sprague-Dawley (assay #3) rats (Simonsen Laboratories, Gilroy, Calif.), weighing between 70 and 100 g when obtained, were housed three to a cage. All were fed a normal diet of powedered Simonsen Rat Maintenance Diet for two days prior to the start of the assay. (This was done in order to allow them to "settle in", establish a regular feeding regimen, and hence insure a fairly constant intake of thiouracil in their feed.)[87] Groups of six rats were used for each reference compound or analog dose. The thiouracil and normal control groups contained six to twelve rats. The normal control received a normal diet and all other animals received 0.3% thiouracil in their feed (thiouracil obtained from the Nutritional Biochemicals Corporation and incorporated into the normal feed by the Pharmaceutical Technology Laboratory, School of Pharmacy, University of California, San Francisco). Diets were begun one day before the injections were started. The normal and thiouracil control groups were injected with normal saline 0.01 N in NaOH. The other groups were injected with the reference compound or analog at the predetermined dose levels. The volume of solution injected was determined on the basis of daily weighings as follows:

Rat Weight (g)	Volume of Solution (ml)
60 - 79	0.075
80 - 99	0.100
100 - 119	0.125
120 - 139	0.150
140 - 159	0.175
160 - 179	0.200
180 - 199	0.225
200 - 219	0.250

After 10 days of injections, the animals were sacrificed by ether-chloroform inhalation on the eleventh day. After determining the body weight, the thyroid glands were excised, kept moist with normal saline on filter paper, cleaned of extraneous tissue under a dissecting microscope, blotted on a filter paper, and immediately weighed to the nearest 0.1 mg. (In any one assay, the same individual performed all of the final cleanings of the thyroid glands in order to insure, as best possible, a consistent degree of extraneous tissue removal from the glands.)

Thyroid weights were converted to mg/100 g body weight and mean values were calculated for the control groups and for each dose level group of the reference compound or of the analog. The statistical analysis of the data was performed as follows:[88,89]

$\overline{X}_{Tu}$ = average thiouracil control thyroid weight/100 g body weight

$\overline{X}_{compd}$ = average compound thyroid weight/100 g body weight

S_{Tu} = standard deviation for $\overline{X}_{Tu}$

S_{compd} = standard deviation for $\overline{X}_{compd}$

N_{Tu} = number of rats used to calculate $\overline{X}_{Tu}$

N_{compd} = number of rats used to calculate $\overline{X}_{compd}$

$$S_p^2 = \frac{(N_{Tu} - 1)\cdot S_{Tu}^2 + (N_{compd} - 1)\cdot S_{compd}^2}{N_{Tu} + N_{compd} - 2}$$

$$S_{diff} = \sqrt{(S_p^2/N_{Tu}) + (S_p^2/N_{compd})}$$

$$= S_p \sqrt{(N_{Tu} + N_{compd})/(N_{Tu}\cdot N_{compd})}$$

$$t_{calcd} = (\overline{X}_{Tu} - \overline{X}_{compd})/S_{diff} \qquad \text{(Eqn. 3-1)}$$

t_{calcd} is compared with tabulated Student's t critical point values to test whether $\overline{X}_{compd} < \overline{X}_{Tu}$ (a "one-tailed" test). (Of course, if $\overline{X}_{compd} \geq \overline{X}_{Tu}$, then there is no need to perform the test.) t_{calcd}, for a particular compound at a particular dose, is compared with t_P^{DF}, the Student's t critical point value with DF = $(N_{Tu} + N_{compd} - 2)$ degrees of freedom at a probability of P. If

$$t_{P1}^{DF} \leq t_{calcd} \leq t_{P2}^{DF}$$

then the level of significane, P_{calcd}, at which it can be stated that $\overline{X}_{compd} < \overline{X}_{Tu}$, is obtained by interpolation linearly with log P:[90]

$$\log P_{calcd} = \log P_1 + (\log P_2 - \log P_1)\left[\frac{t_{calcd} - t_{P1}^{DF}}{t_{P2}^{DF} - t_{P1}^{DF}}\right]$$

(Eqn. 3-2)

If $t_{75\%}^{DF} > t_{calcd}$ or $t_{calcd} > t_{99.9\%}^{DF}$, then the level of significance at which it can be stated that $\overline{X}_{compd} < \overline{X}_{Tu}$ is merely given as < 75% or > 99.9%, respectively. Similarly, statistical analysis can be performed to determine whether $\overline{X}_{compd} > \overline{X}_{NC}$, where:

$\overline{X}_{NC}$ = average normal control thyroid weight/100 g body weight

Recalculation and Standardization of Analog Activities.

The steadily increasing interest in quantitative structure-activity correlation studies of thyroid hormone analogs[3,25,39,40,56,91-95] and the inherent need in such studies of the most accurate and complete

listing of analog activities have prompted us to recompile and re-evaluate (as completely as reasonably possible) the activities of these analogs in the rat antigoiter biological assay. Development of the bioassay procedure can be roughly traced as follows:

1943: The maintenance or restoration of normal thyroid weight by administration of thyroxine to rats simultaneously treated with thiouracil first used as the basis of an assay procedure for T_4.[79]

1949: Estimation of relative activities of analogs as compared T_4.[80]

1957: Use of log dose vs. response curves to estimate activities of analogs.[81]

1962: Standard use of[34]

:thiouracil inclusion in solid food in preference to inclusion in the drinking water, where the bitter taste discourages the rats from drinking and results in uneven thiouracil injection.

: doses of analogs in units of μg (or μmoles)/100 g body weight/day in preference to μg (or μmoles)/ rat/day.

: analog activity estimates based on molar rather than weight ratios.

1974: Use of analog stock solutions containing no base in order to avoid decomposition of particularly labile analogs.[9]

1976: Conversion from L-T_4 to L-T_3 as standard reference compound.[95]

A careful review of the literature revealed that many investigators had based estimates of analog activities on weight rather than molar ratios as compared to T_3 or T_4. In view of the necessity of as accurate as possible molar activities for meaningful quantitative structure-activity correlation studies, we decided to recalculate the rat antigoiter bioassay activities of all known analogs on a molar basis.

Both T_3 and T_4 (D, L, and DL stereoisomers) have been used as standard reference compounds in rat antigoiter bioassays. In addition, literature estimates (some based on molar ratios; others based on weight ratios) of the activity of T_3 have ranged from 300% to 1000% that of T_4.[83,96-110] Conversion of analog activities from the T_3 to the T_4 reference-compound scale and vice versa have characteristically involved a scaling factor of from 5 to 8. Peripheral deiodination from both aromatic rings of iodinated thyroid hormones and analogs has been shown to occur in vivo.[50,111-122] In particular, studies utilizing ^{14}C- or ^{125}I-labeled T_4 have given estimates of about 17% conversion of secreted T_4 to T_3 in the rat[116] and of about 33% conversion of daily T_4 production to T_3 in man.[112]

Metabolic deiodination of T_4 can give rise to formation of the much more active analog T_3, while metabolic deiodination of T_3 leads only to analogs with very low activities. Variations in peripheral deiodination rates in vivo (between rats in the same and in different assays) could hence cause larger variations in the apparent T_4 activity than in the apparent T_3 activity. In a similar comparison, contamination of a reference T_4 sample with small amounts of the much more active T_3 is likely to lead to larger variations in activity than contamination of a reference T_3 sample with small amounts of the less active T_4 and other halogenated thyronines. As a result, we concluded that L-T_3

is a better choice than L-T_4 or DL-T_4 as a standard reference compound and that all of our molar recalculations of rat antigoiter bioassay activities would be with L-T_3 (DL-T_3, where comparison was originally to DL-T_3 or DL-T_4) as the reference compound. After careful examination of the literature, a large number of comparisons of T_3 and T_4 activities were omitted from our recalculation of the relative T_3 and T_4 activities for one or more of the following reasons:

1. Form of T_4 (Na salt pentahydrate; free amino acid; stereochemistry) not specified.
2. Compounds administered orally.
3. Too few dose levels.
4. Purity of compounds possibly questionable (especially possible T_3 contamination of T_4).
5. Method of dosaging unclear or seemingly arbitrary.

On the basis of the remaining T_3 vs. T_4 activity studies[97-99,101,105,108] the corrected molar activity of L-T_3 was calculated to be approximately 553% that of L-T_4. If the activity of L-T_3 is arbitrarily set equal to 100%, then L-T_4 has a relative molar activity of (1/5.53) x 100% = 18.1%. That the apparent T_4 activity in vivo may be increased by peripheral $T_4 \rightarrow T_3$ deiodination is supprted by the fact that the in vitro relative binding affinities of L-T_4 to intact rat hepatic nuclei and to solubilized high affinity rat hepatic nuclear protein "receptors" are 12.5%[24,25] and 13.85%,[43] respectively, of that of L-T_3.

In Tables I-1 through I-11 of Appendix I are presented our recalculated rat antigoiter bioassay molar activities of thyroid hormone analogs. The values were calculated as follows:

1. Activities initially reported on a weight basis were converted to a molar basis.

2. For any assay containing L-T_3 or DL-T_3 as the reference compound or as an analog, the activities of the analogs were calculated as relative to the L-T_3 or DL-T_3 activity.
3. For assays containing L-T_4 or DL-T_4 (and neither L-T_3 or DL-T_3) as the reference compound or as an analog, the activities of the analogs relative to L-T_4 or DL-T_4 were divided by our calculated scaling factor of 5.53 to give estimates of the analog activities relative to L-T_3 or DL-T_3, respectively.
4. When there is more than one literature report of an analog activity, an average value was taken.
5. Certain reported analog activities were omitted in calculating the average molar activities for the same reasons as with the T_3 vs. T_4 comparison and/or if the activity differed unreasonably from two or more other reported activities.
6. If only one activity has been reported for an analog but the reported activity seems questionable based on qualitative and/or quantitative correlations of structure with activity, the value is listed but the question of its reliability is noted.
7. If the activity is merely reported as zero, it is given as such. If the activity is reported as zero at a certain highest dose level, the activity is given as a < X% value; use of such a maximal activity as the actual analog activity in quantitative structure-activity studies is at best risky since in general the activity of the analog is probably well below this maximal value, but was not determined exactly.

The main fact to be noted from these recalculated molar rat antigoiter bioassay activities is that the qualitative picture of correlation of structure with activity for the thyroid hormone analogs remained unchanged. Only the exact quantitative activity estimates necessary for the most accurate quantitative structure-activity correlation studies have been altered.

Although the details of the rat antigoiter bioassay are available,[9,82] the following guidelines are recommended to insure optimal reproducibility and accuracy:

1. Goitrogen supplied in solid food; not in drinking water.
2. Subcutaneous injections, rather than oral route, for analog and reference compound administration.
3. Doses in μg (or μmoles)/100 g body weight/day; not μg (or μmoles)/rat/day.
4. Activity estimated on a molar basis for 50% inhibition of goitrogen-induced goiter; full log dose vs. response curve highly preferable.
5. Use of L-T_3 as standard reference compound should give more accurate and reproducible results.
6. Thin layer chromatographic (and/or other methods of) inspection of analog and reference compound purities; very important, because the routes of synthesis to and the labilities (especially with time) of thyroid hormone analogs often give rise to significant impurities, both hormonally active and inactive.

One additional scaling factor can be used for analogs with alanine side chains so that variations in activity due to the alanine side chain stereochemistry can be eliminated: i.e., so that for an analog with an alanine side chain, the estimated activity of the L-analog relative to

L-T_3 may be calculated. The activity of an L-analog relative to L-T_3 is of course unchanged. The activity of a DL-analog relative to DL-T_3 is assumed to be equal to what the L-analog activity relative to L-T_3 would be. The activity of a DL-analog relative to L-T_3 is divided by 0.59 (activity of DL-T_3 = 59% that of L-T_3[108]) to give the estimated activity of the L-analog relative to L-T_3. This estimate of DL-analogs having about 59% the activity of their L-analogs is (as with all other estimations concerning scaling antigoiter activities) only a fairly approximate value. It should, however, provide a slightly more accurate scaling factor than merely assuming the activity of an L-analog to be 2x that of the corresponding DL-analog in that it does take into account the low, yet significant, in vivo activity of the D-stereoisomer. Such scaling should obviously only be applied to analogs with an alanine side chain in the 1-position.

Results and Discussion.

The compounds tested for their thyromimetic activities in the three rat antigoiter bioassays are listed in Table 3-1. The detailed results of the three assays are presented in Tables 3-2 through 3-4. The corresponding log dose vs. biological response curves are presented in Figures 3-1 through 3-3. A combined summary of the analog activities determined in these assays is presented in Table 3-5. The reference compound was L-T_4 (3-2) in the first two assays and L-T_3 (3-1) in the third assay. All of the compounds in the first assay were underdosed. The approximately 50% reversal with the highest L-T_4 (3-2) dose, however, permitted evaluation of the activities of at least some of the analogs tested. Dose levels were more accurate for assays #2 and #3, complete

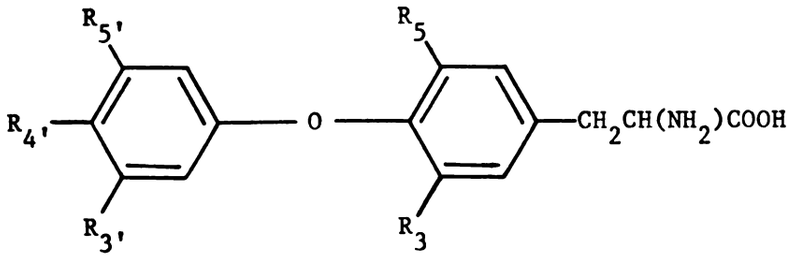
log dose vs. response curves, or at least spanning of 50% goiter reversal, being obtained for most of the analogs. All activities mentioned below in discussion of the assay results are relative to L-T_3 (3-1) = 100% (either by direct comparison in assay #3 or by conversion to the L-T_3 reference compound scale for assays #1 and #2).

The single dose evaluation of the activity of 4'OMe-I_2iPr (3-15) as 19% is consistent with O-demethylation in vivo permitting the analog to demonstrate partial activity as compared to the free 4'-OH analog L-I_2iPr (3-3: 142.1%[123]). The other 4'-OMe analog, L-4'OMe-I_2sBu (3-16), was, like 4'OMe-I_2iPr (3-15), assayed in order to increase the scant number of 4'-OCH_3 analogs for which in vivo activities are available. 3-16 was apparently underdosed in its one testing, and its activity from the assay can only be estimated as < 21%.

L-I_2iPr (3-3) and L-Me_2iPr (3-8) were included in the assays in order to reaffirm their previously determined activities and to provide direct comparison with the corresponding nPr and sBu analogs. The 108% activity determined for L-I_2iPr (3-3) is slightly lower than but consistent with the previous evaluations of the activity of this compound.[123] The values of 4.22% and 3.94% activity determined for L-Me_2iPr (3-8) are consistent with the only other previously determined value of 3.25%[9] and once again reaffirm that halogen is not an essential feature for thyroid hormone activity.

Initial predictions of the activities of L-I_2nPr (3-5) and L-Me_2nPr (3-9) were much too high and only the much higher dose levels of the third assay allowed evalution of the activities of these two compounds as 39.5% and 2.36%, respectively. That the activities of these two compounds are significantly lower than those of their corresponding iPr

Table 3-1. Thyroxine Analogs Tested in the Three Rat Antigoiter Bioassays.



Compound	Abbreviation	$R_3 = R_5$	$R_{3'}$	$R_{5'}$	$R_{4'}$
3-1	L-T_3	I	I	H	OH
3-2	L-T_4	I	I	I	OH
3-3	L-I_2iPr	I	iPr	H	OH
3-4	L-$I_2$$iPr_2$	I	iPr	iPr	OH
3-5	L-I_2nPr	I	nPr	H	OH
3-6	L-I_2sBu	I	±sBu	H	OH
3-7	L-I_2NO_2	I	NO_2	H	OH
3-8	L-Me_2iPr	Me	iPr	H	OH
3-9	L-Me_2nPr	Me	nPr	H	OH
3-10	L-Me_2sBu	Me	±sBu	H	OH
3-11	L-4'H-I_2iPr	I	iPr	H	H
3-12	L-4'H-I_2F	I	F	H	H
3-13	L-4'H-I_2Cl	I	Cl	H	H
3-14	L-4'H-I_2Br	I	Br	H	H
3-15	L-4'OMe-I_2iPr	I	iPr	H	OCH_3
3-16	L-4'OMe-I_2sBu	I	±sBu	H	OCH_3

Table 3-2. Rat Antigoiter Bioassay #1 of Thyroid Hormone Analogs.[a]

Compound Injected	Daily Dose per 100 g (μg)	Molar Ratio	Mean Thyroid Weight per 100 g (mg ± sd)	t_{calcd}[b]	P_{calcd}[c]
Normal Control[d]	---	---	9.48 ± 1.15	17.922	>99.9%
Thiouracil Control	---	---	29.96 ± 2.54	---	---
L-T_4[e]	0.600	0.600	28.57 ± 4.30	0.682	<75%
	1.000	1.000	26.37 ± 5.39	1.476	91.5%
	1.667	1.667	20.08 ± 4.17	4.956	>99.9%
		1.75[f]			
L-I_2iPr	0.064	0.100	29.19 ± 4.40	0.371	<75%
	0.128	0.200	27.88 ± 6.90	0.693	<75%
L-I_2nPr	0.064	0.100	27.24 ± 5.19	1.153	86.5%
	0.128	0.200	25.26 ± 3.13	2.856	99.1%
L-I_2sBu	0.065	0.100	22.52 ± 4.38	3.599	99.8%
	0.131	0.200	32.60 ± 5.55	g	---
L-Me_2iPr	1.545	4.000	28.35 ± 4.45	0.770	77.3%
	3.091	8.000	18.51 ± 3.26	6.786	>99.9%
		7.5[f]			
L-Me_2nPr	1.545	4.000	28.79 ± 3.82	0.625	<75%
	3.091	8.000	31.89 ± 3.72	g	---
L-Me_2sBu	1.608	4.000	29.51 ± 3.37	0.261	<75%
	3.217	8.000	24.04 ± 4.75	2.692	98.9%
		∿11[f]			

Table 3-2. (Continued)

Compound Injected	Daily Dose per 100 g (μg)	Molar Ratio	Mean Thyroid Weight per 100 g (mg ± sd)	t_{calcd}[b]	P_{calcd}[c]
L-4'OMe-I_2iPr	0.981	1.500 1.65[f]	21.11 ± 8.03	2.574	98.6%
L-4'OMe-I_2sBu	1.004	1.500 >1.5[f]	30.13 ± 3.79	g	---
L-I_2NO_2[h]	64.135	100.0 >100[f]	28.31 ± 4.41	0.779	77.5%

[a]Six rats in each control and experimental group.

[b]Calculated using Eqn. 3-1.

[c]Confidence level at which the mean thyroid weight for this dose level or for the untreated control may be considered to be significantly lower than the mean thyroid weight for the thiouracil control. Calculated using Eqn. 3-2.

[d]Untreated control group; all other rats received 0.3% thiouracil in their diets.

[e]Sodium L-thyroxine pentahydrate.

[f]Molar dose ratio required to cause 50% reversal of thiouracil-induced goiter.

[g]Mean thyroid weight at this dose level ≥ mean thyroid weight for the thiouracil control; hence, this may be considered an inactive dose.

[h]Five rats in this group.

Table 3-3. Rat Antigoiter Bioassay #2 of Thyroid Hormone Analogs.[a]

Compound Injected	Daily Dose per 100 g (μg)	Molar Ratio	Mean Thyroid Weight per 100 g (mg ± sd)	t_{calcd}[b]	P_{calcd}[c]
Normal Control[d]	---	---	9.26 ± 2.00	7.093	>99.9%
Thiouracil Control	---	---	29.68 ± 6.78	---	---
L-T_4[e]	1.000	1.000	26.16 ± 5.45	1.060	85.0%
	1.590	1.590	22.09 ± 1.55	2.664	99.0%
	2.520	2.520	12.80 ± 4.11	5.431	>99.9%
	4.000	4.000	8.34 ± 2.01	7.412	>99.9%
		1.60[f]			
L-I_2iPr	0.128	0.200	25.81 ± 3.22	1.292	89.2%
	0.221	0.346	12.78 ± 7.14	4.632	>99.9%
	0.383	0.600	9.17 ± 5.70	6.092	>99.9%
		0.268[f]			
L-I_2nPr	0.128	0.200	29.36 ± 5.54	0.096	<75%
	0.221	0.346	34.07 ± 9.63	g	---
	0.383	0.600	24.03 ± 5.40	1.706	94.4%
L-I_2sBu	0.131	0.200	25.31 ± 2.24	1.508	92.3%
	0.226	0.346	21.80 ± 6.77	2.207	97.7%
	0.392	0.600	16.25 ± 6.09	3.906	>99.9%
		0.362[f]			

Table 3-3. (Continued)

Compound Injected	Daily Dose per 100 g (μg)	Molar Ratio	Mean Thyroid Weight per 100 g (mg ± sd)	t_{calcd} [b]	P_{calcd} [c]
L-Me_2iPr	1.545	4.000	26.05 ± 10.25	0.831	79.4%
	3.090	8.000	18.36 ± 5.31	3.433	99.8%
	3.662	9.480	17.96 ± 11.75	2.464	98.6%
	6.181	16.00	9.37 ± 6.68	5.715	>99.9%
		7.34[f]			
L-Me_2nPr	2.318	6.000	33.66 ± 5.34	g	---
	4.230	10.95	30.06 ± 4.80	g	---
	7.727	20.00	24.65 ± 4.96	1.553	92.8%
L-Me_2sBu	2.413	6.000	25.98 ± 3.44	1.225	88.1%
	4.403	10.95	22.55 ± 6.91	1.980	96.6%
	8.042	20.00	9.35 ± 4.78	6.334	>99.9%
		9.94[f]			

[a]Six rats in normal control and each experimental group.

Nine rats in thiouracil control group.

[b-g]See corresponding footnotes, Table 3-2.

Table 3-4. Rat Antigoiter Bioassay #3 of Thyroid Hormone Analogs.[a]

Compound Injected	Daily Dose per 100 g (μg)	Molar Ratio	Mean Thyroid Weight per 100 g (mg ± sd)	t_{calcd}[c]	P_{calcd}[d]
Normal Control[d]	---	---	11.19 ± 1.27	13.741	99.9%
Thiouracil Control	---	---	39.07 ± 5.22	---	---
L-T_3	0.0625	0.250	33.96 ± 4.93	1.992	96.8%
	0.125	0.500	32.78 ± 5.67	2.345	98.4%
	0.250	1.000	27.21 ± 3.64	4.960	>99.9%
	0.500	2.000	11.87 ± 4.40	10.928	>99.9%
		0.92[f]			
L-T_4[e]	0.912	2.667	32.61 ± 2.34	2.857	99.4%
	2.731	8.000	19.25 ± 5.48	7.476	>99.9%
		5.1[f]			
L-I_2nPr	0.327	1.500	32.64 ± 7.38	2.151	97.6%
	1.307	6.000	7.48 ± 1.84	14.202	>99.9%
		2.33[f]			
L-Me_2nPr	5.934	45.0	22.41 ± 7.67	5.469	>99.9%
	23.74	180	7.85 ± 1.99	13.972	>99.9%
		39[f]			
L-I_2iPr_2[h]	3.509	15.0	38.58 ± 6.63	0.172	<75%
	14.04	60.0	33.54 ± 7.71	1.811	95.6%
L-4'H-I_2iPr	0.095	0.450	41.04 ± 7.43	g	---
	0.381	1.800	38.71 ± 7.10	0.123	<75%

Table 3-4. (Continued)

Compound Injected	Daily Dose per 100 g (μg)	Molar Ratio	Mean Thyroid Weight per 100 g (mg ± sd)	t_{calcd} [c]	P_{calcd} [d]
L-4'H-I_2F	10.12	50.0	28.79 ± 5.05	3.979	>99.9
	40.48	200	14.17 ± 4.09	10.174	>99.9%
		66[f]			
L-4'H-I_2Cl	1.878	9.00	34.94 ± 2.32	1.828	95.7%
	7.514	36.0	9.51 ± 1.81	13.301	>99.9%
		11.83[f]			
L-4'H-I_2Br	1.129	5.00	25.34 ± 2.22	6.099	>99.9%
	4.516	20.0	8.58 ± 0.88	13.997	>99.9%
		5.1[f]			

[a]Six rats in each experimental group, seven rats in normal control group, and twelve rats in thiouracil control group.

[b-g]See corresponding footnotes, Table 3-2.

[h]Compound precipitated from both dose level injection solutions.

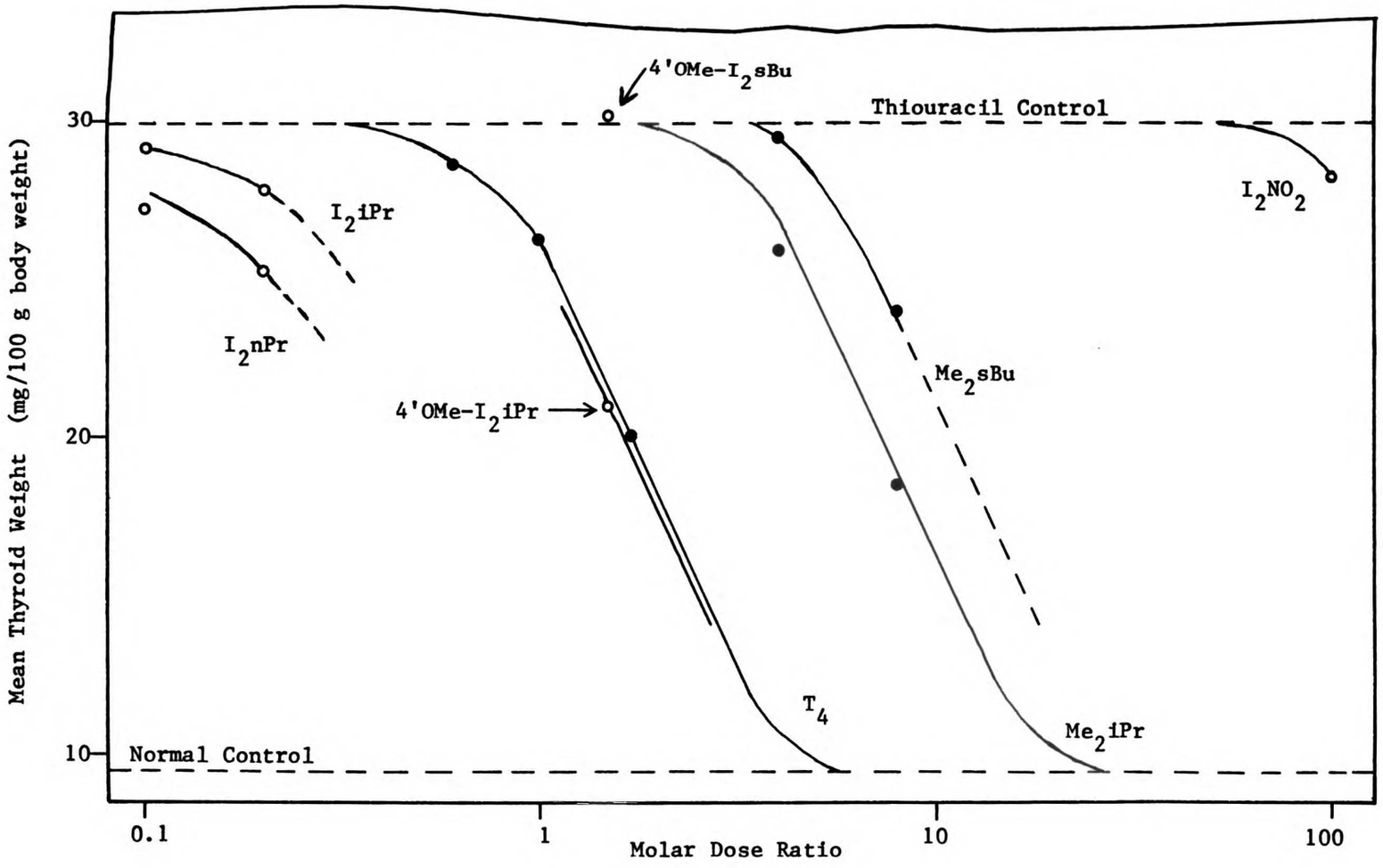


Figure 3-1. Log dose vs. biological response curves: rat antigoiter bioassay #1.

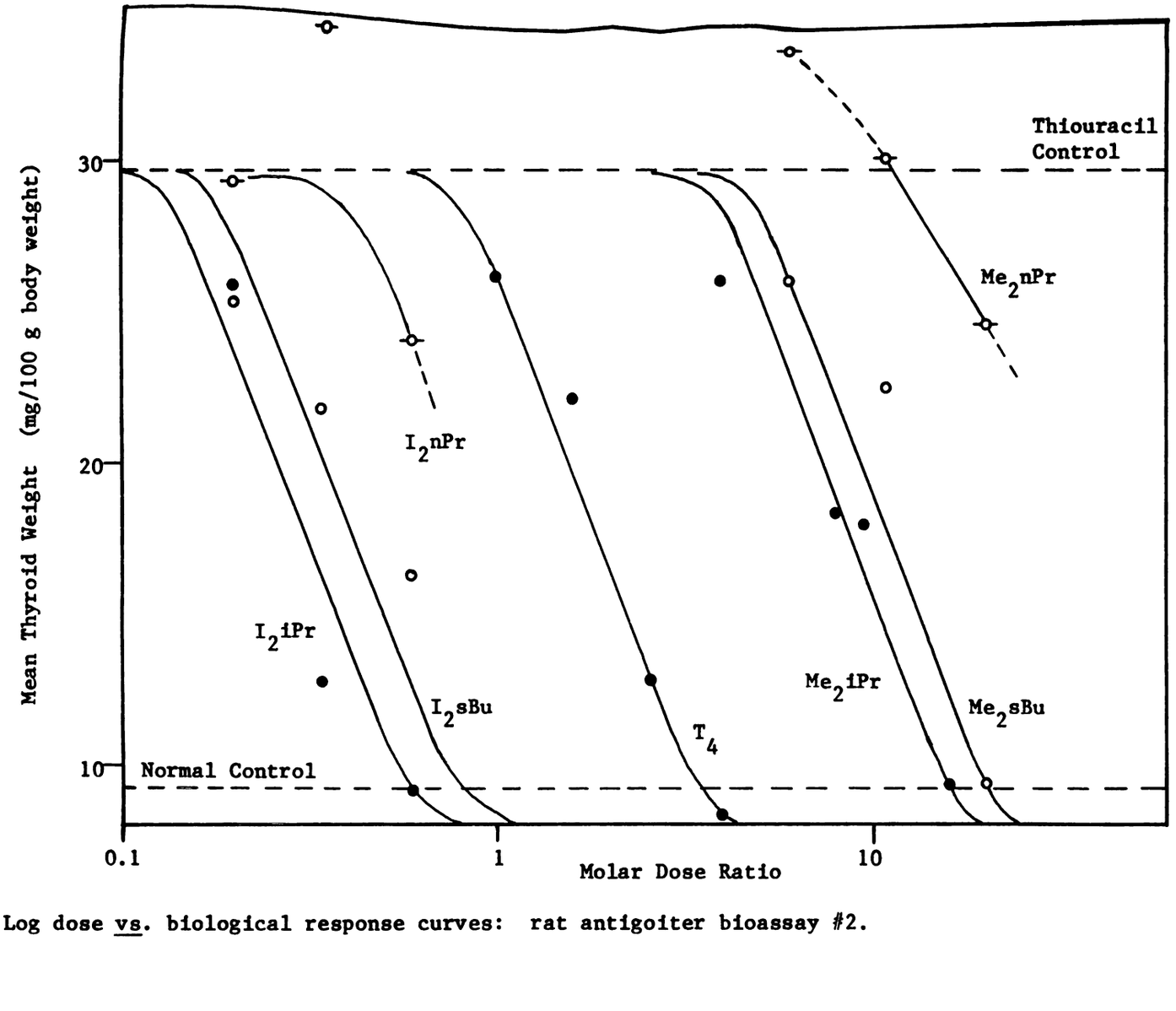


Figure 3-2. Log dose vs. biological response curves: rat antigoiter bioassay #2.

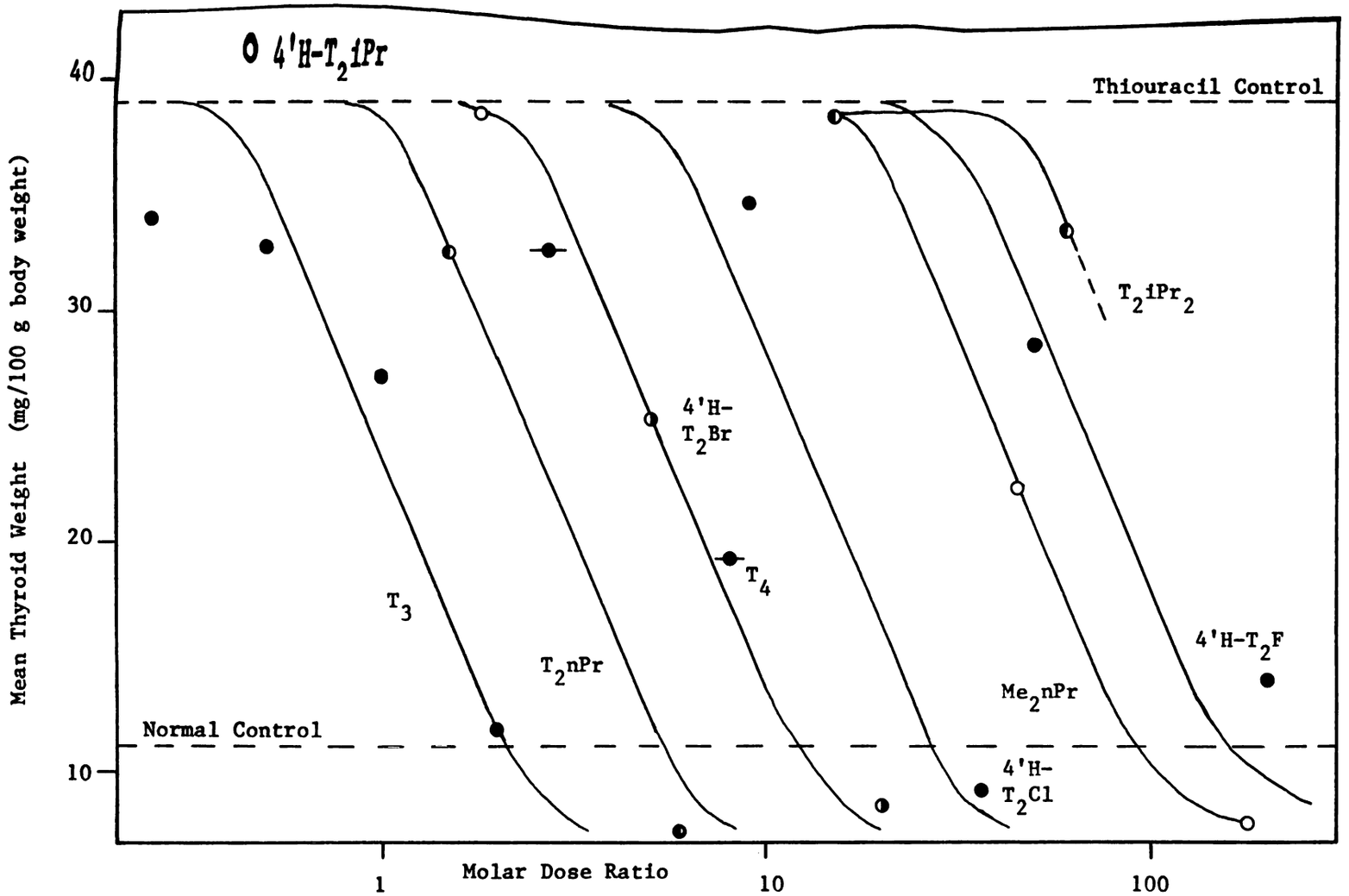


Figure 3-3. Log dose vs. biological response curves: rat antigoiter bioassay #3.

Table 3-5. Summary of Rat Antigoiter Bioassay Results.

Compound	Activity(% L-T_3)	Activity From Assay #
L-T_3(3-1)	100	3
L-T_4(3-2)	18.0	3
L-I_2iPr (3-3)	108[a]	2
L-I_2iPr$_2$(3-4)	b	3
L-I_2nPr (3-5)	39.5	3
L-I_2sBu (3-6)	79.9[a]	2
L-I_2NO_2 (3-7)	<0.32[a]	1
L-Me$_2$iPr (3-8)	4.22,[a]3.94[a]	1,2
L-Me$_2$nPr (3-9)	2.36	3
L-Me$_2$sBu (3-10)	2.88,[a]2.91[a]	1,2
L-4'H-I_2iPr (3-11)	b	3
L-4'H-I_2F (3-12)	1.39	3
L-4'H-I_2Cl (3-13)	7.78	3
L-4'H-I_2Br (3-14)	18.0	3
L-4'OMe-I_2iPr (3-15)	19[a]	1
L-4'OMe-I_2sBu (3-16)	<21[a]	1

[a]Activity determined relative to L-T_4 divided by 5.53 to convert from L-T_4 to L-T_3 reference compound scale.

[b]Assay results inconclusive.

analogs explicitly demonstrates that activity is influenced not only by 3' substituent hydrophobicity but also by the specific conformational size, and steric characteristics of the 3' substituent. Despite its conformational flexibility and its lipophilicity about equal to that of a 3'-iPr substituent, the 3'-nPr substituent apparently also makes a negative steric or hydrophobic contribution to activity due to its extending out from the 3' position considerably further than I or iPr. The activities of L-I_2sBu (3-6) and L-Me_2sBu (3-10) were determined as 79.9% and 2.91%, respectively. This qualitatively suggests that the 3'-sBu substituent bulk or average distance it extends out from the 3' position further than iodine (which will be greater than for even nPr) can be balanced in part by increasing the lipophilicity of the 3' substituent. This point will be more quantitatively developed with structure-activity correlations in Chapter Five. The > 0% activities of L-Me_2nPr (3-9) and L-Me_2sBu (3-10) also reaffirm that halogen is not essential for thyromimetic activity.

Of additional interest is the fact that the activities of both the 3,5-I_2-thyronines and the 3,5-Me_2-thyronines are of the order H < Me < iPr > sBu > nPr for 3' substituents. This suggests that *in vivo* activity is not a function of distribution for these two sets of analogs, which differ considerably in their lipophilicities, but rather is directly related to 3' substituent lipophilicity and inversely related to 3' substituent size or bulk greater than iodine.

The activity of L-$I_2$$iPr_2$ (3-4) was not determined, because upon refrigeration this compound precipitated from the injection solutions. Attempted injection of the suspensions apparently failed due to the inability of most of the suspended solid to be drawn into the injection

syringes. In order to insure solubility, future *in vivo* evalution of this analog should probably utilize: (1) a 100% ethanolic stock solution (with a drop or two of water to effect solution), appropriate fractions being diluted with 100% EtOH to give the injection solutions; and (2) storage of the injection solutions at room temperature.

The < 0.32% activity determined for L-I_2NO_2 (3-7) is consistent with a moderately hydrophilic 3' substituent, a high degree of 4'-OH ionization due to the 3'-NO_2 (favoring TBG binding), and (for the fraction of the analog which has an unionized 4'-OH) a very strong intramolecular hydrogen bond between the 3'-NO_2 and the 4'-OH, preventing 4'-OH hydrogen bond donation to the nuclear receptor.

The activities determined for L-4'H-I_2F (3-12), L-4'H-I_2Cl (3-13), and L-4'H-I_2Br (3-14) (1.39%, 7.78%, and 18.0%, respectively) can be compared with those of the corresponding 4'-OH analogs (1.12%, 4.88%, and 23.78%, respectively[123]). When compared in addition with the activities of several other 4'-H and 4'-OH analogs,[123] it becomes evident that *in vivo* 4'-position hydroxylation of 4'-H analogs can lead to activities ranging from significantly less than to slightly greater than those of the corresponding 4'-OH analogs. Although further study is obviously needed in this area, the data suggests that the degree or rate of 4'-position hydroxylation *in vivo* might be inversely related to the bulk of 3' and 5' substituents, which could provide some steric hindrance to this metabolic transformation. The low *in vivo* activity of L-4'H-I_2iPr(3-11) is consistent with its low *in vitro* binding to solubilized rat hepatic nuclear protein (see below) but is inconsistent with an expected high activity due to metabolic transformation to L-I_2iPr (3-3). Further purification and/or structural verification

of the sample prepared may be necessary to elucidate this ambiguity.

As mentioned earlier, we have proposed the permanent conversion from L-T_4 (3-2) to L-T_3 (3-1) as the reference compound for future rat antigoiter bioassays. The determination of an 18.0% activity for L-T_4 (as directly compared to L-T_3) in the third assay is in excellent agreement with our calculated average estimate from the literature data of 18.1% activity. It thus also provides substantial justification for and validity of our proposal for switching to the L-T_3 reference compound scale.

The dose levels of the analogs tested in the first two bioassays were qualitatively estimated based on the known qualitative structure-activity relationships of the thyroid hormones and analogs. This led to inaccurate dosing for several analogs in these assays. For the third assay, the dose levels of all of the analogs for which activities were obtained were estimated by calculation from correlations of *in vivo* activities with *in vitro* binding to solubilized rat hepatic nuclear protein (using Eqn. 5-42). This permitted fairly accurate determination of the analog activities by choosing dose levels 1/2x and 2x that calculated as necessary for 50% goiter reversal. The success of this method, as demonstrated by the third bioassay results, obviously speaks to its further utiliztion in future assays.

IN VITRO BINDING TO SOLUBILIZED RAT HEPATIC NUCLEAR PROTEIN

Studies[124] have indicated that the thyroid hormones cause sequential increases in nuclear RNA synthesis, nuclear RNA-polymerase activity, cytoplasmic ribosomal and microsomal RNA levels, and finally cytoplasmic enzyme synthesis. As mentioned earlier, most cell nuclei contain non-histone proteins strongly associated with the chromatin, which

possess high-affinity, limited-capacity binding sites for the thyroid hormones and analogs.[24-28] These non-histone nuclear proteins can be solubilized with retention of binding affinity for the thyroid hormones and analogs.[26-28,31,32] The binding affinities of thyroid hormones and analogs to intact rat hepatic nuclei and to solubilized non-histone nuclear protein correlate well with the in vivo thyromimetic activities of the coumpounds, once metabolic effects are taken into account[24-26] (and see Chapter Five). Apparently the binding to such nuclear receptors is the first step in initiating the events which lead to subsequent hormonal expression through enzyme synthesis (Figure 3-4).

Seven of the analogs newly synthesized for this study were tested for their binding affinities to solubilized rat hepatic nuclear protein receptors.[86] A brief assay description and a summary and discussion of the assay results are presented below.

Assay Description.

Rat livers are homogenized and then centrifuged to obtain intact rat hepatic nuclei. The nuclei are lysed by sonication, and high salt concentration (0.2 M $(NH_4)_2SO_4$) is utilized to solubilize the nuclear protein. Centrifugation yields a supernatant containing the desired non-histone nuclear protein. Relative binding affinities of analogs are obtained by determining (by Scatchard analysis) the ability of varying analog concentrations to competitively displace ^{125}I-L-T_3 from the nuclear protein under equilibrium conditions at T = 25°C. Results obtained can be expressed as K_A = analog equilibrium association constant, as $(K_A/K_{T_3}) \times 10^2$ = relative binding affinity of the analog (relative to L-T_3 = 100), or as $\Delta G_A = -RT \ln K_A$ = free energy of analog binding to the nuclear protein receptor (R = ideal gas constant = 1.9872

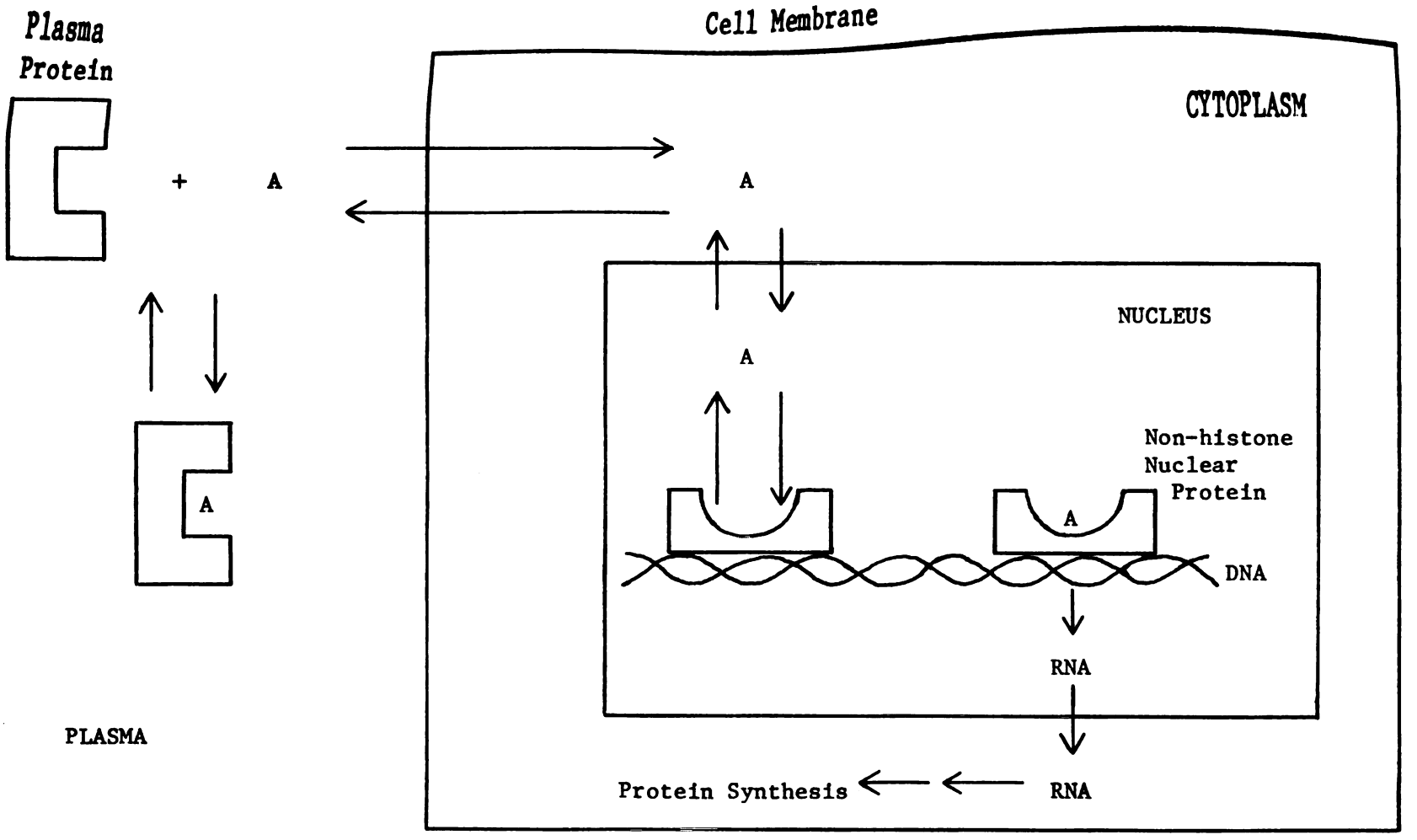
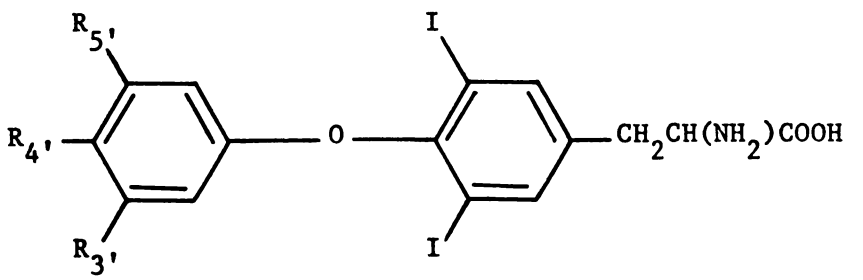


Figure 3-4. Simplified view of events leading to expression of thyromimetic activity; A = thyroid hormone or analog.

Table 3-6. In Vitro Binding Affinities of Thyroid Hormone Analogs to Solubilized Rat Hepatic Non-Histone Nuclear Protein.[a]



Compound	$R_{3'}$	$R_{5'}$	$R_{4'}$	K_A/K_{T3}[b] x 10^2	$-\Delta G_A$[c] (kcal/mole)
L-T_3 (3-1)[d]	I	H	OH	100	12.42
L-I_2iPr$_2$ (3-3)	iPr	iPr	OH	1.10	9.75
L-I_2nPr (3-4)	nPr	H	OH	23.97	11.58
L-I_2iPrCl (3-17)	iPr	Cl	OH	52.56	12.04
L-I_2iPrBr (3-18)	iPr	Br	OH	21.95	11.52
L-I_2iPrI (3-19)	iPr	I	OH	12.41	11.19
L-4'H-I_2iPr (3-11)	iPr	H	H	0.492	9.28
L-4'OMe-I_2iPr (3-15)	iPr	H	OMe	6.820	10.83

[a] At T = 25° C.

[b] Binding affinity relative to L-T_3 = 100. K_{T3} = 1.29 x 10^9 M^{-1}.

[c] $\Delta G_A = -RT \ln K_a$.

[d] Reference compound.

x 10^{-3} kcal/deg·mole).

Results and Discussion.

The results of the binding assays of the analogs newly synthesized for this study to solubilized rat hepatic nuclear protein are presented in Table 3-6.

The binding affinities (relative to L-T_3 = 100) of L-I_2iPr (3-3), L-I_2iPrCl (3-17), L-I_2iPrBr (3-18), L-I_2iPrI (3-19), and L-I_2iPr_2 (3-20) were determined as 89.15,[26] 52.56, 21.95, 12.41, and 1.10, respectively. These results qualitatively support the concept that binding affinity is: (1) indirectly related to 5' substituent size or lipophilicity; and (2) increased by electron withdrawing 3'5' substituents which orientate the 4'-OH toward the 5' position. This will be developed further and more quantitatively in Chapter Five.

The decrease of binding affinity for L-4'OMe-I_2iPr (3-15) (relative binding affinity = 6.820), as compared to L-I_2iPr (3-3) (relative binding affinity = 89.15[26]), is consistent with previous results[26] demonstrating loss of binding upon replacement of 4'-OH with 4'-OMe. The binding affinity of L-4'H-I_2iPr (3-11) is inconsistently low, as compared with other 4'-H analogs,[26] which tend to bind with approximately the same affinity as their 4'-OMe analogs. This once again suggests the need for further evaluation of purity and structure of the sample prepared. The relative binding affinity of 23.97 for L-I_2nPr (3-4) is consistent with: (1) previous results,[26] which generally show relative binding affinities of 3'-alkyl analogs being slightly less than their corresponding rat antigoiter activities; and (2) both 3' substituent hydrophobicity contributing to and "size" detracting from binding affinity.

The binding affinities of these (and other) analogs are used in Chapter Five for the development of quantitative structure-activity relationship correlations of thyroid hormone analogs.

CHAPTER FOUR: MOLECULAR ORBITAL STUDIES

The thyroid hormones and analogs have been the subject of a limited number of extended Hückel[125,126] and CNDO/2[17] molecular orbital studies, which have mainly focused on a qualitative analysis of diphenyl ether ring geometries and of the relative conformer populations. The more precise CNDO/2 calculations,[17] as well as more recent and more extensive CNDO/2 studies[38,41] of this area, have confirmed that: (a) certain 3,5 substituents are capable of "locking" the diphenyl ether thyronine nucleus into the approximately equal energy distal and proximal conformations, which are readily interconvertible at room temperature; and (b) this "locking" ability is directly related to the size of the 3,5 substituents and to thyromimetic activity.

In this chapter are described the results of the molecular orbital calculations undertaken to elucidate the role of the outer ring substituents (3',4', and 5') and of the alanine side chain in determining the *in vivo* and *in vitro* thyromimetic activities of the thyroid hormones ana analogs. These CNDO/2 and *ab initio* molecular orbital calculations

include:

1. A preliminary study, in which the first extensive theoretical examination of the intramolecular hydrogen bonding and interactions in ortho-substituted phenols and thiophenols is provided.[37]
2. Conformational analyses of a number of ortho-alkyl phenols.[37,38]
3. The intermolecular hydrogen bonding of ortho-substituted phenols and phenoxides, as model systems for the binding of the outer ring of thyroid hormones and analogs to nuclear receptors and plasma proteins, respectively.[38]
4. A preliminary conformational analysis of the naturally occurring alanine side chain.[38]

COMPUTATIONAL DETAILS

The CNDO/2 molecular orbital method[127-129] was used in some of these calculations. Except for the halogens the standard atomic parameters were used. Except where noted the halogen parameters employed were those previously used by Kollman, et al.,[17] and only s and p, but no d orbitals were used for F, Cl, Br and I. Unless specifically noted, standard geometrical parameters (selected as suitable average values from available experimental data) were used.[130,131] In particular, if available, bond lengths were taken directly from reference 130 and if not, they were selected from reference 131. For comparison with the CNDO/2 results and experimental data, ab initio molecular orbital calculations using the Gaussian 70 quantum chemistry program[132] with an STO-3G basis set[133] were carried out in selected cases.

As will be shown below, we found that the ability of ortho-iodophenols to form intramolecular hydrogen bonds is best predicted when the iodine

Slater exponent, ζ_I, is given a value of 1.20 instead of the value of 1.09 originally used by Kollman, et al.[17] In order to examine the effects of varying ζ_I, we conducted a series of CNDO/2 calculations on model systems. The results of these studies are presented in Tables 4-1, 4-2, and 4-3.

The first and simplest model system we examined was CH_3-I. CNDO/2 searches for minimum energy C-I bond lengths as a function of ζ_I led to a prediction at ζ_I = 1.09 of a bond distance of 2.07 Å, which only slightly underestimates the experimental value of 2.14 Å. This underestimation increases slightly as ζ_I increases to 1.20. When ζ_I = 1.09 the experimental dipole moment is greatly overestimated. As ζ_I is increased, however, the predicted dipole moment decreases, although it is still somewhat overestimated at ζ_I = 1.20. The atomic population on I varies very little as ζ_I is varied. Variation of ζ_I has little apparent effect on either the overestimated dipole moments or the atomic populations on I of iodobenzene and C_2H_5I.

The rotational barriers in ethane and in C_2H_5I were examined as a function of ζ_I. CNDO/2 predicts the rotational barrier for ethane fairly well. At ζ_I = 1.09 the rotational barrier for C_2H_5I is slightly overestimated. As ζ_I is increased to 1.20 the predicted rotational barrier comes into better agreement with the experimental value. The generally reasonable agreement of these results with the experimental data provides some confidence in our use of ζ_I = 1.20 when examining the intramolecular hydrogen bonding of ortho-iodophenols. This must be viewed with some caution, however, in that the interactions of I in these model systems may be quite different from those involved in intramolecular hydrogen bonding in ortho-iodophenols. It appears, though

that a ζ_I value of 1.2 does reduce the exchange repulsion between I and neighboring atoms enough to bring such properties as rotational barriers and hydrogen bonds, which involve I---H non-bonded interactions, into reasonable agreement with experiment.

A CNDO/2 search for a minimum energy θ_{COH}[137] in phenol led to a prediction of 110°, consistent with the neutron diffraction studies of Frey, et al.,[138] who found C-O-H angles of 111.1° and 113.0° for the phenolic hydroxyls of L-tyrosine and L-tyrosine·HCl, respectively. It is also consistent with the concept of lone pair - lone pair repulsions on the oxygen reducing the magnitude of θ_{COH} from the pure sp^2 value of 120° for a hydroxyl conjugated with an aromatic ring (just as lone pair - lone pair repulsions on oxygen reduce θ_{HOH} of water to 104.52[139] from the pure sp^3 value of 109.47°). Frey, et al.,[138] also found C-O bond lengths of 1.369 Å and 1.378 Å and O-H bond lengths of 0.982 Å and 0.989 Å for the phenolic hydroxyls of L-tyrosine and L-tyrosine·HCl, respectively. These values are close to the values of 1.36 Å and 0.96 Å[130] we used for the phenolic C-O and O-H bond lengths, respectively. Further justification for our use of θ_{COH} = 110° for phenols in all subsequent calculations in these studies was provided when it was found that the CNDO/2 energies of all the ortho-halophenols (cis and trans conformers), of all the unsymmetrical 2,6-dihalophenols ("cis" and "trans" conformers), and of ortho-cresol (cis and trans conformers; all CH_3 rotamers) are all significantly lowered when the phenolic θ_{COH} is decreased from 120° to 110°, as seen in Table 4-4. (Further comment will be made below on the trends in this table.)

Table 4-1. CNDO/2 Geometry Searches for "Best" C-I Bondlength in CH_3I.[a,b]

α_I	$R(C-I)_{calcd}$, Å
1.09	2.07
1.145	2.00
1.20	1.94

[a] R(C-H) = 1.10 Å; all angles tetrahedral.

[b] $R(C-I)_{exptl}$ = 2.14 Å.[134]

Table 4-2. CNDO/2 Rotational Barriers C_2H_5X[a] (kcal/mole).

X	I	ΔE_{calcd}[b]	ΔE_{exptl}[b]
H	---	2.21	2.75[c]
I	1.09	4.16	
	1.145	3.93	
	1.20	3.72	
	---		3.2 +0.5[c]

[a] R(C-H) = 1.10 Å; R(C-C) = 1.54 Å; R(C-I) = 2.14 Å; all angles tetrahedral.

[b] $\Delta E = E_{eclipsed} - E_{staggered}$.

[c] Reference 135.

Table 4-3. CNDO/2 Dipole Moments and Atomic Populations.

Compound	$\mathscr{S}_I$	μ_{calcd}, D	μ_{exptl}, D	Atomic Population On I
CH_3I[a]	1.09	2.47		53.040
	1.145	2.39		53.041
	1.20	2.32		53.042
	---		1.64[e]	
C_2H_5I[b]	1.09	2.56		53.060
	1.145	2.60		53.074
	1.20	2.64		53.087
	---		1.90[e]	
PhI	1.09	2.72[c]		53.067[c]
	1.20	2.72[d]		53.084[d]
	---		1.70[e]	

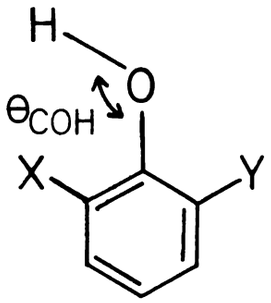
[a] Calculated at minimum energy geometries of Table 4-1.

[b] See footnote a, Table 4-2 for geometries.

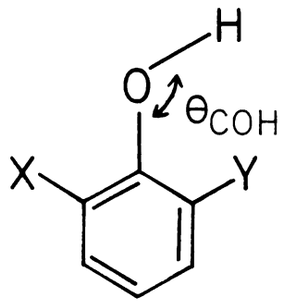
[c] From reference 17 at R(C-I) = 2.086 Å.

[d] R(C—I) = 2.05 Å.

[e] Reference 136.

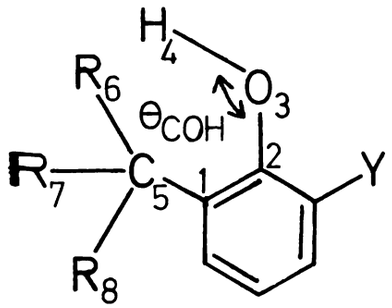


4-1

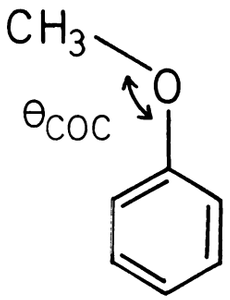


4-2

Because we were also interested in examining the intramolecular hydrogen bonding of ortho-methoxyphenol, we conducted a CNDO/2 search for a minimum energy θ_{COC} for anisole (4-4; CH_3 protons staggered). The resulting prediction of θ_{COC} = 113° was used in subsequent calculations on ortho-methoxyphenol.



4-3



4-4

This value is not far from θ_{COC} = 116.9° found for the two aromatic methoxy groups of l-rotenene in the X-ray crystal study of Arora, et al.[140]

In order to estimate relative populations of different conformations or geometries of certain molecules, classical Boltzmann distribution

Table 4-4. CNDO/2 Energy Dependence of ortho-Substituted Phenols (4-1) and 4-2) on θ_{COH}.

X	Y	ϕ_{2156}[a]	$\Delta E(\theta_{COH} = 110° \rightarrow \theta_{COH} = 120°)_{calcd}$ (kcal/mole) 4-1	4-2
H	H		1.90	---
F	H		2.25	1.98
Cl	H		2.63	1.95
Br	H		2.38	1.92
I[b]	H		1.98	1.87
F	Cl		2.30	2.71
F	Br		2.27	2.46
F	I[b]		2.22	2.07
Cl	Br		2.65	2.43
Cl	I[b]		2.59	2.04
Br	I[b]		2.33	2.00
CH_3	H	0	1.46[c]	1.96[d]
		30	1.96[c]	1.95[d]
		45	2.16[c]	1.94[d]
		60	2.23[c]	1.94[d]

[a] ϕ_{2156} refers to structure 4-3 where $R_6=R_7=R_8=Y=H$.

[b] δ_I = 1.20.

[c] ϕ_{1234} = 0° for structure 4-3.

[d] ϕ_{1234} = 180° for structure 4-3.

partition functions, T = 298° K, and constant entropy contributions between the different conformations or geometries were assumed (see references 141 and 142 for discussion and examples of this type of treatment).

Our assumption that ΔS for the cis → trans conversion for unsymmetrically ortho-substituted phenols and thiophenols is essentially zero is supported by experimental thermodynamic studies on ortho-tert-butylphenol,[143] 2-tert-butyl-6-methylphenol,[143] ortho-bromophenol,[144] 2,4-dibromo-6-tert-butylphenol,[144] and ortho-iodophenol[145]. (See reference 146 for a more complete discussion of this area.)

INTRAMOLECULAR HYDROGEN BONDING IN ORTHO-SUBSTITUTED PHENOLS AND THIOPHENOLS

Although molecular orbital calculations have been carried out on a wide variety of hydrogen bonded systems,[147] relatively few studies have involved molecules with an internal hydrogen bond. The intramolecular hydrogen bonds of the enol forms of malonaldehyde and acetylacetone have been examined by a number of MO methods.[148-153] Murthy, et al.,[148] used EHT and CNDO/2 MO methods to study intramolecular hydrogen bonds and their effects on cis-trans isomerism in ortho-fluorophenol, ortho-nitrophenol, and salicylaldehyde. The influence of intramolecular hydrogen bond formation on the conformation of 1,3-propanediol has been examined by Johansson, et al.,[154] using ab initio MO calculations. Such studies have generally given reasonable estimates of the energy of intramolecular hydrogen bond formation, although most have dealt with systems which form unusually strong intramolecular hydrogen bonds due to internal geometry constraints.[149-153] It is somewhat surprising, however, that

more MO studies of intramolecular hydrogen bonding have not been undertaken in view of the fact that a wide variety of biologically active compounds possess as necessary for activity functional groups capable of forming intramolecular hydrogen bonds.

Some of the first examples of intramolecular hydrogen bonding to be studied experimentally were a number of ortho-substituted phenols.[155,156] Interest in these molecules was stimulated by the observation of two IR O-H stretching bands. From relative IR O-H stretching intensities, the amounts of "cis" hydrogen bonded and "trans" non-hydrogen bonded conformations and hence the energy of the intramolecular hydrogen bond could be estimated. In this manner, for example, Pauling[155] first estimated the intramolecular hydrogen bond energy of ortho-chlorophenol in CCl_4 to be about 1.4 kcal/mole. Since then, the intramolecular hydrogen bonding of a number of ortho-substituted phenols[143,144,155-165] and unsymmetrical 2,6-dihalophenols[160,166] has been similarly examined in various solvents and in the gas phase. O-H torsional frequencies of the cis and trans conformations of a number of ortho-substituted phenols have been used to calculate the enthalpy differences between the two conformations.[157,167] The intramolecular hydrogen bonding of orthotrifluoromethylphenol has been examined by Doddrell, et al.,[168] using EHT and CF_3 ^{19}F and OH ^{1}H chemical shift studies. Schaefer[169] has proposed linear relationships of intramolecular hydrogen bond energies with OH ^{1}H chemical shifts as well as with O-H torsional frequencies. Allan and Reeves[170,171] have also used OH ^{1}H chemical shifts for the study of intramolecular hydrogen bonds in ortho-substituted phenols.

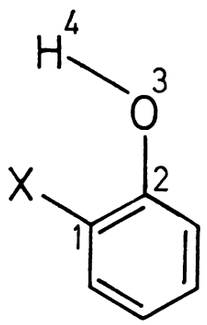
It has been shown that the phenolic 4'-OH of the thyroid hormones and analogs is essential for maximal in vivo and in vitro thyromimetic

activity.[2,24,25,50,54,55,172] The role of this phenolic OH has been logically ascribed to involvement in hydrogen bond formation with some appropriate receptor functional group. Little attention has been paid, however, to ortho-substituent interactions with the phenolic OH group, especially with respect to their effect on both intramolecular and intermolecular hydrogen bond formation. Because of the paucity of theoretical MO studies of intramolecular hydrogen bonding and because of our interest in the thyroxine system, we have undertaken the first extensive theoretical examination of the intramolecular interactions of ortho-substituents with the phenolic OH group of various phenols (as model systems) using CNDO/2 and ab initio MO calculations. We also examined the intramolecular hydrogen bonding of several ortho-substituted thiophenols in order to compare their hydrogen bonding with that of the phenols.

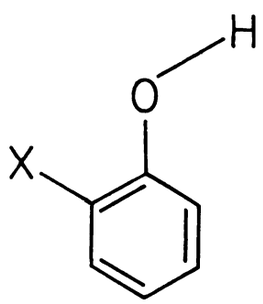
The questions to which we address ourselves in this section are (1) Can we explain the "anomalous" order[158,160,167,170] of intramolecular hydrogen bonding strengths in the ortho-halophenols (Cl $\gtrsim$ F > Br > I)? (2) Can our theoretical model explain the minimum energy conformational and hydrogen bonding energies of other ortho-substituted phenols? (3) Specifically, for ortho-CF_3-phenol, why is the larger hydrogen-bonded peak in the IR shifted to higher frequencies from the free O-H stretching frequency[161] (to our knowledge, the only known hydrogen bond where such an effect occurs)? (4) Can we rationalize the observed far and near infrared absorption spectra using our calculations? and (5) Can we predict the intramolecular hydrogen bond energies and properties of ortho-substituted phenols and thiophenols not yet determined experimentally?

Intramolecular Hydrogen Bonding in ortho-Halophenols.

The first ortho-substituted phenols we examined were the ortho-halophenols. CNDO/2 and ab initio cis → trans energy differences and data from a number of sources and representative of a variety of experimental and theoretical methods are presented in Table 4-5. A CNDO/2 value of 1.09 for the iodine exponent $\mathcal{S}_I$ does not adequately predict the intramolecular hydrogen bond strengths of ortho-iodoohenols, as seen in Table 4-6. Increasing $\mathcal{S}_I$ to 1.20 significantly improves



4-5



4-6

the agreement of the CNDO/2 results with the experimental data. Because of this improvement, we elected to use a value of 1.20 for $\mathcal{S}_I$ in all subsequent calculations. Our CNDO/2 calculations predict that the order of intramolecular hydrogen bond strengths for the ortho-halophenols is Cl > Br > F > I, while the ab initio calculations, although lacking the Br and I compounds, suggest the order Cl $\gtrsim$ F > Br > I. The various experimental data give the order to be either Cl > Br > F > I or Cl > F > Br > I, depending on the experimental method of study and the solvent used. Our calculations should relate most directly to the gas phase where the order appears to be Cl $\gtrsim$ F > Br > I.[167]

Table 4-5. Experimental[a] and Theoretical Values for Intramolecular Hydrogen Bond Strengths of the ortho-Halophenols (4-5 and 4-6).

X	ΔE(4-5 → 4-6) (kcal/mole)	Method of Study[b]	Solvent	Ref.
F	0.2	A	---	148
	1.1	B	---	148
	1.44	C	Cyclohex.	167
	1.63	C	Vapor	167
	1.37	B	---	*
	1.68	D	---	*
Cl	1.62	C	Cyclohex.	167
	1.63	C	Vapor	167
	1.44	E	CCl_4	160
	2.38	E	CCl_4	158
	2.36	F	CS_2	170
	2.30	B	---	*
	1.77	D	---	*
Br	1.57	C	Cyclohex.	167
	1.53	C	Vapor	167
	1.21	E	CCl_4	160
	2.15	E	CCl_4	158
	2.14	F	CS_2	170
	1.68	B	---	*

Table 4-5. (Continued)

X	ΔE(4-5 → 4-6) (kcal/mole)	Method of Study[b]	Solvent	Ref.
I	1.45	C	Cyclohex.	167
	1.32	C	Vapor	167
	1.08	E	CCl_4	160
	1.54	E	CCl_4	158
	1.65	F	CS_2	170
	0.75[c]	B	---	*

[a] ΔE values estimated, if necessary, from the experimental data.

[b] A = EHT MO calculations; B = CNDO/2 calculations; C = IR: OH torsional frequencies; D = Ab Initio MO calculations; E = IR: OH stretching frequencies: F = NMR: ^1H chemical shifts.

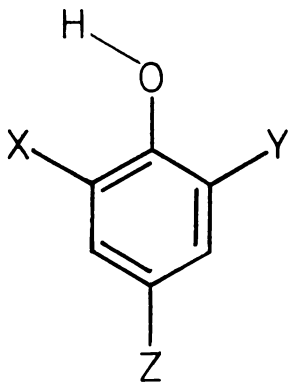
[c] $\mathcal{S}_I$ = 1.20.

*This study.

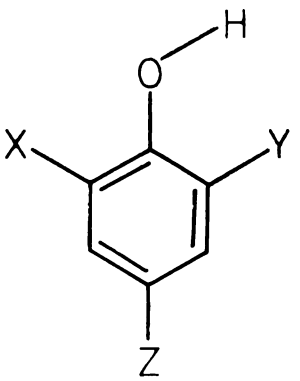
Table 4-6. Dependence on $\mathcal{J}_I$ of CNDO/2 Intramolecular Hydrogen Bond Strengths of ortho-Iodophenols (4-1 and 4-2).

X	Y	$\mathcal{J}_I$	ΔE_{calcd} (4-1 → 4-2) (kcal/mole)
I	H	1.09	-0.86
		1.145	0.11
		1.20	0.75
F	I	1.09	0.02
		1.20	0.68
Cl	I	1.09	3.14
		1.20	1.52
Br	I	1.09	2.53
		1.20	0.90

Table 4-7. Experimental, CNDO/2, and Ab Initio Intramolecular Hydrogen Bond Strengths of Unsymmetrical 2,6-Dihalophenols (4-7 and 4-8).



4-7



4-8

X	Y	Z	ΔE(4-7 → 4-8) (kcal/mole)		
			CNDO/2[a]	Ab Initio[a]	Exptl.[b]
Cl	F	H	0.89	0.04	0.18[d]
Cl	F	Cl			0.25[e]
Br	F	H	0.27		0.10[d]
Br	F	Br			0.08[d]
Cl	Br	H	0.61		0.19[e]
Cl	Br	Cl			0.28[e]
F	I	H	0.68[c]		0.36[d]
F	I	I			0.33[d]
					0.40[e]
Cl	I	H	1.52[c]		0.55[d]
Cl	I	Cl			0.56[d]
Br	I	H	0.90[c]		0.70[e]
Br	I	Br			0.47[e]

[a]This study.

[b]Method of Study = A, footnote b, Table 4-5: Solvent CCl_4.

[c]α_I = 120. [d]Ref. 160. [e]Ref. 166.

In either case, the magnitudes of the calculated internal hydrogen bond strengths are both reasonable and in moderately good agreement with the experimental data. Considering the relative electronegativities, one might expect the order of intramolecular hydrogen bond strengths to be F > Cl > Br > I. Yet both the theoretical calculations and the experimental data are in agreement with the fact that ortho-fluorophenol forms a weaker internal hydrogen bond than expected. This finding will be discussed in more detail below.

In order to further investigate this "anomalous" trend in the intramolecular hydrogen bond strengths of the ortho-halophenols, we next examined the intramolecular hydrogen bonding of the unsymmetrical 2,6-dihalophenols. The CNDO/2 and ab initio results are summarized in Table 4-7. Experimental data on some unsymmetrial 2,6-dihalophenols is also presented for comparison. The CNDO/2 calculations predict the intramolecular hydrogen bond strength order of the halogens of the unsymmetrical 2,6-dihalophenols as Cl > Br > F > I. Again, although lacking the Br- and I-containing compounds, the ab initio calculations suggest the order to be Cl $\gtrsim$ F > Br > I. The CNDO/2 calculations generally slightly overestimate the differences between the relative hydrogen bond strengths. Once again, however, both the experimental data and the theoretical calculations predict an anomalously weak intramolecular hydrogen bond for F when located ortho to a phenolic OH group.

This "anomalous" ordering of the intramolecular hydrogen bond strengths of the ortho-halophenols has been attributed to differences in:

1. Interactions of halogens with solvent.[143,160,162,163,173,174]
2. Tendencies to dimerize.[167]
3. Deviations from optimal hydrogen bonding geometries[158,163,166,167,171,173]
4. Intrinsic hydrogen bonding capabilities of the halogens.[167]
5. Inductive and mesomeric capabilities of the halogens.[158,166]
6. Repulsive halogen--oxygen and halogen--hydrogen "interorbital" interactions.[158]

Since both the experiments and ab initio calculations find the "anomalous" order (Cl $\gtrsim$ F > Br > I) to hold for gas phase intramolecular hydrogen bonding, we feel explanations 1 and 2 can not be used to explain the anomalous order.

The next possible explanation we examined was geometric. The optimal strength for an intermolecular O-H- - H hydrogen bond should occur when $\theta_{HOX} = 0$.[147,175] Hence, the magnitude of the deviation of θ_{HOX} from 0° should be reflected in a corresponding deviation in the H- - X hydrogen bond strength. Although there is a full 10° variation in θ_{HOX} in the ortho-halophenols (Table 4-8), the difference in θ_{HOX} between X=F (50.54°) and X=Cl (44.26°) is only 6.28°, too small a change to account by itself for the weaker than expected H - - F hydrogen bond. Based on the angles alone, ortho-iodophenol (θ_{HOX} = 39.11°) should form the strongest hydrogen bond. Besides the hydrogen bond angle, the intramolecular geometry constraints might cause some repulsions which do not folow the same order as electronegativity. By comparing $R(H- - H)_{calcd}$ and the sums of Van der Waals radii for H + X (Table 4-8), one can see that the degree of overlap of the Van der Waals radii of H and X is in the order F < Cl < Br < I. In particular, while there is significant

overlap for Cl, Br and I, for F the overlap is considerably less. This suggest that the weaker than expected intramolecular hydrogen bond of F in ortho-fluorophenol may be partially due to the unfavorably (as compared to the halogens of the other ortho-halophenols) large H - - F internuclear distance. This is qualitatively supported by the fact that the CNDO/2 calculated energy dependence on θ_{COH} is apparently essentially independent of the ortho-substituent for the trans conformers of the ortho-monosubstituted phenols, but not for the cis conformers (Table 4-4). As θ_{COH} is decreased from 120° to 110° for the cis conformers of the ortho-halophenols, R(H - - X) decreases and the overlap of the H and X Van der Waals radii should increase. For ortho-fluorophenol, the minimal overlap of the H and F Van der Waals radii at θ_{COH} = 120° is not significantly changed at θ_{COH} = 110°. For ortho-chlorophenol, the overlap of the H and Cl Van der Waals radii significant (and perhaps nearly optimal) and hence $\Delta E(\theta_{COH} = 110° \rightarrow \theta_{COH} = 120°)$ is much larger than for ortho-fluorophenol. Thus, decreasing θ_{COH} from 120° to 110° causes a small increase in hydrogen bonding for ortho-fluorophenol and a larger increase for ortho-chlorphenol, in addition to the inherent stabilization seen in phenol. This causes the $\Delta E(\theta_{COH} = 110° \rightarrow \theta_{COH} = 120°)$ ordering to be Cl > F > H for X. While the overlap of the Van der Waals radii also increases significantly for X = Br and I, apparently the overlap is greater than the optimal value and H - - X repulsive interactions also increase significantly as θ_{COH} is decreased, to the point where $\Delta E(\theta_{COH} = 110° \rightarrow \theta_{COH} = 120°)$ is only slightly greater for ortho-iodophenol than for phenol itself. These trends for the cis ortho-halophenols are mirrored by parallel trends for the unsymmetrical 2,6-dihalophenols (Table 4-4).

Table 4-8. Geometrical Parameters and Halogen Electronegativities for the cis ortho-Halophenols (4-5).

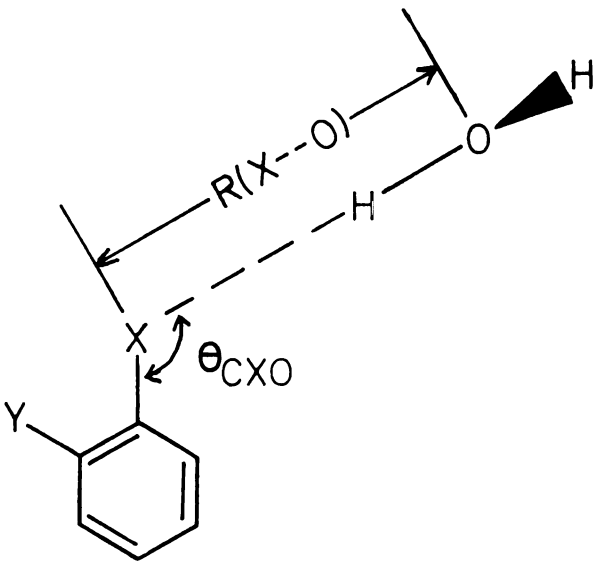
X	Electronegativity of X[a]	θ_{HOX}[b] (degrees)	R(H - - X)[b] (Å)	R(O - - X)[b] (Å)	ΣVan der Waals Radii (Å) H+X	X+O
F	4.0	50.54	2.26	2.75	2.67	2.99
Cl	3.0	44.26	2.35	2.94	2.97	3.29
Br	2.8	41.96	2.41	3.03	3.08	3.40
I	2.5	39.11	2.49	3.16	3.27	3.59

[a]Reference 178.

[b]Calculated based on the geometries used in our calculations.

[c]Reference 179 ; Van der Waals radii values used were: H, 1.20 Å; O, 1.52 Å; S, 1.80 Å; F, 1.47 Å; Cl, 1.77 Å; Br, 1.88 Å; I, 2.07 Å.

Geometrical constraints clearly do not provide a complete and satisfactory explanation for the order of the intramolecular hydrogen bond strengths. To determine what the "intrinsic" hydrogen bond acceptor capabilities of aromatically substituted halogens are, we carried out a series of CNDO/2 and ab initio model calculations on the intermolecular hydrogen strengths of the four different H_2O/halobenzene dimers. This was done so that we might examine the deviations of the hydrogen bond strengths and the halogen - - H and halogen - - O internuclear distances of the ortho-halophenols from the "ideal" equilibrium values of these model systems. The model system geometry (4-9: Y = H) was defined as follows. The halogen, oxygen, and proton involved in the



4-9

hydrogen bond are colinear since this geometry should give maximal hydrogen bond strength.[147,175] The O-H bond involved in the hydrogen bond lies in the plane of the halobenzene in order to best approximate

the cis geometry of the respective ortho-halophenol. The second O-H bond of H_2O lies in a plane perpendicular to the halobenzene ring plane in order to minimize any interactions of this second H_2O proton with the halobenzene. With $\theta_{COX} = 180°$, a geometry search for the minimum energy R(X - - O) was conducted for each halobenzene (see 4-9). Then, at this minimum energy R(X- - O), a geometry search (30° variations in θ_{CXO} to 90°) for the minimum energy θ_{CXO} was conducted. The results are presented in Table 4-9 and help a great deal in explaining the CNDO/2 and *ab initio* orders for the intramolecular hydrogen bond strengths of the ortho-halophenols. The calculated orders of equilibrium intermolecular hydrogen bond energies for the H_2O/halobenzene dimers are Cl > F > Br > I (CNDO/2) and F > Cl > Br > I (extrapolating for the Br and I points, *ab initio*). So the *ab initio* calculated intermolecular hydrogen bond energies apparently are of the same order as the halogen electronegativities, as expected. With CNDO/2, however, the F value is anomalously out of line with the trend expected for the halogen electronegativities. The *ab initio* ordering appears to be the correct one for these intermolecular hydrogen bonds, since experimentally the ordering for intermolecular hydrogen bond strengths is F > Cl > Br > I for the phenol/cyclohexyl halide dimers[181,182] and for the phenol/n-pentyl halides dimers.[183] The CNDO/2 intramolecular hydrogen bond strengths of the ortho-halophenols (Table 4-1) range from 35% to 45% of the intermolecular hydrogen bond strengths of the respective H_2O/halobenzene dimer model systems. These decreases from the theoretically "optimal" intermolecular hydrogen bond strengths appear with CNDO/2 to be due to the geometrical constraints of the ortho-halophenols. On the other hand, the *ab initio* intramolecular hydrogen bond strength of ortho-fluorophenol is approximately equal to the intermolecular hydrogen bond strength of the respective H_2O/fluorobenzene

Table 4-9. CNDO/2 and Ab Initio Hydrogen Bond Energies and Geometrical Parameters for H_2O/Halobenzene Dimers (4-9)[a].

X	Y	ΔE_{calcd}[b,c] (kcal/mole)	θ_{CXO} (degrees)	$R(X - - O)_{calcd}$[c,d] (Å)	$R(X - -H)_{calcd}$[e] (Å)	
F	H	3.82	180	2.56	1.60	CNDO/2
		3.92	120[g]	2.56	1.60	
Cl	H	5.50	180[g]	2.97	2.01	
Br	H	3.73	180[g]	3.20	2.24	
I	H	1.94	180[g]	3.47	2.51	
F	H	1.52	180	2.91	1.95	Ab Initio
		2.09	120[g]	2.91	1.95	
Cl	H	0.66	180	3.93	2.97	
		0.91	120[g]	3.93	2.97	
F	OH[f]	2.21	120[h]	2.91[h]	1.95[h]	
Cl	OH[f]	0.97	120[h]	3.93[h]	2.97[h]	

[a] H_2O geometry: see note 180.

[b] ΔE = hydrogen bond strength.

[c] ΔE and $R(X - - O)$ calculated exactly with CNDO/2. With ab initio, $R(X - - O)$ calculated to $\pm$ 0.07 Å; ΔE and $R(X - - O)$ then estimated by a three point quadratic fit.

[d] $R(X - - O)_{calcd}$ = minimum energy $R(X - - O)$ value at $\theta_{CXO} = 180°$.

[e] $R(X - - H)_{calcd} = R(X - - O)_{calcd} - 0.96$ Å.

[f] O-H trans to X.

[g] θ_{CXO} = minimum energy θ_{CXO} value for 30° variations in θ_{CXO} from 180° to 90°.

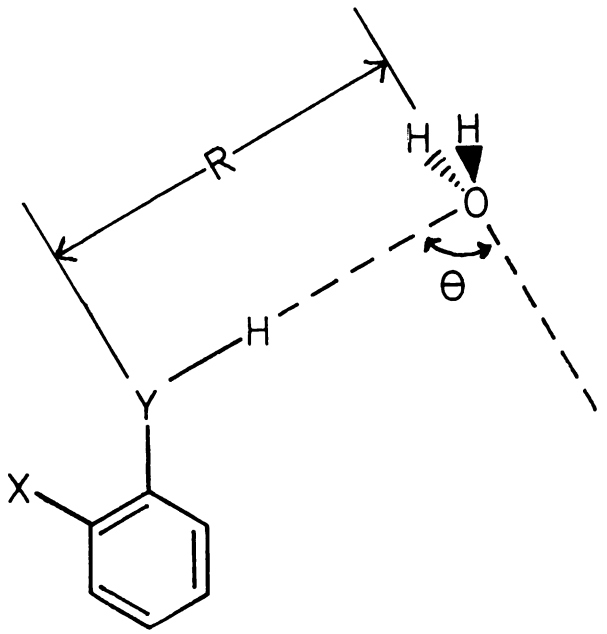
[h] Minimum energy geometry for H_2O/C_6H_5X dimer.

dimer model system, and the ab initio intramolecular hydrogen bond strength of ortho-chlorophenol is actually significantly greater than the intermolecular hydrogen bond strength of the respective H_2O/chlorobenzene dimer. It appears reasonable that a significant amount of this greater intramolecular hydrogen bond strength in ortho-chlorophenol might be due to relief of O - - Cl repulsions existing in the trans isomer (4-6: X = Cl) rather than to the intrinsic H - - Cl hydrogen bond strength. The overlap of the halogen and O Van der Waals radii is approximtely equal for all 4 of the H_2O/halobenzene dimers (Table 4-9) but increases dramatically as a function of halogen size from very little for ortho-fluorophenol to quite significant for ortho-iodophenol (Table 4-8). This, coupled with the ab initio ΔE(cis → trans) for ortho-chlorophenol being much greater than ΔE for the H_2O/chlorobenzene dimer, suggests that the intramolecular hydrogen bond energies of the ortho-halophenols may be in part due to the phenols attempting to relieve halogen - - O repulsion and in part due to specific H - - halogen attractions. O - - halogen repulsive bond orders increase F < Cl < Br < I and H - - halogen attractive bond orders increase I > Br > Cl > F (see below) and, thus, it is not surprising that the observed hydrogen bond strengths do not folow the order of electronegativity. The order of intramolecular hydrogen bond strengths parallels the intermolecular order in the CNDO/2 calculations because this method is known to generally underestimate interatomic repulsions. So, the intrinsic H - - halogen attractions play the dominant role in this series.

The ab initio intermolecular hydrogen bonding results for the H_2O/halobenzene dimers are not totally definitive because we have used a model hydrogen bond (HOH - - X - C_6H_5) to represent the intramolecular

hydrogen bonds and have compared them with the actual ortho-X-C_6H_4OH intramolecular hydrogen bond. We thus calculated and compared the hydrogen bond energies of H_2O/ortho-halophenol dimers (4-9: Y = OH trans to X) with the intermolecular hydrogen bond energies of the corresponding H_2O/halobenzene dimer (4-9: Y = H). In order to enable direct comparison of hydrogen bond energies, the hydrogen bonding geometries of the H_2O/ortho-halophenol dimers were taken to be the minimum energy geometries calculated for the corresponding H_2O/halobenzene dimers. As can be seen from Table 4-9, an ortho-hydroxyl substituent slightly increased the intrinsic hydrogen bonding capabilities of fluorobenzene and chlorobenzene by essentially the same percentages: 5.9% and 6.5%, respectively.

Reversing the situation, the effect of an ortho-halo substituent on the intrinsic hydrogen bonding capability of phenol was next examined. For C_6H_5OH as a proton donor to H_2O (4-10: X = H; Y = O), a "linear"



4-10

dimer was assumed with the two monomeric units lying in perpendicular planes and with the two O atoms and the H involved in the hydrogen bond colinear.[147,175] θ (see 4-10) was taken as 57° (from the STO-3G H_2O dimer value[184]). Only R (the O - - O internuclear distance) was varied in our geometry search. The *ab initio* minimum energy geometry determined for the H_2O/C_6H_5OH dimer was then assumed (in order to enable direct comparison of hydrogen bond energies) for calculating the intermolecular hydrogen bond energies of the H_2O/ortho-halophenol dimers (4-10: X = F or Cl; Y = O). As seen in Table 4-10, an electron-withdrawing, ortho-halogen increases the intermolecular hydrogen bonding energy of phenol as a proton donor. The increases, however, are of the opposite order (F with a 3.8% increase < Cl with a 12.5% increase) as the halogen electronegativities (F > Cl), probably as a result of the ability of F to more easily (than Cl) donate electron density by resonance back into the aromatic ring and hence to the OH group. Mulliken populations show that in both the halobenzenes and the ortho-halophenols, the fluoro compound donates $\sim$ 0.04 more π electrons into the ring than the chloro. (The $\sigma + \pi$ charge of fluorine (-0.130) is slightly more negative than that of Cl(-0.117).) The partial positive charge on the proton is also consistent with the relative strength of O-H as a proton donor, being + 0.217 (phenol), + 0.220 (trans *o*-F phenol) and + 0.224 (trans *o*-Cl phenol). This suggests that Cl in ortho-chlorophenol may have a greater effect than F in ortho-fluorophenol in reinforcing the intramolecular hydrogen bond. Estimating the intrinsic hydrogen bond energy for the dimer ortho-F-C_6H_4OH - - - F-C_6H_5-ortho-OH by (hydrogen bond energy for ortho-OH-C_6H_4F - - - H-O-H dimer) x (hydrogen bond energy ratio for ortho-F-C_6H_4OH - - - OH_2 dimer

Table 4-10. CNDO/2 and *Ab Initio* Hydrogen Bond Energies (ΔE)[a] and Geometries[b] for H_2O/*o*-X-C_6H_4YH (Y= O or S) Dimers (4-10 and 4-31).

Dimer Structure	X		Y = S		Y = O	
			CNDO/2	*Ab Initio*	CNDO/2	*Ab Initio*
4-31	H	R	2.90	3.35	2.54	2.79
		ΔE	2.31	2.57	6.12	4.31
4-10	H	R	2.76	3.56	2.56	2.64
		ΔE	10.54	1.12	6.04	8.97
4-10	F	R				2.64[c]
		ΔE				9.31
4-10	Cl	R				2.64[c]
		ΔE				10.10

[a]Energies in kcal/mole.

[b]R = Y - - O internuclear distance in Å.

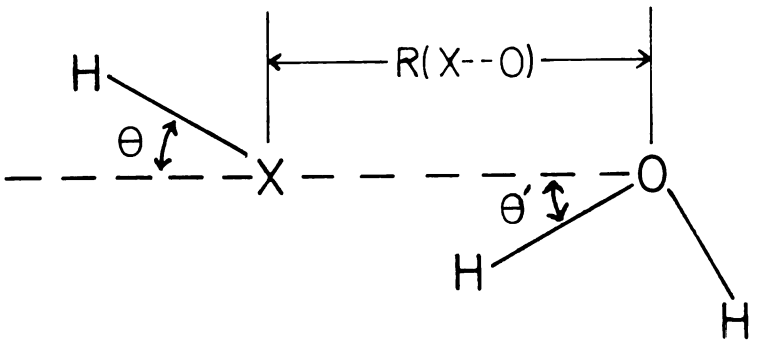
[c]Minimum energy R value for H_2O/C_6H_5OH dimer 4-10.

vs. H-O-H - - - OH_2 dimer) = (2.21 kcal/mole) x (9.31 kcal/mole/5.88 kcal/mole[184]), a value of 3.50 kcal/mole is obtained.

That this is significantly larger than the actual ab initio calculated intramolecular hydrogen bond energy of ortho-fluorophenol of 1.68 kcal/mole can, as before, be primarily ascribed to deviations in ortho-fluorophenol from the "optimal" hydrogen bonding geometry. A similar estimation of the intramolecular hydrogen bond energy of ortho-chlorophenol yields a value of (0.97 kcal/mole) x (10.0 kcal/mole/5.88 kcal/mole) = 1.67 kcal/mole, which is still slightly less than the actual ab initio calculated intramolecular hydrogen bond energy of ortho-chlorophenol of 1.77 kcal/mole. Taking into account the non-optimum geometry of the intramolecular Cl - - H-O hydrogen bond, it seems very surprising that this estimated intermolecular hydrgoen bond is weaker than the intramolecular hydrogen bond. Evidently there is an apparent enhancement of intramolecular hydrogen bond energy in ortho-chlorophenol that is not reflected in these simple model systems.

It is clear from this comparison of the intra- and intermolecular hydrogen bonding capabilities of the halogens, however, that the mesomeric, inductive, and intrinsic hydrogen bond properties of the halogens (possible explanations 4 and 5) are not the major reason for the fact that the intramolecular hydrogen bond in ortho-chlorophenol is stronger than the corresponding bond in ortho-fluorophenol.

As an additional model system for the intramolecular hydrogen bonding of the cis ortho-halophenols (4-5), we conducted ab initio examinations of the intermolecular hydrogen bonding of the H-X - - - H-O-H dimers (4-11: X = F or Cl). The results of these studies are presented in Table 4-11. A "linear" dimer was assumed with the two monomer units



4-11

lying in one plane. For the initial energy-minimization, the X, O, and H atoms involved in the hydrogen bond were assumed to be colinear[147,175] (4-11: $\theta' = 0°$). Geometry searches were conducted simultaneously for both R(X - - O) and θ (see 4-11). The θ value obtained was used in all subsequent calculations. As expected, the dimer hydrogen bond was greater for HF (4.79 kcal/mole) than for HCl (1.19 kcal/mole). Hydrogen bond energies were then calculated for the dimers upon changing either the minimum energy R(X - - O) distance to the R(X - - O) distance in the corresponding ortho-X-phenol and/or the minimum energy θ' angle (0°) to the θ_{HOX} angle of the corresponding cis ortho-X-phenol. R(F - - O) for the minimum energy H-F - - - H-O-H dimer is only slightly less than R(F - - O) for ortho-fluorophenol. Hence, the dimer energy is only slightly decreased upon changing the R(F - - O) distance of the minimum energy dimer to that of cis ortho-fluorophenol. Variation of θ' from 0° to θ_{HOX} for cis ortho-fluorophenol, however, results in over 80% loss of hydrogen bond strength. Simultaneous variation of R(F - - O) and

Table 4-11. Geometries and Ab Initio Hydrogen Bond Energies (ΔE) of H-X - - - H_2O Dimers (4-11).[a]

X	R(X - - O) (Å)	θ[b] (degrees)	θ'[b] (degrees)	ΔE (kcal/mole)
F[c]	2.65[e]	70[e]	0[e]	4.79
	2.65	70	50.54[g]	0.91
	2.75[f]	70	0	4.54
	2.75[f]	70	50.54[g]	1.01
Cl[d]	3.60[e]	77[e]	0[e]	1.19
	3.60	77	44.26[g]	0.66
	2.94[f]	77	0	-4.51
	2.94[f]	77	44.26[g]	-1.81

[a] H_2O experimental geometry: see note 180.

[b] See 4-11.

[c] HF experimental geometry: R(H-F) = 0.9170 Å: from reference 131.

[d] HCl experimental geometry: R(H-Cl) = 1.2745 Å: from reference 131.

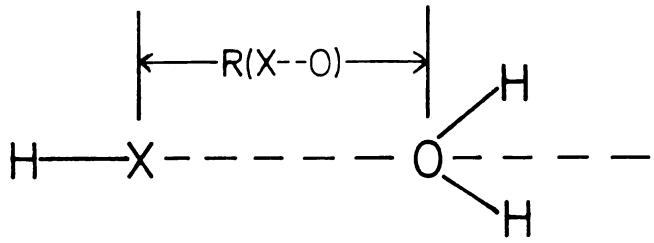
[e] Minimum energy dimer geometry; θ was calculated to 1° and R(X - - O) was calculated to 0.01 Å.

[f] Equal to R(X - - O) for ortho-X-phenol: see Table 4-8.

[g] Equal to θ_{HOX} for cis ortho-X-phenol: see Table 4-8.

θ' to the cis ortho-fluorophenol values results in a large hydrogen bond energy bond energy loss dominated by the θ' change but slightly compensated for by the increase in R(F - - O). The situation is quite different, however, for the H-Cl - - - H-O-H dimer. Changing θ' from 0° to the cis ortho-chlorophenol θ_{HOX} value results in loss of 50% of the hydrogen bond strength, much less than for the fluorine case. R(Cl - - O) in the H-Cl - - - H-O-H minimum energy dimer is much greater than R(Cl - - O) for the ortho-chlorophenol. Changing R(Cl - - O) to the ortho-chlorophenol value, therefore, causes a large Cl - - H repulsion that results in large net dimer repulsion. Simultaneously changing θ' and R(Y - - C) to the corresponding cis ortho-chlorophenol values results in a net dimer repulsion dominated by the Cl - - H repulsion but significantly compensated for in part by allowing the O-H bond to move off te Cl - - O axis. Although this is a very simplified model system for the intramolecular hydrogen bonding of the ortho-fluoro- and ortho-chloro-phenols, these results suggest that the deviations of the intramolecular hydrogen bond strengths from values which should be intrinsically possible may be due primarily to deviation of θ_{HOX} from 0° for ortho-fluorophenol and to H - - Cl repulsion (due to a small R(Cl - - O)) in ortho-chlorophenol. In the latter case, deviation of θ_{HOX} to larger angles actually might relieve this H - - Cl repulsion. On this basis one might intuitively predict that θ_{COH} for the cis ortho-halophenols should increase F < Cl < Br < I, although experimental evidence is not available to test this hypothesis. As a corresponding model for the X - - O repulsions in the trans ortho-halophenols (4-6), we calculated the repulsion energies for the H-X - - - OH_2 dimers (4-12: X = F or Cl: all atoms coplanar: H-X bond bisecting

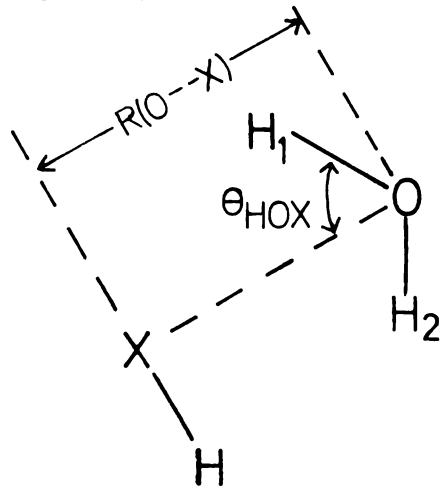
H-O-H angle: R(X - - O) = R(X - - O) for the ortho-X-phenols). The ab initio calculations surprisingly predict the repulsive energies for



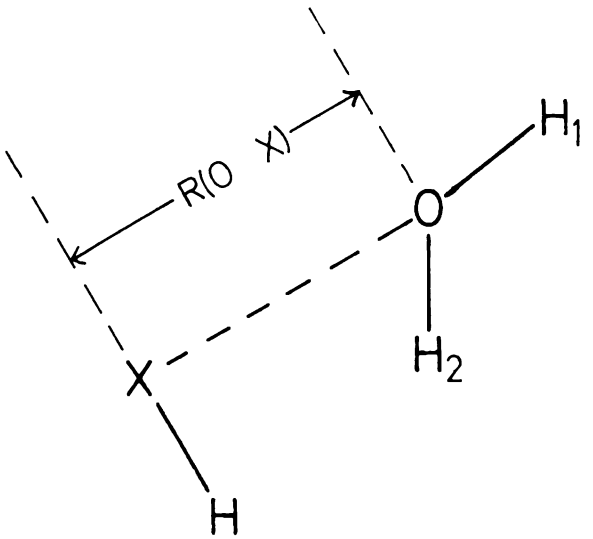
4-12

the two dimers to be essentially the same: 1.54 kcal/mole for HF - - OH_2 and 1.53 kcal/mole for HCl - - - OH_2. This simple model system, however, is unable to reflect any influence that F and Cl might have on the X - - O repulsion by inductive and resonance electronic effects.

We examined one further set of geometries for the HX/H_2O dimers in order to attempt to approximate the ortho-X-phenol geometries more exactly. The geometries of the HX/H_2O dimers (4-13 and 4-14) were chosen such that R(X - - O) = R(X - - O) for the corresponding ortho-X-phenol (Table 4-8), the X-H, $O-H_1$, and $O-H_2$ bonds have the same vectorial orientations as the F-C, O-H and O-C bonds for the corresponding ortho-X-phenol, and the H-O - - - X angle of 4-13 =



4-13



4-14

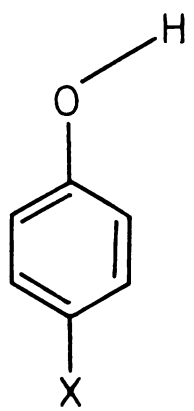
θ_{HOX} for the corresponding cis ortho-X-phenol (4-5) (Table 4-8). (4-13 and 4-14 reflect the geometries of the corresponding cis and trans conformers, respectively, of the corresponding ortho-X-phenols (4-5 and 4-6) and should reflect more accurately (than 4-11 and 4-12) the spatial distributions of electron densities of the H, X, and O atoms of these phenols. This requires an H_2O θ_{HOX} of 110° and R(O-H) of 0.96 Å.) ΔE (4-13 → 4-14) values can then be used to estimate the intramolecular hydrogen bond strengths of the corresponding ortho-X-phenols. The ab initio calculations gave ΔE(4-13 → 4-14) values of -4.75 kcal/mole and -1.08 kcal/mole for X=F and Cl, respectively. However, the total energies for these structures was ∿ 300 kcal/mole above those for the isolated monomers, so it may be that this difference only reflects a relief of H - - H repulsions in the trans conformation.

We also tried to estimate the physical forces behind these H-bonds by comparing the energies of the para-halophenols with those of the cis and trans conformations of the corresponding ortho-halophenols. These

results are presented in Table 4-12. It appears that one can not use the para-ortho energy comparison as support for the importance of O - - halogen repulsion effects, since the inductive effect of two electronegative groups ortho is more destabilizing for the fluoro (more inductively withdrawing than the chloro). The mesomeric effect of F might also contribute to lowering the ortho-halophenol (cis-trans) energy difference since it makes F more positive. However, this is not the dominant effect since our calculations find that fluorobenzene is still capable of forming stronger intermolecular bonds than chlorobenzene. Additional insight into the nature of intramolecular hydrogen bonding of the ortho-halophenols and unsymmetrical 2,6-dihalophenols comes from examining the atomic populations and bond orders of these compounds (Tables 4-13 and 4-14). In all cases electron density is shifted from the phenolic proton to both the phenolic oxygen and the halogen upon hydrogen bond formation. With the ab initio calculations the majority of the electron density shift is from the phenolic proton to the proton-accepting halogen, with much less of the shift from the phenolic proton to the proton-donating oxygen. These same qualitative charge density shifts have been noted before.[148,154] The H, O, and halogen charge densities, as well as the H - - halogen and O - - halogen bond orders, are affected very little by the substitution of a second halogen ortho to the phenolic OH for both the CNDO/2 and ab initio calculations. This is supported by the fact that the difference between the intramolecular hydrogen bond strengths (CNDO/2 or ab initio) of any two of the ortho-halophenols (Table 4-5) is in each case almost equal to the energy difference between the two conformations of the corresponding unsymmetrical 2,6-dihalophenol (Table 4-7). These observations somewhat surprisingly

Table 4-12. CNDO/2 and Ab Initio Relative Energies of ortho-Halophenols (4-5 and 4-6) and para-Halophenols (4-15).

X	ΔE_{calcd} (4-5 → 4-15) (kcal/mole)		ΔE_{calcd} (4-6 → 4-15) (kcal/mole)	
	CNDO/2	Ab Initio	CNDO/2	Ab Initio
F	-0.17	-0.51	-1.54	-2.19
Cl	2.67	0.32	0.37	-1.45
Br	2.31		0.63	
I	1.09		0.33	



4-15

suggest that the two halogen substituents ortho to a phenolic OH interact essentially independently with the OH group.

In summary, it seems that a combination of explanations 3 and 6 is the major cause of this "anomalous" hydrogen bond order in the ortho-halophenols. The fact that ortho-fluorophenol is further from an optimal hydrogen bond geometry than ortho-chlorophenol makes the hydrogen bond in the F-compound weaker than one might expect. However, there also appear to be significant repulsions in the trans conformation (4-6) of ortho-chlorophenol which make its $\Delta E(cis \rightarrow trans)$ unusually large when compared to the ΔE for forming an intermolecular Cl - - - H-O hydrogen bond.

Intramolecular Hydrogen Bonding in Other ortho-Substituted Phenols.

We next chose to examine the intramolecular hydrogen bonds in other ortho-substituted phenols in order to compare their properties with those of the ortho-halophenols. While CNDO/2 calculations provide reasonable CH_3 rotational barriers for both the cis and trans conformers of ortho-CH_3-phenol (Table 4-15), they predict that the most stable cis conformer is 0.84 kcal/mole more stable than the most stable trans conformer (each with the CH_3 group staggered with respect to the OH group). This is in contrast to the repulsive interaction that one might expect to exist between the CH_3 and OH groups in the cis conformer. Experimental evidence (Table 4-16) confirms the existence of this repulsion in that it shows that the trans conformer is slightly more stable than the cis for ortho-methyl-phenol. Also, $\Delta E(cis \rightarrow trans)$ for ortho-tert-butyl phenol increases to a sligthly less negative value upon substitution of a CH_3 group in the other ortho position.

Table 4-13. CNDO/2 and Ab Initio Atomic Populations of Phenol, ortho-Halophenols, and Unsymmetrical 2,6-Dihalophenols (4-1 and 4-2).

		Total Atomic Populations							
		4-1				4-2			
X	Y	H	O	X	Y	H	O	X	Y
H	H	0.885	8.250	---	---	0.885	8.250	---	---
F	H	0.844	8.242	9.202	---	0.853	8.235	9.196	---
Cl	H	0.847	8.255	17.086	---	0.854	8.245	17.082	---
Br	H	0.846	8.259	35.058	---	0.855	0.248	35.053	---
I	H	0.843	8.259	53.081	---	0.854	8.245	53.074	---
F	Cl	0.843	8.237	9.199	17.072	0.844	8.240	9.194	17.076
F	Br	0.844	8.240	9.200	35.042	0.844	8.245	9.194	35.046
F	I	0.843	8.238	9.199	53.060	0.841	8.244	9.193	53.066
Cl	Br	0.846	8.253	17.083	35.048	0.845	8.255	17.078	35.053
Cl	I	0.845	8.251	17.081	53.068	0.842	8.254	17.076	53.075
Br	I	0.845	8.255	35.052	53.069	0.842	8.256	35.047	53.077
H	H	0.783	8.300	---	---	0.783	8.300	---	---
F	H	0.778	8.299	9.143	---	0.780	8.297	9.130	---
Cl	H	0.776	8.300	17.134	---	0.776	8.290	17.117	---
F	Cl	0.770	8.288	9.135	17.109	0.763	8.297	9.122	17.126

CNDO/2

Ab Initio

Table 4-14. CNDO/2 and Ab Initio Bond Orders of Phenol, ortho-Halophenols, and Unsymmetrical 2,6-Dihalophenols (4-1 and 4-2).

		Bond Orders							
		4-1				4-2			
X	Y	H- - X	O- - X	O- - Y	O-H	H- - Y	O- - X	O- - Y	O-H
H	H	----	----	----	0.5371	----	----	----	0.5371
F	H	0.0021	-0.0006	----	0.5367	----	-0.0003	----	0.5371
Cl	H	0.0175	-0.0050	----	0.5343	----	-0.0013	----	0.5371
Br	H	0.0237	-0.0065	----	0.5335	----	-0.0014	----	0.5369
I	H	0.0247	-0.0065	----	0.5337	----	-0.0013	----	0.5368
F	Cl	0.0021	-0.0006	-0.0013	----	0.0175	-0.0003	-0.0051	----
F	Br	0.0021	-0.0006	-0.0014	----	0.0237	-0.0003	-0.0066	----
F	I	0.0021	-0.0006	-0.0014	----	0.0247	-0.0003	-0.0066	----
Cl	Br	0.0173	-0.0049	-0.0014	----	0.0235	-0.0013	-0.0065	----
Cl	I	0.0173	-0.0050	-0.0014	----	0.0245	-0.0014	-0.0065	----
Br	I	0.0234	-0.0065	-0.0014	----	0.0244	-0.0015	-0.0064	----
H	H	----	----	----	0.5375	----	----	----	0.5375
F	H	0.0046	-0.0010	----	0.5356	----	0.0000	----	0.5390
Cl	H	0.0076	-0.0077	----	0.5404	----	0.0000	----	0.5395
F	Cl	0.0045	-0.0010	-0.0006	0.5373	0.0077	0.0000	-0.0078	0.5416

CNDO/2

Ab Initio

Our ab initio results (Table 4-15) agree with the experimental data; with the CH_3 group staggered with respect to the OH group, the trans conformer is found to be more stable by 1.53 kcal/mole. The spurious attractive interaction between OH and CH_3 in the CNDO/2 calculations (apparently an artifact of the approximations of the method) can be seen from the decrease in the CNDO/2 intramolecular hydrogen bond strengths of ortho-chloro- and ortho-iodo-phenol (Table 4-5) upon addition of a CH_3 group in the other ortho position (Table 4-15). After a complete CNDO/2 geometry search of ortho-methylphenol (15° variations in the CH_3 rotation and 30° variations in the OH rotation), it was found (assuming a Boltzmann distribution between all conformers) that the net energy of the cis conformers) that the net energy of the cis conformers (4-5: $X=CH_3$: $-90° < \phi_{1234} < 90°$) was still 0.57 kcal/mole less than the net energy of the trans conformers (4-5: $X=CH_3$ $90° < \phi_{1234} < 270°$).

CNDO/2 calculations on ortho-isopropylphenol and ortho-tert-butylphenol (Table 4-15) gave more reasonable results than were obtained for ortho-methylphenol, especially with respect to phenolic OH/ortho-alkyl repulsive interactions. The most stable trans OH conformer of the ortho-isopropylphenol was found to be 1.56 kcal/mole more stable than the most stable cis OH conformer; for ortho-tert-butylphenol, this energy difference is 4.48 kcal/mole. Assuming a Boltzmann distribution between the various cis and trans isopropyl rotamers of Table 4-15, it was found that the net energy of the trans-OH isopropyl rotamers is 1.23 kcal/mole less than the net energy of the cis-OH isopropyl rotamers. The corresponding energy difference in the t-butyl case was 4.31 kcal/mole. Thus, even though CNDO/2 underestimates repulsions, it still gives the correct sign for $\Delta E(\text{cis} \rightarrow \text{trans})$ for ortho-tert-butyl- and ortho-isopropyl-phenol.

Table 4-15. CNDO/2 and Ab Initio Conformational Dependence of Energies of ortho-Alkylphenols (4-3) (Relative Energies in kcal/mole).

R_6	R_7	R_8	Y	ϕ_{2156}[a]	CNDO/2 ϕ_{1234}[a]		Ab Initio ϕ_{1234}[a]	
					0	180	0	180
H	H	H	H	0	1.92	0.95	---	---
				30	0.72	0.91	---	---
				60	0.00	0.84	1.53	0.00
H	CH_3[b]	CH_3[b]	H	0	1.64	0.60	---	---
				30	1.56	1.69	---	---
				60	3.80	2.10	---	---
				90	81.2	4.81	---	---
				120	51.5	3.42	---	---
				150	79.2	2.47	---	---
				180	4.52	0.00	---	---
CH_3[b]	CH_3[b]	CH_3[b]	H	0	49.8	1.65	---	---
				30	79.0	2.39	---	---
				60	4.48	0.00	---	---
H	H	H	Cl	60	1.48	0.00	---	---
H	H	H	I	60	0.00	0.04	---	---
F	F	F	H	0	2.97	2.50	3.09	---
				25	---	---	3.02	---
				30	0.02	2.53	0.04	---
				33	0.00	---	---	---
				35	0.02	---	0.00	---
				40	---	---	0.15	---
				45	0.32	---	---	---
				60	0.64	2.50	0.84	0.12

[a] In degrees.

[b] CH_3 protons in staggered conformation.

Table 4-16. CNDO/2 , Ab Initio, and Experimental Intramolecular Hydrogen Bond Strengths of ortho-Substituted Phenols (4-1 and 4-2).

		ΔE(4-1 → 4-2) (kcal/mole)		
X	Y	CNDO/2	Ab Initio	Exptl.
CH_3	H	0.84[f]	-1.53	-0.86[g]
				-0.29[h,i]
				-0.51[j,k]
iPr	H	-1.56[f]	---	
tBu	H	-4.48[f]	---	-1.38[h,i]
				-1.57[j,k]
				-1.38[l]
				-3.04[g]
tBu	CH_3	---	---	-1.05[h,i]
				-1.06[j,k]
				-2.22[g]
CF_3	H	2.50[f]	0.12[f]	>0 and <∿ 2.5[j,m]
NO_2[a]	H	8.29	---	6.65°
				2.1[p]
				4.7[q]
OH[b]	H	1.37	3.27	2.29[h,r]
OCH_3[c]	H	1.32	---	2.00[h,i]
C_6H_5[d]	H	1.66	---	2.73[h,i]
				1.45[s,j]
CHO[e]	H	6.02	7.44	7.09°
				1.8[p]
				3.6[q]
CN	H	2.01	---	1.73[h,i]

[a]Structure 4-17.

[b]Structure 4-16.

[c]Structure 4-18: CH_3 group staggered.

[d]Structure 4-19: $\phi_{2156} = 90°$

[e]Structure 4-20.

[f]See Table 4-15.

[g]Reference 185: Method of study = molecular mechanics force field calculation

Table 4-16. (Continued)

[h]Method of study = IR OH torsional frequency: Cyclohexane solution.

[i]Reference 157.

[j]Method of Study = IR OH stretching frequency: CCl_4 solution.

[k]Reference 164.

[l]Reference 185: Method of study = dipole moment: CCl_4 solution

[m]Estimated from reference 161.

[n]Reference 186.

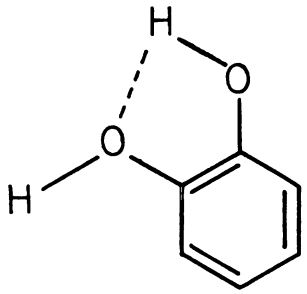
[o]Reference 169: Method of study = OH ^1H chemical shift: CCl_4 solution.

[p]Reference 149: Method of study = EHT.

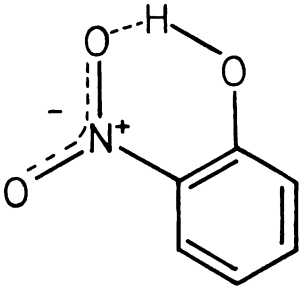
[q]Reference 149: Method of study = CNDO/2

[r]Estimated from references 169 and 176.

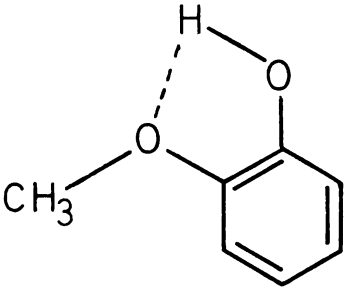
[s]Reference 187.



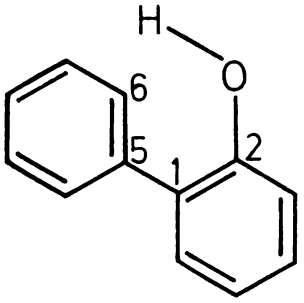
4-16



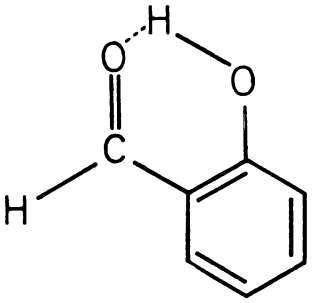
4-17



4-18



4-19



4-20

That the alkyl and phenolic OH groups of the ortho-alkylphenols interact in a repulsive and not an attractive manner is emphasized by the $\Delta\nu_{OH}$ values of various ortho-substituted phenols (Table 4-17). Ortho-substituents that are capable of forming intramolecular hydrogen bonds with the phenolic OH cause ν_{OH} of the cis conformer to shift to lower frequencies, ν_{OH} for the trans conformer being relatively unaffected. For ortho-alkyl substituents which have repulsive interactions with the phenolic OH, ν_{OH} for the cis conformer is shifted to higher frequencies, ν_{OH} for the trans conformer being relatively unaffected. An intramolecular hydrogen bond should lengthen the O-H bond, decreasing the O-H bond energy and consequently ν_{OH}. Conversely, it has been suggested[164] that steric interactions between an ortho-alkyl substituent and the phenolic OH narrow the potential energy well of the O-H stretching mode by repelling the phenolic proton, cause the O-H bond to shorten, and increase ν_{OH}.

Ortho-CF_3-phenol is an unusual case in which there are apparently both attractive and repulsive interactions between the CF_3 and phenolic OH groups. Ortho-CF_3-phenol displays two ν_{OH} bands[161] (Table 4-17): a more intense band (3624.6 cm^{-1}) shifted to higher frequency from ν_{OH} for phenol (and apparently corresponding to the cis conformer) and a less intense band (3605 cm^{-1}) at about ν_{OH} for phenol (and apparently corresponding to the trans conformer). These assignments are confirmed by the two ν_{OH} bands for 2-Br-6-CF_3-phenol[161] (Table 4-17): one at a ν_{OH} (3510.4 cm^{-1}) about equal to ν_{OH} for the cis conformer of ortho-bromophenol (Table 4-17) and one of less intensity (3616.9 cm^{-1}) at about the ν_{OH} assigned to the cis conformer for ortho-CF_3-phenol. Konovalov, et al.,[186] also observed a ν_{OH} doublet at 3605 and 3626 cm^{-1} for ortho-CF_3-phenol

Table 4-17. Experimental ν_{OH} and $\Delta\nu_{OH}$ Values[a] for ortho-Substituted Phenols (4-1 and 4-2).

		$\nu_{OH}(cm^{-1})$			
X	Y	4-1	4-2	$\Delta\nu_{OH}(cm^{-1})$[b]	Ref.
H	H	3610.5	---	---	163
F	H	3591	c	c	163
Cl	H	3545	3608	63	163
Br	H	3522	3604	82	163
I	H	3499	3600	101	163
CH_3	H	3614	c	c	157
tBu	H	3647	3607	-40	157
tBu	CH_3	3649	3610	-39	157
iPr	H	3614	c	c	164
C_6H_5	H	3564.9	3606.9	40	185
CF_3	H	3624.6	3605	-19.6	161
CH_3	Br	3616.9	3510.4	-106.5	161

[a] All determined in CCl_4.

[b] $\Delta\nu = \nu_{OH}$ (4-2) - ν_{OH} (4-1).

[c] ν_{OH} (4-1) $\approx$ ν_{OH}(4-2).

and assigned the higher frequency to the cis conformer. They also found that with increasing temperature the intensity of the 3605 cm^{-1} band increased while that of the 3625 cm^{-1} band decreased. This study gave a ΔH(cis → trans) value of 0.9 kcal/mole. In contrast to the above studies,[161,186] Marler and Hopkins[188] assigned the less intense 3606 cm^{-1} ν_{OH} and the more intense 3624 cm^{-1} ν_{OH} of ortho-CF_3-phenol to the cis and trans conformers, respectively. In addition, they found that the ratio of the integrated intensities of the 3624 cm^{-1} band to the 3606 cm^{-1} band <u>increased</u> with increasing temperature. Their data yields values of ΔH (cis → trans) = 1.4 kcal/mole and ΔS(cis → trans) = ∿ 6 cal./deg/mole. This rather large ΔS for intramolecular hydrogen bond formation is inconsistent with experimental ΔS values that are essentially zero for other ortho-substituted phenols[143-145] and with theoretical considerations.[146] We support the assignment of the higher ν_{OH} frequency to the cis conformer based on our calculated ΔE(cis → trans) values (see below), ΔS(cis → trans) considerations,and the more abundant (although still scant) experimental evidence supporting this assignment. Further experimental studies on ortho-CF_3-substituted phenols are certainly indicated for the resolution of the previous experimental ambiguities.

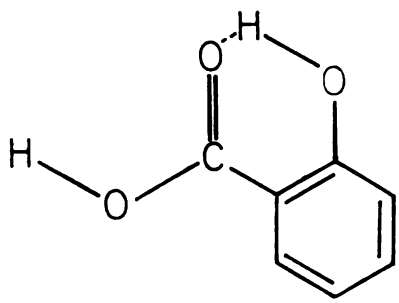
That ortho-CF_3-phenol does form intramolecular hydrogen bonds is also supported by ^{19}F NMR solvent shift and dilution studies.[168] Hence, like the ortho-alkylphenols there is apparently a repulsive steric interaction between the CF_3 and the phenolic OH, causing ν_{OH} for the cis conformer to be shifted to a higher frequency. And like the ortho-halophenols there is also an attractive hydrogen bond interaction between the CF_3 and the phenolic OH, causing the cis conformer to be more

stable relative to the trans conformer. The results of our CNDO/2 and ab initio calculations on ortho-CF_3-phenol are presented in Table 4-15. For the cis conformer, both methods of calculation predict an identical energy minimum (with one of the F cis to the OH and rotated 33° up from the ring plane), as well as nearly identical CF_3 rotational potentials. For the trans conformer, the CNDO/2 calculations predict a shallow rotational potential about 2.50 kcal/mole less stable than the cis conformer energy minimum, whereas the ab initio calculations predict the trans conformer to be only 0.12 kcal/mole less stable than the cis conformer energy minimum. The fact that the ν_{OH} intensity for the ortho-CF_3-phenol cis conformer is "several times" that for the trans conformer and the fact that the ν_{OH} intensity for the 2-Br-6-CF_3-phenol conformer with the OH cis to the Br is greater than that of the conformer with the OH cis to the CF_3[161] suggests that the actual intramolecular hydrogen bond strength of ortho-CF_3-phenol lies somewhere between our CNDO/2 value of about 2.50 kcal/mole and our ab initio value of 0.12 kcal/mole. This is supported by the $\Delta H(cis \rightarrow trans)$ value of 0.9 kcal/mole of Konavalov, et al.,[186] for ortho-CF_3-phenol.

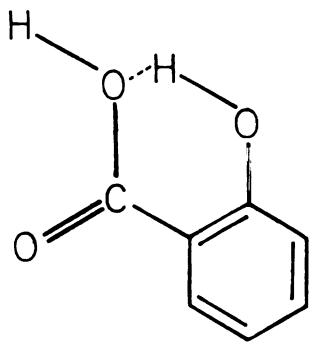
We also carried out calculations on the cis-trans isomerism of ortho-substituted phenols for a selection of ortho-substituents capable of forming intramolecular hydrogen bonds with the phenolic OH. The CNDO/2 and ab initio results are presented in Table 4-16. The CNDO/2 results are in fairly good agreement with the experimental data for the chelated ortho-NO_2- and ortho-CHO-phenols, for the weak intramolecular hydrogen bonding of the ortho-OH- and ortho-OCH_3-phenols, and for the weak O-H - - - π intramolecular interactions of the ortho-CN- and ortho-C_6H_5-phenols. The ab initio hydrogen bond strengths for the ortho-OH- and ortho-CHO-phenols are in slightly better agreement with the experi-

mental data than the CNDO/2 results.

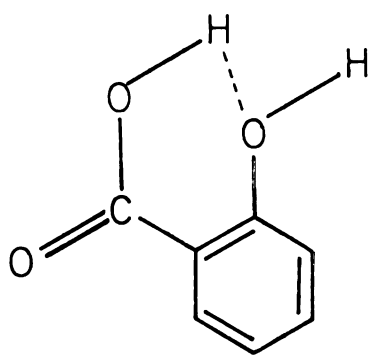
A CNDO/2 study was also conducted on the intramolecular hydrogen bonding of salicylic acid with the relative energies of conformers 4-21, 4-22, and 4-23 predicted as 0.00, 1.38, and 4.62 kcal/mole, respectively. That 4-21 is actually the intramolecularly hydrogen bonded conformer that predominates is supported by the IR studies of Mori, et al.[189]



4-21



4-22



4-23

Thus, both CNDO/2 and minimal basis ab initio methods are capable of qualitatively reproducing almost all of the experimental data (except for the ortho-CH_3 with CNDO/2) for intramolecular hydrogen bonding of

ortho-substituted phenols. The ab initio calculations yield semi-quantitative agreement with experiment in the molecules studied and rank correctly the intramolecular "hydrogen bond" strengths in the series $CHO > OH > Cl \gtrsim F > CF_3 > H > CH_3$.

Intramolecular Hydrogen Bonding in ortho-Substituted Thiophenols.

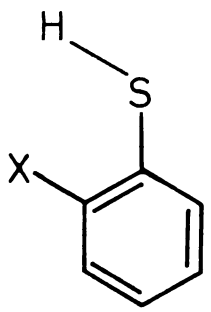
Having examined the ability of CNDO/2 and ab initio calculations to predict the interactions of the phenolic OH with the ortho-substituents of ortho-substituted phenols, we next examined the abilities of the two computational methods to predict the intramolecular hydrogen bonding capabilities of ortho-substituted thiophenols. A CNDO/2 geometry search for a minimum energy θ_{CSH} in thiophenol yielded a value of 98°, which was used in all subsequent CNDO/2 and ab initio calculations on the thiophenols. As a parallel to the phenol studies, the intramolecular hydrogen bonding of the ortho-halothiophenols was examined first and the results are presented in Table 4-18. The CNDO/2 and ab initio calculations are in agreement with ortho-fluorothiophenol forming an intramolecular hydrogen bond about half the strength of the intramolecular hydrogen bond of ortho-fluorophenol. For the ortho-halothiophenols, however, the CNDO/2 calculations predict intramolecular hydrogen bond strengths about 150% of those for the corresponding ortho-halophenols. These CNDO/2 results for ortho-chloro-, ortho-bromo-, and ortho-iodo-thiophenol are inconsistent with one's intuition based on pKa values[190] that SH is a poorer proton donor than OH. In contrast, the ab initio calculations predict the trans conformer of ortho-chlorothiophenol to be 2.79 kcal/mole more stable than the cis conformer. Experimentally two ν_{SH} bands are actually observed[191,192] for ortho-chloro- and

Table 4-18. CNDO/2 and Ab Initio Energy Calculations on Thiophenol and ortho-Halothiophenols (4-24, 4-25, and 4-26).

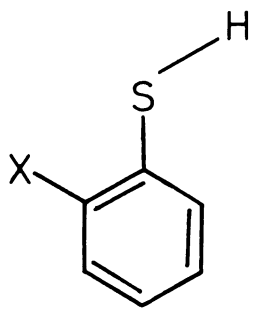
	$V(90°)^a$ (kcal/mole)	V_1 = ΔE(4-24 → 4-25) (kcal/mole)		$V_2{}^b$ (kcal/mole)
X	CNDO/2	CNDO/2	Ab Initio	CNDO/2
H	0.35	0.00	0.00	0.35
F	0.89	0.75	0.82	0.51
Cl	4.00	3.53	-2.79	2.23
Br	4.16	3.46	---	2.43
I	2.28	1.38	---	1.59

[a] V(90°) = ΔE(4-25 → 4-26).

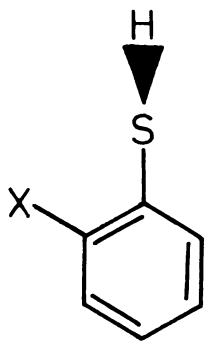
[b] Calculated from V(90°), V_1, and Eqn. 4-3.



4-24



4-25



4-26

ortho-bromo-thiophenol. The ν_{SH} band for the cis conformer was found[192] to represent no more than about 20% of the population for the ortho-chloro- and ortho-bromo-thiophenols, in agreement with the ab initio but not the CNDO/2 calculations. Some insight into the source of this discrepancy between the CNDO/2 and ab initio calculations is provided by the bond orders, atomic populations, and geometrical parameters found in these calculations (Tables 4-19, 4-20, and 4-21). Upon the trans to cis conformational transition, there is a rise in the S - - X repulsive bond order (especially for the ab initio calculations). For the cis conformer there is a positive H - - X attractive bond order in the CNDO/2 calculations on all four ortho-halothiophenols and in the ab initio calculations on ortho-fluorothiophenol. Rather dramatically, however, a large repulsive ab initio H - - X bond order occurs for the cis ortho-chlorothiophenol conformer. As the halogen size increases R(H - - X) increases very little for the ortho-halothiophenols while the sum of the Van der Waals radii for H + X increases significantly. The amount of H - - X overlap of the Van der Waals radii is about the same for the ortho-halophenols and ortho-halothiophenols for each X. Just as R(O - - X) increases more slowly for the ortho-halophenols than the sum of the Van der Waals radii for O + X as the halogen size increases, so also R(S - - X) increases more slowly for the ortho-halothiophenols than the sum of the Van der Waals radii for S + X. The amount of S - - X overlap repulsion (i.e., the amount R(S - - X) is less than the sum of Van der Waals radii of S and X) for the ortho-halothiophenols is significantly greater, however, than the amount of O - - X repulsion for the ortho-halophenols for each X. The Cl > Br > I> F CNDO/2 attractive

Table 4-19. Geometrial Parameters of the cis ortho-Halothiophenols (4-24).

X	$\theta_{HSX_{calcd}}$[a] (degrees)	$R(H- - X)_{calcd}$[a] (Å)	$R(S- - X)_{calcd}$[a] (Å)	ΣVan der Waals radii[b] (Å)	
				H + S	S + X
F	45.05	2.23	2.96	2.67	3.27
Cl	38.79	2.25	3.13	2.97	3.57
Br	36.45	2.27	3.20	3.08	3.68
I	33.50	2.32	3.31	3.27	3.87

[a]Calculated based on the geometries used in our calculations.
[b]See footnote c, Table 4-8.

Table 4-20. CNDO/2 and Ab Initio Bond Orders for the ortho-Halothiophenols (4-24 and 4-25).

		Bond Order		
	4-24		4-25	
X	H- - X	S- - X	S- - X	
F	0.0023	-0.0010	-0.0005	CNDO/2
Cl	0.0285	-0.0086	-0.0015	CNDO/2
Br	0.0423	-0.0121	-0.0015	CNDO/2
I	0.0556	-0.0151	-0.0015	CNDO/2
F	0.0023	-0.0030	0.0000	Ab Initio
Cl	-0.0075	-0.0222	-0.0043	Ab Initio

Table 4-21. CNDO/2 and Ab Initio Atomic Populations for Thiophenol and the ortho-Halothiophenols (4-24 and 4-25).

		Total Atomic Populations					
	4-24			4-25			
X	H	S	X	H	S	X	
H	---	---	---	0.988	16.100	---	CNDO/2
F	0.962	16.091	9.200	0.981	16.077	9.198	
Cl	0.958	16.117	17.071	0.986	16.093	17.080	
Br	0.953	16.128	35.037	0.988	16.098	35.051	
I	0.946	16.133	53.040	0.989	16.097	53.056	
H	---	---	---	1.029	15.854	---	Ab Initio
F	1.017	15.855	9.141	1.027	15.845	9.136	
Cl	0.994	15.862	17.130	1.026	15.829	17.126	

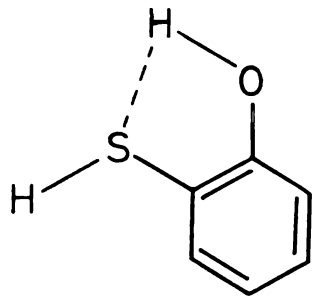
intramolecular hydrogen bond strengths for the ortho-halothiophenols reflect these trends, but apparently, as with the O - - X repulsions in the phenols, the CNDO/2 calculations tend to poorly represent the S - - X repulsions. The ab initio attractive intramolecular hydrogen bond strength for ortho-fluorothiophenol and repulsive intramolecular interactions for ortho-chlorothiophenol reflect not only these trends but also the ability of the ab initio calculations to correctly represent and weight the H - - X attraction and S - - X repulsion. However, they are not always completely successful in this weighting (see below).

As a direct comparison of the intramolecular hydrogen bonding capabilities of the ortho-substituted phenols and thiophenols, we next looked at ortho-hydroxythiophenol. Based on the intensities of the ν_{SH} and ν_{OH} bands of the various possible conformers (4-27, 4-28, and 4-29) of ortho-hydroxythiophenol, David and Hallam[191] suggest that the conformations 4-27 and 4-29 are present in about equal amounts in dilute CCl_4 solution. As seen from Table 4-22 the CNDO/2 results appear to agree with the experimental results[191] concerning the relative stabilities of the conformers, but the ab initio results do not. While bond orders (Table 4-23) do not directly reflect these differences, they do suggest that the differences in the CNDO/2 and ab initio results are due not so much to their differences in handling the H - - S and H - - O interactions but more to their differences in handling changes in the O - - S repulsive interactions.

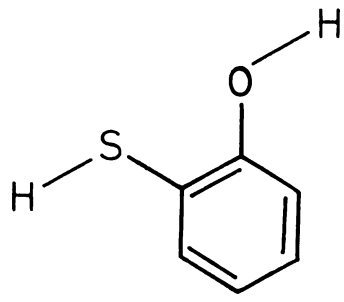
Because of these differences of the CNDO/2 and ab initio calculations in representing the intramolecular attractive and repulsive interactions of the ortho-halo- and ortho-hydroxy-thiophenols, we conducted CNDO/2 and ab initio calculations on the intermolecular hydrogen bonding of the

Table 4-22. CNDO/2 and Ab Initio Relative Energies for the Conformers of ortho-Hydroxythiophenol (4-27, 4-28, and 4-29).

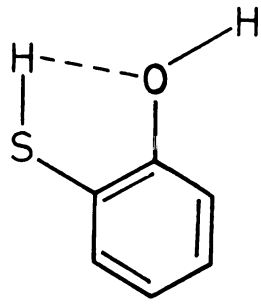
Conformer	Relative Energies (kcal/mole)	
	CNDO/2	Ab Initio
4-27	0.07	2.14
4-28	0.98	1.20
4-29	0.00	0.00



4-27



4-28



4-29

Table 4-23. CNDO/2 and Ab Initio Bond Orders for the Conformers of ortho-Hydroxythiophenol (4-27, 4-28, and 4-29).

	Bond Orders					
	CNDO/2			Ab Initio		
Conformer	H- - O	H- - S	O - -S	H - - O	H - - S	O - - S
4-27	---	0.0057	-0.0016	---	0.0064	-0.0056
4-28	---	---	-0.0006	---	---	-0.0002
4-29	0.0070	---	-0.0014	0.0025	---	-0.0048

Table 4-24. Ab Initio Hydrogen Bond Energies (ΔE)[a] and Geometrical Parameters[b] for H_2O/H_2S Dimers (4-30).

	Proton Donor		
	H_2O	H_2S	
E	2.54	1.86	STO-3G basis set[c]
R	3.33	3.37	
θ	76	46	
E	3.9	3.8	431G basis set[d]
R	3.66	3.59	
θ	78	22	

[a] Energies in kcal/mole.

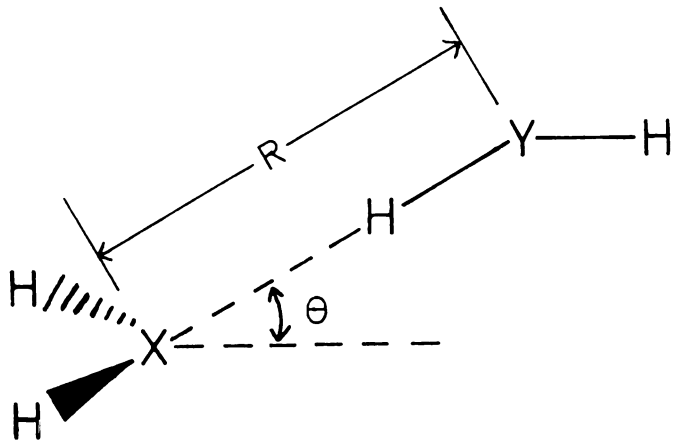
[b] R = O - - S internuclear distance (in Å); θ in degrees: see 4-30.

[c] These results from this study.

[d] These results from reference 194.

H_2O/H_2S, H_2O/phenol, and H_2O/thiophenol dimers. This was done in order to provide some reference points with which to compare the CNDO/2 and ab initio calculations on the intramolecular interactions of the ortho-substituted thiophenols.

For the H_2O/H_2S intermolecular hydrogen bonding, a "linear" dimer (4-30: X=S and Y=O, or vice versa) was assumed, with the two monomer units lying in perpendicular planes with the X, Y, and H atoms involved in the hydrogen bond colinear.[180,193] A geometry search was conducted simultaneously for both R(the O - - S internuclear distance) and θ (see 4-30).

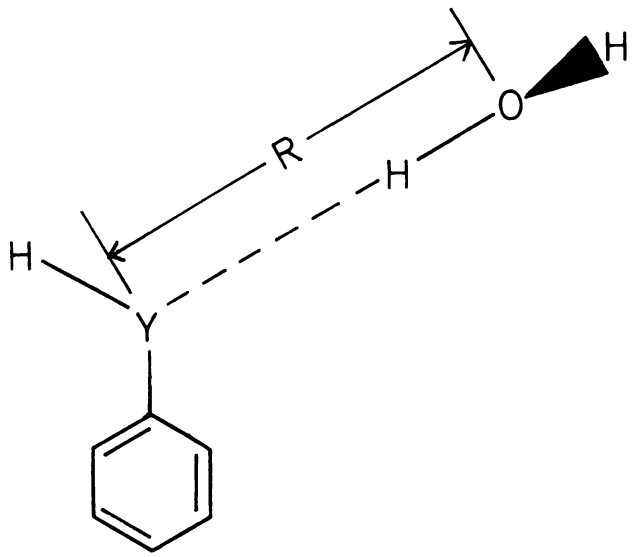


4-30

The results of our ab initio calculations with an STO-3G basis set and a previous ab initio study[194] using a 431G basis set are presented in Table 4-24. As noted before,[194] the STO-3G basis set predicts ΔE values 1-2 kcal/mole less than the 431G basis set ΔE values. Except for slightly shorter R values, the STO-3G geometries are very similar to the 431G geometries for the H_2O/H_2S dimers. As would be expected, ΔE for H_2S as the proton donor is significantly less than for H_2O as the proton donor.

We next examined with CNDO/2 and ab initio calculations the intermolecular hydrogen bonding of the H_2O/C_6H_5SH dimers. For C_6H_5YH

(Y=S or O) as a proton acceptor (4-31), a "linear" dimer was assumed with Y - - H-O lying in the ring plane and on a line bisecting θ_{CYH} of C_6H_5YH. Y - - H-O-H all lie in a plane perpendicular to the ring plane in order to minimze the interactions of the second H_2O proton



4-31

with C_6H_5YH. For C_6H_5YH as the proton donor (4-10: X=H) a "linear" dimer was assumed with the two monomer units lying in perpendicular planes and with the Y, O, and H atoms involved in the hydrogen bond colinear. θ (see 4-10) was taken as before as 57° for Y=O (from the STO-3G H_2O dimer value)[184] and as 46° for Y=S (from the STO-3G H_2S/H_2O dimer value for H_2S as proton donor: Table 4-24). Only R (the Y - - O internuclear distance) was varied in our geometry searches. Several interesting observations can be made from the CNDO/2 and ab initio results, which are presented in Table 4-10. The CNDO/2 calculations on the H_2O/C_6H_5OH dimer with phenol both as the proton donor and as the proton acceptor yield ΔE and R values very close to the CNDO/2 ΔE (5.9 kcal/mole) and R (3.3 Å) values for the H_2O/H_2O dimer (4-30:

X=Y=O with θ = 0° and H_2O experimental geometry).[195,196] Our *ab initio* calculations on the H_2O/C_6H_5OH dimer give a ΔE that is more than twice as large for phenol as the proton donor than for H_2O as the proton donor. This larger ΔE is accompanied by a smaller R value. Conversely, our *ab initio* calculations on the H_2O/C_6H_5SH dimer give a ΔE that is more than twice as small for thiophenol as the proton donor than for H_2O as the proton donor. This smaller ΔE is accompanied by a larger R value. Our *ab initio* calcuations predict C_6H_5SH to be both a poorer proton donor and a poorer proton acceptor than phenol. In each case ΔE and R correspond well with ΔE and R for the corresponding H_2O/H_2S dimer (Table 4-24). While our CNDO/2 calculations give a reasonable approximation for ΔE for thiophenol as a proton acceptor, they grossly overestimate the ΔE for thiophenol as a proton donor. Apparently this same error is reflected in our CNDO/2 calculations predicting very attractive hydrogen bonds for the ortho-halothiophenols (Table 4-18).

That the *ab initio* calculations predict relative instability for conformer 4-27 of ortho-hydroxythiophenol can possibly be rationalized inspecting the geometries of all three ortho-hydroxythiophenol conformers (4-27, 4-28, and 4-29) (Table 4-25). The O - - H Van der Waals radii overlap for 4-29 is significantly less than the rather large S - - H overlap for 4-27. In contrast to CNDO/2 underestimating O - - halogen repulsions in the ortho-halophenols, apparently *ab initio* may overestimate the S - - H repulsions for 4-27. However, the observed relative intensities of the S-H and O-H stretches make a precise estimate of the amount of conformers 4-27-4-29 in o-hydroxythiopenol ambiguous.

Table 4-25. Geometrial Parameters of ortho-Hydroxythiophenol Conformers (4-27, 4-28, and 4-29).

Conformer	$R(O- -S)^{a}_{calcd}$ (Å)	$R(S- -H)^{a}_{calcd}$ (Å)	$R(O- -H)^{a}_{calcd}$ (Å)	Σ Van der Waals radii[b] O+S	S+H	O+H
4-27	2.97	2.37	---	3.32	3.00	---
4-28	2.97	---	---	3.32	---	---
4-29	2.97	---	2.23	3.32	---	2.72

[a] Calculated based on the geometries used in our calculations.

[b] See footnote c, Table 4-8.

Infrared Spectral Properties of ortho-Substituted Phenols.

The ortho-halophenols (except for ortho-fluorophenol) exhibit in "inert" solvents two O-H stretching frequencies, ν_{OH}: one approximately equal to ν_{OH} of phenol and corresponding to the trans non-hydrogen bonded conformer; the other shifted to a lower frequency and corresponding to the cis intramolecularly hydrogen bonded conformer. (Ortho-fluorophenol exhibits only a single (but broad) ν_{OH} because ν_{OH} for the cis conformer $\tilde{\nu}\nu_{OH}$ for the trans conformer). The difference ($\Delta\nu_{OH}$) between the two frequencies is of the order F < Cl < Br < I (Table 4-17). Both the experimental data and our CNDO/2 and ab initio calculations indicate that the order of intramolecular hydrogen bond strengths of the ortho-halophenols is most likely Cl $\tilde{\nu}$ F > Br > I or Cl > Br > F > I (depending on which studies are cited). This is in conflict with the Badger-Bauer rule[197] which states that $\Delta\nu_{OH}$ (the shift to lower frequencies upon hydrogen bond formation) is directly proportional to the hydrogen bond strength. This discrepany has been attributed[160,166,173] to these intramolecular hydrogen bonds being highly bent from an ideal colinear geometry for O-H - - X and to the H - - X distances being fixed by the molecular geometry of the phenols at values not necessarily equal to the preferred interacting distances.[198] It appears[166] that for the ortho-substituted phenols $\Delta\nu_{OH}$ is a measure of the amount of H - - X overlap and not the net energy of the OH and X interactions, which, for example, will include the O - - X repulsion. Both our CNDO/2 and ab initio calculations support this hypothesis. The cis conformer H - - X bond orders (Table 4-14), providing some measure of the H - - X interaction, correlate well with the $\Delta\nu_{OH}$ values

(Table 4-17) but not with the intramolecular hydrogen bond strengths. The experimental $\Delta\nu_{OH}$ shifts to lower frequencies should be paralleled by similar decreases in the ortho-halophenol O-H bond orders upon hydrogen bond formation. The CNDO/2 phenol and trans ortho-halophenol O-H bond orders are all essentially the same, just as the experimental phenol and trans ortho-halophenol $\Delta\nu_{OH}$ values are essentially the same. In addition, the CNDO/2 calculated O-H bond order decreases upon hydrogen bond formation for the ortho-halophenols closely parallel the corresponding experimental $\Delta\nu_{OH}$ values (except for the I O-H bond order which is slightly out of line). While the ab initio calculated O-H bond orders are also fairly constant for phenol and the trans ortho-halophenols, the ab initio O-H bond order decreases of the two ortho-halophenols upon hydrogen bond formation do not correlate with the corresponding experimental ν_{OH} values.

In order to see whether either the CNDO/2 or ab initio calculations could predict the experimental $\Delta\nu_{OH}$ values for the ortho-halophenols, we conducted geometry searches for the minimum energy O-H bond lengths for phenol and the ortho-halophenols. Assuming a harmonic oscillator model for changes in energy with R(O-H) variation near the minimum energy R(O-H), force constants (k) and hence the ν_{OH} values were calculated for the O-H stretch (Table 4-26). The CNDO/2 calculations overestimate the "expected" equilibrium R(O-H), k, and ν_{OH} values. The CNDO/2 ortho-halophenol ν_{OH} values, even though slightly overestimated, are in reasonable agreement with the experimental data both in magnitude and ordering (except for I which is slightly out of line). The CNDO/2 cis ortho-halophenol equilibrium R(O-H) values vary in essentially the same manner as $\Delta\nu_{OH}$ for the halogens. The ab initio calculations give

reasonable estimates for the ortho-halophenol $R(O-H)_{min}$ values and ν_{OH} values that are less overestimated than for the CNDO/2 calculations. However, the ab initio calculations do very poorly in predicting the magnitude of $\Delta\nu_{OH}$ for the ortho-halophenols.

Because of this insensitivity of the ab initio calculations to $\Delta\nu_{OH}$ for the ortho-halophenols, we decided to investigate this area further. Assuming again a harmonic oscillator model for the O-H stretch:

$$1/2\ k \left\langle x^2 \right\rangle_n = \left\langle V \right\rangle_n$$

$$= 1/2\ E_n$$

where: $\left\langle x^2 \right\rangle_n$ = the expectation value of x^2 of the nth O-H stretching energy level

$x = \left| R(O-H) - R(O-H)_{min} \right|$

$\left\langle V \right\rangle_n$ = expectation value for V (the potential energy of the O-H bond) in the nth O-H stretching energy level

E_n = energy of the nth O-H stretching energy level

$= (n + 1/2)\ h\nu_{OH}$

Then: $\left\langle x^2 \right\rangle_n = E_n/k$ (Eqn. 4-1)

Assuming $\nu_{OH} \tilde{\approx} 3600\ cm^{-1}$ gives:

$k = 7.65 \times 10^5\ ergs/cm^2$

$E_o = 1/2\ h\nu_{OH} = 1800\ cm^{-1}$

$E_1 = 3/2\ h\nu_{OH} = 5400\ cm^{-1}$

Eqn. 4-1 then gives

$$\left\langle x^2 \right\rangle_o^{1/2} = 0.068\ \text{Å}$$

$$\left\langle x^2 \right\rangle_1^{1/2} = 0.118\ \text{Å}$$

Table 4-26. CNDO/2 and Ab Initio O-H Stretching Minimum Energy Bond Lengths, Force Constants, Frequencies, and Frequency Shifts for Phenol and Various ortho-Substituted Phenols (4-5).

	$R(O-H)_{min}$ (Å)		$k(\times 10^6$ ergs/cm$^2)$		$\nu_{OH}{}^c(cm^{-1})$		$\Delta\nu_{OH}{}^c(cm^{-1})$	
X	CNDO/2[a]	Ab Initio[b]	CNDO/2[a]	Ab Initio	CNDO/2	Ab Initio	CNDO/2	Ab Initio
H	1.0323	0.985	1.6662	0.9992	5318	4118	---	---
F	1.0328	0.987	1.6524	1.0003	5296	4121	22	-3
Cl	1.0378	0.985	1.5841	0.9996	5185	4119	133	-1
Br	1.0387	---	1.5664	---	5156	---	162	---
I	1.0374	---	1.5881	---	5192	---	126	---
CH_3[d]	1.0347	---	1.6224	---	5248	---	70	---
CHO[e]	1.048	0.990	1.4234	0.9541	4915	4024	403	94
OH[f]	1.033	0.986	1.6406	0.9977	5277	4115	41	3
CF_3[g]	1.0381	---	1.6043	---	5218	---	100	---
CF_3[h]	---	0.981	---	0.9984	---	4117	---	1

[a]Calculated using least squares quadratic fit using 7 to 14 points evenly spaced from R(O-H) = 1.01 Å to 1.06Å.

[b]Calculated using a three point quadratic fit (R(O-H) = 0.98, 0.99 and 1.00 Å). Least squares quadratic fits to 5 points at 0.01 Å intervals of R(O-H) from 0.96 to 1.01 Å gave poor results due to anharmonicity over this R(O-H) range.

Table 4-26. (Continued)

[c] ν_{OH} (phenol was assumed to be equal to ν_{OH} for the trans conformers (4-6) of each of the ortho-halophenols. The ν_{OH} values in this table therefore refer to the ortho-halophenol cis conformers. Also, $\Delta\nu_{OH} = \nu_{OH}$ (trans) - ν_{OH} (cis) $\approx\nu_{OH}$ (phenol) - ν_{OH} (cis).

[d] Structure 4-3: $R_6 = R_7 = R_8 = Y = H$; $\phi_{2156} = 60°$.

[e] Structure 4-20.

[f] Structure 4-16.

[g] Structure 4-3: $R_6 = R_7 = R_8 = F$; Y=H; $\phi_{2156} = 33°$.

[h] Structure 4-3: $R_6 = R_7 = R_8 = F$; Y = H; $\phi_{2156} = 30°$.

We expected that for a phenol the energy difference between these two R(O-H) geometries might give a better indication of the H - - X interactions than the harmonic oscillator ν_{OH} for two reasons. First, the comparison occurs on a portion of the O-H stretch curve that is steeper and hence more sensitive to variations in the environment around the O-H bond, compared to the less steep portion of the curve around $R(O-H)_{min}$. Secondly, by comparing energies at the respective E_0 and E_1 values of x correspondings to $\langle x^2 \rangle_n^{1/2}$, we would be looking at the portions of the curve where the the proton spends a good portion of its time in the ground and vibrationally excited state. The results, given as $\Delta E(\langle x^2 \rangle_0^{1/2} \rightarrow \langle x^2 \rangle_1^{1/2})$, are presented in Table 4-27. Although the ab initio results do qualitatively suggest shifts to lower O-H stretching frequencies for the cis ortho-fluoro- and ortho-chloro-phenols, the sensitivity of the model is poor and it does not predict the correct order for X=H, F, and Cl.

As stated earlier, the ortho-CF_3-phenol is unusual in that the hydrogen bonded (cis) O-H stretching peak is larger than the trans, but shifted to higher frequencies from the trans O-H stretch. A CNDO/2 geometry search for a minimum energy O-H bond length for ortho-CF_3-phenol (Table 4-26) gives a $R(O-H)_{min}$ value which is longer than the CNDO/2 calculated $R(O-H)_{min}$ value for phenol. From the data points used in the search, a ν_{OH} was obtained which incorrectly predicts a shift to lower frequency for ν_{OH} for the cis conformer of about 100 cm^{-1}. A similar ab initio geometry search yielded a $R(O-H)_{min}$ value that is shorter than the ab initio calculated $R(O-H)_{min}$ value for phenol and a ν_{OH} for the cis conformer that is about equal to the ab initio calculated ν_{OH} value for phenol (Table 4-26). For ortho-CF_3-phenol the ab initio $\Delta E(\langle x^2 \rangle_o \rightarrow \langle x^2 \rangle_1^{1/2})$ value, in contrast to the ortho-fluoro-

and chlorophenol cases, is larger than for phenol, suggesting a steeper O-H bond stretch potential for ortho-CF_3-phenol than for phenol. This in turn suggests that this steric repulsion between the hydrogen-bonded CF_3 and OH groups causes the ν_{OH} shift to higher frequency observed for the cis conformer of ortho-CF_3-phenol.

From the geometry of our calculations, the internuclear distance between the phenolic proton and the closest F of the CF_3 group for the minimum energy CF_3 rotamer of the cis conformer of ortho-CF_3-phenol is calculated to be only 1.70 Å. Comparison of this value with the H- - F internuclear distances for the cis conformer of ortho-fluorophenol (2.26 Å) (Table 4-8) and for the equilibrium H_2O/fluorobenzene dimer (1.50 Å, ab initio: Table 4-9) and with the sum of Van der Waals radii for H + F (2.67 Å) (Table 4-8) suggests that for ortho-CF_3-phenol: 1) The intramolecular hydrogen bond strength is due to the expected attractive interaction of the pehnolic proton with the F atom; 2) The repulsive interaction of the phenolic proton with the F atoms is due to the latter being forced (because of geometrical constraints) into very close proximity with the former in order to maximize the attractive interaction of the two; and 3) This repulsive interaction due to the H - - F internuclear distance being forced to be so small is reflected in our ab initio calculations by a shortened O-H bond length, an increased O-H bond stretching force constant, and a shift in ν_{OH} to higher frequency. This shortened H - - F internuclear distance for the closest F atom in the minimum energy cis-CF_3-phenol conformer is reflected in the H - - F bond order (0.0169, CNDO/2; 0.0264; ab initio), which is significantly greater than the value for the cis ortho-fluorophenol

Table 4-27. Ab Initio Calculations on the O-H Stretching Potential Energy Curve for Various Phenols (4-5).

X	$\Delta E(\langle x^2 \rangle_0^{1/2} \rightarrow \langle x^2 \rangle_1^{1/2})$[a] (kcal/mole)
H	5.32
F	5.25
Cl	5.30
CF_3[b]	5.40
CHO[c]	4.76
OH[d]	5.30

[a] See text for derivation of $\langle x^2 \rangle_n^{1/2}$ values.

[b] Structure 4-3: $R_6 = R_7 = R_8 = F$; Y = H; $\phi_{2156} = 30°$.

[c] Structure 4-20.

[d] Structure 4-16.

conformer (0.0021, CNDO/2; 0.0046, ab initio). In addition, θ_{HOF} for the most stable cis ortho-CF_3-phenol conformer is only 23.4°, a much more favorable value than for the ortho-halophenols.

In a series of articles[157,167,174] Fateley, et al., have assigned phenolic OH torsional frequencies to the cis and trans conformations of a number of ortho-substituted phenols and then used these to calculate the enthalpy difference between the two conformations for each (Table 4-28). They assumed the potential associated with the internal rotation of the phenolic OH to be adequately represented by the Fourier cosine series

$$V(\alpha) = 1/2 \sum_n V_n \quad (1-\cos n \ \alpha) \qquad \text{(Eqn. 4-2)}$$

which could be truncated at n = 2 for most ortho-substituted phenols. (Approximate calculations have shown that higher terms are negligibly small.[200]) Equating $V(\alpha)$ with our ϕ_{1234} of 4-1,

$$V(\phi_{1234}) = V_1(1-\cos \phi_{1234})/2 + V_2(1-\cos 2 \ \phi_{1234})/2 \qquad \text{(Eqn. 4-3)}$$

For an ortho-substituted phenol, V_1 is equal to the energy difference between the cis and trans conformations and V_2 corresponds to the OH rotational barrier with the V_1 (cis/trans) contribution factored out; i.e., V_2 is essentially the energy required to rotate the phenolic OH out of conjugation with the aromatic ring; differences in V_2 reflect differences in the inductive and resonance interactions of the different ortho-substituents with the phenolic OH. From our MO calculations we can derive values for V_1 and V_2, which, together with corresponding experimental values, are presented in Table 4-28. For phenol itself the ab initio calculations overestimate the experimental V_2 rotational

Table 4-28. CNDO/2, Ab Initio, and Experimental OH Rotational Energies for Phenols and ortho-Substituted Phenols (4-1, 4-2, and 4-32).

		V(90°)(kcal/mole)		V_1[b] (kcal/mole)			V_2[c] (kcal/mole)		
X	Y	CNDO/2	Ab Initio	CNDO/2	Ab Initio	Exptl.	CNDO/2	Ab Initio	Exptl.
H	H	2.88	5.13	0[l]	0[l]	0[l]	2.88	5.13	3.56[o,u]
F	H	3.52	5.59	1.37	1.68	1.63[m,n]	2.84	4.75	4.72[m,n]
						1.44[n,o]			4.44[n,o]
Cl	H	4.65	6.15	2.30	1.77	1.63[m,n]	3.50	5.26	5.46[m,n]
						1.62[n,o]			5.16[n,o]
Br	H	4.26	---	1.68	---	1.53[m,n]	3.42	---	5.40[m,n]
						1.57[n,o]			5.15[n,o]
I	H	3.53	---	0.75	---	1.32[m,n]	3.15	---	4.97[m,n]
						1.45[n,o]			4.47[n,o]
CH_3[d]	H	3.53	---	0.84	-1.53	-0.86[p]	3.11	---	≈3.29[o,q]
						-0.29[o,q]			
						-0.85[o,q]			
CH_3[d]	CH_3[e]	3.28	---	0[l]	---	0[l]	3.28	---	3.41[o,q], 2.31[p]
CF_3[f]	H	5.12	---	≈2.50	---	≈0.9[r]	≈3.87	---	≈3.34[o,p]
CHO[g]	H	8.84	---	6.02	7.44	7.09[s]	5.83	---	---
NO_2[h]	H	11.05	---	8.29	---	6.65[s]	6.91	---	---
OH[i]	H	3.28	---	1.37	3.27	≈2.29[o,t]	2.60	---	---
OCH_3[j]	H	3.37	---	1.32	---	2.00[o,q]	2.71	---	5.94[o,q]
C_6H_5[k]	H	4.09	---	1.66	---	2.73[o,q]	3.26	---	4.52[o,q]

Table 4-28. (Continued)

---

[a] V (90°) = ΔE (4-1 → 4.32).

[b] V_1 = ΔE(4-1 → 4-2).

[c] Calculated from V(90°), V_1, and Eqn. 4-3.

[d] Structure 4-3: $R_6 = R_7 = R_8 = H$; $\phi_{2156} = 60°$.

[e] 2nd CH_3 staggered the same as the first with respect to the OH.

[f] Structure 4-3: $R_6 = R_7 = R_8 = F$; Y=H; $\phi_{2156} = 33°$.

[g] Structure 4-20.

[h] Structure 4-17.

[i] Structure 4-16.

[j] Structure 4-18: CH_3 group staggered.

[k] Structure 4-19: $\phi_{2156} = 90°$, i.e. with the two rings perpendicular.

[l] By definition.

[m] Method of study = IR OH torsional frequency: Vapor state.

[n] Reference 167.

[o] Method of study = IR OH torsional frequency: Cyclohexane solution.

[p] Reference 185: Method of study = Force field molecular mechanics calculations.

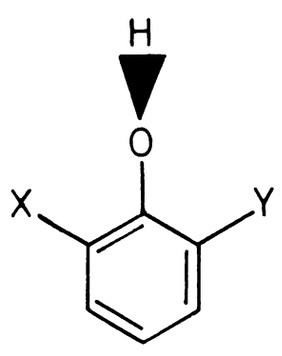
[q] Reference 157.

[r] Reference 186: Method of study = IR OH stretching frequency; CCl_4 solution.

[s] Reference 169: Method of study = OH ^1H chemical shift; CCl_4 solution.

[t] Estimated from references 169 and 174.

[u] Reference 199.



4-32

barrier by a much larger amount than the CNDO/2 underestimation of V_2. The ab initio calculations, however, agree extremely well with both the magnitudes and ordering of V_1 and V_2 for ortho-fluoro- and ortho-chloro-phenol in the vapor state, but are higher than the values of V_1 and V_2 in cyclohexane. This suggests an explanation for part of the overestimation by ab initio of V_2 for phenol itself since this experimental value was also measured in cyclohexane solution and not in the vapor state. The agreement of the CNDO/2 calculations with the experimental V_1 values for the four ortho-halophenols is moderately good. Interestingly, the ordering of V_2 for the ortho-substituents of the ortho-halophenols and phenol itself is Cl > Br > I > H > F for CNDO/2, Cl > H > F for ab initio, and Cl > Br > I > F > H experimentally. Apparently, changes in V_2 of the ortho-halophenols (as compared to phenol) are determined by two factors. First, the greater the electronegativity of the ortho substituent (F > Cl > Br > I > H), the better if is able to inductively withdraw electron density from the phenolic ring, causing the phenolic oxygen to donate electron density into the ring, thus increasing the phenolic C-O double bond character and hence V_2. Second, the greater the ability of the ortho substituent (F >> Cl, Br, I > H) to donate lone pair electron density by resonance into the aromatic ring, the better it is able to oppose the delocalization of the phenolic oxygen lone pair electrons into the ring, hence decreasing V_2. The interpretation of our results are consistent with a study[199] on para-fluorophenol in which it was found that the para-fluoro-substituent actually decreased V_2 (as compared to phenol) by 0.60 and 0.53 kcal/mole in experimental and ab initio studies, respectively. As seen in Table 4-28, the CNDO/2 and ab initio calculations generally give reasonable predictions for V_2

for the ortho-OH-, NO_2-, CN-, CHO-, C_6H_5-, CF_3-, and OCH_3-phenols, although the CNDO/2 results tend to underestimate the experimental V_2 values (where available for comparison) slightly more than the ab initio results.

Fateley and Carlson[157] found that the phenolic OH torsional frequency region was more complicated than at first expected for ortho-methylphenol. This apparently is due to the CH_3 rotational potential being superimposed upon the phenolic OH rotational potential. With the CH_3 group staggered with respect to the phenolic OH, CNDO/2 gave a value for V_2 of 3.11 kcal/mole (Table 4-28) in good agreement with Fateley and Carlson's experimental value of 3.29 kcal/mole.[157] With the addition of a second ortho-methyl group, the CNDO/2 V_2 value rises slightly to 3.28 kcal/mole, paralleling a rise in the experimental V_2 value to 3.41 kcal/mole.[157]

In order to test the validity of truncating the Fourier cosine series, Eqn. 4-2, at n = 2 to give Eqn. 4-3 for the potential associated with the internal rotation of the phenolic OH group, we calculated with CNDO/2 the variation of the energy of phenol and of ortho-chlorophenol at 15° increments of rotation of the OH group. Multiple least squares linear regression analyses were then conducted, using the CNDO/2 energies with Eqn. 4-2, including various combinations of the higher order terms. The results (Table 4-29) lead to the following conclusions concerning these calculations. Although the inclusion of v_n terms with $n > 2$ is statistically justifiable, the changes in V_1 and V_2 induced by these inclusions are small enough so that quite accurate approximations of V_1 and V_2 can be obtained from Eqn. 4-3. As theoretically expected, the

changes induced in V_1 and V_2 by the inclusion of V_n terms with n > 2 is such that V_1 is affected only by inclusion of V_n terms with n odd and V_2 is affected only by inclusion of V_n terms with n even. The magnitude of any V_{n+2} term is only between 5% and 20% of the V_n term.

Assuming a Boltzmann distribution between the conformers used for these regressions, it was found that the net energy of the cis conformers (4-5: $-90° < \phi_{1234} < 90°$) of ortho-chlorophenol was 2.33 kcal/mole less than the net energy of the trans conformers (4-5: $90° < \phi_{1234} < 270°$), in very good agreement with the value of 2.30 kcal/mole obtained for V_1 considering only ΔE(cis → trans). That this complete an analysis gives almost identical results with the simple ΔE(cis → trans) type of analysis provides additional justification for the latter's use in analyzing such cis/trans conformational changes. However, this ΔE (2.3 kcal/mole) is not the same as the V_1 derived from the least squares fit to the Fourier series (1.93 kcal/mole). Thus, the experimental values derived from the V_1 (the torsional frequencies in the IR) may not be quantitatively comparable to the ΔE values derived by other methods, such as relative intensities of the O-H stretching peaks in the near IR.

CNDO/2, ab initio, and experimental dipole moments for a number of compounds examined in this study are presented in Table 4-30. Both CNDO/2 and ab initio qualitatively predict trends in the dipole moments for families of compounds in the table. Although there are a few relatively large devitions from the experimental values, most of the calculated dipole moments are in reasonable agreement with the experimental values.

Table 4-29. Least Squares Fit of CNDO/2 OH Rotational Potential to Eqn. 4-2 for Phenol and ortho-Chlorophenol (4-5).

		Equation 4-2:[a,b] $V(\alpha) = \sum_n V_n (1-\cos n)/2$						
X	Eqn.	Intercept	V_1	V_2	V_3	V_4	V_5	V_6
H	4-4	0.075 (±0.071)		2.897 (±0.120)				
H	4-5	0.000 (±0.000)		2.873 (±0.004)		0.182 (±0.004)		
H	4-6	0.000 (±0.000)		2.872 (±0.000)		0.182 (±0.000)		0.005 (±0.000)
Cl	4-7	0.288 (±0.250)	1.979 (±0.322)	3.534 (±0.324)				
Cl	4-8	0.143 (±0.208)	1.931 (±0.237)	3.534 (±0.236)	0.339 (±0.237)			
Cl	4-9	0.163 (±0.223)	1.979 (±0.255)	3.492 (±0.258)		0.313 (±0.258)		
Cl	4-10	0.017 (±0.029)	1.931 (±0.030)	3.492 (±0.030)	0.339 (±0.030)	0.313 (±0.030)		
Cl	4-11	0.004 (±0.011)	1.926 (±0.011)	3.492 (±0.011)	0.335 (±0.011)	0.313 (±0.011)	0.036 (±0.011)	
Cl	4-12	0.000 (±0.001)	1.926 (±0.001)	3.490 (±0.001)	0.345 (±0.001	0.312 (±0.001)	0.036 (±0.001)	0.012 (±0.001)

Table 4-29. (Continued)

---

[a] $\alpha = \phi_{1234}$ for 4-5.

[b] For both phenol and ortho-chlorophenol, 13 data points at 15° intervals from $\alpha = 0°$ to $\alpha = 180°$ were used in the regressions. V_n values given for each equation are for the particular values of n included in that regression. Values in parentheses are 95% confience intervals. The inclusion of each additional higher order term ($n \rightarrow 2$) is (by F-test analysis) statistically (at $\geq$ the 95% confidence level) significant.

Table 4-30. CNDO/2, Ab Initio, and Experimental Dipole Moments (D).

A. Halobenzenes (C_6H_5X)

X	μ_{calcd}(CNDO/2)	μ_{calcd}(Ab Initio)	μ_{exptl}[a]	Solvent
F	1.68	0.93	1.57-1.66	Vapor
			1.39-1.51	C_6H_6
Cl	2.00	2.28	1.75	Vapor
			1.58	C_6H_6
Br	2.08	---	1.70-1.79	Vapor
			1.50-1.57	C_6H_6
I	2.72	---	1.70-1.71	Vapor
			1.25-1.39	C_6H_6

Table 4-30. (Continued)

B. Ortho-Substituted Phenols (o-X-C_6H_5OH: 4-5 and 4-6).

	μ_{calcd}(CNDO/2)			μ_{calcd}(Ab Initio)			μ_{exptl}[a]	Solvent
X	4-5	4-6	Avg.[b]	4-5	4-6	Avg.[b]		
H	1.73	---	1.73	1.22	---	1.22	1.55[c]	---
F	1.39	3.39	1.57	0.74	2.11	0.81	1.32	C_6H_6
Cl	1.53	3.64	1.57	1.42	3.41	1.51	2.19	Vapor
							1.24-1.43	C_6H_6
Br	1.56	3.72	1.68	---	---	---	1.27-1.39	C_6H_6
I	1.84[m]	4.16[m]	3.72[m]	---	---	---	1.54	C_6H_6
	1.90[n]	4.24[n]	2.97[n]					
	1.96[o]	4.32[o]	2.48[o]					
NO_2[d]	4.21	7.06	4.21	---	---	---	3.13-3.22	C_6H_6
CHO[e]	3.30	3.92	3.30	2.24	2.61	2.24	2.86-2.91	C_6H_6
OH[f]	3.00	2.66	2.97	2.15	1.50	2.14	2.60-2.64	C_6H_6

Table 4-30. (Continued)

	μ_{calcd} (CNDO/2)			μ_{calcd} (Ab Initio)			μ_{exptl}[a]	Solvent
X	4-5	4-6	Avg.[b]	4-5	4-6	Avg.[b]		
OCH_3[g]	2.86	2.57	2.83	---	---	---	2.37-2.44	C_6H_6
CN	2.49	5.00	2.57	---	---	---		
C_6H_6[h]	1.75	1.77	1.75	---	---	---		
CH_3[i]	1.80[p]	1.57[p]	1.76	1.45[p]	1.00[p]	1.03	1.42-1.55	C_6H_6
	1.82[q]	1.56[q]	1.73[q]					
iPr[j]	1.90	1.53[u]	1.55	---	---	---		
	1.95[q]	1.52[q]	1.57[q]					
tBu[k]	1.89[p]	1.61[p]	1.61	---	---	---	1.35[w]	C_6H_6
	1.89[q]	1.61[q]	1.61[q]					
CF_3[l]	1.93[s]	4.39[p]	1.97	0.92[r]	2.84[p]	1.82		
	1.92[t]	4.39[t]	1.98[t]					
COOH	----	---	2.06[v]	---	---	---	2.65	Dioxane

Table 4-30. (Continued)

	μ_{calcd}(CNDO/2)			μ_{calcd}(Ab Initio)			$\mu_{exptl.}$[a]	Solvent
C. Para-Halophenols (p-X-C_6H_4OH: 4-15)								
X	4-15			4-15				
F	1.95			1.43			2.10-2.17	C_6H_6
Cl	2.15			2.51			2.22-2.4	C_6H_6
Br	2.21			---			2.14-2.28	C_6H_6
I	2.71			---			2.21	C_6H_6
D. Ortho-Substituted Thiophenols (o-X-C_6H_4SH: 4-24 and 4-25)								
X	4-24	4-25	Avg.[x]	4-24	4-25	Avg.[x]		
H	2.26	---	2.26	0.75	---	0.75	1.18-1.34	C_6H_6
F	2.47	3.72	2.75	0.60	1.65	0.81		
Cl	2.73	4.00	2.73	1.55	2.94	2.92	1.98	C_6H_6
Br	2.77	4.05	2.77	---	---	---	1.96	C_6H_6
I	3.02	4.55	3.16	---	---	---		
OH	---	---	3.32[y]	---	---	1.78[y]		

Table 4-30. (Continued)

E. H_2X: Experimental Geometry	μ_{calcd}(CNDO/2)	μ_{calcd}(Ab Initio)	$\mu_{exptl.}$[a]	Solvent
X	H_2X	H_2X		
O[z]	2.15	1.73	1.82[c]	Vapor
S[aa]	---	1.02	0.90-1.01	Vapor

[a] From reference 136 unless otherwise noted.
[b] Calculated as a simple weighted average assuming a Boltzmann distribution between 4-5 and 4-6.
[c] A "best" average experimental value.
[d] Structure 4-17.
[e] Structure 4-20.
[f] Structure 4-16.
[g] Structure 4-18.
[h] Structure 4-19; $\phi_{2156} = 90°$.
[i] Structure 4-3: $R_6 = R_7 = R_8 = Y = H$.
[j] Structure 4-3: $R_5 = Y = H$; $R_6 = R_7 = CH_3$ (staggered).
[k] Structure 4-3: $R_5 = R_6 = R_7 = CH_3$ (staggered); $Y = H$.
[l] Structure 4-3: $R_5 = R_6 = R_7 = F$; $Y = H$.
[m] $\mathcal{Y}_I = 1.09$.
[n] $\mathcal{Y}_I = 1.145$.

Table 4-30. (Continued)

---

[o] $\mathcal{I}_I$ = 1.20.

[p] ϕ_{2156} = 60°.

[q] Calculated assuming a Boltzmann distribution between the conformers of Table 4-15.

[r] ϕ_{2156} = 30°.

[s] ϕ_{2156} = 33°.

[t] Calculated assuming a Boltzmann distribution between conformers of Table 4-15.

[u] ϕ_{2156} = 180°.

[v] Calculated assuming a Boltzmann distribution between conformers 4-21, 4-22, and 4-23.

[w] From reference 185.

[x] Calculated as a simple weighted average assuming a Boltzmann distribution between conformers 4-24 and 4-25.

[y] Calculated assuming a Boltzmann distribution between conformers of Table 4-22.

[z] Experimental geometry: see note 180.

[aa] Experimental geometry: see note 193.

Conclusions.

With a few exceptions, the CNDO/2 MO method generally does a fairly reasonable job in reproducing the experimental intramolecular interactions (and in particulr the experimental ΔE values) for ortho-substituted phenols. The probable underestimation by CNDO/2 of the intramolecular hydrogen bond strength for ortho-fluorophenol can be viewed as only a minor deficiency in the method. That the cis conformer of ortho-methylphenol is found by CNDO/2 to be more stable than the trans is, however, a more serious error. For most of the phenols studied in this article, the ab initio MO method does at least as well and often better than CNDO/2 in reproducing the experimental intramolecular interactions of ortho-substituted phenols, especially of ortho-fluoro- and ortho-methylphenol. Both the CNDO/2 and ab initio calculations do well in predicting and providing some insight into the phsyical origins of the "anomalous" ordering of the experimental intramolecular hydrogen bond strengths of the ortho-halophenols. These studies suggest that the intramolecular interactions of the ortho-halophenols are mainly determined by a competition between the attractive and repulsive H - - halogen interactions in the cis conformer as well as the O - - halogen repulsions in the cis and trans conformers. In addition, these interactions are a strong function of the H-O - - X angle. The calculations suggest that the magnitudes of the ν_{OH} shifts for the cis conformers of the ortho-halophenols are determined by the magnitude of the H - - halogen interactions, which do not necessary reflect the net intramolecular hydrogen bond energies. Similar physical effects are apparently operative in ortho-CF_3-phenol. In this compound, the H - - F distance is foced to be sufficiently close in the more stable

cis conformer, so that one observes a shift to higher frequency for the O-H stretch, as well as predicts (ab initio) a shortened O-H bond length in this conformer. Low temperature neutron diffraction might be used to test this prediction.

The ab initio calculations are successful in reproducing the limited experimental data for the ortho-halothiophenols. The importance of considering X - - S repulsion effects, which should be greater than X - - O repulsion, is clear from these studies and experiments. The prediction that ortho-fluorothiophenol favors the cis hydrogen-bonded conformer and ortho-chlorothiophenol the trans (the opposite trend in cis stability from the ortho-halophenols) is a clear indication of the greater repulsive forces involved in ortho-interactions when both substituents are from the second row.

Our calculations indicate that the theoretical methods we have employed are capable of yielding a better understanding of the important forces determining the near and far IR properties of ortho-substituted phenols and thiophenols.

This study suggests a number of avenues for further work:

1. In the area of semi-empirical MO theory, we have done a very limited variation of the parameters of the I atom, but clearly more systematic variations in the spirit of Dewar, et al.,[201] are possible and would likely lead to a set of parameters which can better predict both intramolecular and intermolecular hydrogen bonding effects of I (as well as Br, Cl, and F) than the parameters we used in this study.

2. We have also used model intermolecular H-bond potentials to gain insight into the nature of intramolecular H-bonding. This should

prove of utility in studying intramolecular H-bonded systems with more conformational flexibility, such as 1,3 propanediol,[154] where one can use such calculations to separate conformation and H-bond effects in determining the final minimum energy structure.

3. The ab initio calculations with a minimal basis set do a very good job in predicting the hydrogen-bonding properties of the ortho-substituted phenols and thiophenols, with the possible exception of ortho-hydroxythiophenol. The Mulliken populations for the thiophenols are strange (S δ^+ and H δ^-), but the net dipole moment has the opposite direction, indicating that the wave function represents the polarity in a satisfactory manner. The intermolecular potential surfaces for the H_2O/C_6H_5SH dimers (thiophenol as the proton donor or acceptor) also support the net S δ^-/H δ^+ polarity in the wave function. However, more accurate calculations on ortho-hydroxythiophenol would probably be instructive, in order to see if they can reproduce the qualitative relative conformational stabilities determined by David and Hallam.[191]

4. We have also examined some of these ortho-substituted phenols and thiophenols for which there is no direct quantitative $\Delta E(cis \rightarrow trans)$ data. So we hope the calculations presented here will be an impetus for further experimental physical chemical studies.

5. Our results suggest that to reproduce $\Delta\nu$ for the X-H stretch in H-bonded systems may require a careful analysis of the anharmonic part of the proton potential. For a crude estimate of relative frequency shifts, $E(\langle x^2 \rangle_1^{1/2}) - E(\langle x^2 \rangle_o^{1/2})$ may be a better guide than directly calculated $\Delta\nu_{XH}$.

INTERMOLECULAR HYDROGEN BONDING IN ORTHO-SUBSTITUTED PHENOLS AND PHENOXIDES

There are numerous plausible nuclear receptor and plasma protein binding sites one could imagine in order to explain the outer ring structure-activity relationships of the thyroid hormones and analogs. Therefore, we decided to use experimental and theoretically calculated inter- and intramolecular hydrogen bond strengths of ortho-substituted phenols and phenoxides in order to examine:

1. Likely orientations for possible nuclear receptor and plasma protein proton donor and proton acceptor groups.

2. The relative biological activities for different 4' substituents.

3. Possible reasons for the the "non-additivity" effects in binding for 3', 4' substituted analogs.

Following the extensive theoretical CNDO/2 and ab initio examinations of the intramolecular hydrogen bonding and interactions of ortho-substituted phenols, we examined the intermolecular hydrogen bonding of ortho-substituted phenols as a model system for the unionized phenolic ring binding to nuclear receptors and the intermolecular hydrogen bonding of ortho-substituted phenoxides as a model system for the ionized phenolic ring binding to TBG. H_2O was used both (a) as a model proton acceptor in the nuclear receptor when the unionized phenol is functioning as a proton donor (4-33) and (b) as a model proton donor in the nuclear receptor when the un-ionized phenol is functioning as a proton acceptor (4-34) or in TBG when the ionized phenol is functioning as a proton acceptor (4-35).

The geometry of the model system for proton donation of a phenol to H_2O (4-33) was defined as follows. The two monomeric units lie in

perpendicular planes with the phenolic OH and O of the H_2O coplanar with the aromatic ring and with the aromatic ring plane bisecting θ_{HOH} of the H_2O. Variations in hydrogen bond energies (ΔE) were then determined with respect to variations in R, ϕ, and θ for phenol (4-33: X=Y=H) and are presented in Table 4-31 and Figs. 4-1 and 4-2.

With $\phi = 0°$, R and θ were simultaneously varied to give minimum energy values for R, θ, and ΔE of 2.63 Å, 47° and -8.93 kcal/mole (CNDO/2) and 2.54 Å, 12°, and -6.63 kcal/mole (*ab initio*), respectively. With R fixed at these values, variations in the minimum energy values of ΔE and θ were then determined as a function of θ. As can be seen in Table 4-31, the variations in θ with respect to ϕ are more reasonable for *ab initio* (formation of a bifurcated hydrogen bond as ϕ increases) than for CNDO/2 (H_2O protons practically directed at the phenolic O at $\phi = 70°$). Hence, all further CNDO/2 and *ab initio* calculations were performed using the *ab initio* minimum energy values of R and of θ as a function of ϕ. Although the *ab initio* are greater than the CNDO/2 ΔE values, the shapes of the *ab initio* and CNDO/2 ΔE *vs.* ϕ curves for phenol are approximately the same and all minimize at $\phi = 0°$ (O-H--H colinear) (Fig. 4-1 and 4-2).

The geometry of the model system for H_2O as proton donor to a phenol (4-34) was defined as follows. The two monomeric units lie in perpendicular planes with the phenolic OH and the H-O of the H_2O involved in the hydrogen bond coplanar with the aromatic ring. The O - - H-O involved in the hydrogen bond are colinear since this geometry should give maximal hydrogen bond strength.[147] The second O-H bond of the H_2O lies in a plane perpendicular to the aromatic ring plane in order to minimize any interactions of this second H_2O proton with the phenol.

Variations in CNDO/2 and *ab initio* ΔE values were then determined for simultaneous minimum energy variations of R and θ_{COO} for phenol (4-34: X=Y=H) and are presented in Table 4-32 and Fig. 4-3 and 4-4. In contrast to phenol as the proton donor to H_2O, the CNDO/2 are greater than the *ab initio* ΔE values, although the shapes of the *ab initio* and CNDO/2 ΔE *vs.* θ_{COO} curves are approximately the same and both minimize at $\theta_{COO} \approx 125°$ (hydrogen bond approximately bisecting θ_{COH} of phenol). All further CNDO/2 and *ab initio* calculations were performed utilizing these CNDO/2 and *ab initio*, respectively, minimum energy values of R as a function of θ_{COO}.

The geometry of the model system for H_2O as a proton donor to a phenoxide (4-35) was defined as above for H_2O as a proton donor to a phenol (4-34), except the phenolic proton is left out and the phenoxide R(C-O) shortened to 1.33 Å.[202] Variations in CNDO/2 ΔE values were then determined for simultaneous variations of R and θ_{COO} for phenoxide (4-35: X=Y=H) and are presented in Table 4-33 and Fig.4-5. The minimum energy R values are essentially invarinat (and are considerably shorter than for 4-34) as θ_{COO} is varied. As θ_{COO} is varied, ΔE varies very little, minimizing at $\theta_{COO} \approx 125°$ and is about 3.5 to 4 times ΔE for the corresponding phenol--HOH dimer (4-34). All further CNDO/2 calculations were performed utilizing these CNDO/2 minimum energy values of R as a function of θ_{COO}.

Utilizing the reference geometries for 4-33, 4-34, and 4-35, as defined above for X=Y=H for minimum energy, R, ∅, θ, and θ_{COO} values, ΔE values were calculated for a variety of variations of X, Y, and geometrical

Table 4-31. CNDO/2 and Ab Initio Hydrogen Bond Energies (ΔE in kcal/mole) of Phenol as Proton Donor to H_2O (4-33: X = Y = H).

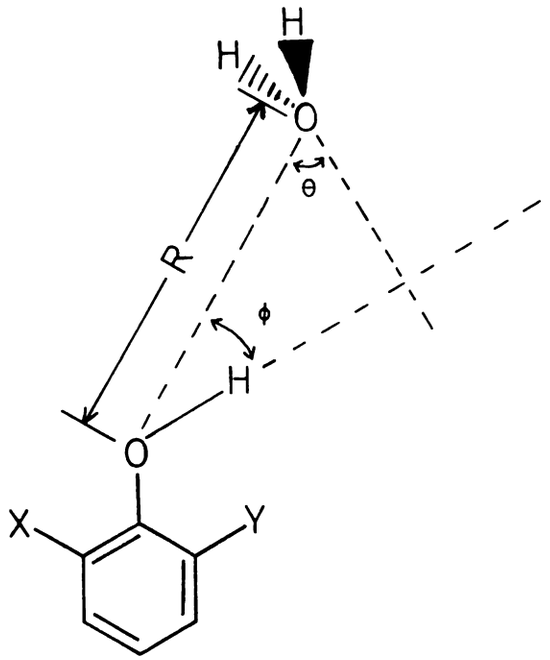
			ϕ (°)					
	R(Å)		0	10	20	30	50	70
Ab Initio	2.63[a]	ΔE	8.93	8.70	7.36	5.62	2.87	1.89
		θ[b]	47	58	67	77	98	126
CNDO/2	2.63[a]	ΔE	6.05	5.64	4.72	3.49	1.51	1.06
		θ[b]	47	58	67	77	98	126
	2.54[c]	ΔE	6.63	6.42	5.52	4.18	1.69	1.52
		θ[d]	12	11	17	24	73	157

[a] Minimum energy (Ab Initio) value for phenol at ϕ = 0°.

[b] Minimum energy (Ab Initio) values for phenol at R = 2.63 Å.

[c] Minimum energy (CNDO/2) value for phenol at ϕ = 0°.

[d] Minimum energy (CNDO/2) values for phenol at R = 2.54 Å.



4-33

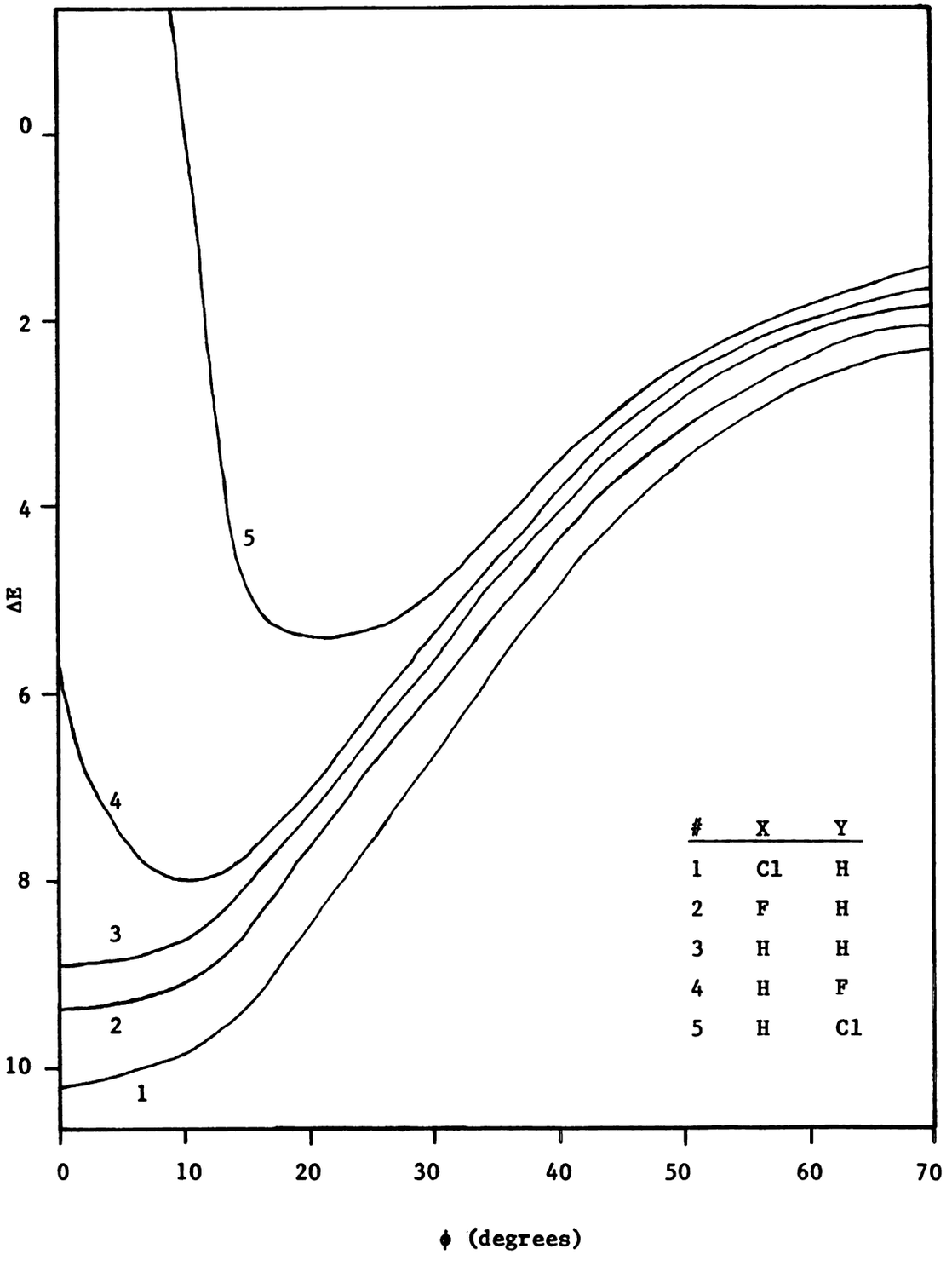


Figure 4-1. Ab Initio hydrogen bond energies (ΔE: kcal/mole) of various phenols as proton donors to H_2O (4-33).

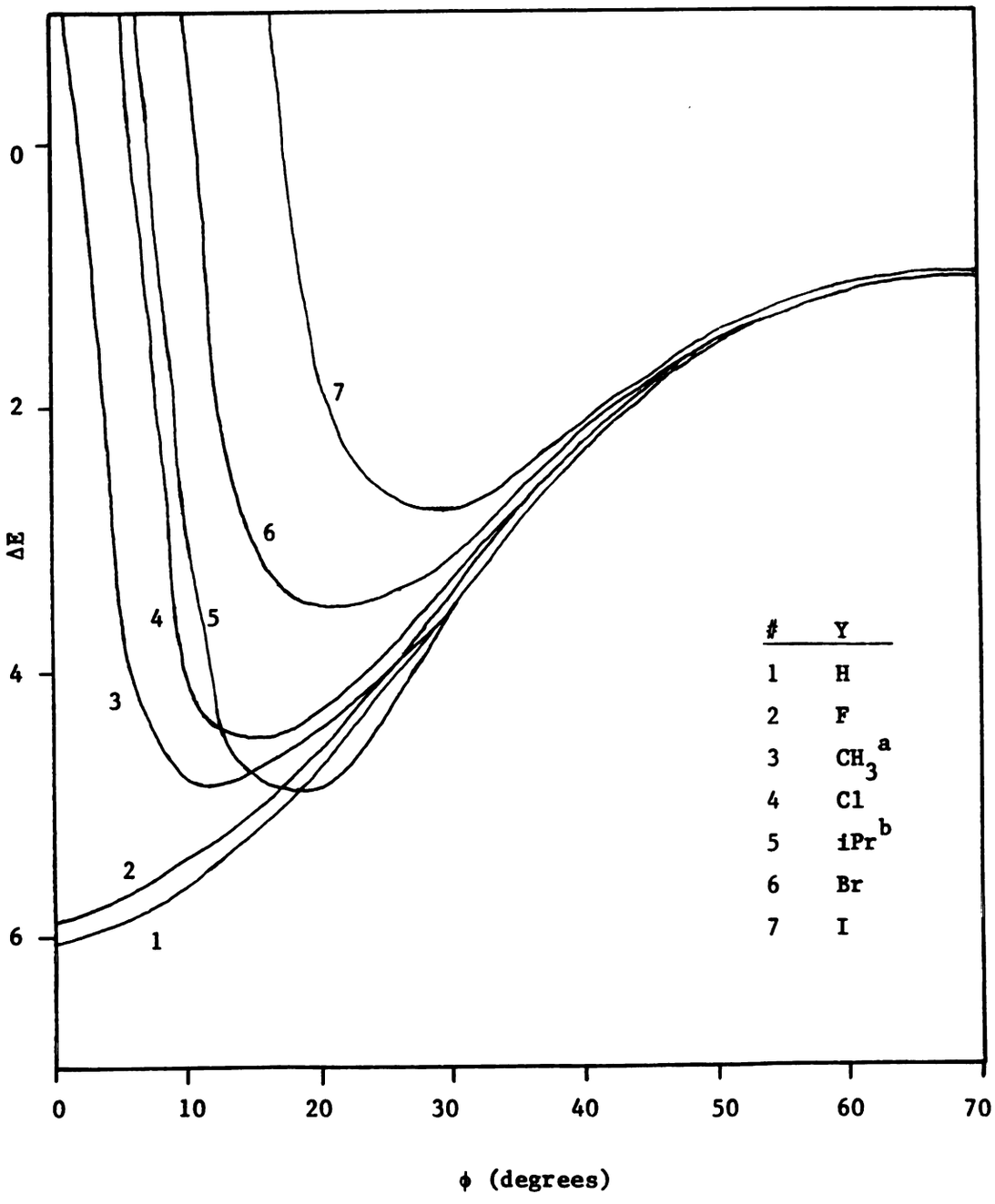


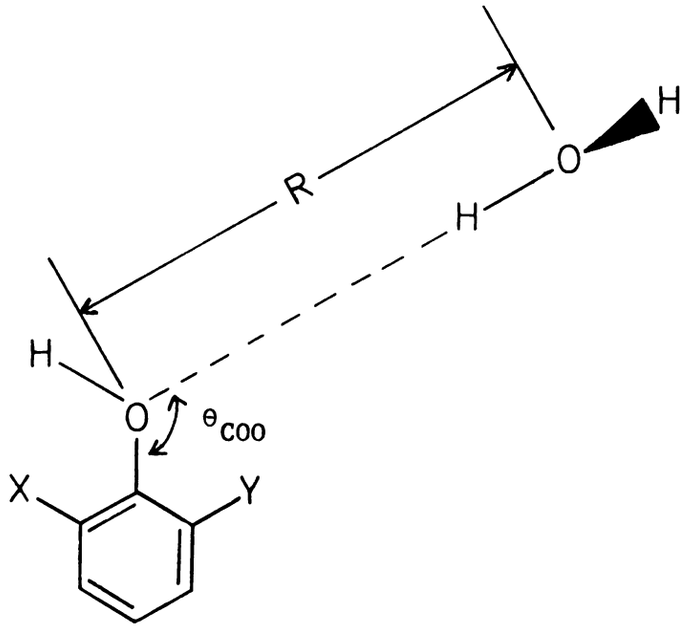
Figure 4-2. CNDO/2 hydrogen bond energies (ΔE: kcal/mole) of various phenols as proton donors to H_2O (4-33: X = H); [a] CH_3 group staggered; [b] iPr CH_3 groups pointed away from OH and staggered.

Table 4-32. CNDO/2 and Ab Initio Hydrogen Bond Energies (ΔE in kcal/mole) of H_2O as Proton Donor to Phenol (4-34: X = Y = H).

		θ_{COO}						
		110	120	125	130	140	160	180
Ab Initio	ΔE	3.33	4.14	4.31	4.30	4.05	2.53	0.46
	R(Å)[a]	2.83	2.80	2.79	2.79	2.80	2.90	3.37
CNDO/2	ΔE	5.94	6.16	6.12	6.06	5.83	4.77	2.83
	R(Å)[b]	2.57	2.56	2.54	2.56	2.56	2.60	2.74

[a]Minimum energy (Ab Initio) values for phenol at each θ_{COO}.

[b]Minimum energy (CNDO/2) values for phenol at each θ_{COO}.



4-34

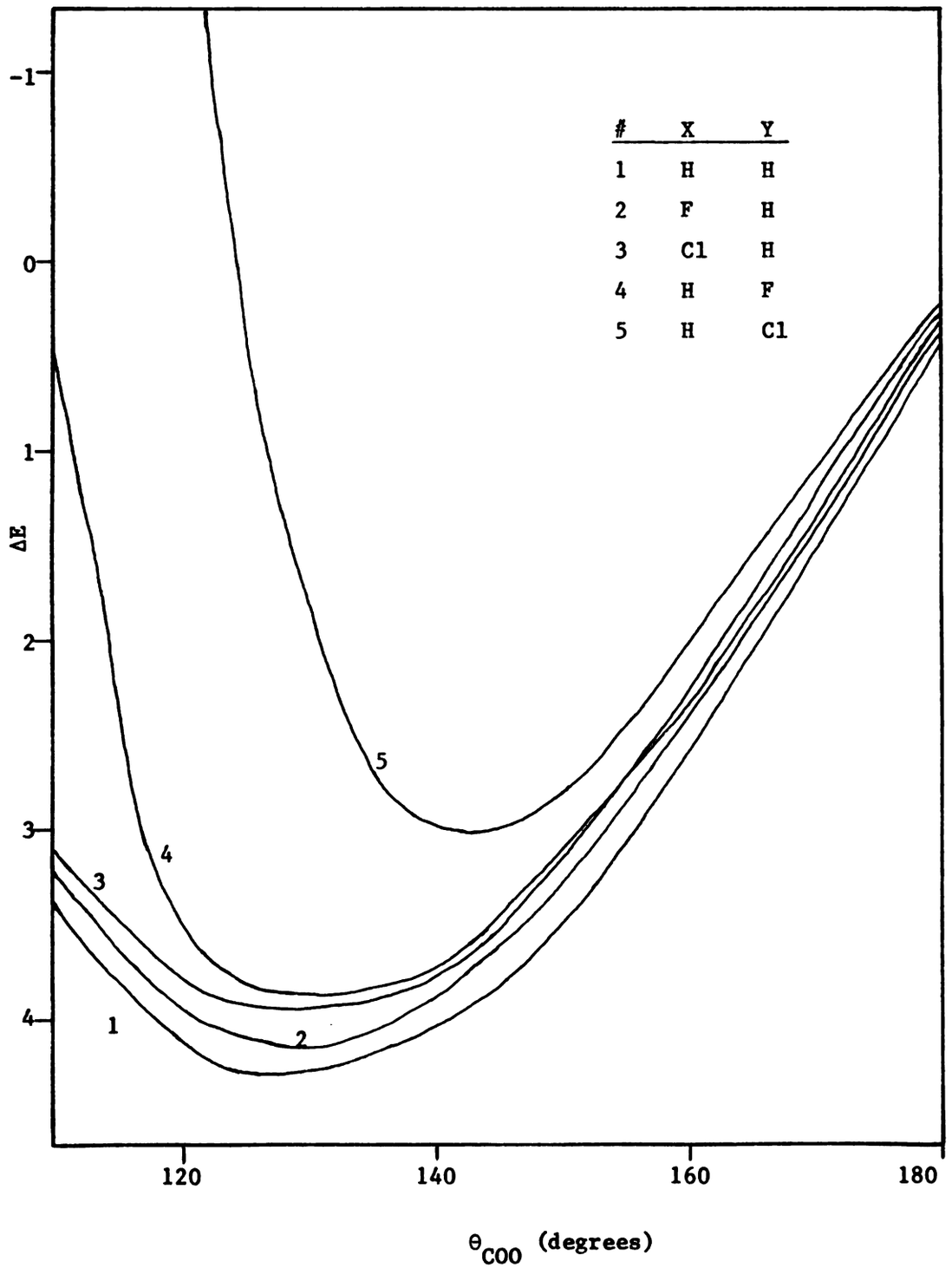


Figure 4-3. Ab Initio hydrogen bond energies (ΔE: kcal/mole) of H_2O as proton donor to various phenols (4-34).

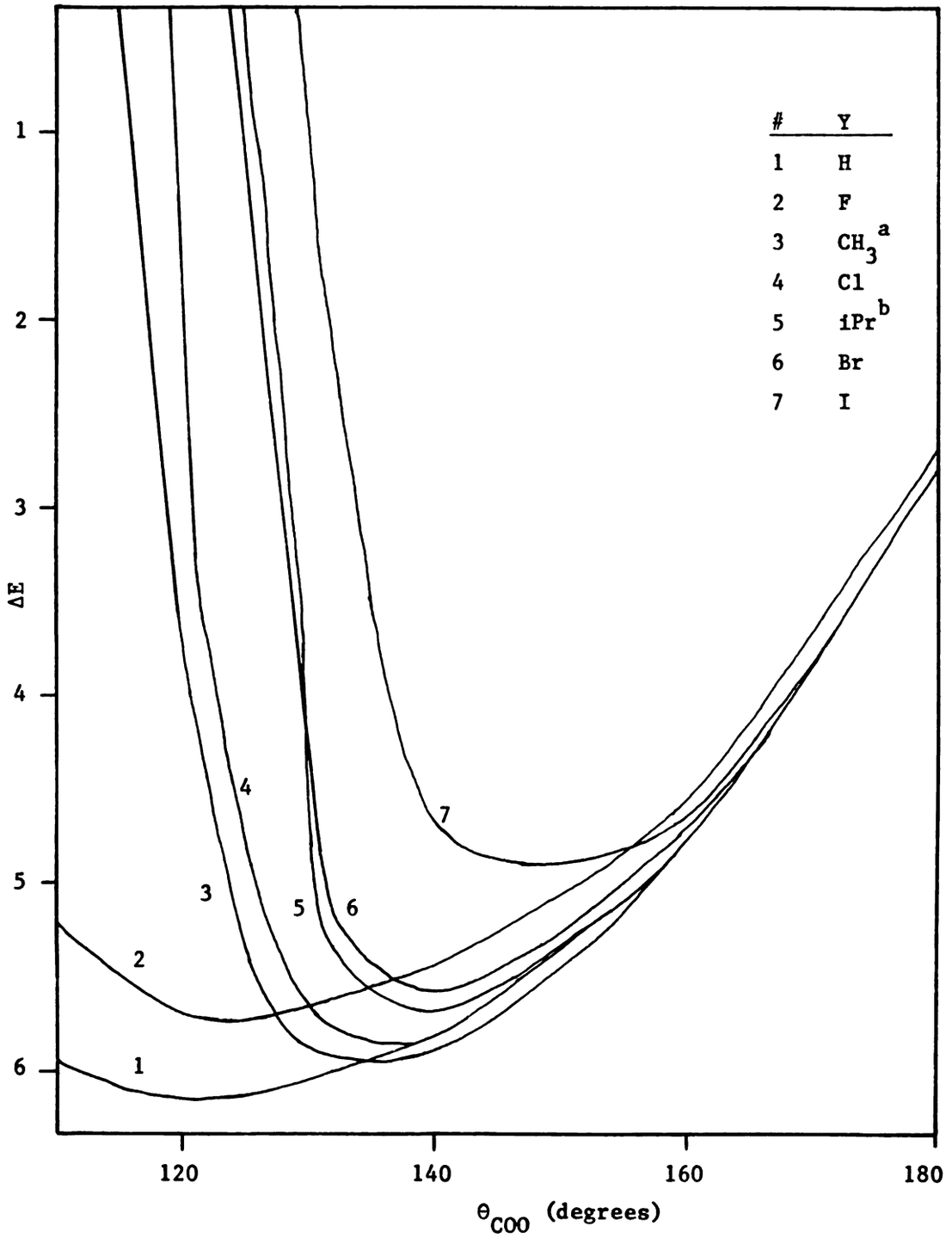


Figure 4-4. CNDO/2 hydrogen bond energies (ΔE: kcal/mole) of H_2O as proton donor to various phenols (4-34: X = H); $^{a}CH_3$ group staggered; ^{b}iPr CH_3 groups pointed away from OH and staggered.

parameters and are presented graphically in Fig. 4-1 through 4-5 and/or are described below.

It was found that the CNDO/2 ΔE values were only very slightly affected ($\pm$ 0.2 kcal/mole maximum) for Y=H and X=H, F, Cl, Br, I, CH_3, iPr, or OH for 4-35. In contrast, the change in the ab initio ΔE values for Y=H and X=H, F, or Cl for 4-33 (see Fig. 4-1) and 4-34 (see Fig. 4-3) are much larger. In almost all cases, the direction of change of ΔE as X varies is as expected: i.e., electron-withdrawing X substituents reduce the electron density on the phenolic O, decreasing ΔE for 4-34 and 4-35 and increasing ΔE for 4-33; electron-donating X substituents of course have the opposite effect.

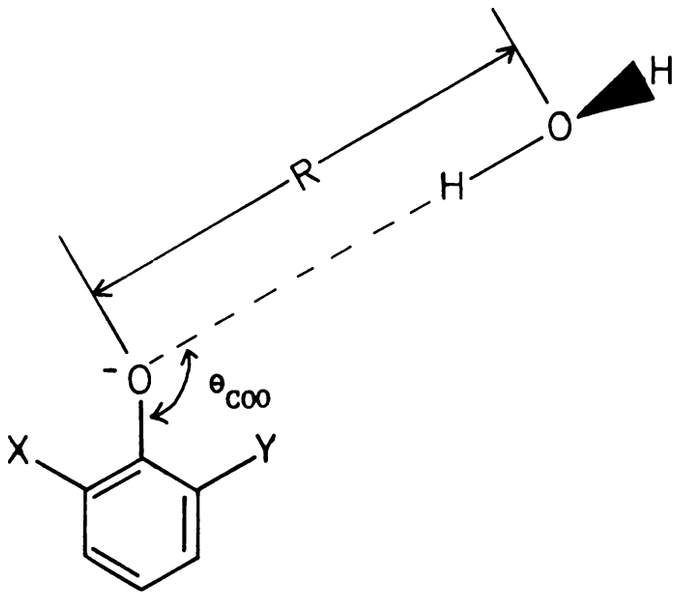
The situation is quite different when the Y substituent is varied in that it is capable of sterically interacting with the H_2O molecule. From the CNDO/2 and ab initio ΔE potentials for 4-33, 4-34, and 4-35 as functions of ϕ or θ_{COO} and of Y, as presented in Fig. 4-1 through 4-5,the following observations can be made. Strong repulsive interactions result between Y and the H_2O molecule for small values of ϕ for 4-33. These repulsive interactions decrease as ϕ increases, until the ΔE for the Y-substituted phenol equals that of unsubstituted phenol once the H_2O and Y-substituent are no longer within contact distance. Obviously, the larger Y is, the larger the repulsive H_2O/Y interactions are and the larger the ϕ value must be before ΔE returns to the unsubstituted phenol value. The same hold for 4-34 and 4-35, replacing ϕ with θ_{COO}. The CNDO/2 H_2O/Y repulsive ΔE potentials are much "harder" than the corresponding ab initio ones.

It was also found for both the CNDO/2 and ab initio calculations that the electronic effects of the X substituent and the electronic and

Table 4-33. CNDO/2 Hydrogen Bond Energies (ΔE in kcal/mole) of H_2O as Proton Donor to Phenol (4-35: X = Y = H).

	θ_{COO}							
	110	115	120	125	130	140	160	180
ΔE	22.52	23.11	23.67	23.74	23.68	23.31	22.20	21.74
R[a]	2.36	2.35	2.35	2.35	2.35	2.35	2.35	2.35

[a]Minimum energy (CNDO/2) values for phenoxide at each θ_{COO}.



4-35

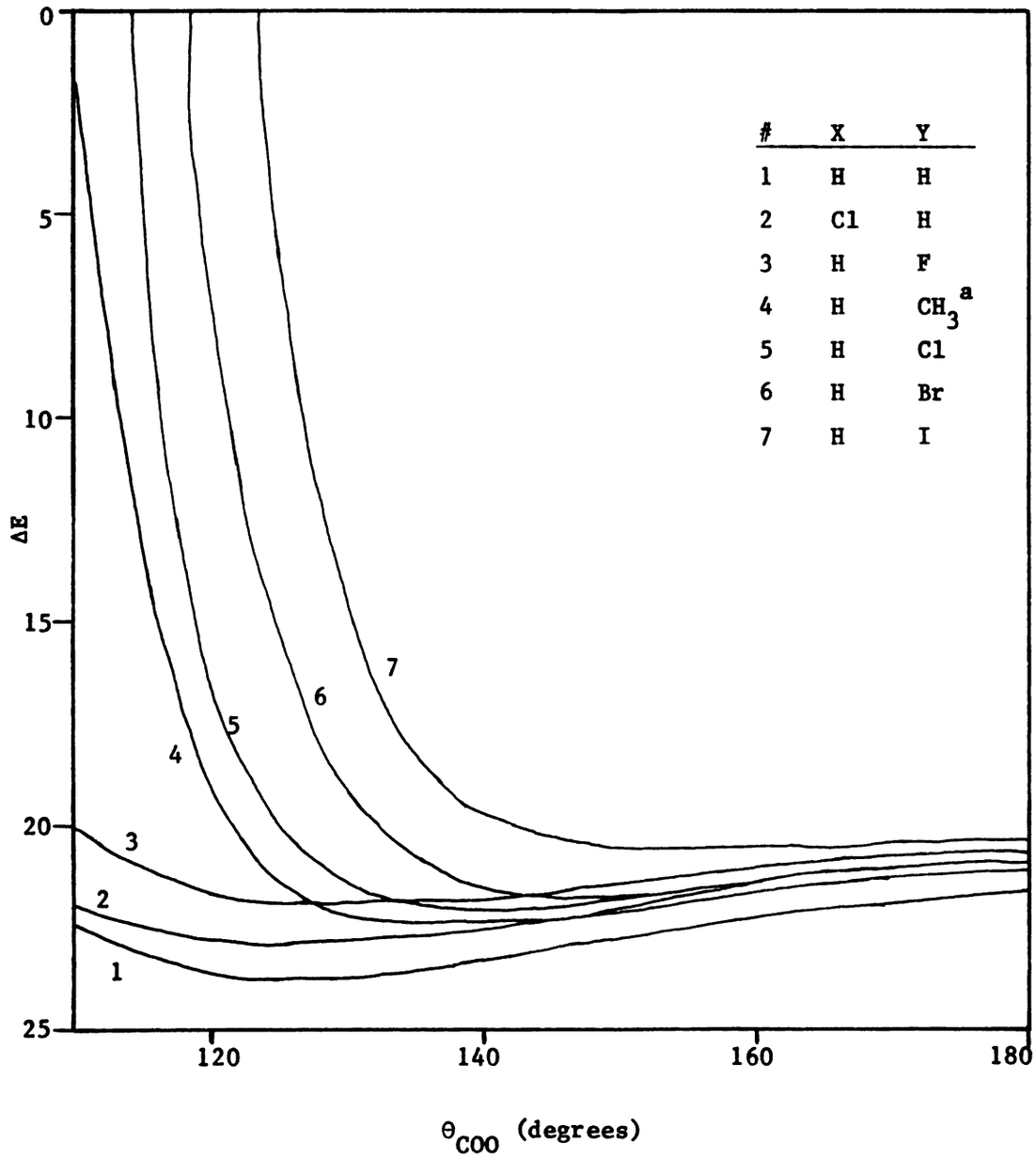


Figure 4-5. CNDO/2 hydrogen bond energies (ΔE: kcal/mole) of H_2O as proton donor to various phenoxides (4-35); $^{a}CH_3$ group staggered.

steric effects of the Y substituent on the ΔE of intermolecular hydrogen bond formation are essentially additive for 4-33, 4-34, and 4-35. That is, X and Y act essentially independently in their influence on intermolecular hydrogen bond formation.

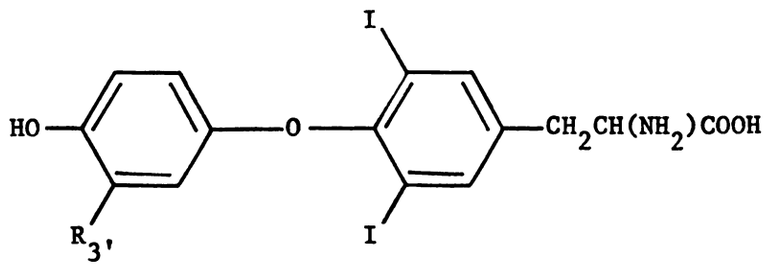
For 4-33, 4-34, and 4-35, for various X and Y substituents, the magnitudes but not the general shapes of the CNDO/2 ΔE potential curves were changed upon slightly increasing the R values.

Variation of ΔE upon movement of the H_2O molecule oxygen out of the phenol or phenoxide plane was also examined with CNDO/2. Maintaining $\theta_{COO} = 110°$ for 4-33 and the phenolic hydroxyl coplanar with the aromatic ring, movement of the H_2O out of the ring plane and hence loss of O-H - - O colinearity resulted in a large hydrogen bond energy loss for phenol (4-33: X=Y=H) for more than about 20° movement of the H_2O oxygen out of the phenol plane. For 2,6-diiodophenol (4-33: X=Y=I) a similar movement results in loss of H_2O/I repulsion but no sgnificant overall phenol/H_2O attraction. Maintaining $\theta_{COO} = 125°$ for 4-34 and the phenolic hydroxyl coplanar with the aromatic ring, movement of the H_2O out of the aromatic plane retains O-H - - O colinearity. Hence, such a movement has almost no effect (loss of $\Delta E \sim 0.3$ kcal/mole) for phenol (4-34: X=Y=H) and results in a significant ($\Delta E \sim 4.5$ kcal/mole) phenol/H_2O attraction for 2,6-diiodophenol (4-34: X=Y=I). A similar result was found for movement of H_2O out of plane for the phenoxide/H_2O model system (4-35) [as for the phenol/H_2O model system (4-34)] for phenoxide (4-35: X=Y=H) and 2,6-diiodophenoxide (4-35: X=Y=I). On the basis of these calculations and analysis of various binding activities, some tentative conclusions can be drawn concerning the probable nature of the intermolecular hydrogen bonds which are formed between thyroid hormone analogs and the

plasma proteins and nuclear receptors to which they bind. The binding affinity of L-T_3 to TBG is only 9% that of L-T_4.[28] Although this binding difference can be explained primarily on the basis of the pK_a's of the 4'-hydroxyl group of these two analogs, the geometrical orientation of the proton donor on the TBG molecule must be such that the 3' and 5' iodines provide little if any steric interference to this hydrogen bond formation. If this were not the case, then, the binding affinity of L-T_4 would not be expected to be so much greater than that of L-T_3. This suggests, using 4-35 as a model system and examining Fig. 4-5, that the geometrical orientation of the TBG proton donor probably is such that either the $C\text{-}O^-$ - - proton donor angle is substantially > 125° and/or that the proton donor's approach to the phenoxide ion is substantially out of the phenoxide ring plane.

For binding to the nuclear receptor, the situation is considerably more complicated. First, the 4'-phenolic hydroxyl could be functioning either as a proton donor (model system 4-33) or as a proton acceptor (model system 4-34). Second, the inverse correlation of binding affinity with the size of the 5'-substituent could be due either to direct steric interaction of the 5'-substituent with the receptor and/or to steric interference of the 5'-substituent with intermolecular hydrogen bond formation between the 4'-hydroxyl and the receptor. Third, whether the 4'-hydroxyl is functioning as a proton donor or acceptor, the phenolic hydroxyl could be directed either "cis" or "trans" to the 3'-position. This question of 4'-hydroxyl functioning as a proton donor or acceptor to receptor is then best approached by examining relative *in vitro* binding potencies of analogs to intact rat hepatic nuclei[24,25] and to solubilized rat hepatic nuclear protein receptors.[26,43] (This

eliminates complications due to differences in analog metabolism and binding to plasma proteins *in vivo*.) In particular, if the 4'-hydroxyl were functioning as a proton acceptor, then one would expect (as with TBG) the binding affinities of analogs to be inversely proportional to the pKa of the 4'-hydroxyl. The *in vitro* binding studies have shown however, that the binding of analogs to these nuclear receptors is only slightly affected by the electronic interactions of the 3'- and 5'-substituents with the 4'-hydroxyl. In addition, the binding to nuclear receptors is approximately equal for analogs with 3'-alkyl or 3'-halo substituents of approximately equal lipophilicities. The relative binding affinities and physical properties of 3',4'-substituted analogs support the model of the 4'-hydroxyl functioning as a proton donor which is directed "trans" to the 3'-position. The contribution of the 4'-hydroxyl to the ΔG of binding to solubilized rat hepatic nuclear protein receptors can be calculated (Eqn. 2-2) as the difference in ΔG values for binding of a 4'-deoxy analog and the corresponding 4'-hydroxy compound. Such ΔG values have been calculated[26,43] (for 4-36) as -1.23, -1.61, and -1.91 kcal/mole for $R_{3'}$ = H, CH_3, and tBu, respectively.



4-36

This is in agreement with the results that would be expected if the increasing bulk of these 3'-alkyl substituents were orienting the 4'-hydroxyl toward the 5'-position and a proton acceptor on the receptor. Similarly, the calculated ΔG contribution of the 4'-hydroxyl group with 3'-monohalo substitution is inversely related to the strength of the intramolecular hydrogen bond that would have to be broken in order to orient the 4'-hydroxyl toward the 5'-position (ΔG (4'-OH) = -1.47, -2.03, -2.48, and -3.60 kcal/mole for 3'-F, 3'-Cl, 3'-Br, and 3'-I, respectively). That the 3'-halogen substituents (especially Br and I) enhance the ΔG for the 4'-OH group suggests that the 3'-halogens are significantly enhancing the 4'-OH proton donor ability (consistent with the ab initio results). A receptor proton acceptor on the 3' side is unlikely because bulky 3' substituents (e.g., Br or I) would certainly interfere with any hydrogen bond between the 4'-OH and such a receptor group. (One can not exclude the possibility that the hydrophobic interactions of the 3'-substituent with the receptor are inducing small conformational changes in the receptor such that there is a direct cooperativity in binding between the 3'-hydrophobic interaction and the 4'-hydrogen bond.) With this model, it becomes plausible to ascribe at least part of the intolerance of the nuclear receptor to 3', 5' disubstituted compounds to steric interference of the 5'substituent with the 4'-OH - - - receptor hydrogen bond. Our theoretical studies predict strongly repulsive potentials for bulky 5'-substituents if $\emptyset$ (see 4-33) = 0° (i.e., for a colinear hydrogen bond). Although 5'-bulk is detrimental for in vivo activity and in vitro binding, even a group as large I causes only a 5 to 8 fold loss in activity (a ΔG loss for binding of $\sim$ 1.27 kcal/mole). This suggests that the receptor proton

acceptor is probably so oriented that steric interaction with 5'-substituents is somewhat reduced, e.g., by orientation either slightly out of the phenolic ring plane and/or with $\phi > 0°$ (see 4-33). It appears that a 5'-substituent sterically interferes with either hydrogen bond formation between the 4'-hydroxyl and receptor and/or also decreases activity by direct repulsive steric interaction with the receptor.

In order for the model of the 4'-hydroxyl as a proton donor directed towards the 5'-postion to be acceptable, it must also be able to account for the low *in vitro* binding affinities of various 4'-OCH_3, 4'-H, and 4'-NH_2 analogs.[24,26] Benzene/H_2O dimers (to model the 4'-H substituent) (optimal phenol/H_2O geometries for 4-33 and 4-34 with H then replacing OH) gave (CNDO/2 calculations) only very small repulsive interactions for H_2O as proton donor (-0.69 kcal/mole) or as proton acceptor (0.27 kcal/mole). With CNDO/2, anisole (to model the 4'-methoxy substituent) was found to be as good a proton acceptor as phenol. Then using 4-33, (X=Y=H), replacing the OH proton with a staggered CH_3 group, the dimer gives rise to repulsive interactions, the magnitude of which depend on R and the CH_3 conformation. Experimentally 4'-OCH_3 analogs are generally found to bind with affinities similar to those of the corresponding 4'-H analogs, in agreement with a 4'-OH donor model.

The ΔE for aniline (to model the 4'-NH_2 substituent) as proton donor to H_2O was calculated by CNDO/2 to be -3.85 kcal/mole with the NH_2 group coplanar with the aromatic ring and with $\theta_{HNH} = 120°$. The microwave spectrum of aniline,[203] however, predicts that the NH_2 group adopts an out-of-plane angle of $\sim 37.5°$ with $\theta_{HNH} \tilde{\sim} 113.1°$. With the microwave spectrum geometry for aniline[203] and with H_2O as a proton

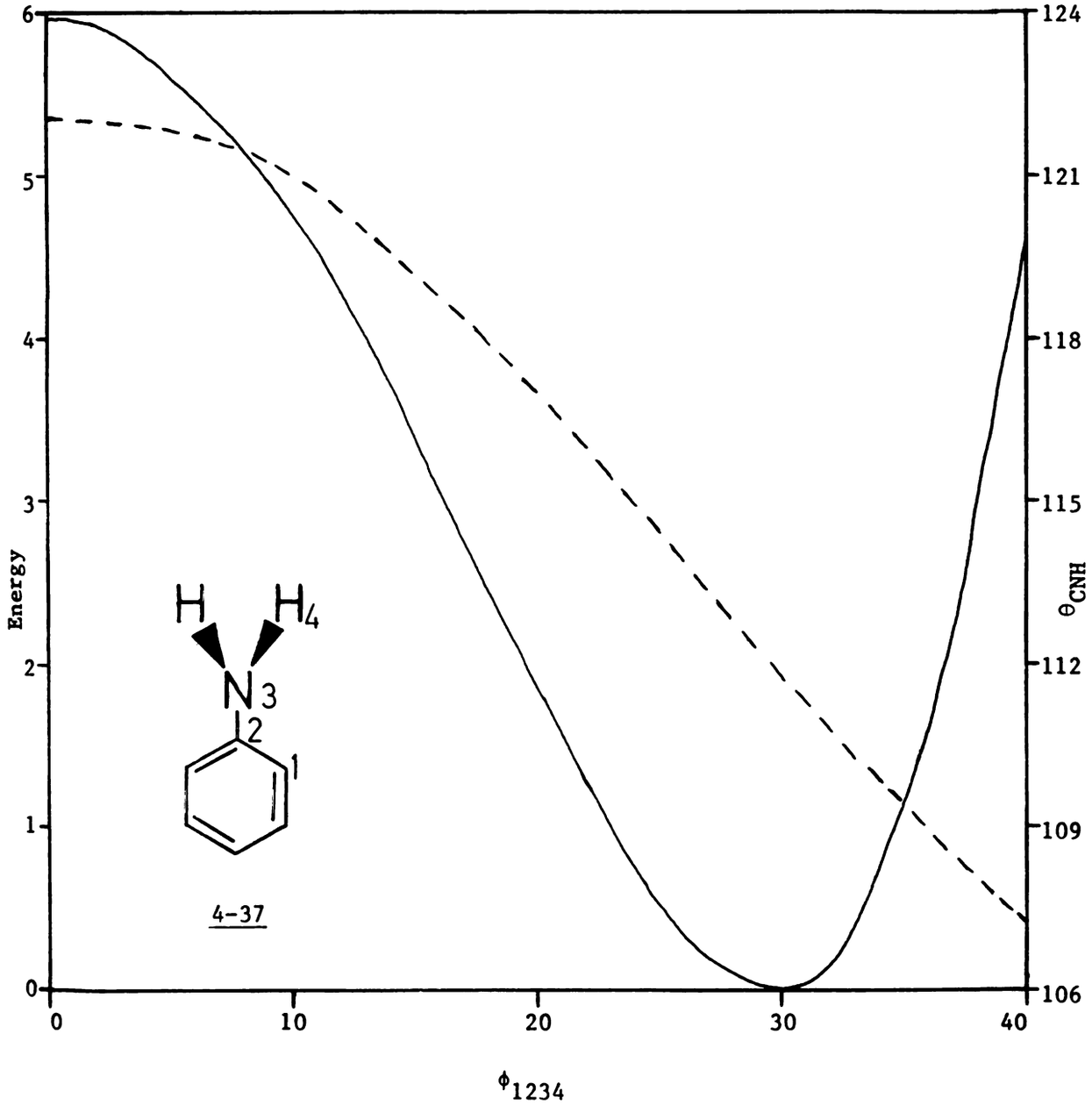


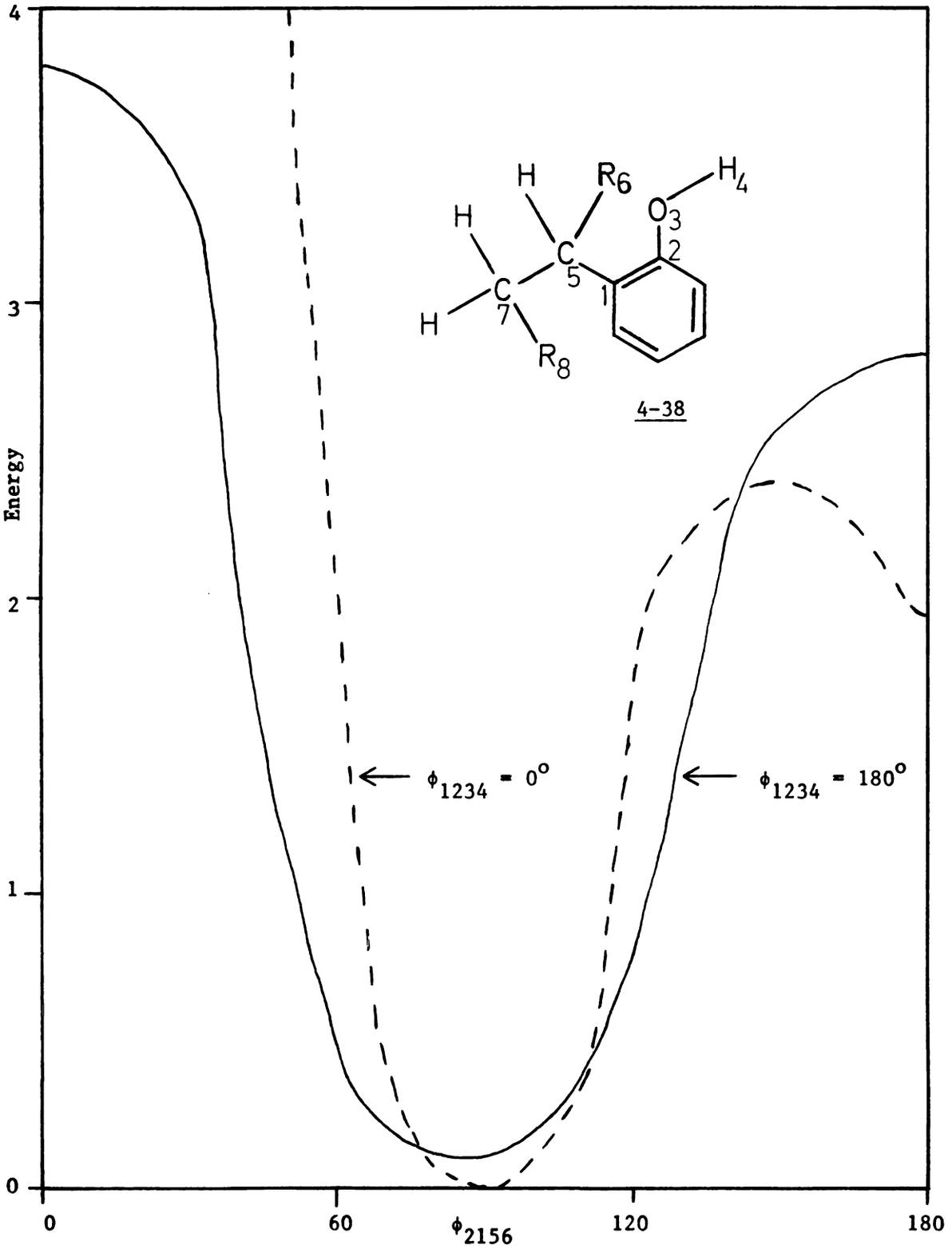
Figure 4-6. CNDO/2 dependence of Energy (relative, in kcal/mole; ———) and θ_{CNH} (in degrees; -----) of aniline (4-37) on ϕ_{1234} (in degrees).

acceptor (but with the H_2O oxygen coplanar with the ring plane and with the closes N-H bond), CNDO/2 predicts a ΔE for the dimer of only -2.56 kcal/mole. The CNDO/2 energy and θ_{CNH} profile for inversion of the NH_2 group of aniline is given in Fig. 4-6. Assuming the 4'-OH proton donor model, it is clear why the 4'-NH_2 binds less strongly than 4'-OH, but it is not clear at this point why the 4'-NH_2 compound should bind <u>less</u> strongly than the 4'-OH or 4'-OCH_3 to nuclear receptors. A reasonable model pictures a receptor proton acceptor location which is out of the outer ring plane. This allows a weak O-H - - - H bond, but an H . . X distance which is too short for 4'-NH_2,resulting in steric repulsion.

Steric bulk in the 3'-position is disadvantageous for nuclear or TBG binding, but only for substituents which extend out from the molecule a distance greater than about the group size of I.[25,26,56] This suggests that the 3'-substituents do not sterically interfere with an interaction of the 4'-position with receptor or TBG, but rather that the 3'-substituent binds in a size-limited pocket approximately the size of I. However, since the 3'-substituent might influence the cis-trans isomerism of the 4'-OH, we carried out a conformational analysis for several ortho-alkyl-phenols. The results of these conformational analyses are presented in Fig. 4-7 through 4-9 (see also the intramolecular hydrogen bonding studies above). The most striking feature of the conformational energy maps and our previous calculations on the ortho-alkylphenols is that the lower energy conformations prefer $\phi_{2156} > 90° \pm 30°$ (although favoring $\phi_{2156} > 90°$) with ϕ_{1578} fairly unrestricted within this ϕ_{2156} range. Branching on the carbon alpha to the ring tends to increase the energetic favorability of conformations with $\phi_{2156} > 90°$, as

would be expected. In general, the alkyl groups tend to extend up and away from the hydroxyl group, although no great preference is seen for fully extended conformations. With the OH cis to the alkyl group ($\phi_{1234} = 0°$), conformations with $\phi_{2156} < 60°$ tend to be completely excluded. The conformational preference of alkyl chains to orient up and away from the phenolic hydroxyl could be detrimental with respect to binding in that these very orientations could be such that the chains would sterically be interacting with the 3'-pocket receptor surface. Any 3'-substituent with more than a 2 carbon chain (e.g., 3'-n-propyl) is significantly less tightly bound to nuclear receptor than those of similar total size (3'-iPr), with only 2 carbon extensions from the 3'position.

Further experiments will be interesting in order to test this "picture" of the receptor. One might test 4'-H or F, 5'-OH or CH_2OH compounds in order to ascertain whether 5'-substituents interfere with 4'-H-bonding or because there is steric repulsion with the receptor. The fact that 4'-OCH_3 is not less tightly bound than 4'-H suggests the receptor H bond acceptor can move to relieve steric interactions with the methyl group. The relative inactivity of the 4'-NH_2 group is surprising, and it might be interesting to test 4'-N(alkyl)$_2$ to completely remove the proton donor functionality. If our reason for the inactivity of the 4'-NH_2 is correct, 4'-$N(CH_3)_2$ might not further decrease binding.



4-38

Figure 4-7. Conformational CNDO/2 Energy (relative: in kcal/mole) dependence for o-ethylphenol (4-38: $R_6 = R_8 = H$: CH_3 staggered).

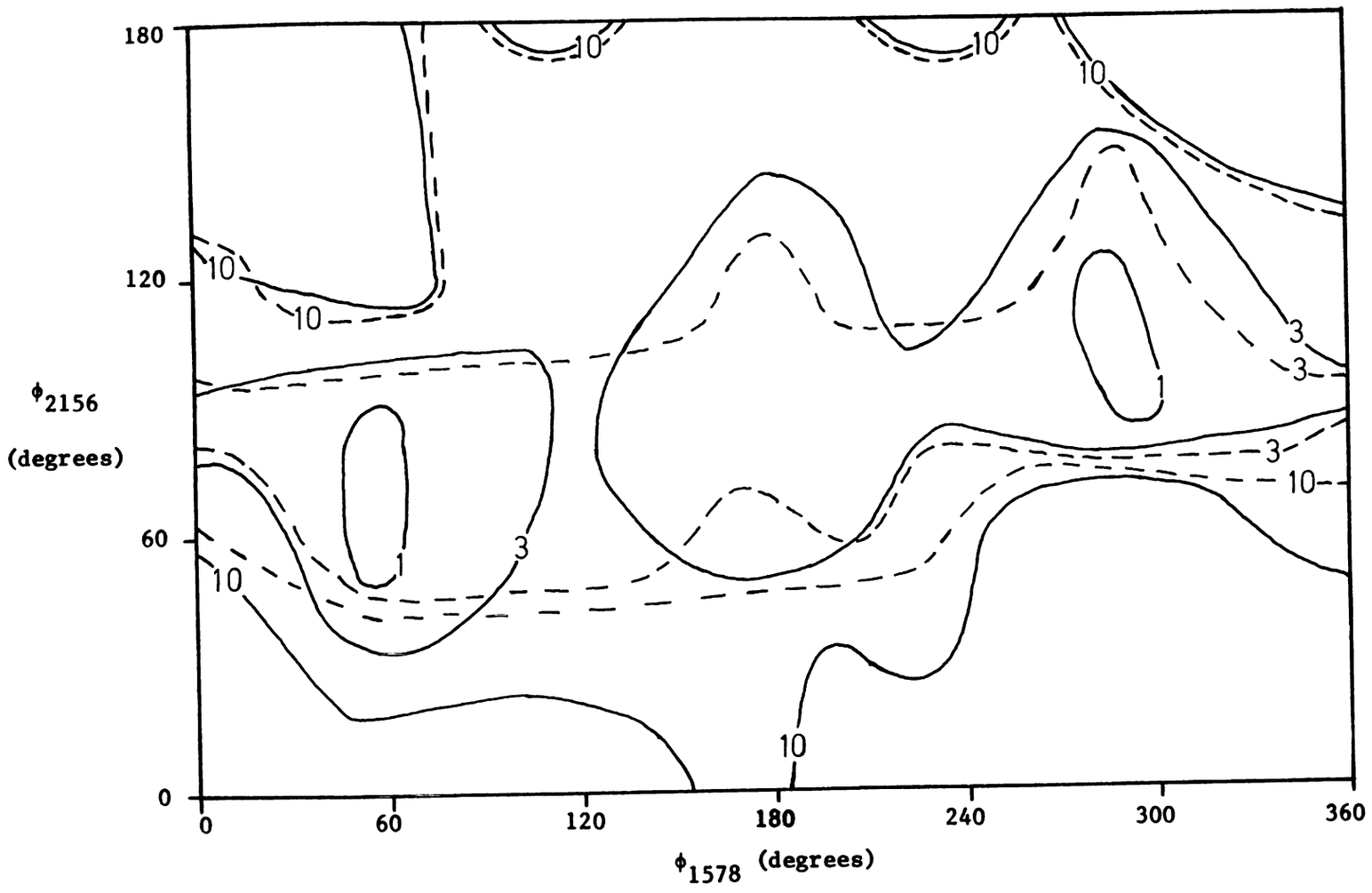


Figure 4-8. Conformational CNDO/2 energy map for ortho-n-propylphenol (4-38: R_6 = H; R_8 = staggered CH_3). $\phi_{1234} = 0^o$ = -----. $\phi_{1234} = 180^o$ = ——. Energies in kcal/mole.

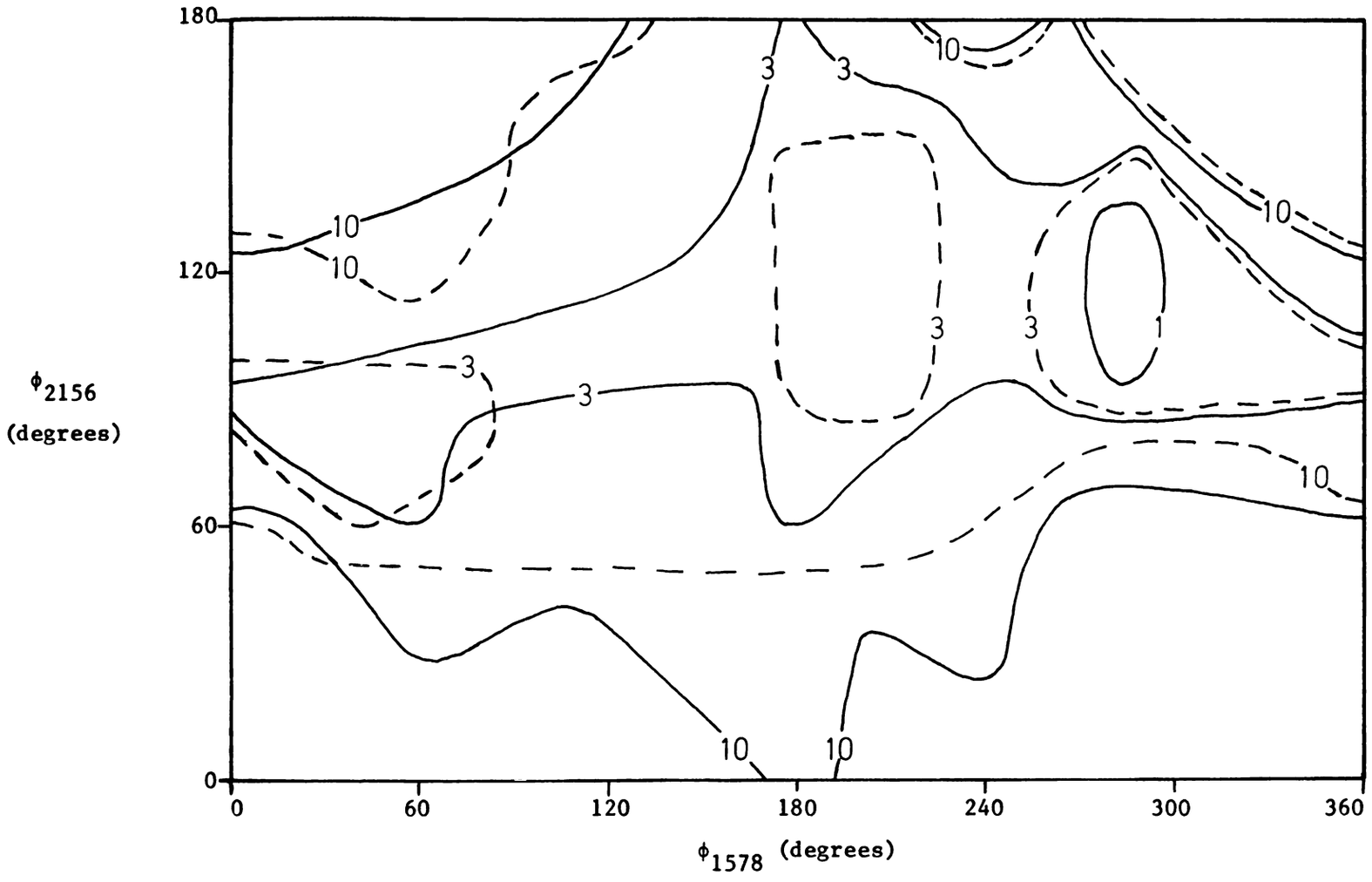


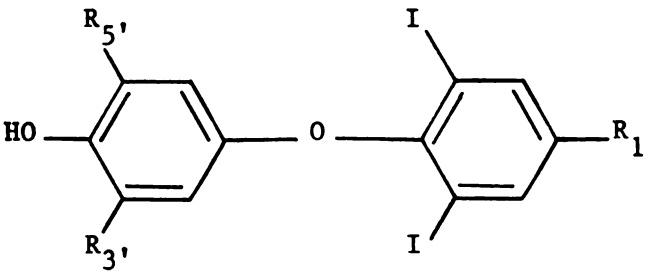
Figure 4-9. Conformational CNDO/2 energy map for ortho-sec-butylphenol (4-38: $R_6 = R_8$ = staggered CH_3). $\phi_{1234} = 0^o$ = -----. $\phi_{1234} = 180^o$ = ———. Energies in kcal/mole.

CONFORMATION OF THE ALANINE SIDE CHAIN

The length, stereochemistry, and associated charges of the 1-position side chain of the thyroid hormones and analogs are extremely important in determining in vivo (see Tables I-1, I-8 and 1-9 of Appendix I) and in vitro (see Table 4-34) thyromimetic activities. Many of the in vivo activity differences can be ascribed[8,85] to differences in metabolic susceptibilities and clearance rates of the different side chain variations. However, the in vitro activities (Table 4-34) clearly indicate that there are inherent differences in the binding of the various side chains of thyroid hormone analogs to nuclear receptors and to TBG. Hence a preliminary CNDO/2 conformational analysis study of the naturally occurring alanine side chain was undertaken.

For analysis of the amino acid side chain conformation, 4-OCH_3-3,5-I_2-L-phenylalanine (4-40) was used as a model system. The CH_3 and NH_4^+ groups were assumed to be staggered.[205] Conformation studies were performed utilizing variations in $\phi_1 = \phi C_6C_1C_7C_8$, $\phi_2 = \phi C_1C_7C_8N_9$, and $\phi_3 = \phi N_9C_8C_{10}O_{11}$. Taking ϕ_1 = 270° and ϕ_2 = 180° (the fully extended "transoid" conformer with the least expected steric repulsions of the NH_3^+ and COO^- with the aromatic ring), 15° variations in ϕ_3 led to a minimum energy at 345°. ϕ_3 was taken as 345° in all further calculations. 15° variations in ϕ_1 and ϕ_2 led to the energy local minima listed in Table 4-35. Besides these minima, there are two very steep minima at ϕ_1 = 60°/240° and ϕ_2 = 210° which are about 20 kcal/mole more stable than the local minimum ϕ_1 = 75°, ϕ_2 = 300°. Bonds orders show that this stabilization is due to an unexplainable attractive interaction between the carboxyl group and C_2-H and C_6-H and is apparently an artifact of the approximations of the CNDO/2 MO method. From the calculations it can be concluded that for the model system (4-40)

Table 4-34. Dependence of In Vitro Thyromimetic Activities of Thyroid Hormones and Analogs on 1-Position Side Chain.

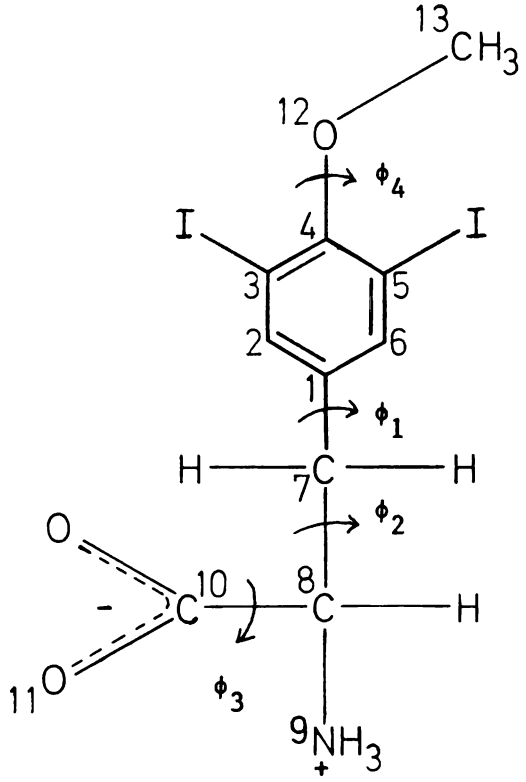


4-39

R_1	$R_{3'}$	$R_{5'}$	Binding to Intact Rat Hepatic Nuclei (% L-T_3)[a]	Binding to TBG (% L-T_4)[b]
CH_2COOH	I	H	100	---
CH_2COOH	I	I	5	3.6
CH_2CH_2COOH	I	H	---	0.3
CH_2CH_2COOH	I	I	---	1.7
D-$CH_2CH(NH_2)COOH$	I	H	70	---
L-$CH_2CH(NH_2)COOH$	I	H	100	9.0
D-$CH_2CH(NH_2)COOH$	I	I	---	54
L-$CH_2CH(NH_2)COOH$	I	I	12.5	100
L-$CH_2CH(NHCOCH_3)COOH$	I	I	---	25.0

[a]References 24 and 204.

[b]References 30.



4-40

studied: (1) there is no great preference for either a cisoid ($\phi_1 \overset{\sim}{\sim} 90°$) or a transoid ($\phi_1 \overset{\sim}{\sim} 270°$) conformation; (2) ϕ_2 may assume any of the expected staggered values of approximately 60°, 180°,or 300°; (3) $\phi_3 \overset{\sim}{\sim} 345°$; and (4) the various ϕ_1, ϕ_2 conformers are readily interconvertible: barriers between ϕ_1 conformers $\lesssim$ 7.5 kcal/mole and between ϕ_2 conformers $\lesssim$ 4 kcal/mole. These theoretical results are in agreement with a large number of X-ray crystallographic studies[10-14,206-208] of aromatic amino acids and thyroid hormone analogs, with the small (and not unexpected) exception that the transoid conformation is usually (but not always) observed. The MO and X-ray studies emphasize the fact that the amino acid side chain can probably assume whatever conformation is required to maximize binding interactions. Further CNDO/2 studies, attempting to lock the side chain conformation by use of 2-CH_3 or 2,6-$(CH_3)_2$ groups were also performed. Unfortunately, again because of the spurious COO^-/C_2 and C_6 attractions, it was impossible to estimate quantitatively from these calculations the ability of these CH_3 groups to lock or restrict the side chain conformations (except where direct steric interactions result). Despite this, it would be useful to synthesize and test the binding activities of 2- and 6-substituted analogs, especially with regards to the ability of these substituents to restrict the side chain conformations.

Table 4-35. CNDO/2 Conformational Energy Local Minima of Alanine Side Chain (4-40) (Energy in kcal/mole).

ϕ_1	ϕ_2	E
105	60	3.12
285	60	2.87
90	180	$\sim 0^a$
270	180	$\sim 0^a$
75	300	0.00
255	300	0.19

[a]See text.

CHAPTER FIVE: QUANTITATIVE STRUCTURE-ACTIVITY CORRELATIONS

Although the structure-activity relationships of the thyroid hormones and analogs have been investigated most extensively over the last 30 years through the synthesis and testing of approximately 500 analogs,[52] with very few exceptions such studies have involved qualitative rather than quantitative evaluation of activities. This lack of quantitative structure-activity studies has apparently been due to: (1) lack of a substantial number of consistently reliable activities; (2) enormous variability of assay types; and (3) the complexity and number of physico-chemical properties which affect the activities of the thyroid hormone analogs. The first two of these deficiencies have been in part eliminated by the extensive compilation and re-evaluation of *in vivo* activities (reference 2 and 8, and Appendix I), as well as by the recent extensive determination of accurate *in vitro* binding affinities to nuclear receptors[24-26,43] and plasma proteins.[30] The

third set of problems has step-by-step been clarified by series of experimental, classical analog, and theoretical studies (as outlined in Chapters One and Two), and hopefully in part by this study.

A large number of quantitative structure-activity relationship correlation studies of the various activities of thyroid hormones and analogs were undertaken in the course of these studies. For reasons of clarity and of space limitations, however, only the results of those studies representing final QSAR models or developmental stages of the correlations are presented in this chapter. Listings and origins (where appropriate) of substituent parameters used in the correlation studies are presented first. Rat antigoiter bioassay activity correlations are the first examined, since all previous QSAR studies of the thyroid hormone analogs have involved *in vivo* activities. After presentation of the previously uninvestigated area of using QSAR methods to analyze various *in vitro* binding affinities, correlations between *in vivo* activities and *in vitro* binding affinities are presented.

SUBSTITUENT PARAMETERS AND COMPUTATIONAL DETAILS

Values for the substituent parameters utilized in these quantitative structure-activity relationship studies are presented in Table 5-1.

For an electronic parameter for 3' and 5' substituents, σ_p was utilized, under the assumption[92] that the electronic effect of a substituent on an ortho position should be comparable to that on the para. $\sigma_{3',5'} = \sigma_{3'} + \sigma_{5'}$. As the electronic effects of 3' and 5' substituents were assumed to be expressed through the ionization and/or hydrogen bonding of the 4'-OH, $\sigma_{3',5'}$ was set equal to 0.0 for 4'-H and 4'-OCH_3 analogs for *in vitro* assays. This was not done for

correlations involving in vivo activities, since it was assumed 4'-H and 4'-OCH_3 analogs would be metabolized to the corresponding 4'-OH analogs in vivo.

In agreement with the results of Kubinyi,[56] choice of a system for π values was not found to be crucial to the overall results of the correlations; using π values from different systems did not substantially change the equations derived. The number of known π substituent constants is largest for the benzene system (π_{BZ}), thus requiring the fewest estimations of unknown values. Hence, unless otherwise noted, $\pi = \pi_{BZ}$ was used for all 3, 5, 3', and 5' substituents. $\pi_{35} = \pi_3 + \pi_5$. $\pi_{3'5'} = \pi_{3'} + \pi_{5'}$.

I2' = an indicator variable for 2' substitutions (there was not enough variation in 2' substituents for use of a steric or hydrophobic parameter)

= 0 for 2' substituent = H

= 1 for 2' substituent not = H (including 2',3'-$(CH)_4$)

I4'H = an indicator variable for 4'-H analogs

= 1 for 4' substituent = H

= 0 for 4' substituent not = H

I4'OCH_3 = an indicator variable for 4'-OCH_3 analogs

= 1 for 4' substituent = OCH_3

= 0 for 4' substituent not = OCH_3

INTERACT is a parameter derived from experimental data and MO calculations and is an estimate of the free energy change (in kcal/mole) for orientation of the 4'-OH from cis to the 3' substituent to cis to the 5' substituent. Values of $INTERACT_{3'}$, listed in Table 5-1 are for a 3' substituent with the 5' substituent = H. $INTERACT_{5'}$ (i.e., for a 5' substituent with the 3' substituent = H) = $-INTERACT_{3'}$.

Table 5-1. Substituent Parameters Used in Structure-Activity Correlations

Substituent	σ[a]	π_{3PA}[b]	π_{BZ}[c]	$INTERACT_{3'}$
H	0.00	0.00	0.00	0.00
F	0.06	---	0.14	1.37[d]
Cl	0.23	---	0.71	2.30[d]
Br	0.23	---	0.86	1.68[d]
I	0.18	---	1.12	0.75[d]
OH	-0.37	-0.49	-0.67	---
NO_2	0.78	---	-0.28	8.29[d]
CH_3	-0.17	0.51	0.56	-0.51[e]
C_2H_5	-0.15	0.97	1.02	---
i-C_3H_7	-0.15	1.30	1.53	-0.99[f]
n-C_3H_7	-0.13	---	1.55	-0.72[f]
i-C_4H_9	-0.12[g]	1.81[h]	2.00[h]	---
s-C_4H_9(±)	-0.12[g]	---	2.00[i]	-1.01[f]
t-C_4H_9	-0.20	1.68	1.98	-1.57[e]
c-C_6H_{11}	-0.22	---	2.51	---
C_6H_5	-0.01	1.89	1.96	---
CF_3	0.54	---	0.88	---
2',3'-$(CH)_4$	0.04	---	0.99[j]	---

[a] σ_p values from reference 57 unless otherwise noted.

[b] From the 3-substituted phenoxyacetic acid system; from reference 209 unless otherwise noted.

[c] From the benzene system; from reference 57 unless otherwise noted.

Table 5-1. (Continued)

Substituent	Es^k	3'SIZE > I
H	1.24	0.0
F	0.78	0.0
Cl	0.27	0.0
Br	0.08	0.0
I	-0.16	0.0
OH	---	0.0
NO_2	---	0.0
CH_3	0.0[1]	0.0
C_2H_5	---	0.127
i-C_3H_7	0.47[1]	0.253
n-C_3H_7	---	0.405
i-C_4H_9	---	1.160
s-C_4H_9(±)	---	0.707
t-C_4H_9	---	0.920
c-C_6H_{11}	---	2.46
C_6H_5	---	2.22
CF_3	---	0.0
2',3'-$(CH)_4$	---	0.0

[d] CNDO/2 estimate; see Tables 4-5 and 4-16.

[e] Experimental value; see Table 4-16.

[f] CNDO/2 interpolation between Me and tBu experimental values; see Table 4-16.

Table 5-1 (Continued)

---

[g]From reference 210.

[h]Estimated.

[i]Estimate from reference 56.

[j]0.99 = 3/4(1.32) for 2',3'-$(CH)_4$ since it was assumed that only approximately 3 of the 4 carbons could be fitting into the 3' substituent hydrophobic pocket.

[k]From reference 211 unless otherwise noted.

[l]From reference 212.

Thus, INTERACT (3',5'disubstitution) = $INTERACT_{3'}$ + $INTERACT_{5'}$. Since the "INTERACT" effects of 3' and 5' substituents were assumed to be expressed through their influencing the hydrogen bonding capabilities of the 4'-OH by virtue of their orienting capabilities, INTERACT was set equal to 0.0 for 4'-H and 4'-OCH_3 analogs for in vitro correlations.

The parameter 3'SIZE > I (based on bond distances, Van der Waals radii, and conformational considerations) is an estimate of the average distance a 3' substituent extends out from the 3' position further than iodine. Iodine and smaller 3' substituents were assigned values of 0.0 for this parameter. Utilizing the bond distances used in our MO calculations (Chapter Four), an estimate of 2.0 Å [178] for the Van der Waals radii of a CH_2 or CH_3 (r_{VW,CH_3}) group, and Van der Waals radii of Bondi,[179] 3'SIZE > I values were calcualted as follows. First, the distance (r_z) was calculated to the furthest out non-hydrogen atom from the 3' carbon atom. For 3' substituent = H, r_z = 1.09 Å. For 3' substituent = OH, r_z = distance to H. For 3' substituent = cHex, r_z = distance to $C_{4''}$ for a $C_{1''}$-equitorially-substituted cHex. For 3' substituent = Ph, r_z = distance to H on $C_{4''}$ for a $C_{1''}$-substituted Ph. The appropriate heteroatom, H, or CH_2 Van der Waals radius was then added to r_z to give $r_{3'}$ = approximate average Van der Waals size of a 3' substituent extending out from the 3' carbon. 3'SIZE > I values were calculated as follows:

$$3'SIZE > I = r_{3'} - r_{3'}(I)$$
$$= r_{3'} - 4.12 \text{ Å}$$
$$= 0.0 \text{ if } r_{3'} - 4.12 \text{ Å} < 0.0$$

For acyclic 3' alkyl substituents, 3'SIZE > I values were calculated in a slightly different manner in order to take into account conformational flexibility:

$r_z(n)$ = Calculated distance to the furthest carbon for an alkyl chain n carbons long in a fully extended, staggered conformation

$r_{3'}(n)$ = Calculated $r_{3'}$ value for an alkyl chain n carbons long in a fully extended, staggered conformation

$= r_z(n) + r_{VW,CH_3}$

$= r_z(n) + 2.0$ Å

3'SIZE > I(n) = Calculated 3'SIZE > I value for an alkyl chain n carbons long in a fully extended, staggered conformation

$= r_{3'}(n) - r_{3'}(I)$

$= r_{3'}(n) - 4.12$ Å

$= 0.0$ if $r_{3'} - 4.12$ Å ≤ 0.00

In particular, calculated values are:

3'SIZE > I(n=1) = 0.00 Å

3'SIZE > I(n=2) = 0.38 Å

3'SIZE > I(n=3) = 1.74 Å

A three-fold conformational rotation for branching at the carbon α to the 3' carbon was then examined in order to calculate the <u>average</u> distance a particular 3' acyclic alkyl substituent extends out from the 3' carbon. For example, a 3' CH_3 substituent, for the three staggered α-carbon rotamers, extends out (in a particular direction) as a carbon chain n=1 carbons long for all three rotamers. Hence,

3'SIZE > I(CH_3) = [3'SIZE > I(n=1) + 3'SIZE > I(n=1) +3'SIZE > I(n=1)]/3

= (0.00 Å + 0.00 Å + 0.00 Å)/3

= 0.00 Å

A 3' Et substituent, for the three staggered α-carbon rotamers, extends out (in a particular direction) as a carbon chain n=1 carbons long for 2/3 of the rotamers and as a carbon chain n=2 carbons long for 1/3 of the rotamers. Hence:

$$3'\text{SIZE} > I(\text{Et}) = [3'\text{SIZE} > I(n=1) + 3'\text{SIZE} > I(n=1) + 3'\text{SIZE} > I(n=2)]/3$$

$$= (0.00\ \text{Å} + 0.00\ \text{Å} + 0.38\ \text{Å})/3$$

$$= 0.127\ \text{Å}$$

A 3'iPr substituent extends out as a carbon chain n=1 carbons long for 1/3 of the rotamers and as a carbon chain n=2 carbons long for 2/3 of the rotamers. Hence:

$$3'\text{SIZE} > I(\text{iPr}) = [3'\text{SIZE} > I(n=1) + 3'\text{SIZE} > I(n=2) + 3'\text{SIZE} > I(n=2)]/3$$

$$= (0.00\ \text{Å} + 0.38\ \text{Å} + 0.38\ \text{Å})/3$$

$$= 0.253\ \text{Å}$$

Similarly:

$$3'\text{SIZE} > I(\text{nPr}) = [3'\text{SIZE} > I(n=1) + 3'\text{SIZE} > I(n=1) + 3'\text{SIZE} > I(n=3)]/3$$

$$= (0.00\ \text{Å} + 0.00\ \text{Å} + 1.74\ \text{Å})/3$$

$$= 0.580\ \text{Å}$$

$$3'SIZE > I(sBu) = [3'SIZE > I(n=1) + 3'SIZE > I(n=2) + 3'SIZE > I(n=3)]/3$$

$$= (0.00\ \text{Å} + 0.38\ \text{Å} + 1.74\ \text{Å})/3$$

$$= 0.707\ \text{Å}$$

$$3'SIZE > I(iBu) = [3'SIZE > I(n=1) + 3'SIZE > I(n=1) + 2*3'SIZE > I(n=3)]/3$$

$$= (0.00\ \text{Å} + 0.00\ \text{Å} + 2*1.74\ \text{Å})/3$$

$$= 1.160\ \text{Å}$$

For 3'-iBu, the 3'SIZE > I(n=3) value was multiplied by two to take into account the branching at the β-carbon atom

$$3'SIZE > I(tBu) = [3'SIZE > I(n=2) + 3'SIZE > I(n=2) + r_{VW,CH_3}]/3$$

For 3'-tBu, r_{VW,CH_3} = 2.0 Å was used in place of 3'SIZE > I(n=2) for the third α-carbon branch. The 3'-tBu is the only 3'-alkyl substituent with a third non-hydrogen α-carbon branch, and apparently this extra steric bulk, by interaction with the receptor and/or with the 4'-OH, adds an extra negative steric influence to this group. Although this is admittedly a "fudging" factor to account for the apparent extra steric bulk of the 3'-tBu substituent, the necessity

of its inclusion to "fit" the relative activity of this substituent provides additional insight into the strict conformational and size requirements of 3' substituents; i.e., the third non-hydrogen α-carbon branch of this substituent does add an additional negative steric interaction beyond that attributable to the average distance the 3' substituent extends out from the 3' carbon further than I (see below).

All regression correlations of this study were performed with PROGRAM QSAR47 (Appendix II), utilizing standard multiple regression techniques. Details of the computational methods are presented in the program documentation (Appendix II). For presentation of the regression equations:

Values in parentheses after regression coefficients are 95% confidence intervals.

R = the multiple least squares regression coefficient.

N = the number of data points used in the calculation of the regression equation.

S = the overall standard deviation of the regression.

$F_{DFN,DFD}$(calcd) = the calculated F statistic value for DFN degrees of freedom in the numerator and DFD degrees of freedom in the denominator.

(Z%) = the % confidence level at which $F_{DFN,DFD}$(calcd) is significant.

= (<75%) for F(calcd) < F(75%).

= (>99.9%) for F(calcd) > F(99.9%).

For F(X%) $\leq$ F(calcd) < F(Y%), (Z%) is obtained by extrapolation linearly with log (%):[90]

$$\log(Z) = \log(Y) + [\log(X) - \log(Y)][\frac{F(calcd) - F(Y)}{F(X) - F(Y)}]$$

On the basis of the qualitative in vivo structure-activity relationships of the thyroid hormone analogs (in particular, only 3' substituent "size" > I decreases activity, all 5' substituent "size" decreases activity), for analogs with $R_{3'} \neq R_{5'}$, the larger substituent was assumed to be the 3' substituent and the smaller substituent was assumed to be the 5' substituent.

RAT ANTIGOITER BIOASSAY ACTIVITIES

Except for the correlations of this study, there have been few other[56,91-94] quantitative SAR studies of in vivo activities of the thyroid hormones and analogs and none using in vitro activities. These previous studies, however, did not specifically restrict analysis to data from a single assay type in a single animal type: the studies examined thyroxine-like activity in amphibia,[91] mammalia,[91] rodents,[92,93] and the rat.[56,94] Although various metabolic, antigoiter, and metamorphosis activities are often similar, significant deviations do occur between specific assay types,[2,8] In the previous studies, DL corrections were simply L = 2x DL or were not made at all, and DL- or L-T_4 was used as the reference compound. In addition (and most certainly unknown to the previous investigators) activities had not been corrected from a weight to a molar basis. In order to conduct our studies with the largest, most accurate possible set of experimental data for a single assay type, we conducted our in vivo quantitative SAR correlations solely with rat antigoiter bioassay activities. Unless otherwise noted,

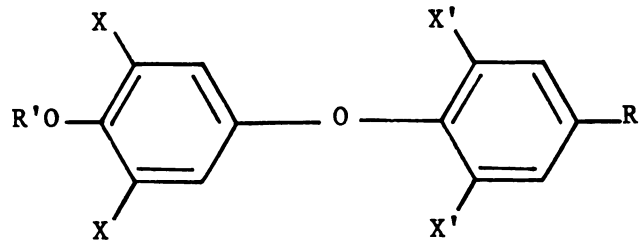
all activities were corrected to a molar basis, and for comparison of L-analogs with L-T_3 as the reference compound (see Chapter Three for details).

The first attempt to quantitatively study thyromimetic activity (and one of the early quantitative studies of structure-activity relationships) was a study of Bruice, et al.,[91] who derived equations relating thyroxine-like activity in amphibia and mammalia of the type of Eqn. 5-1:

$$\log (\% \text{ thyroxine-like activity}) = k \Sigma f + c \qquad \text{(Eqn. 5-1)}$$

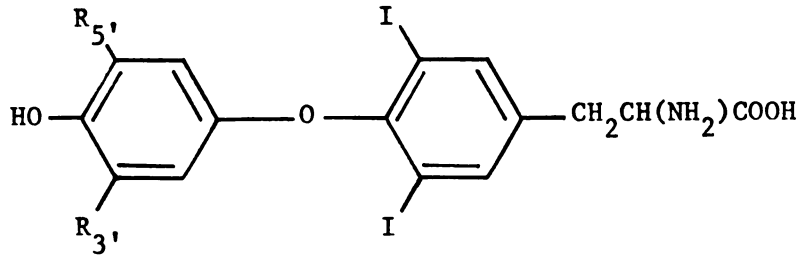
where $\Sigma f = f_X + f_X + f_{X'} + f_{X'} + f_{OR'}$

and f_X, $f_{X'}$, and $f_{OR'}$ are empirical constants for 5-1.



5-1

For the action of thyroxine analogs on rodents, Hansch and Fujita[92] developed Eqn. 5-2 for structure 5-2. R_3, and R_5, = various halogen



5-2

combinations. π values are from the the 2-substituted phenol system. $\sigma = \sigma_p$ values. Thyroxine-like activity = A, relative to L-T_4 = 100.

$$\log (A) = -1.134 (\pi_{3',5'})^2 + 7.435 \pi_{3',5'} - 16.323 \sigma_{3',5'} - 0.287$$

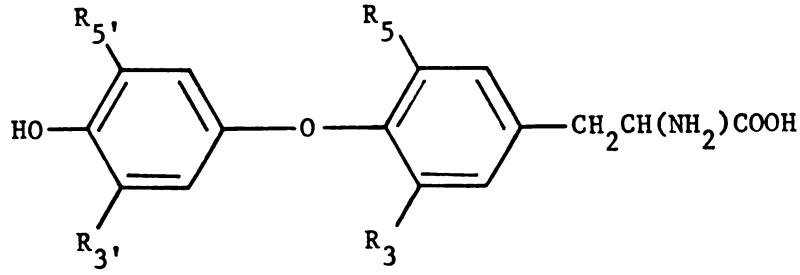
(Eqn. 5-2)

N = 9 R = 0.884 S = 0.660

Although the accuracy of biological data available and our conception of the SAR of thyroid hormone analogs have both changed immensely since this study, it was in part on the basis of Eqn. 5-2 (predicting activity to be optimal for moderately lipophilic, electron-donating 3',5' substituents) that more extensive examination of the thyromimetic activities of 3',5' alkyl-substituted analogs was encouraged.

Much more recently Kubinyi[56,94] developed a large series of equations considering thyroxine-like activity of thyroxine analogs in the rat, as an example of a mixed approach to quantitative structure-activity relationships based on Hansch and Free-Wilson analysis. For

analogs of the type 5-3, they utilized: π values from the benzene system: σ_p values for 3',5' substituents



5-3

: $ES_{3'}{}^{corr} = Es_{3'} - Es_I$

$= 0$ for values > 0 (i.e., for substituents with $Es_{3'} > E_I$)

= an approximate measure of 3' substituent size > I (see later discussion).

: $Es' = Es_{5'} + Es_{3'}{}^{corr}$

= sum of 3' and 5' steric influences on activity.

: [I] and $[CH_3]$ = Free-Wilson parameters for group contributions of I and CH_3, respectively, based on $a_{Br} = 0.00$.

With these parameters, their results can be summarized with Eqns. 5-3 through 5-5 for a wide variety of substituent types.

$$\log(A) = +1.673\ (\pm 0.324)\ \pi_{3'5'} - 1.242\ (\pm 0.969)\ \sigma_{3'5'} + 1.714\ (\pm 0.600)\ Es_{3'}{}^{corr} + 0.856 \qquad \text{(Eqn. 5-3)}$$

N = 10 R = 0.984 S = 0.201

$$\log(A) = +\ 1.908\ (\pm 0.517)\ \pi_{3',5'} - 2.151\ (\pm 1.517)\ \sigma_{3',5'} + 1.871\ (\pm 0.700)\ Es_{5'} - 1.598 \qquad \text{(Eqn. 5-4)}$$

$$N = 13 \qquad R = 0.946 \qquad S = 0.347$$

$$\log(A) = +\ 1.569\ (\pm 0.251)\ \pi_{3',5'} - 1.582\ (\pm 0.555)\ \sigma_{3',5'} + 1.493\ (\pm 0.299)\ Es' + 0.176\ (\pm 0.159)\ [I] - 0.563\ (\pm 0.195)\ [CH_3] - 1.348 \qquad \text{(Eqn. 5-5)}$$

$$N = 23 \qquad R = 0.965 \qquad S = 0.250$$

All three of these equations predict *in vivo* thyromimetic activity to be proportional to the sum of 3' and 5' substituent lipophilicities and electron-donating capabilities. In addition it is predicted that:

(1) Utilizing Eqn. 5-3 and 5-3 ($R_3 = R_5 = I$; $R_{5'} = H$), 3' substituent steric bulk greater than iodine (estimated by $Es_{3'}^{corr}$) reduces activity.

(2) Utilizing Eqn. 5-4 and 5-3 ($R_3 = R_5 = I$; $R_{3'}$ = substituents not sterically "larger" than I), any 5' substituent bulk (estimated by $Es_{5'}$) reduces activity.

(3) Utilizing Eqn. 5-5 and 5-3, activity is reduced by the sum of steric bulk of 3' substituents larger than iodine and of 5' substituents (estimated by Es') and is of the order I > Br > CH_3 for R_3 and R_5 substituents.

The QSAR studies of Hansch[92] and Kubinyi[56,94] are important and represent the evolving understanding of *in vivo* thyromimetic activity structure-activity relationships. The choice and significance of the substituent parameters utilized in these previous studies will be examined below.

As an examination of the possibility of a parabolic dependence of *in vivo* thyromimetic activity on lipophilicity, a preliminary study[25] of *in vivo* rat antigoiter bioassay activities (BA) of 3,5-diiodo-thyronines (5-4: Table 5-2) yielded Eqn. 5-6, utilizing π values derived from the 3-substituted phenoxyacetic acid system, a simple L = 2x DL correction factor, L-T_4 = 100% as reference compound, BA values *not* corrected to a molar basis, and $R_{3'}$ = substituents all with approximately the same electronic contributions in order to restrict analysis to an inspection of the π/π^2 parabolic relationship.

$$\log(BA) = +1.358\ (\pm 0.541) + 2.405\ (\pm 1.076)\ \pi_{3'} - 1.192\ (\pm 0.652)\ \pi_{3'}^2 \qquad \text{(Eqn. 5-6)}$$

N = 8 R = 0.936 S = 0.383

log(BA) maximized for ideal $\pi_{3'}$ = 1.01

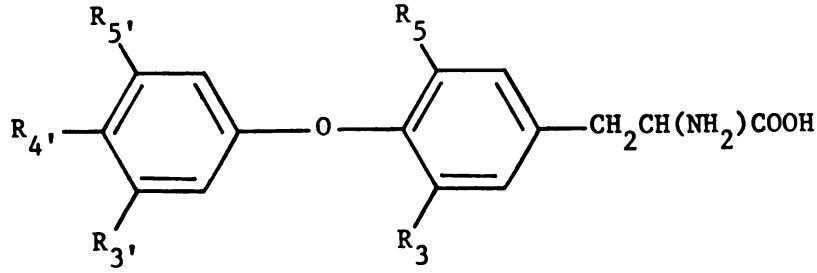
Squared independent variable cross correlation matrix

$\pi_{3'}$ -- $\pi_{3'}^2$ element = 0.847

Development of Eqn. 5-6 is presented in Table 5-3. Eqn. 5-6 is highly significant and supports the study of Hansch and Fujita,[92] which predicts a parabolic dependence of activity on compound lipophilicity.

Two factors, however, raise the possibility of whether this π/π^2 relationship truely represents a distribution phenomenon or rather a steric effect for large 3' substituents: (1) As first noted at the time of this study,[25] activity rises linearly with $\pi_{3'}$, up to $\sim\pi_{3'iPr}$, but then sharply drops for larger $\pi_{3'}$ values. This can be seen in Table 5-2 from the deviations of $\log(BA)_{calcd}$ from $\log(BA)_{obsd}$; (2) As to be shown quantitatively below, and as was observed at the time of this study,[25] binding of analogs to intact rat hepatic nuclei (where distribution should not be a factor) also peaks at$\sim \pi_{3'iPr}$ and then sharply decreases for larger $\pi_{3'}$ values.

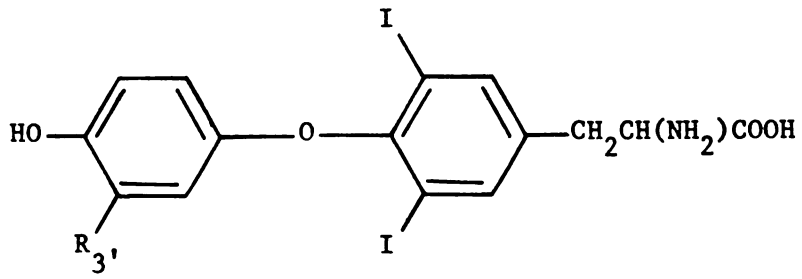
After a series of developmental studies, the essentially equivalent Eqns. 5-8 through 5-11 were developed as the simplest, most general equations for predicting in vivo thyromimetic rat antigoiter bioassay activities (BA) for structures of the type 5-5 (Table 5-4). Biological



5-5

data used for calculating the equations is presented in Table 5-5. The representative stepwise development of Eqn. 5-8 is presented in Table 5-6. The independent variable squared cross correlation matrix for the variables utilized in these equations is presented in Table 5-7.

Table 5-2. Data Used in the Formulation of Eqn. 5-6 Correlating Rat Antigoiter Activities (BA) for Thyroid Hormone Analogs (5-4).



5-4

Data Point #	$R_{3'}$	BA(%) Obsd.[a]	log(BA) Obsd.	log(BA) Calcd.[b]	log(BA) Dev.
1	OH	1.5	0.176	-0.107	0.283
2	H	10	1.000	1.358	-0.358
3	Me	85	1.929	2.274	-0.345
4	Et	550	2.740	2.569	0.172
5	iPr	1000	3.000	2.469	0.531
6	tBu	120	2.079	2.033	0.046
7	iBu	60	1.778	1.805	-0.027
8	Ph	22	1.342	1.644	-0.302

[a]Uncorrected to a molar basis. $L\text{-}T_4$ = 100%. Assuming L = 2x DL.

[b]Calculated with Eqn. 5-6.

Table 5-3. Stepwise Development of Eqn. 5-6.

Eqn. #			Eqn. and Statistical Data
5-7	log (BA)	=	+ 1.186 + 0.594 π_3,
			R = 0.572 S = 0.813
			$F_{1,6}$ = 2.91 (83.6% vs. mean)
5-6			$F_{2,5}$ = 17.62 (99.4% vs. mean)
			$F_{1,6}$ = 22.09 (99.4% vs. Eqn. 5-7)

Table 5-4. Structures of Thyroid Hormone Analogs (5-5) Used in Deriving Eqns. 5-8 through 5-11.

Data Point #	Abbreviation	R_3	R_5	$R_{3'}$	$R_{5'}$	$R_{4'}$
1	3Me-3'I-T	Me	H	I	H	OH
2	Me3-T	Me	Me	Me	H	OH
3	Me4-T	Me	Me	Me	Me	OH
4	3'iPr-Me2-T	Me	Me	iPr	H	OH
5	3'nPr-Me2-T	Me	Me	nPr	H	OH
6	3'sBu-Me2-T	Me	Me	sBu	H	OH
7	3'I-Me2-T	Me	Me	I	H	OH
*8	5Me-T1	Me	I	H	H	OH
9	5Me-33'-T2	Me	I	I	H	OH
*10	C13-T	Cl	Cl	Cl	H	OH
11	55'C12-33'-T2	I	Cl	I	Cl	OH
12	3'iPr-Br2-T	Br	Br	iPr	H	OH
13	Br3-T	Br	Br	Br	H	OH
*14	Br4-T	Br	Br	Br	Br	OH
15	3'I-Br2-T	Br	Br	I	H	OH
16	3'5'I2-Br2-T	Br	Br	I	I	OH
17	55'Br2-33'-T2	I	Br	I	Br	OH
*18	5Br-33'-T2	I	Br	I	H	OH
*19	33'-T2	I	H	I	H	OH
20	T2	I	I	H	H	OH
21	3'Me-T2	I	I	Me	H	OH
22	3'5'Me2-T2	I	I	Me	Me	OH
23	3'Et-T2	I	I	Et	H	OH

Table 5-4. (Continued)

Data Point #	Abbreviation	R_3	R_5	$R_{3'}$	$R_{5'}$	$R_{4'}$
24	3'iPr-T2	I	I	iPr	H	OH
25	3'iPr-355'-T3	I	I	iPr	I	OH
26	3'nPr-T2	I	I	nPr	H	OH
27	3'iBu-T2	I	I	iBu	H	OH
28	3'sBu-T2	I	I	sBu	H	OH
29	3'tBu-T2	I	I	tBu	H	OH
30	3'Ph-T2	I	I	Ph	H	OH
31	3'NO2-T2	I	I	NO_2	H	OH
32	3'OH-T2	I	I	OH	H	OH
33	3'F-T2	I	I	F	H	OH
*34	3'5'F2-T2	I	I	F	F	OH
*35	5'F-T3	I	I	I	F	OH
36	3'Cl-T2	I	I	Cl	H	OH
37	3'5'Cl2-T2	I	I	Cl	Cl	OH
38	3'Br-T2	I	I	Br	H	OH
*39	3'5'Br2-T2	I	I	Br	Br	OH
40	T3	I	I	I	H	OH
41	T4	I	I	I	I	OH
42	4'OCH3-3'iPr-T2	I	I	iPr	H	OCH_3
43	4'OCH3-3'tBu-T2	I	I	tBu	H	OCH_3
44	4'OCH3-T3	I	I	I	H	OCH_3

*Not used in calculating Eqns. 5-8 through 5-11.

Table 5-5. Data Used in the Formulation of Eqns. 5-8 Through 5-11 Correlating Rat Antigoiter Bioassay Activities (BA) for Thyroid Hormone Analogs.

Data[a] Point #	BA_{obsd}[b]	log(BA)		
		Obsd.	Calcd.[c]	Dev.
1	0.12	-0.921	-0.598	-0.322
2	0.54	-0.268	-0.398	0.130
3	0.36	-0.444	-0.461	0.017
4	3.60	0.556	0.493	0.063
5	2.36	0.373	0.326	0.047
6	2.91	0.464	0.520	-0.056
7	0.90	-0.046	0.102	-0.148
*8	0.093	-1.032	-0.487	-0.545
9	6.24	0.795	0.802	-0.007
*10	0.091	-1.041	-0.059	-0.982
11	2.27	0.356	0.662	-0.306
12	30.0	1.477	1.244	0.233
13	4.63	0.666	0.502	0.163
*14	0.065	-1.187	0.132	-1.319
15	16.87	1.227	0.853	0.375
16	1.97	0.294	0.437	-0.142
17	2.83	0.452	0.807	-0.356
*18	71.98	1.857	1.178	0.679
*19	0.056	-1.252	0.102	-1.354
20	0.81	-0.092	0.214	-0.305
21	14.47	1.160	1.003	0.157

Table 5-5. (Continued)

Data Point # [a]	BA_{obsd} [b]	log(BA)		
		Obsd.	Calcd. [c]	Dev.
22	9.04	0.956	0.940	0.016
23	93.5	1.971	1.412	0.559
24	142.1	2.153	1.894	0.258
25	55.36	1.743	1.478	0.265
26	39.5	1.597	1.727	-0.131
27	7.74	0.889	1.382	-0.493
28	79.9	1.902	1.921	-0.018
29	21.7	1.336	1.687	-0.351
30	3.50	0.544	0.011	0.534
31	0.18	-0.745	-0.568	-0.177
32	0.27	-0.569	-0.411	-0.157
33	1.12	0.049	0.354	-0.305
*34	0.43	-0.366	0.281	-0.648
*35	6.03	0.780	1.430	-0.650
36	4.88	0.688	0.967	-0.278
37	3.80	0.580	0.638	-0.059
38	23.78	1.376	1.153	0.224
39	1.58	0.199	0.782	-0.584
40	100.0	2.000	1.503	0.497
41	18.1	1.258	1.087	0.171
42	19.0	1.279	1.132	0.147

Table 5-5. (Continued)

Data Point # [a]	BA_{obsd} [b]	log(BA)		
		Obsd.	Calcd. [c]	Dev.
43	2.35	0.371	0.925	-0.554
44	11.25	1.051	0.741	0.310

[a] See Table 5-4.

[b] See Appendix I. Corrected to a molar basis. Assuming L = DL/0.59.

$L\text{-}T_3$ = 100 = reference compound.

[c] Calculated using Eqn. 5-8.

* Not used in calculating Eqns. 5-8 through 5-11.

Table 5-6. Stepwise Development of Eqn. 5-8.

Eqn. #	Eqn. and Statistical Data
5-12	log(BA) = + 0.056 + 0.620 $\pi_{3'}$
	R = 0.511 S = 0.698
	$F_{1,34}$ = 12.01 (99.8% vs. mean)
5-13	log(BA) = - 1.722 + 0.898 π_{35} + 0.674 $\pi_{3'}$
	R = 0.754 S = 0.541
	$F_{1,33}$ = 23.61 (>99.9% vs. Eqn. 5-12)
5-14	log(BA) = - 2.319 + 1.082 π_{35} + 1.106 $\pi_{3'}$
	- 0.933 3'SIZE > I
	R = 0.862 S = 0.424
	$F_{1,32}$ = 21.60 (>99.9% vs. Eqn. 5-13)
5-15	log(BA) = - 2.518 + 1.171 π_{35} + 1.191 $\pi_{3'}$
	- 0.988 3'SIZE > I - 0.632 I4'OCH_3
	R = 0.892 0.384
	$F_{1,31}$ = 8.16 (99.2% vs. Eqn. 5-14)

Table 5-6. (Continued)

Eqn. #		Eqn. and Statistical Data
5-16	log(BA)	$= -2.546 + 1.217\ \pi_{35} + 1.196\ \pi_{3'} - 1.115\ 3'\text{SIZE} > \text{I} - 0.667\ \sigma_{3'5'} - 0.696\ \text{I4'OCH}_3$
		$R = 0.924 \qquad S = 0.332$
		$F_{1,30} = 11.51$ (99.7% vs. Eqn. 5-15)
5-8		$F_{1,29} = 4.77$ (96.0% vs. Eqn. 5-16).

Table 5-7. Independent Variable Squared Cross Correlation Matrix for Eqns. 5-8 through 5-11.

	$\pi_{3'}$	$\pi_{5'}$	$\pi_{3'5'}$	$\sigma_{3'5'}$	$Es_{5'}$	3'SIZE > I	π_{35}	I4'OCH_3
$\pi_{3'}$	1.000	0.002	0.745	0.051	0.008	0.411	0.006	0.041
$\pi_{5'}$		1.000	0.218	0.118	0.949	0.052	0.004	0.022
$\pi_{3'5'}$			1.000	0.001	0.170	0.204	0.001	0.011
$\sigma_{3'5'}$				1.000	0.059	0.097	0.003	0.011
$Es_{5'}$					1.000	0.059	0.001	0.024
3'SIZE > I						1.000	0.024	0.009
π_{35}							1.000	0.040
I4'OCH_3								1.000

$$\log(BA) = -2.588\ (\pm 0.546) + 1.251\ (\pm 0.243)\ \pi_{35} + 1.240\ (\pm 0.233)\ \pi_{3'} - 1.189\ (\pm 0.340)\ 3'\text{SIZE} > \text{I} - 0.282\ (\pm 0.321)\ \pi_{5'} - 0.557\ (\pm 0.453)\ \sigma_{3'5'} - 0.762\ (\pm 0.414)\ \text{I4'OCH}_3 \quad \text{(Eqn. 5-8)}$$

$N = 36 \qquad R = 0.935 \qquad S = 0.312$

$$\log(BA) = -2.790\ (\pm 0.616) + 1.241\ (\pm 0.243)\ \pi_{35} + 1.225\ (\pm 0.231)\ \pi_{3'} - 1.183\ (\pm 0.339)\ 3'\text{SIZE} > \text{I} + 0.190\ (\pm 0.222)\ \text{Es}_{5'} - 0.606\ (\pm 0.442)\ \sigma_{3'5'} - 0.756\ (\pm 0.415)\ \text{I4'OCH}_3 \quad \text{(Eqn. 5-9)}$$

$N = 36 \qquad R = 0.934 \qquad S = 0.313$

$$\log(BA) = -2.588\ (\pm 0.546) + 1.251\ (\pm 0.243)\ \pi_{35} + 1.240\ (\pm 0.233)\ \pi_{3'5'} - 1.189\ (\pm 0.340)\ 3'\text{SIZE} > \text{I} - 1.522\ (\pm 0.436)\ \pi_{5'} - 0.557\ (\pm 0.453)\ \sigma_{3'5'} - 0.762\ (\pm 0.414)\ \text{I4'OCH}_3 \quad \text{(Eqn. 5-10)}$$

$N = 36 \qquad R = 0.935 \qquad S = 0.312$

$$\log(BA) = -3.561\ (\pm 0.761) + 1.194\ (\pm 0.256)\ \pi_{35} + 1.137\ (\pm 0.231)\ \pi_{3'5'} - 1.132\ (\pm 0.356)\ 3'\text{SIZE} > \text{I} + 0.951\ (\pm 0.301)\ \text{Es}_{5'} - 0.804\ (\pm 0.472)\ \sigma_{3'5'} - 0.719\ (\pm 0.440)\ \text{I4'OCH}_3 \quad \text{(Eqn. 5-11)}$$

$N = 36 \qquad R = 0.925 \qquad S = 0.334$

On the basis of Eqns. 5-8 and 5-9, we arrive at the following conclusions concerning in vivo rat antigoiter bioassay activities:

(1) Activity is enhanced by bulky, lipophilic 3 and 5 substituents (π_{35}). This is consistent with the concept of thyromimetic activity being directly related to the ability of the 3 and 5 substituents (by virtue of their size or bulk) to constrain the diphenyl ether thyronine nucleus to the two approximately equal energy, readily interconvertible proximal and distal conformers. Because of the near colinearity of π and group size for 3 and 5 substituents, however, it is not possible to rule out or confirm an inherent hydrophobic effect for the 3 and 5 substituents.

(2) Although directly related to 3' substituent lipophilicity ($\pi_{3'}$), activity is also decreased by 3' substituent steric bulk which extends out from the 3' position further than iodine (3'SIZE > I). Kubinyi[56] used $Es_{3'}^{corr}$ derived from $Es_{3'}$ (see above and Eqn. 5-3) as an estimate of 3' substituent size or steric bulk greater than iodine. Es is a measure of the steric effect of a substituent on a reaction site or binding position located ortho to or on the next atom to the substituent. For essentially symmetrical substituents (e.g., H, F, Br, I, CH_3, tBu), Es is also a good measure of how far a substituent can extend out from say the 3' position. For substituents with conformational flexibility to move by internal rotations away from the "ortho" position (e.g., nPr, cHex, iBu, Ph), however, Es will not reflect substituent "size" extending out from the position. Indeed, Kubinyi[56] was forced to exclude analogs with 3'-iBu and 3'-Ph substituents from his correlations utilizing $Es_{3'}^{corr}$. The ability of 3'SIZE > I to account for the negative, "greater than iodine" steric effects of Et, iPr, nPr, tBu, iBu, Ph, and sBu 3' substituents indicates that this parameter more accurately (than

$Es_{3',corr}$) represents this effect.

(3) Activity is enhanced by electron donating 3' and 5' substituents ($\sigma_{3',5'}$), as previously observed by both Hansch[92] and Kubinyi.[56]

(4) Activity is decreased by 5' substituent lipophilicity ($\pi_{5'}$) or bulk ($Es_{5'}$). The almost complete lack of orthogonality between $\pi_{5'}$ and $Es_{5'}$ (see Table 5-7) allows prediction of the detrimental effect of 5' substituents by either parameter (Eqns. 5-8 and 5-9), although in most correlations $\pi_{5'}$ was found to be a slightly better predictor of the negative 5' substituent effects than $Es_{5'}$.

(5) Activity correlates well with an indicator variable ($I4'OCH_3$) for the less active 4'-OCH_3 analogs which are metabolized to the naturally occurring 4'-OH analogs <u>in vivo</u>.

Following the example of Kubinyi[56] (Eqns. 5-3 through 5-5), Eqns. 5-10 and 5-11 utilize $\pi_{3',5'}$ instead of $\pi_{3'}$ alone. Inspection of the equations shows, however, that Eqn. 5-10 is merely a linear combination of the variables of Eqn. 5-8 and hence (since $\pi_{5'}$ and $Es_{5'}$ are so well correlated) Eqn. 5-11 is essentially equivalent to Eqn. 5-9; i.e.,

$$\log(BA) = a(\pi_{3',5'}) - b(\pi_{5'})$$

is equivalent to

$$\log(BA) = a(\pi_{3'}) - (b-a)(\pi_{5'})$$

and due to the $\pi_{5'}$--$Es_{5'}$ lack of orthogonality

$$\log(BA) = c(\pi_{3',5'}) + d(Es_{5'})$$

is essentially equivalent to

$$\log(BA) = e(\pi_{3'}) + f(Es_{5'})$$

As only negative effects on activity are observed for 5' substituents, the $+\pi_{3'}/-\pi_{5'}$, or $+Es_{5'}$, model, and not the $+\pi_{3',5'}/-\pi_{5'}$, or $+Es_{5'}$, model, makes

more intuitive sense and is favored by the principle of parsimony; all things being equal, one accepts the simplest model.[213] Synthesis and testing of analogs with 5' substituents which are considerably more orthogonal in $\pi_{5'}$ and $Es_{5'}$ should help to resolve this ambiguity.

Because the coefficients of the Es terms are nearly identical in Eqns. 5-3 and 5-4, both Es values were combined by Kubinyi[56] as $Es' = Es_{5'} + Es_{3'}^{corr}$. Combination of originally separate variables in this manner can be misleading since this implies that the magnitudes and mechanisms of the effects described by the two original variables are equivalent and additive. Such equivalence may not hold, however, for analogs not yet studied. In addition, if the original model is wrong, then the relative importance of the variables may be different than in the original model. This is especially evident from our <u>in vivo</u> studies which show that Kubinyi's[56] $\pi_{3',5'}/Es_{3'}^{corr}/Es_{5'}$ model is essentially equivalent to a $\pi_{3'}/Es_{3'}^{corr}/Es_{5'}$ model: the former model, but not the latter model, would allow the $Es' = Es_{3'}^{corr} + Es_{5'}$ alteration.

The poor prediction of the activities of the analogs in Tables 5-4 and 5-5 which were not included in the regression calculation of Eqns. 5-8 through 5-11 can be almost entirely ascribed to questionable synthesis and/or to questionable activity determinations. The correlations of Eqns. 5-8 through 5-11 can really be considered quite good correlations considering that the synthesis and testing of the analogs were conducted by a large number of different investigating groups during an over 30 year period.

For the correlations presented below for <u>in vitro</u> binding of analogs to rat hepatic intact nuclei and solubilized nuclear protein, essentially identical equations could be obtained in almost every case for a $\pi_{3'}/\pi_{5'}$

model, for a $\pi_{3'}/Es_{5'}$ model, for a $\pi_{3',5'}/\pi_{5'}$ model, or for a $\pi_{3',5'}/Es_{5'}$ model (just as described above for the correlations involving *in vivo* antigoiter activities). Since in most cases the best correlations were obtained with the $\pi_{3'}/\pi_{5'}$ model, it is the model for which equations are presented, although the other models (because of lack of $\pi_{5'}$ and $Es_{5'}$ orthogonality) can not be ruled out (but see concluding remarks at the end of this chapter).

An unsuccessful attempt was made to expand Eqns. 5-8 through 5-11 to include 4'-deoxy analogs by inclusion of the I4'H indicator variable. This failure is apparently due to: (1) uncertainty in the antigoiter activities of some of the 4'-deoxy analogs; and (2) unequal *in vivo* hydroxylation of different 4'-deoxy analogs (possibly because of varying 3' and 5' substituent bulk affecting the ease of *in vivo* hydroxylation). The antigoiter activities of 4'-deoxy analogs certainly deserves further study, especially with respect to 3' and 5' substituent influences on *in vivo* hydroxylation.

BINDING TO INTACT RAT HEPATIC NUCLEI

Utilizing the analogs of Table 5-8 and data of Table 5-9, the correlation Eqn. 5-17 for *in vitro* binding of analogs to intact rat hepatic nuclei (BN) was derived for structures of type 5-6. The appropriate stepwise development of Eqn. 5-17 and independent variable squared cross correlation matrix are presented in Tables 5-10 and 5-11, respectively.

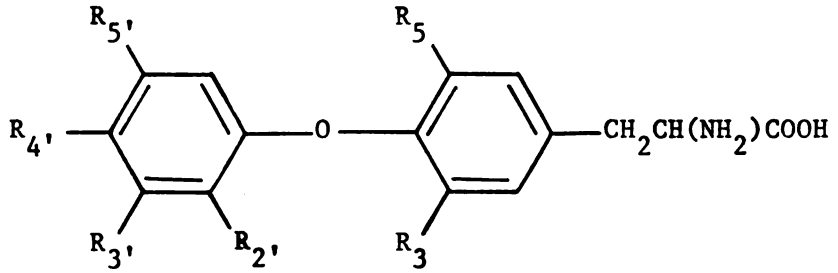
$$\log(BN) = -3.292\ (\pm 0.660) + 1.680\ (\pm 0.290)\ \pi_{35}$$
$$+ 1.147\ (\pm 0.362)\ \pi_{3'} - 1.218\ (\pm 0.307)\ 3'\text{SIZE} > \text{I}$$
$$- 0.873\ (\pm 0.289)\ \pi_{5'} - 0.920\ (\pm 0.432)\ \text{I2}'$$
$$- 2.049\ (\pm 0.411)\ \text{I4}'\text{H} \qquad \text{(Eqn. 5-17)}$$

$$N = 25 \qquad R = 0.969 \qquad S = 0.280$$

Just as was found for the correlations of *in vivo* rat antigoiter activities, binding of analogs to intact rat hepatic nuclei is enhanced by large, lipophilic 3,5 substituents (π_{35}) and is decreased by 5' substituent size or lipophilicity (estimated here by $\pi_{5'}$) and by 2' substitution (I2'). Of interest is that the 3' substituent apparently binds in a hydrophobic pocket ($\pi_{3'}$) approximately the size of iodine (3'SIZE > I). That this same size-limited, 3' substituent hydrophobic effect is observed for *in vivo* and *in vitro* activities suggests that for the former it is reflecting receptor binding and not distribution. The inherent loss of 4'-OH binding for 4'-deoxy analogs to the intact nuclei can be seen from the indicator variable (I4'H). In contrast to the *in vivo* quantitative SAR, addition of a $\sigma_{3',5'}$ parameter is *not* significant.

QSAR studies represent extrathermodynamic linear free energy correlations of activity with the physico-chemical properties of the analogs. *In vitro* equilibrium binding affinities, K_a, such as measured with binding to intact nuclei or proteins, permit direct correlation of the apparent free energy of binding (ΔG) with the physico-chemical properties of the analogs:

Table 5-8. Structures of Thyroid Hormone Analogs (5-6) Used in Deriving Eqns. 5-17 and 5-24.



5-6

Data Point #	Abbreviation	R_3	R_5	$R_{3'}$	$R_{5'}$	$R_{4'}$	$R_{2'}$
*1	T2	I	I	H	H	OH	H
2	3'Cl-T2	I	I	Cl	H	OH	H
3	3'Me-T2	I	I	Me	H	OH	H
4	3'Et-T2	I	I	Et	H	OH	H
5	3'iPr-T2	I	I	iPr	H	OH	H
6	3'tBu-T2	I	I	tBu	H	OH	H
7	3'iBu-T2	I	I	iBu	H	OH	H
8	3'Ph-T2	I	I	Ph	H	OH	H
9	3'cHex-T2	I	I	cHex	H	OH	H
10	2'3'Me2-T2	I	I	Me	H	OH	Me
11	2'5'Me2-T2	I	I	H	Me	OH	Me
*12	2'Me-5'I-T2	I	I	H	I	OH	Me
13	Napth-T2	I	I	a	H	OH	a
14	T3	I	I	I	H	OH	H
15	T4	I	I	I	I	OH	H

Table 5-8. (Continued)

Data Point #	Abbreviation	R_3	R_5	$R_{3'}$	$R_{5'}$	$R_{4'}$	$R_{2'}$
16	3'5'C12-T2	I	I	Cl	Cl	OH	H
17	3'5'Me2-T2	I	I	Me	Me	OH	H
18	3'5'iPr2-T2	I	I	iPr	iPr	OH	H
19	4'H-T3	I	I	I	H	H	H
20	4'H-3'CF3-T2	I	I	CF_3	H	H	H
21	4'H-3'Me-T2	I	I	Me	H	H	H
22	3'iPr-Br2-T	Br	Br	iPr	H	OH	H
23	Me3-T	Me	Me	Me	H	OH	H
24	Me4-T	Me	Me	Me	Me	OH	H
25	3'iPr-Me2-T	Me	Me	iPr	H	OH	H
26	33'-T2	I	H	I	H	OH	H
27	R-T3	I	H	I	I	OH	H

[a] $2',3'-(CH)_4$

*Not used in calculating Eqns. 5-17 and 5-24.

Table 5-9. Data Used in the Formulation of Eqn. 5-17 Correlating *In Vitro* Binding to Intact Rat Hepatic Nuclei (BN) for Thyroid Hormone Analogs.

Data Point # [a]	BN_{obsd} [b]	log(BN)		
		Obsd.	Calcd. [c]	Dev.
*1	0.3	-0.523	0.472	-0.995
2	6.2	0.792	1.286	-0.494
3	13.5	1.130	1.114	0.016
4	21.0	1.322	1.487	-0.165
5	104.0	2.017	1.918	0.099
6	38.5	1.586	1.622	-0.037
7	20.0	1.301	1.353	-0.052
8	2.0	0.301	0.016	0.285
9	1.4	0.146	0.355	-0.209
10	1.1	0.041	0.194	-0.153
11	0.1	-1.000	-0.937	-0.063
*12	0.3	-0.523	-1.426	0.903
13	8.0	0.903	0.687	0.216
14	100.0	2.000	1.756	0.244
15	12.5	1.097	0.778	0.319
16	4.5	0.653	0.666	-0.013
17	6.2	0.792	0.625	0.167
18	1.4	0.146	0.582	-0.436
19	0.4	-0.398	-0.293	-0.105
20	0.2	-0.699	-0.568	-0.131
21	0.2	-0.699	-0.935	0.236

Table 5-9. (Continued)

Data Point # [a]	BN_{obsd} [b]	log(BN)		
		Obsd.	Calcd. [c]	Dev.
22	36.0	1.556	1.045	0.512
23	0.1	-1.000	-0.768	-0.232
24	0.1	-1.000	-1.257	0.257
25	0.7	-0.155	0.036	-0.191
26	0.5	-0.301	-0.126	-0.175
27	0.1	-1.000	-1.104	0.104

[a] See Table 5-8.

[b] From references 24 and 25. No DL/L correction. On a molar basis. BN = $(K_A/K_{T3}) \times 100$, where K_A = equilibrium association constants for analog A. BN = relative binding affinity (relative to $BN(T_3) = 100$ as reference compound).

[c] Calculated using Eqn. 5-17.

* Not used in calculating Eqn. 5-17.

Table 5-10. Stepwise Development of Eqn. 5-17.

Eqn. #	Eqn. and Statistical Data
5-18	$\log(BN) = -1.753 + 1.070\ \pi_{35}$
	$R = 0.499 \quad S = 0.870$
	$F_{1,23} = 7.63$ (98.8% vs. mean)
5-19	$\log(BN) = -2.003 + 1.284\ \pi_{35} - 1.471$ I4'H
	$R = 0.697 \quad S = 0.736$
	$F_{1,22} = 10.11$ (99.5% vs. Eqn. 5-18)
5-20	$\log(BN) = -2.370 + 1.215\ \pi_{35} + 0.436\ \pi_{3'}$
	$- 1.323$ I4'H
	$R = 0.742 \quad S = 0.704$
	$F_{1,21} = 3.08$ (90.4% vs. Eqn. 5-19)
5-21	$\log(BN) = -3.598 + 1.475\ \pi_{35} + 1.411\ \pi_{3'}$
	$- 1.121$ 3'SIZE > I $- 1.509$ I4'H
	$R = 0.872 \quad S = 0.526$
	$F_{2,20} = 11.52$ (>99.9% vs. Eqn. 5-19)
	$F_{1,21} = 17.54$ (>99.9% vs. Eqn. 5-20)

Table 5-10. (Continued)

Eqn. #	Eqn. and Statistical Data
5-22	$\log(BN) = -3.422 + 1.503\ \pi_{35} + 1.453\ \pi_{3'} - 1.288\ 3'SIZE > I - 0.778\ \pi_{5'} - 1.782\ I4'H$
	$R = 0.934 \qquad S = 0.396$
	$F_{1,19} = 16.41$ (>99.9% vs. Eqn. 5-21)
5-17	$F_{1,18} = 20.05$ (>99.9% vs. Eqn. 5-22)

Table 5-11. Independent Variable Squared Cross Correlation Matrix for Eqns. 5-17 and 5-24.

	$\pi_{3'}$	$\pi_{5'}$	π_{35}	3'SIZE > I	I2'	I4'H
$\pi_{3'}$	1.000	0.019	0.007	0.621	0.145	0.028
$\pi_{5'}$		1.000	0.006	0.047	0.003	0.043
π_{35}			1.000	0.042	0.040	0.040
3'SIZE > I				1.000	0.031	0.031
I2'					1.000	0.018
I4'H						1.000

$$\Delta G = -RT \ln(K_A) \qquad \text{(Eqn. 5-23)}$$

Using Eqn. 5-23, $K_{T3} = 6.1 \times 10^{-8}\ M^{-1}$ at T = 310° K,[33] and the data of Table 5-9, Eqn. 5-17 can be converted (with the resulting data of Table 5-12) to the equivalent Eqn. 5-24:

$$\begin{aligned} -\Delta G = &+ 4.955\ (\pm 0.936) + 2.384\ (\pm 0.412)\ \pi_{35} \\ &+ 1.626\ (\pm 0.514)\ \pi_{3'} - 1.727\ (\pm 0.435)\ 3'\text{SIZE} > I \\ &- 1.239\ (\pm 0.409)\ \pi_{5'} - 1.305\ (\pm 0.612)\ I2' \\ &- 2.906\ (\pm 0.584)\ I4'H \end{aligned} \qquad \text{(Eqn. 5-24)}$$

N = 25 R = 0.969 S = 0.396

Eqn. 5-24 is extremely interesting in that it now allows estimation of the kcal/mole contributions to the free energy of binding of various analog structural features and physico-chemical properties. For example, for each π unit of the 3' and of the 3,5 substituents, Eqn. 5-24 predicts 1.63 and 2.38 kcal/mole contributions, respectively, to the free energy of binding.

Of particular interest is that the 4'-OH apparently contributes ∿2.91 kcal/mole to the free energy of binding, a reasonable value for net hydrogen bond formation between the 4'-OH and a nuclear receptor.

Table 5-12. Data Used in the Formulation of Eqn. 5-24 Correlating *In Vitro* Free Energy of Binding to Intact Rat Hepatic Nuclei (ΔG) for Thyroid Hormone Analogs.

Data Point # [a]	$-\Delta G$(kcal/mole) [b]		
	Obsd. [c]	Calcd. [d]	Dev.
*1	8.883	10.294	-1.411
2	10.749	11.449	-0.700
3	11.228	11.205	0.023
4	11.500	11.734	-0.234
5	12.486	12.346	0.140
6	11.874	11.926	-0.052
7	11.470	11.544	-0.074
8	10.052	9.648	0.404
9	9.832	10.128	-0.296
10	9.684	9.900	-0.217
11	8.206	8.296	-0.089
*12	8.883	7.602	1.281
13	10.906	10.600	0.306
14	12.462	12.116	0.346
15	11.181	10.728	0.452
16	10.551	10.569	-0.018
17	10.749	10.511	0.237
18	9.832	10.450	-0.618
19	9.060	9.209	-0.149
20	8.633	8.819	-0.186

Table 5-12. (Continued)

Data Point # [a]	-ΔG (kcal/mole) [b]		
	Obsd. [c]	Calcd. [d]	Dev.
21	8.633	8.298	0.335
22	11.832	11.106	0.726
23	8.206	8.536	-0.329
24	8.206	7.842	0.364
25	9.405	9.676	-0.271
26	9.198	9.446	-0.249
27	8.206	8.059	0.148

[a] See Table 5-8.

[b] See text for derivation.

[c] From data of Table 5-9.

[d] Calculated using Eqn. 5-24.

* Not used in calculating Eqn. 5-24.

BINDING TO SOLUBILIZED RAT HEPATIC NUCLEAR PROTEIN

Utilizing the analogs of Table 5-13 and the data of Table 5-14, the correlation Eqns. 5-25 and 5-26 for *in vitro* binding of analogs to solubilized rat hepatic nuclear protein (BS) were derived for analogs of structure 5-7. The appropriate stepwise development of Eqns. 5-25 and 5-26 and the independent variable squared cross correlation matrix are presented in Tables 5-15 and 5-16, respectively.

$$\log(BS) = -0.304\ (\pm 0.380) + 1.675\ (\pm 0.420)\ \pi_{3'}$$
$$- 2.118\ (\pm 0.876)\ 3'\text{SIZE} > \text{I} - 0.634\ (\pm 0.406)\ \pi_{5'}$$
$$- 1.540\ (\pm 0.375)\ \text{I4'H} - 1.347\ (\pm 0.489)\ \text{I4'OCH}_3$$

(Eqn. 5-25)

N = 31 R = 0.946 S = 0.400

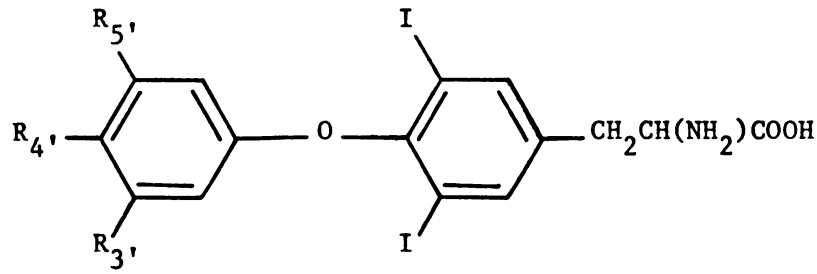
$$\log(BS) = -0.222\ (\pm 0.430) + 1.546\ (\pm 0.445)\ \pi_{3'}$$
$$- 1.904\ (\pm 0.825)\ 3'\text{SIZE} > \text{I} - 0.780\ (\pm 0.391)\ \pi_{5'}$$
$$- 1.553\ (\pm 0.366)\ \text{I4'H} - 1.323\ (\pm 0.447)\ \text{I4'OCH}_3$$
$$+ 0.958\ (\pm 0.793)\ \sigma_{3',5'} - 0.114\ (\pm 0.109)\ \text{INTERACT}$$

(Eqn. 5-26)

N = 31 R = 0.960 S = 0.364

Utilizing Eqn. 5-25, we can draw the following conclusions concerning *in vitro* binding of analogs to solubilized rat hepatic nuclear protein:

Table 5-13. Structures of Thyroid Hormone Analogs (5-7) Used in Deriving Eqns. 5-25, 5-26 and 5-33.



5-7

Data Point #	Abbreviation	$R_{3'}$	$R_{5'}$	$R_{4'}$
1	4'H-T2	H	H	H
2	4'H-3'Me-T2	Me	H	H
*3	4'H-3'iPr-T2	iPr	H	H
4	4'H-3'tBu-T2	tBu	H	H
5	4'H-3'F-T2	F	H	H
6	4'H-3'Cl-T2	Cl	H	H
7	4'H-3'Br-T2	Br	H	H
8	4'H-T3	I	H	H
*9	T2	H	H	OH
10	3'Me-T2	Me	H	OH
11	3'iPr-T2	iPr	H	OH
12	3'nPr-T2	nPr	H	OH
13	3'tBu-T2	tBu	H	OH
14	3'F-T2	F	H	OH

Table 5-13. (Continued)

Data Point #	Abbreviation	$R_{3'}$	$R_{5'}$	$R_{4'}$
15	3'Cl-T2	Cl	H	OH
16	3'Br-T2	Br	H	OH
17	T3	I	H	OH
18	4'OCH3-3'Me-T2	Me	H	OCH_3
19	4'OCH3-3'iPr-T2	iPr	H	OCH_3
20	4'OCH3-3'tBu-T2	tBu	H	OCH_3
21	4'OCH3-T3	I	H	OCH_3
22	3'NO2-T2	NO_2	H	OH
23	3'sBu-T2	sBu(±)	H	OH
24	4'H-3'5'Me2-T2	Me	Me	H
25	T4	I	I	OH
26	3'5'iPr2-T2	iPr	iPr	OH
27	3'5'Cl2-T2	Cl	Cl	OH
28	3'5'Br2-T2	Br	Br	OH
29	3'iPr-5'Cl-T2	iPr	Cl	OH
30	3'iPr-5'Br-T2	iPr	Br	OH
31	3'iPr-355'-T3	iPr	I	OH
32	4'H-3'NO2-T2	NO_2	H	H
33	3'5'Me2-T2	Me	Me	OH
*34	4'OCH3-3'5'Me2-T2	Me	Me	OCH_3
*35	4'OCH3-3'sBu-T2	sBu(±)	H	OCH_3

*Not used in the calculation of Eqns. 5-25, 5-26, and 5-27.

Table 5-14. Data Used in the Formulation of Eqns. 5-25 and 5-26 Correlating In Vitro Binding to Solubilized Rat Hepatic Nuclear Protein Receptors (BS) for Thyroid Hormone Analogs.

Data Point #	BS_{obsd}[b]	log (BS)				
		Obsd.	Calcd.[c]	Dev.	Calcd.[d]	Dev.
1	0.01	-2.000	-1.844	-0.156	-1.775	-0.225
2	0.225	-0.648	-0.906	0.258	-0.910	0.262
*3	0.492	-0.308	0.183	-0.491	0.108	-0.416
4	0.335	-0.475	-0.476	0.001	-0.467	-0.008
5	0.0136	-1.866	-1.610	-0.257	-1.559	-0.308
6	0.118	-0.928	-0.654	-0.274	-0.678	-0.250
7	0.24	-0.620	-0.403	-0.217	-0.446	-0.174
8	0.23	-0.638	0.032	-0.671	-0.044	-0.594
*9	0.082	-1.086	-0.304	-0.782	-0.222	-0.864
10	3.30	0.518	0.634	-0.116	0.538	-0.020
11	89.15	1.950	1.723	0.227	1.630	0.321
12	23.97	1.380	1.435	-0.055	1.359	0.020
13	8.45	0.927	1.064	-0.138	1.073	-0.146
14	0.164	-0.785	-0.069	-0.716	-0.105	-0.681
15	3.73	0.572	0.886	-0.314	0.834	-0.262
16	15.89	1.201	1.137	0.064	1.136	0.065
17	100.0	2.000	1.572	0.428	1.596	0.404
18	0.17	-0.770	-0.713	-0.057	-0.680	-0.090
19	6.82	0.834	0.376	0.458	0.338	0.496
20	0.27	-0.569	-0.283	-0.286	-0.237	-0.332

Table 5-14. (Continued)

Data Point # [a]	BS_{obsd} [b]	log (BS)				
		Obsd.	Calcd. [c]	Dev.	Calcd. [d]	Dev.
21	1.29	0.111	0.225	-0.115	0.186	-0.075
22	0.225	-0.648	-0.773	0.125	-0.852	0.204
23	78.29	1.894	1.549	0.345	1.522	0.371
24	0.145	-0.839	-1.261	0.422	-1.347	0.508
25	13.85	1.141	0.863	0.279	0.980	0.162
26	1.10	0.041	0.754	-0.713	0.179	-0.138
27	3.71	0.569	0.436	0.134	0.762	-0.192
28	5.07	0.705	0.592	0.113	0.876	-0.172
29	52.56	1.721	1.274	0.447	1.624	0.097
30	21.95	1.341	1.178	0.163	1.436	-0.095
31	12.41	1.094	1.014	0.080	1.080	0.014
32	0.038	-1.420	-2.313	0.893	-2.208	0.788
33	0.845	-0.073	0.280	-0.353	-0.120	0.046
*34	0.335	-0.475	-1.068	0.593	-1.117	0.642
*35	1.29	0.111	1.549	-1.438	1.522	-1.412

[a] From Table 5-13.

[b] From references 26 and 43. No DL/L correction. On a molar basis. BS = relative binding affinity (relative to $BS(L\text{-}T_3)$ = 100 as reference compound) = $(K_A/K_{T3}) \times 100$, where K_A = equilibrium association constant for analog A.

[c] Calculated using Eqn. 5-25.

[d] Calculated using Eqn. 5-26.

* Not used in calculating Eqns. 5-25 and 5-26.

Table 5-15. Stepwise Development of Eqns. 5-25 and 5-26.

Eqn. #	Eqn. and Statistical Data
5-27	log(BS) = + 0.689 - 1.737 I4'H
	R = 0.709 S = 0.812
	$F_{1,29}$ = 29.25 (>99.9% vs. mean)
5-28	log(BS) = -0.165 + 0.760 $\pi_{3'}$ - 1.360 I4'H
	R = 0.820 S = 0.669
	$F_{1,28}$ = 14.65 (>99.9% vs. Eqn. 5-27)
5-29	log(BS) = - 0.036 + 0.829 $\pi_{3'}$ - 1.533 I4'H
	- 1.138 I4'OCH_3
	R = 0.884 S = 0.556
	$F_{1,27}$ = 13.51 (99.9% vs. Eqn. 5-28)
5-30	log(BS) = - 0.393 + 1.433 $\pi_{3'}$ - 1.634 3'SIZE > I
	- 1.388 I4'H - 1.086 I4'OCH_3
	R = 0.923 S = 0.467
	$F_{1,26}$ = 12.41 (99.8% vs. Eqn. 5-29)
5-25	$F_{1,25}$ = 10.35 (99.6% vs. Eqn. 5-30)

Table 5-15. (Continued)

Eqn. #	Eqn. and Statistical Data
5-31	log(BS) = - 0.429 + 1.755 $\pi_{3'}$ - 2.114 3'SIZE > I - 0.651 $\pi_{5'}$ - 1.465 I4'H - 1.327 I4'OCH_3 + 0.523 $\sigma_{3'5'}$ R = 0.951 S = 0.391 $F_{1,24}$ = 2.24 (82.9% vs. Eqn. 5-25)
5-32	log(BS) = - 0.182 + 1.567 $\pi_{3'}$ - 2.038 3'SIZE > I - 0.678 $\pi_{5'}$ - 1.599 I4'H - 1.352 I4'OCH_3 - 0.045 INTERACT R = 0.948 S = 0.402 $F_{1,24}$ = 0.81 (<75% vs. Eqn. 5-25)
5-26	$F_{1,23}$ = 4.67 (95.7% vs. Eqn. 5-31) $F_{1,23}$ = 6.26 (97.6% vs. Eqn. 5-32) $F_{2,23}$ = 3.62 (95.5% vs. Eqn. 5-25)

Table 5-16. Independent Variable Squared Cross Correlation Matrix For Eqns. 5-25, 5-26, and 5-33.

	$\pi_{3'}$	$\pi_{5'}$	$\sigma_{3'5'}$	3'SIZE > I	I4' H	I4'OCH_3	INTERACT
$\pi_{3'}$	1.000	0.038	0.119	0.602	0.121	0.036	0.291
$\pi_{5'}$		1.000	0.006	0.007	0.083	0.052	0.099
$\sigma_{3'5'}$			1.000	0.103	0.020	0.007	0.329
3'SIZE > I				1.000	0.025	0.025	0.086
I4'H					1.000	0.061	0.000
I4'OCH_3						1.000	0.000
INTERACT							1.000

(1) Just as was found for binding of analogs to intact rat hepatic nuclei, the 3' substituent apparently binds in a size-limited (3'SIZE > I), hydrophobic ($\pi_{3'}$) pocket, while any 5' substituent bulk or lipophilicity (estimated here by $\pi_{5'}$) is detrimental to binding.

(2) Indicator variables again reflect the inherent loss of 4'-OH hydrogen bonding with the receptor due to replacement with a 4'-H (I4'H) or with a 4'-OCH_3 (I4'OCH_3).

Addition of a $\sigma_{3'5'}$ term alone or of an INTERACT term alone to Eqn. 5-25 is not significant (see Table 5-15). Simultaneous inclusion of both terms, however, (to give Eqn. 5-26) is significant at the 95.5% confidence level. The signs of the regression coefficients indicate that binding is enhanced by electron-withdrawing substituents which orient the 4'-OH towards the 5' position. This suggests that the 4'-OH donates a hydrogen bond to the 5' side of the nuclear receptor. This also suggests that the negative effect of 5' substitution might be due to interference with 4'-OH hydrogen bond formation with the receptor and/or to direct steric interaction of the 5' substituent with the receptor. These results are consistent with the results of our MO studies of Chapter Four, which support the model of 3' and 5' substituents interacting with and affecting the hydrogen bonding of the 4'-OH to the nuclear receptor.

As was done for the binding of analogs to intact nuclei, we converted Eqn. 5-26 (using Eqn. 5-23, $K_{T3} = 1.29 \times 10^9\ M^{-1}$ at T = 298° K,[26] and the data of Table 5-14) with the resulting data of Table 5-17 to the equivalent Eqn. 5-33:

Table 5-17. Data Used in the Formulation of Eqn. 5-33 Correlating *In Vitro* Free Energy of Binding to Solubilized Rat Hepatic Nuclear Protein (ΔG) for Thyroid Hormone Analogs.

Data Point # [a]	-ΔG(kcal/mole) [b]		
	Obsd. [c]	Calcd. [d]	Dev.
1	6.969	7.275	-0.306
2	8.812	8.455	0.357
*3	9.276	9.843	-0.567
4	9.048	9.059	-0.011
5	7.151	7.570	-0.419
6	8.430	8.771	-0.341
7	8.851	9.088	-0.237
8	8.825	9.636	-0.810
*9	8.215	9.392	-1.178
10	10.403	10.430	-0.027
11	12.355	11.918	0.437
12	11.577	11.549	0.028
13	10.960	11.159	-0.199
14	8.625	9.553	-0.928
15	10.475	10.832	-0.357
16	11.334	11.245	0.089
17	12.423	11.872	0.551
18	8.646	8.769	-0.122
19	10.833	10.156	0.677
20	8.920	9.372	-0.452
21	9.846	9.949	-0.102

Table 5-17. (Continued)

Data Point # [a]	-ΔG(kcal/mole) [b]		
	Obsd. [c]	Calcd. [d]	Dev.
22	8.812	8.534	0.278
23	12.278	11.772	0.506
24	8.552	7.859	0.693
25	11.252	11.032	0.220
26	9.752	9.940	-0.188
27	10.472	10.734	-0.262
28	10.657	10.891	-0.234
29	12.042	11.910	0.132
30	11.525	11.654	-0.129
31	11.187	11.168	0.020
32	7.759	6.685	1.074
33	9.596	9.533	0.063
*34	9.048	8.173	0.875
*35	9.896	11.772	-1.875

[a] See Table 5-13.

[b] See text for derivation.

[c] From data of Table 5-14.

[d] Calculated using Eqn. 5-33.

* Not used in calculating Eqn. 5-33.

$$-\Delta G = +\ 9.392\ (\pm 0.587) \quad +\ 2.108\ (\pm 0.607)\ \pi_{3'}$$

$$-\ 2.597\ (\pm 1.125)\ 3'\text{SIZE} > \text{I} \quad -\ 1.064\ (\pm 0.533)\ \pi_{5'}$$

$$-\ 2.117\ (\pm 0.499)\ \text{I4'H} \quad -\ 1.804\ (\pm 0.610)\ \text{I4'OCH}_3$$

$$+\ 1.306\ (\pm 1.080)\ \sigma_{3'5'} \quad -\ 0.155\ (\pm 0.149)\ \text{INTERACT} \qquad \text{(Eqn. 5-33)}$$

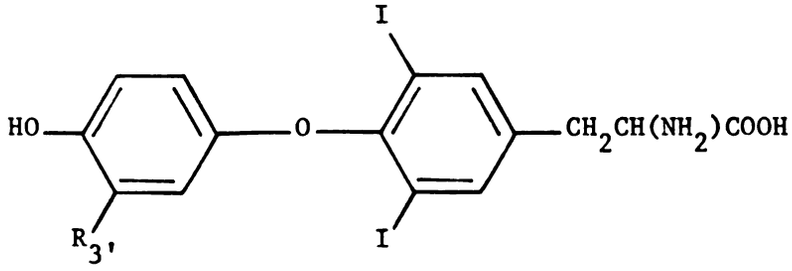
$$N = 31 \qquad R = 0.960 \qquad S = 0.496$$

Comparing Eqns. 5-24 and 5-33, there are quantitative differences in the ΔG contributions of the outer ring substituents to the free energy of binding of analogs to intact nuclei and to solubilized nuclear protein receptors, although the qualitative picture remains unchanged. Of particular interest, Eqn. 5-33 predicts a 4'-OH net hydrogen bond of $\sim$2.1 kcal/mole with the receptor. The accuracy and level of significance of the binding data is generally better than for binding to intact nuclei. On this basis, the ΔG substituent contributions might be considered to be more accurate for Eqn. 5-33 than for Eqn. 5-24, although the differences could also be due to actual differences between the two assay systems.

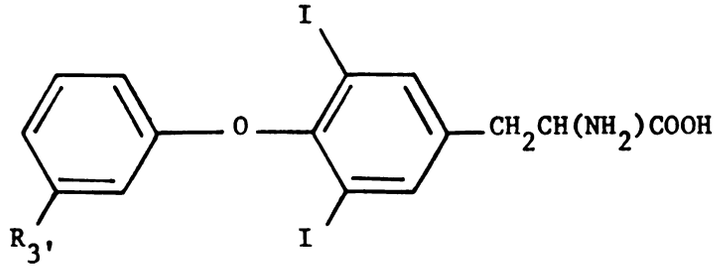
These free energies for binding of analogs to solubilized rat hepatic nuclear protein receptors provide an additional means of more closely inspecting the interactive effects of the 3' and 5' substituents on the hydrogen bonding of the 4'-OH with nuclear receptors. Utilizing the procedure outlined in Chapter Two for the partitioning of substituent contributions to the free energy of binding (using Eqn. 2-2 and the data of Table 5-17), $-\Delta G(OH)$ values can be calculated from Eqn. 5-34:

$$-\Delta G(OH) = -\Delta G(\underline{5\text{-}8}) \quad - \quad -\Delta G(\underline{5\text{-}9}) \qquad \text{(Eqn. 5-34)}$$

For a variety of 3' substituents, it can be seen (from the $-\Delta G(OH)$ values presented in Table 5-18) that the free energy of binding of the 4'-OH to nuclear receptors is not constant from compound to compound: i.e., there are interactive effects between the 4'-OH and the 3' substituent which affect the value of $-\Delta G(OH)$.



5-8



5-9

Using $\sigma_{3',5'}$ and INTERACT (for 5-8) as substituent parameters which should reflect the interactive effects of the 3' and 5' substituents

Table 5-18. Data Used in the Formulation of Eqns. 5-35 Through 5-37 Correlating $-\Delta G(OH)$ for Thyroid Hormone Analogs.

Data Point #	$R_{3'}$[a]	$-\Delta G(OH)$ (kcal/mole)[b]		
		Obsd.[c]	Calcd.[d]	Dev.
*1	H	1.25	2.37	1.12
2	Me	1.59	1.33	-0.26
*3	iPr	3.08	2.01	-1.07
4	tBu	1.91	2.14	0.23
5	F	1.47	1.51	0.04
6	Cl	2.05	2.12	0.07
7	Br	2.48	2.76	0.28
8	I	3.60	3.26	-0.34
9	NO_2	1.05	1.03	-0.02

[a] See structures 5-8 and 5-9.

[b] See text and Eqn. 5-34 for derivation.

[c] From data of Table 5-17.

[d] Calculated using Eqn. 5-37.

* Not used in calculating Eqns. 5-35 through 5-37.

with the 4'-OH and its hydrogen bonding to nuclear receptors, Eqns. 5-35 through 5-37 were derived:

$$-\Delta G(OH) = + 2.095\ (\pm 0.982) \quad - 0.465\ (\pm 2.871)\ \sigma_{3',5'}$$

$$N = 7 \qquad R = 0.183 \qquad S = 0.896$$

(Eqn. 5-35)

$$F_{1,5} = 0.17\ (<75\%\ \underline{vs}.\ \text{mean})$$

$$-\Delta G(OH) = + 2.208\ (\pm 0.944) \quad - 0.106\ (\pm 0.276)\ \text{INTERACT}$$

(Eqn. 5-36)

$$N = 7 \qquad R = 0.404 \qquad S = 0.833$$

$$F_{1,5} = 0.098\ (<75\%\ \underline{vs}.\ \text{mean})$$

$$-\Delta G(OH) = + 2.370\ (\pm 0.357) \quad + 9.206\ (\pm 4.118)\ \sigma_{3',5'}$$

$$- 1.028\ (\pm 0.425)\ \text{INTERACT}$$

(Eqn. 5-37)

$$N = 7 \qquad R = 0.960 \qquad S = 0.286$$

$$F_{1,4} = 45.09\ (99.6\%\ \underline{vs}.\ \text{Eqn. 5-35})$$

$$F_{1,4} = 38.50\ (99.6\%\ \underline{vs}.\ \text{Eqn. 5-36})$$

$$F_{2,4} = 23.40\ (99.3\%\ \underline{vs}.\ \text{mean})$$

It can be seen that neither $\sigma_{3',5'}$ or INTERACT alone correlates very well with $-\Delta G(OH)$. Simultaneous inclusion of both terms is highly

significant, however (Eqn. 5-37). This result is consistent with Eqn. 5-33, which also predicts the 4'-OH hydrogen bond to be enhanced by electron-withdrawing 3' and 5' substituents which tend to orient the 4'-OH towards the 5' position. The magnitudes of the regression coefficients of Eqn. 5-37 are significantly larger than the corresponding ones of Eqn. 5-33. This could be due to: (1) the high degree of correlation between $\sigma_{3'5'}$ and INTERACT for the compounds used to derive Eqn. 5-37; independent variable squared cross correlation matrix element = 0.942 for $\sigma_{3'5'}$--INTERACT; and (2) the partial colinearity of $\sigma_{3'5'}$ and INTERACT with the other parameters for the compounds used to derive Eqn. 5-33 (see Table 5-16). Further studies in this area obviously call for a set of compounds with $\sigma_{3'5'}$ and INTERACT, as well as the rest of the other variables, a great deal more orthogonal in order to provide better acceptability for the + $\sigma_{3'5'}$/-INTERACT model and to avoid the possibility of chance correlations (see comments at the end of this chapter). Eqn. 5-37 does predict that the 4'-OH, without the interactive effects of the 3' and 5' substituents, is apparently forming a hydrogen bond with the nuclear receptor with a net intrinsic hydrogen bond strength of 2.37 kcal/mole. The deviation of the $-\Delta G(OH)$ value for data point #1 of Table 5-18 from this value is consistent with the questionable purity of the analog tested. The same holds for the large discrepancy between the observed and calculated $-\Delta G(OH)$ values for data point #3 of Table 5-18. A similar type analysis for analogs which contain 5' substituents could provide some insight into whether the 5' substituent, in addition to electronically and orientationally affecting the 4'-OH hydrogen bonding with the receptor, exerts its negative influence on thyromimetic

activity by: (1) merely sterically interacting with the receptor; (2) sterically interfering with 4'-OH hydrogen bonding with the receptor (see the MO calculations of Chapter Four); or (3) a combination of (1) and (2).

BINDING TO THYROXINE BINDING GLOBULIN

As shown above, *in vivo* rat antigoiter bioassay activities are enhanced by electron-donating 3' and 5' substituents. In contrast, *in vitro* binding to nuclear receptors is apparently enhanced by electron-withdrawing 3' and 5' substituents, as they influence the association of the 4'-OH with the nuclear receptor. This can be rationalized in part by inspecting the quantitative SAR for the *in vitro* binding of thyroid hormone analogs of structure 5-10 to the plasma protein thyroxine binding globulin (TBG), the principal transport and storage site for the thyroid hormones in human plasma. Utilizing the analogs and data of Table 5-19, the correlation Eqn. 5-38 was derived. The appropriate stepwise development of Eqn. 5-38 and the independent variable squared cross correlation matrix are presented in Tables 5-20 and 5-21, respectively.

$$\log(\text{TBG}) = -0.098\ (\pm 0.704) + 0.563\ (\pm 0.496)\ \pi_{3'5'} - 1.100\ (\pm 0.482)\ \text{3'SIZE} > \text{I} + 2.366\ (+1.676)\ \sigma_{3'5'} \quad \text{(Eqn. 5-38)}$$

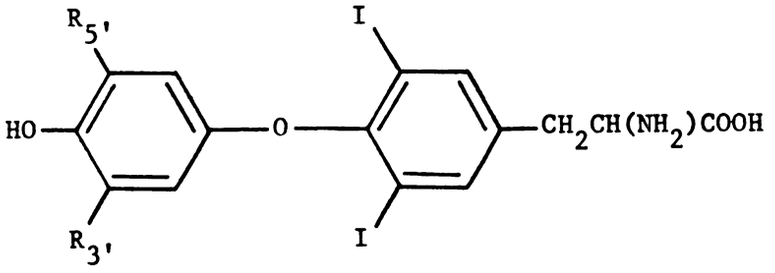
$$N = 10 \qquad R = 0.962 \qquad S = 0.355$$

Hydrophobic bonding by 3' and 5' substituents ($\pi_{3'5'}$), size-limited at least for the 3' substituent (3'SIZE > I), contributes only moderately

to binding. Of greater importance, however, is the apparent binding of the 4'-phenoxide to this plasma protein, as indicated by the large, positive $\sigma_{3',5'}$ regression coefficient. It thus appears that *in vivo* activity is enhanced by electron-donating 3' and 5' substituents, which discourage plasma protein binding and encourage passage of the unionized 4'-OH across cell membranes.

Preliminary reports[85] indicate that halogen-free analogs are capable, upon maternal administration, of effectively transversing placental barriers and mediating thyroid hormone effects in the fetus. This could be especially important in treating fetal hypothyroidism (which has been implicated as preventing normal fetal development, expressed, for example, by respiratory-distress syndrome at birth[214]) by administration of a compound which has low maternal activity but high fetal activity. The high fetal activity is probably due in part to the lack of susceptibility of the halogen-free analogs to fetal deiodinases. Also of importance, however, is the fact that the halogen-free analogs readily cross the placental barriers, while the natural hormones T_3 and T_4 do not. This can not be a distribution phenomenon based on compound lipophilicity, since T_3 and T_4 are more lipophilic than any of the active halogen-free analogs. It appears instead that the degree of *in vivo* 4'-OH ionization is the determining factor for whether an analog can cross the placental barriers. T_3 and T_4, with their much more highly ionized 4'-OH's, are more tightly bound to plasma protein and will resist passage across the placental barrier while ionized (even with an unionized 4'-OH, these compounds are apparently too large to cross

Table 5-19. Analog Structures and Data Used in the Formulation of Eqn. 5-38 Correlating In Vitro Binding to Purified Thyroxine Binding Globulin (TBG) for Thyroid Hormone Analogs (5-10).



5-10

Data Point #	Abbreviation	$R_{3'}$	$R_{5'}$	TBG[a] Obsd.	log(TBG) Obsd.	log(TBG) Calcd.[b]	log(TBG) Dev.
1	T4	I	I	100.0	2.000	2.016	-0.016
2	T3	I	H	9.0	0.954	0.959	-0.005
*3	T2	H	H	0.07	-1.155	-0.098	-1.057
4	3'5'Me2-T2	Me	Me	0.29	-0.538	-0.272	-0.266
5	3'OH-T2	OH	H	0.06	-1.222	-1.351	0.129
6	3'Me-T2	Me	H	0.28	-0.553	-0.185	-0.368
7	3'Et-T2	Et	H	1.59	0.201	-0.018	0.220
8	3'tBu-T2	tBu	H	0.67	-0.174	-0.468	0.294
9	3'iBu-T2	iBu	H	0.10	-1.000	-0.532	-0.468
10	3'Ph-T2	Ph	H	0.04	-1.398	-1.461	0.063
11	3'iPr-T2	iPr	H	3.53	0.548	0.131	0.417

Table 5-19. (Continued)

---

[a]From reference 30. TBG = relative binding affinity to TBG (relative to TBG (T_4) = 100) = (K_A/K_{T4}) x 100, where K_A = equilibrium association constant for analog A. No DL/L correction. On a molar basis.

[b]Calculated using Eqn. 5-38.

*Not used in calculating Eqn. 5-38.

Table 5-20. Stepwise Development of Eqn. 5-38.

Eqn. #	Eqn. and Statistical Data
5-39	$\log(TBG) = + 0.220 + 3.481\ \sigma_{3'5'}$
	$R = 0.731$ $S = 0.770$
	$F_{1,8} = 9.17$ (98.1% vs. mean)
5-40	$\log(TBG) = + 0.592 - 0.777$ 3'SIZE > I
	$+ 3.571\ \sigma_{3'5'}$
	$R = 0.911$ $S = 0.497$
	$F_{1,7} = 12.24$ (99.0% vs. Eqn. 5-39)
5-38	$F_{1,6} = 7.72$ (96.5% vs. Eqn. 5-40)

Table 5-21. Independent Variable Squared Cross Correlation Matrix for Eqn. 5-38.

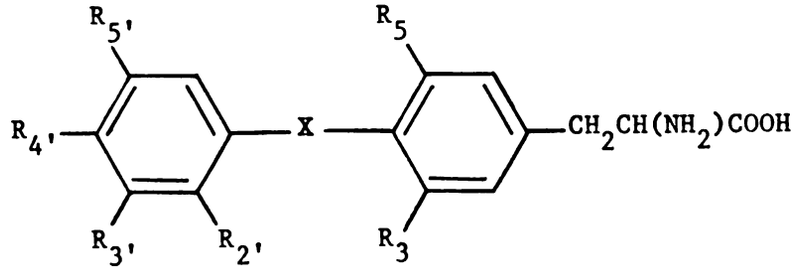
	$\pi_{3'5'}$	$\sigma_{3'5'}$	3'SIZE > I
$\pi_{3'5'}$	1.000	0.316	0.259
$\sigma_{3'5'}$		1.000	0.001
3'SIZE > I			1.000

the barrier, due to the bulk of the I atoms). In contrast, the 3,5-dimethyl-3'-alkyl-thyronines, with the smaller Me groups and essentially unionized 4'-OH, are weakly plasma protein bound and readily cross the placental barriers. It thus appears that the degree of 4'-OH ionization, as determined by the electronic influences of the 3' and 5' substituents, affects not only the degree of palsma protein binding, but also the ease with which analogs are able to cross cell membranes and the placental barriers.

CORRELATIONS BETWEEN *IN VIVO* AND *IN VITRO* ACTIVITIES

Utilizing the analogs of Table 5-22 and the data of Table 5-23, Eqn. 5-41 correlating *in vivo* antigoiter activities (BA) with *in vitro* binding to intact rat hepatic nuclei (BN) was derived.[26] The appropriate stepwise development of Eqn. 5-41 and the independent variable squared cross correlation matrix are presented in Tables 5-24 and 5-25, respectively. Eqn. 5-42 was originally derived[26] for correlation of *in vivo* antigoiter activities (BA) with *in vitro* binding to solubilized rat hepatic nuclear protein (BS). Utilizing the analogs of Table 5-26 and the data of Table 5-27 (improved and expanded from the data used in the derivation of Eqn. 5-42), Eqn. 5-43 was derived for correlation of *in vivo* antigoiter activities (BA) with *in vitro* binding to solubilized rat hepatic nuclear protein (BS). The appropriate stepwise development of Eqn. 5-43 and the independent variable squared cross correlation matrix are presented in Tables 5-24 and 5-28, respectively.

Table 5-22. Structures of Thyroid Hormone Analogs (5-11) Used in Deriving Eqn. 5-41.



5-11

Data Point #	Abbreviation	R_3	R_5	$R_{3'}$	$R_{5'}$	$R_{4'}$	$R_{2'}$	X
1	T3	I	I	I	H	OH	H	O
2	T2	I	I	H	H	OH	H	O
3	3'Me-T2	I	I	Me	H	OH	H	O
4	3'Et-T2	I	I	Et	H	OH	H	O
5	3'iPr-T2(L)	I	I	iPr	H	OH	H	O
6	3'tBu-T2	I	I	tBu	H	OH	H	O
7	3'iBu-T2	I	I	iBu	H	OH	H	O
8	3'Ph-T2	I	I	Ph	H	OH	H	O
*9	3'cHex-T2	I	I	cHex	H	OH	H	O
10	3'Cl-T2	I	I	Cl	H	OH	H	O
11	3'5'Cl2-T2	I	I	Cl	Cl	OH	H	O
12	3'5'Me2-T2	I	I	Me	Me	OH	H	O
13	T4	I	I	I	I	OH	H	O
14	3'iPr-T2(DL)	I	I	iPr	H	OH	H	O

Table 5-22. (Continued)

Data Point #	Abbreviation	R_3	R_5	$R_{3'}$	$R_{5'}$	$R_{4'}$	$R_{2'}$	X
*15	3'5'iPr2-T2	I	I	iPr	iPr	OH	H	O
16	4'H-3'Me-T2	I	I	Me	H	H	H	O
17	4'H-3'CF3-T2	I	I	CF_3	H	H	H	O
18	4'H-T3	I	I	I	H	H	H	O
19	Me3-T	Me	Me	Me	H	OH	H	O
20	Me4-T	Me	Me	Me	Me	OH	H	O
21	3'iPr-Me2-T	Me	Me	iPr	H	OH	H	O
22	2'3'Me2-T2	I	I	Me	H	OH	Me	O
23	2'5'Me2-T2	I	I	H	Me	OH	Me	O
24	2'Me-5'I-T2	I	I	H	I	OH	Me	O
25	Napth-T2	I	I	a	H	OH	a	O
26	4'NH2-T2	I	I	H	H	NH_2	H	O
27	33'-T2	I	H	I	H	OH	H	O
28	33'5'-T3	I	H	I	I	OH	H	O
29	3'iPr-Br2-T	Br	Br	iPr	H	OH	H	O
*30	3'I-iPr2-T	iPr	iPr	I	H	OH	H	O
31	MB-T3	I	I	I	H	OH	H	CH_2
*32	MB-T4	I	I	I	I	OH	H	CH_2
*33	SB-T2	I	I	H	H	OH	H	S
34	SB-T3	I	I	I	H	OH	H	S

[a] 2', 3'-$(CH)_4$

*Not used in calculating Eqn. 5-41.

Table 5-23. Data Used in the Formulation of Eqn. 5-41 Correlating In Vivo Antigoiter Activities (BA) with In Vitro Binding to Intact Rat Hepatic Nuclei (BN) for Thyroid Hormone Analogs.

Data Point # [a]	BA [b] Obsd.	BN [c] Obsd.	log(BN) Obsd.	log (BA) Obsd.	log (BA) Calcd. [d]	log (BA) Dev.
1	100.00	100.00	2.000	2.000	1.701	0.299
2	0.81	0.30	-0.523	-0.092	-0.140	0.049
3	14.47	13.5	1.130	1.160	1.067	0.094
4	40.8	21.0	1.322	1.611	1.207	0.404
5	142.1	104.0	2.017	2.153	1.714	0.439
6	21.7	38.5	1.586	1.337	1.399	-0.062
7	7.74	20.0	1.301	0.889	1.191	-0.302
8	2.03	2.0	0.301	0.308	0.461	-0.154
*9	----	1.4	0.146	----	0.348	----
10	4.88	6.2	0.792	0.688	0.820	-0.131
11	3.80	4.5	0.653	0.580	0.718	-0.138
12	9.04	6.2	0.792	0.956	0.820	0.136
13	18.1	12.5	1.097	1.258	1.042	0.216
14	83.4	100.0	2.000	1.916	1.701	0.214
* 15	----	1.4	0.146	----	0.348	----
16	2.71	0.2	-0.699	0.433	0.847	-0.414
17	13.56	0.2	-0.699	1.132	0.847	0.286
18	27.12	0.4	-0.398	1.433	1.066	0.367
19	0.54	0.1	-1.000	-0.268	-0.489	0.221
20	0.36	0.1	-1.000	-0.444	-0.489	0.045

Table 5-23. (Continued)

Data[a] Point #	BA[b] Obsd.	BN[c] Obsd.	log(BN) Obsd.	log (BA) Obsd.	log (BA) Calcd.[d]	log (BA) Dev.
21	3.60	0.7	-0.155	0.556	0.128	0.428
22	9.04	1.10	0.041	0.956	0.271	0.685
23	0.18	0.1	-1.000	-0.745	-0.489	-0.256
24	0.36	0.30	-0.523	-0.444	-0.140	-0.303
25	18.08	8.0	0.903	1.257	0.901	0.357
26	0.27	0.76	-0.119	-0.569	0.154	-0.723
27	0.25	0.50	-0.301	-0.602	0.021	-0.624
28	0.125	0.1	-1.000	-0.903	0.489	-0.414
29	30.0	36.0	1.556	1.477	1.378	0.100
*30	----	0.2	-0.699	----	-0.269	----
31	54.25	250.0	2.398	1.734	1.992	-0.258
*32	----	2.6	0.415	----	0.544	----
*33	----	1.3	0.114	----	0.324	----
34	13.82	100.0	2.000	1.140	1.701	-0.561

[a] See Table 5-23.

[b] From Appendix I. No DL correction. On a molar basis. Relative to L-T_3 = 100.

[c] See footnote b, Table 5-9.

[d] Calculated using Eqn. 5-41.

* Not used in calculating Eqn. 5-41.

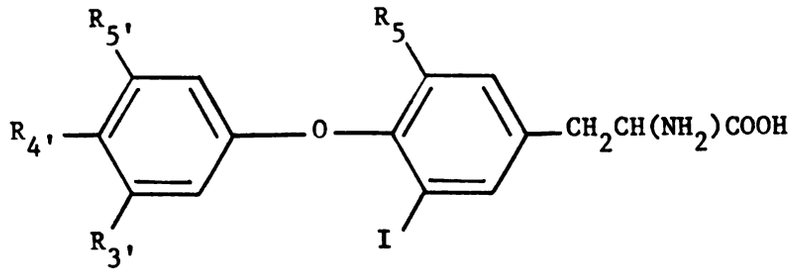
Table 5-24. Stepwise Development of Eqns. 5-41 and 5-43.

Eqn. #	Eqn. and Statistical Data
5-44	$\log(BA) = + 0.404 + 0.635 \log (BN)$
	$R = 0.820 \qquad S = 0.522$
	$F_{1,27} = 55.43$ (>99.9% vs. mean)
5-41	$F_{1,26} = 27.64$ (>99.9% vs. Eqn. 5-44)
5-45	$\log(BA) = + 0.597 + 0.548 \log (BS)$
	$R = 0.779 \qquad S = 0.692$
	$F_{1,24} = 37.08$ (>99.9% vs. mean)
5-46	$\log(BA) = + 0.383 + 0.720 \log (BS) + 0.984$ I4'H
	$R = 0.915 \qquad S = 0.454$
	$F_{1,23} = 32.75$ (>99.9% vs. Eqn. 5-45)
5-43	$F_{1,22} = 7.38$ (98.6% vs. Eqn. 5-46)

Table 5-25. Independent Variable Squared Cross Correlation Matrix for Eqn. 5-41.

	log(BN)	I4'H
log (BN)	1.000	0.095
I4'H		1.000

Table 5-26. Structures of Thyroid Hormone Analogs (5-12) Used in Deriving Eqn. 5-43.



5-12

Data Point #	Abbreviation	R_5	$R_{3'}$	$R_{5'}$	$R_{4'}$
1	4'H-T2	I	H	H	H
*2	T2	I	H	H	OH
3	4'NH2-T2	I	H	H	NH_2
4	4'H-T3	I	I	H	H
5	T3	I	I	H	OH
6	4'OCH3-T3	I	I	H	OCH_3
7	4'H-3'Me-T2	I	Me	H	H
8	3'Me-T2	I	Me	H	OH
*9	4'OCH3-3'Me-T2	I	Me	H	OCH_3
10	3'iPr-T2	I	iPr	H	OH
11	4'OCH3-3'iPr-T2	I	iPr	H	OCH_3
12	3'sBu-T2	I	±sBu	H	OH
*13	4'H-3'tBu-T2	I	tBu	H	H
14	3'tBu-T2	I	tBu	H	OH

Table 5-26. (Continued)

Data Point #	Abbreviation	R_5	$R_{3'}$	$R_{5'}$	$R_{4'}$
15	4'OCH3-3'tBu-T2	I	tBu	H	OCH_3
16	4'H-3'5'Me2-T2	I	Me	Me	H
*17	4'OCH3-3'5'Me2-T2	I	Me	Me	OCH_3
*18	4'H-3'NO2-T2	I	NO_2	H	H
19	3'NO2-T2	I	NO_2	H	OH
20	4'NH2-3'Me-T2	I	Me	H	NH_2
21	4'NH2-3'5'Me2-T2	I	Me	Me	NH_2
22	33'-T2	H	I	H	OH
23	3'F-T2	I	F	H	OH
24	3'Cl-T2	I	Cl	H	OH
25	3'Br-T2	I	Br	H	OH
26	T4	I	I	I	OH
*27	4'H-3'iPr-T2	I	iPr	H	H
28	3'nPr-T2	I	nPr	H	OH
29	4'H-3'Br-T2	I	Br	H	H
30	4'H-3'Cl-T2	I	Cl	H	H
31	4'H-3'F-T2	I	F	H	H
*32	3'5'iPr-T2	I	iPr	iPr	OH
*33	4'OCH3-3'sBu-T2	I	$\pm$sBu	H	OCH_3
34	3'5'Cl2-T2	I	Cl	Cl	OH
*35	3'5'Br2-T2	I	Br	Br	OH
*36	3'iPr-5'Cl-T2	I	iPr	Cl	OH
*37	3'iPr-5'Br-T2	I	iPr	Br	OH
*38	3'iPr-355'-T3	I	iPr	I	OH

*Not used in calculating Eqn. 5-43.

Table 5-27. Data Used in the Formulation of Eqn. 5-43 Correlating *In Vivo* Antigoiter Activities (BA) with *In Vitro* Binding to Solubilized Rat Hepatic Nuclear Protein (BS) for Thyroid Hormone Analogs.

Data Point # [a]	BA [b] Obsd.	BS [c] Obsd.	log(BS) Obsd.	log(BS) Obsd.	log(BS) Calcd. [d]	log(BS) Dev.
1	1.24	0.01	-2.000	0.093	-0.082	0.175
*2	0.81	0.082	-1.086	-0.092	-0.548	0.457
3	0.036	0.0031	-2.502	-1.444	-1.620	0.176
4	27.12	0.230	-0.638	1.433	0.949	0.484
5	100.0	100.0	2.000	2.000	1.788	0.212
6	11.25	1.29	0.111	1.051	0.981	0.070
7	2.71	0.225	-0.648	0.433	0.942	-0.509
8	14.47	3.30	0.518	1.160	0.666	0.494
*9	----	0.17	-0.770	---	0.314	---
10	142.1	89.15	1.950	2.153	1.750	0.402
11	19.0	6.82	0.834	1.279	1.528	-0.249
12	79.9	78.29	1.894	1.902	1.708	0.195
*13	---	0.335	-0.475	---	1.073	---
14	21.7	8.45	0.927	1.336	0.976	0.361
15	2.35	0.27	-0.569	0.371	0.466	-0.095
16	0.054	0.145	-0.839	-1.268	-0.361	-0.907
*17	---	0.335	-0.475	---	0.537	---
*18	---	0.038	-1.420	---	0.358	---
19	0.18	0.225	-0.648	-0.745	-0.216	-0.528
20	0.036	0.031	-1.509	-1.444	-0.868	-0.576

Table 5-27. (Continued)

Data Point # [a]	BA [b] Obsd.	BS [c] Obsd.	log(BS) Obsd.	log(BA) Obsd.	log(BA) Calcd. [d]	log(BA) Dev.
21	0.13	0.041	-1.387	-0.886	-0.776	-0.110
22	0.25	0.688	-0.162	-0.602	0.151	-0.753
23	0.65	0.164	-0.785	-0.187	-0.320	0.133
24	4.88	3.73	0.572	0.688	0.707	-0.018
25	23.78	15.89	1.201	1.376	1.183	0.193
26	18.1	13.85	1.141	1.258	1.138	0.120
*27	---	0.492	-0.308	---	1.199	---
28	39.5	23.97	1.380	1.597	1.318	0.278
29	18.0	0.24	-0.620	1.255	0.963	0.292
30	7.78	0.118	-0.928	0.891	0.730	0.161
31	1.39	0.0136	-1.866	0.143	0.020	0.123
*32	---	1.10	0.041	---	0.305	---
*33	---	1.29	0.111	---	0.981	---
34	3.80	3.71	0.569	0.580	0.705	-0.125
*35	1.58	5.07	0.705	0.199	0.808	-0.609
*36	---	52.56	1.721	---	1.576	---
*37	---	21.95	1.341	---	1.289	---
*38	---	12.41	1.094	---	1.102	---

[a] See Table 5-26.

[b] See footnote b, Table 5-23.

[c] See footnote b, Table 5-14.

[d] Calculated using Eqn. 5-43.

* Not used in calculating Eqn. 5-43.

Table 5-28. Independent Variable Squared Cross Correlation Matrix for Eqn. 5-43.

	log(BS)	I4'H	$I4'OCH_3$
log(BS)	1.000	0.210	0.001
I4'H		1.000	0.064
$I4'OCH_3$			1.000

$$\log(BA) = +\ 0.241\ (\pm 0.170) + 0.730\ (\pm 0.134)\ \log(BN) + 1.116\ (\pm 0.484)\ I4'H$$

(Eqn. 5-41)

N = 29 R = 0.917 S = 0.371

$$\log(BA) = +\ 0.261\ (\pm 0.207) + 0.853\ (\pm 0.152)\ \log(BS) + 1.164\ (0.466)\ I4'H + 0.609\ (\pm 0.437)\ I4'OCH_3$$

(Eqn. 5-42)

N = 22 R = 0.943 S = 0.410

$$\log(BA) = +\ 0.274\ (\pm 0.192) + 0.757\ (\pm 0.135)\ \log(BS) + 1.159\ (\pm 0.362)\ I4'H + 0.623\ (\pm 0.387)\ I4'OCH_3$$

(Eqn. 5-43)

N = 26 R = 0.937 S = 0.402

It thus appears that in vivo antigoiter activities (BA) correlate well with in vitro binding to intact rat hepatic nuclei (BN) and with in vitro binding to solubilized rat hepatic nuclear protein (BS), with adjustments made for in vivo metabolism of 4'-deoxy ($I4'H$) and 4'-OCH_3 ($I4'OCH_3$) analogs. (Data Point #16 = 4'H-3'5'Me2-T2 of Tables 5-26 and 5-27 was assigned a value of 0 for I4'H in that the two 3' and 5' CH_3 groups apparently sterically inhibit the in vivo hydroxylation of this analog.) The deviations of the intercepts from 0.0 and of the log(BN) and log(BS) regression coefficients from 1.0 reflect not only the minimal accuracy of some of the in vivo data, but also differences in analog plasma protein binding, metabolism, elimination, and 4'-OH ionization effects in vivo.

Addition of a $\sigma_{3',5'}$ term to Eqn. 5-43 was not significant, even though log(BA) and log(BS) correlate with $-\sigma_{3',5'}$ (Eqn. 5-8) and with $+\sigma_{3',5'}$ (Eqn. 5-26), respectively. This is not completely surprising, however, since the $\sigma_{3',5'}$ values used for deriving Eqns. 5-8 and 5-26 differ for the 4'-deoxy and 4'-OCH_3 analogs (see the discussion above concerning substituent constant choices). Further examination of this area will certainly be of interest.

CONCLUSIONS

The results of all of these quantitative structure-activity relationship correlations can be briefly summarized as follows:

(1) The correlations confirm that both in vivo antigoiter activity and in vitro binding to nuclear receptors are enhanced by bulky, lipophilic 3 and 5 substituents and by size-limited, lipophilic 3' substituents and are decreased by any 5' substituent bulk or lipophilicity.

(2) In vivo activity is enhanced by electron donating 3' and 5' substituents, which prevent plasma protein binding and encourage movement of the analog into cells.

(3) In vitro binding probably involves hydrogen bond donation of the 4'-OH to the 5' side of the nuclear receptor.

(4) The good correlations between in vivo antigoiter activities and in vitro nuclear receptor binding reflect that the latter is probably the first step in initiating subsequent hormonal responses (as expressed through protein synthesis).

(5) Except for 3' and 5' substituent influences on the degree of 4'-OH ionization (and hence on the degree of plasma protein binding),

distribution, at least within the range of analog lipophilicities studied, does not play a major role in determining whole animal activities.

(6) Binding of analogs to TBG is strongly influenced by the degree of 4'-OH ionization.

(7) The success of our use of the 3'SIZE > I and INTERACT parameters in these QSAR studies indicate that they are reasonable estimates of the physical properties they were designed to predict. Hopefully these, or similarly derived parameters, may be of use in future QSAR studies of the thyroid hormone analog or other systems for which the "traditional" substituent parameters are unable to account for particular physico-chemical properties of the analogs.

A diagramatic representation of the overall picture of thyroid hormone analog binding to nuclear receptor is presented in Figure 5-1.

As mentioned above, one of the difficulties involved in the QSAR examination of thyromimetic activities is the lack of orthogonality between some of the substituent constants. This is especially evident for: (1) π_5, and Es_5,; (2) $\sigma_{3',5'}$ and INTERACT; and (3) π_{35} and Es_{35}. The complete synthesis and testing of each new thyroid hormone analog requires considerable effort, time, and expense. In order to obtain the maximal amount of information for the fewest number of new analogs, therefore, it will be necessary to insure that future design of new analogs takes into account the orthogonality of substituent constants. Pre-synthesis substituent constant cluster analysis[216] and examination of substituent constant squared cross correlation matrices should certainly be utilized in such analysis. Some "intuitive" suggestions

(based on the known qualitative and quantitative SAR's for the various thyromimetic activities of the thyroid hormone analogs, previously observed (see above) lack of orthogonality between substituent constants, and possible ease of synthesis) can be made, however, for a number of new substituent combinations:

(1) 3'-COR, where R = H, CH_3, C_2H_5, C_3H_7, etc.: strong intramolecular hydrogen bond with the potential for varying the substituent "size" and lipophilicity.

(2) 3'-I, 5'-Me; 3'-Br, 5'-Me: $\sigma_{3'5'} \tilde{\approx} \sigma_{3'5'}$ as for 3'-iPr, 5'-halogen, but INTERACT $\tilde{\approx}$ -INTERACT as for 3'-iPr, 5'-halogen.

(3) 3' and/or 5' substituents with π and Es substituent constants not correlated: e.g. NH_2, NMe_2, NEt_2, CH_2OH, C_2H_5OH.

(4) 3'-nBu; 3'-nAmyl; 3'-$CH(CH_3)C_3H_7$; 3'-$CH(C_2H_5)_2$: for further investigation of the π and "size" 3'-substituent requirements for thyromimetic activity.

(5) 3,5-$(CF_3)_2$ and other 3,5 substituents with greater orthogonality of π_{35} and Es_{35}. Although thyromimetic activities have been examined for a wide variety of 3,5 substituents, finite quantitative activities are available for only a limited number of 3,5 substituents (F, Cl, Br, I, CH_3) and for these π_{35} and Es_{35} are well correlated. In addition to thyromimetic activity being influenced by the ability of the 3,5 substituents to "lock" the diphenyl ether nucleus into the proximal and distal conformations, 3,5 substituents larger than iodine (e.g., iPr, SPh) (larger than bromine for TBG binding[30]) also exert a negative steric influence on activity (see Appendix I). Determination of finite quantitative

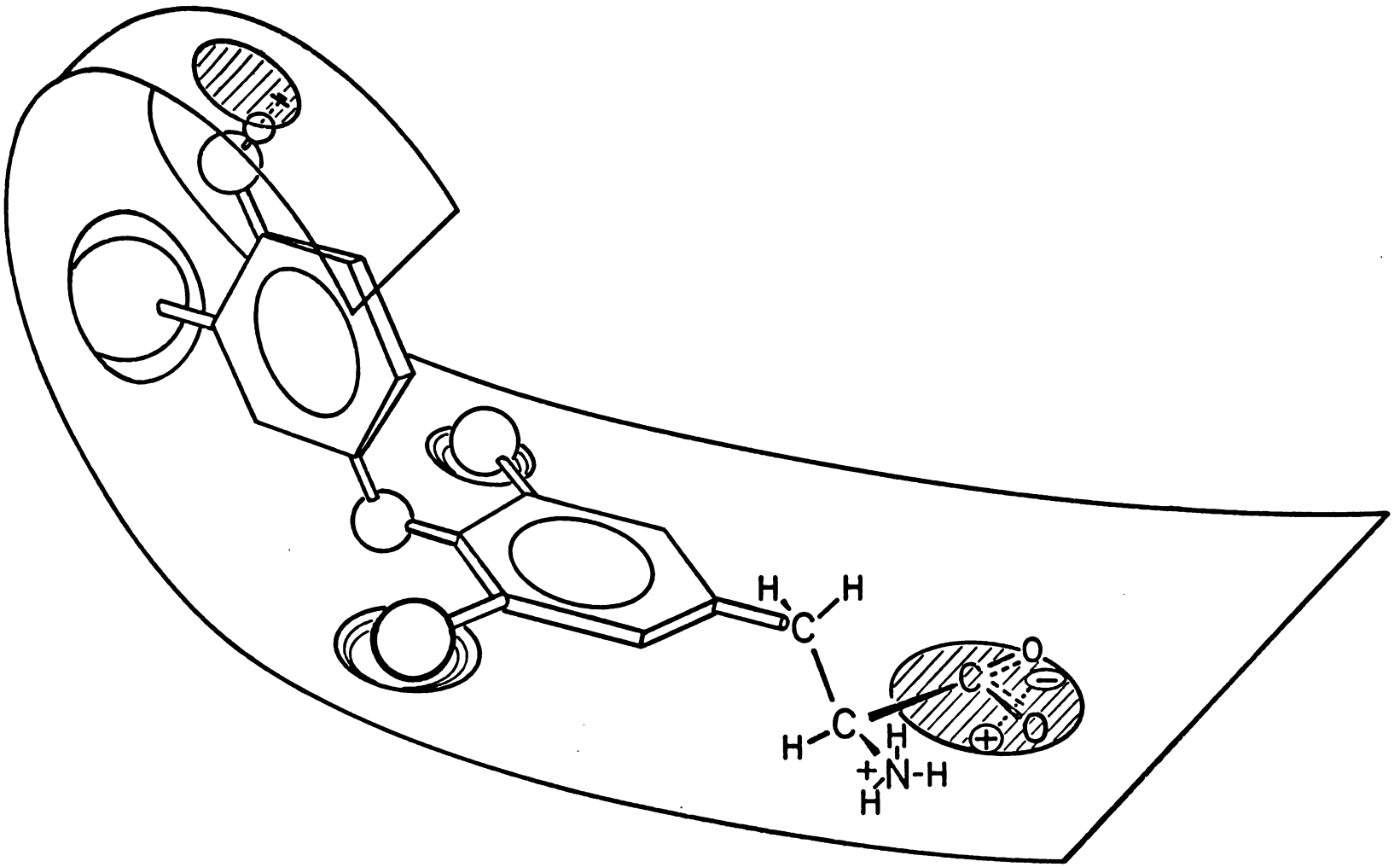


Figure 5-1. Diagrammatic representation of thyroid hormone analog binding to nuclear receptor.[42,215]

thyromimetic activities for a wider variety of 3,5 substituents is most necessary for determining the relative importance in influencing thyromimetic activities of 3,5 substituent lipophilicity, "size" or "bulk", and conformational flexibility (as this affects the diphenyl ether conformation).

CHAPTER SIX: CONCLUDING REMARKS

The thyroid hormones and analogs have proven to be an excellent system for the experimental and theoretical studies presented here. The results of the synthetic and bioassay investigations, coupled with the molecular orbital studies and quantitative structure-activity correlations, have helped to further define and to provide a better understanding of the structural requirements, physico-chemical properties, and interactive effects of the outer ring substituents, as they influence the various thyromimetic activities. The ready accessibility of thyroid hormone nuclear receptors, as well as other assay systems (e.g., in vivo antigoiter assay, in vitro binding to plasma proteins, etc.) provide opportunities for further research. The partitioning into substituent contributions of the free energy of interaction with macromolecules, coupled with available and potentially available X-ray structures for these macromolecules, will provide an opportunity for examining

the physical origins of these interactions. Further theoretical studies (e.g., MO, QSAR, electrostatic potential models) will most certainly play a role in the examination of these interactions. Additional experimental studies (e.g., determination of ΔH and ΔS of binding to macromolecules, further thyromimetic activity determinations, fetal studies, biochemical investigations of nuclear receptors) will be equally important. Of special interest, at least in retrospect from these studies, will be the design, synthesis, and testing of analogs with physico-chemical properties which are considerably more orthogonal than those presently available. This alone should help to provide a better understanding of the physical origins of specific analog--macromolecule interactions. The thyroid hormones and anlogs truly do provide the potential for molecular-level investigation of the structural, physical, and mechanistic properties of the until-recently-elusive "receptor".

REFERENCES AND NOTES

1. R. Pitt-Rivers and J. R. Tata, "The Thyroid Hormones", Pergamon Press, Oxford, England, 1959, Chapter 5.

2. A review: E. C. Jorgensen, "Thyroid Hormones and Antithyroid Drugs", in "Medicinal Chemistry", A. Burger, Ed., 3rd Ed., Part II, John Wiley & Sons, Inc., New York, 1970.

3. G. Cilento and M. Berenholc, Biochim. Biophys. Acta, 94, 271 (1965).

4. Szent-Györgi, "Bioenergetics", Academic Press, New York, 1957, pp. 24, 27.

5. P. A. Lehman, J. Med. Chem., 15, 404 (1972).

6. C. Niemann, Fortsch. Chem. Org. Naturst.,7, 167 (1950).

7. C. Niemann and J. F. Mead, J. Amer. Chem. Soc., 63, 2685 (1941).

8. E. C. Jorgensen, "Structure Activity Relationships of Thyroxine Analogs",in "International Encyclopedia of Pharmacology and Therapeutics", Section 43, "Pharmacology of the Thyroid Gland", in press.

9. E. C. Jorgensen, W. J. Murray, and P. Block, Jr., J. Med. Chem., 17, 434 (1974).

10. V. Cody, J. Med. Chem., 18, 126 (1975).

11. V. Cody, J. Amer. Chem. Soc., 96, 6720 (1974).

12. V. Cody and W. L. Duax, Biochem. Biophys. Res. Commun., 52, 430 (1973).

13. V. Cody and W. L. Duax, Science, 181, 757 (1973).

14. V. Cody, W. L. Duax,and D. A. Norton, Acta Cryst., B28, 2244 (1972).

15. P. A. Lehman and E. C. Jorgensen, Tet., 21, 363 (1965).

16. J. C. Emmett and E. S. Pepper, Nature, 257, 334 (1975).

17. P. A. Kollman, W. J. Murray, M. E. Nuss, E. C. Jorgensen, and S. Rothenberg, J. Amer. Chem. Soc., 95, 8518 (1973).

18. E. C. Jorgensen, Proc. Mayo Clin., 39, 560 (1964).

19. N. Zenker and E. C. Jorgensen, J. Amer. Chem. Soc., 81, 4643 (1959).

20. E. C. Jorgensen and P. N. Kaul, J. Amer. Pharm. Assoc., 48, 653 (1959).

21. E. C. Jorgensen, N. Zenker, and C. Greenberg, J. Biol. Chem., 235, 1732 (1960).

22. E. C. Jorgensen and P. A. Lehman, J. Org.. Chem., 26, 894 (1961).

23. E. C. Jorgensen, P. A. Lehman, C. Greenberg, and N. Zenker, J. Biol. Chem., 237, 3832 (1962).

24. D. Koerner, H. L. Schwartz, M. I. Surks, J. H. Oppenheimer, and E. C. Jorgensen, J. Biol. Chem., 250, 6417 (1975).

25. E. C. Jorgensen, S. W. Dietrich, D. Koerner, M. I. Surks, and J. H. Oppenheimer, Proc. West. Pharmacol. Soc., 18, 389 (1975).

26. E. C. Jorgensen, M. B. Bolger, and S. W. Dietrich, in press, Excerpta Medica, Eighth International Thyroid Conference, Hamburg, Germany, July, 1976.

27. H. H. Samuels, J. S. Tsai, and J. Casanova, Science, 184, 1188 (1974).

28. H. H. Samuels, J. S. Tsai, J. Casanova, and F. Stanley, J. Clln. Invest., 54, 853 (1974).

29. S. M. Snyder, R. R. Cavalieri, I. D. Goldfine, E. C. Jorgensen, and S. H. Ingbar, Excerpta Medica No. 361, International Congress Series, Seventh International Thyroid Conference, Boston, June, 1975.

30. S. M. Snyder, R. R. Cavalieri, I. D. Goldfine, S. H. Ingbar, and E. C. Jorgensen, accepted for publication, J. Biol. Chem.

31. M. I. Surks, D. Koerner, W. Dillman, and J. H. Oppenheimer, J. Biol. Chem., 248, 7066 (1973).

32. M. I. Surks, D. H. Koerner, and J. H. Oppenheimer, J. Clin. Invest., 55, 50 (1975).

33. D. Koerner, M. I. Surks,and J. H. Oppenheimer, J. Clin. Endocrinol. Metab., 38, 706 (1974).

34. K. Sterling, Proc. Mayo Clin., 39, 586 (1964).

35. C. C. F. Blake, M. J. Geisow, I. D. A. Swan, C. Rerat, and B. Rerat, J. Mol. Biol., 88, 1 (1974).

36. S. W. Dietrich, M. B. Bolger, T. A. Andrea, and E. C. Jorgensen, to be submitted for publication, J. Med. Chem.

37. S. W. Dietrich, E. C. Jorgensen, P. A. Kollman, and S. Rothenberg, J. Amer. Chem. Soc., in press, 1976.

38. T. A. Andrea, S. W. Dietrich, W. J. Murray, E. C. Jorgensen, P. A. Kollman, and S. Rothenberg, to be submitted for publication, J. Amer. Chem. Soc.

39. S. W. Dietrich, P. A. Kollman, and E. C. Jorgensen, Abstract #55, Amer. Chem. Soc. Division of Medicinal Chemistry, 172nd ACS National Meeting, San Francisco, August, 1976.

40. S. W. Dietrich, M. B. Bolger, E. C. Jorgensen, and P. A. Kollman, to be submitted for publication, J. Med. Chem.

41. T. A. Andrea, P. A. Kollman,and E. C. Jorgensen, Abstract #54, Amer. Chem. Soc. Division of Medicinal Chemistry, 172nd ACS National Meeting, San Francisco, August, 1976.

42. E. C. Jorgensen and T. A. Andrea, Excerpta Medica, Amsterdam, 1976; Proceedings International Thyroid Conference, Boston, June, 1975.

43. M. B. Bolger and E. C. Jorgensen, unpublished results, 1976.

44. C. C. Schussler and T. Y. Wang, Excerpta Medica No. 361, International Thyroid Conference, Boston, June, 1975.

45. S. F. Nilsson and P. A. Peterson, J. Biol. Chem., 250, 8543 (1975).

46. S. F. Nilsson and P. A. Peterson, J. Biol. Chem., 216, 6098 (1971).

47. S. F. Nilsson, L. Rask, and P. A. Peterson, J. Biol. Chem., 250, 8554 (1975).

48. R. A. Pages, J. Robbins, and H. Edelhoch, Biochem., 12, 2773 (1973).

49. Personal communications from C. C. F. Blake and S. J. Oatley.

50. S. B. Barker and M. Shimada, Proc. Mayo Clin., 39, 609 (1964).

51. E. C. Jorgensen and S. J. Feinglass, unpublished results.

52. E. C. Jorgensen, unpublished estimate.

53. B. Blank, F. R. Pfeiffer, C. M. Greenberg, and J. F. Kerwin, J. Med. Chem., 6, 554 (1963).

54. E. C. Jorgensen and J. A. W. Reid, J. Med. Chem., 8, 533 (1965).

55. K. Tomita, H. A. Lardy, D. Johnson and A. Kent, J. Biol. Chem., 236, 2981 (1961).

56. H. Kubinyi, J. Med. Chem., 19, 587 (1976).

57. C. Hansch, A. Leo, S. H. Unger, K. H. Kin, D. Nikaitani, and E. J. Lien, J. Med. Chem., 16, 1207 (1973).

58. J. H. Barnes, E. T. Borrows, J. Elks, B. A. Hems, and A. G. Long, J. Chem. Soc., 2824 (1950).

59. L. F. Fieser, "Experiments in Organic Chemistry", 3rd Ed., Revised, D. C. Heath and Company, Boston, 1957, p. 310

60. K. S. Dhami and J. B. Stothers, Can. J. Chem., 44, 2855 (1966).

61. T. H. Coffield, A. H. Filbey, G. G. Ecke, and A. J. Kolka, J. Amer. Chem. Soc., 79, 5019 (1957).

62. P. Block, Jr. and D. H. Coy, J. Chem. Soc., Perkins Trans. 1, 633 (1972).

63. J. R. Chalmers, G. T. Dickson, J. Elks, and B. A. Hems, J. Chem. Soc., 3424 (1949).

64. T. A. Andrea, personal communication.

65. "Dictionary of Organic Compounds", 4th Ed., I. Heilbron, Ed., Oxford University Press, New York, 1965.

66. W. F. Zimmer, Jr., U. S. Patent 3,028,410 (April 3, 1962).

67. The conditions for the modified workup of this reaction were worked out with T. A. Andrea.

68. Additional experimental data and suggestions for improving the yields for synthesis of the substituted di-(p-anisyl)-iodonium iodides were provided by B. Blank, P. Block, Jr., and D. H. Coy.

69. Initial GC studies conducted by R. J. Weinkam; final GC studies conducted by W. Vinson.

70. J. H. Barnes, R. C. Cookson, G. T. Dickson, J. Elks, and V. D. Poole, J. Chem. Soc., 1448 (1953).

71. J. H. Gaddum, J. Physiol., 64, 246 (1928).

72. T. C. Bruice, R. J. Winzler, and N. Kharasch, J. Biol. Chem., 210, 1 (1954).

73. A. Wahlborg, C. Bright, and E. Frieden, Endocrinology, 75, 561 (1964).

74. E. Frieden and G. W. Westmark, Science, 133, 1487 (1961).

75. H. P. Ashley, P. Katti, and E. Frieden, Devel. Biol., 17, 293 (1968).

76. E. Frieden and K. Yoshizato, Endocrinology, 95, 188 (1974).

77. S. B. Barker, C. S. Pittman, J. A. Pittman, Jr., and S. R. Hill, Jr., Ann. N. Y. Acad. Sci., 86, 545 (1960).

78. S. B. Barker, M. Shimada, M. Makiuchi, Endocrinology, 76, 115 (1965).

79. E. W. Dempsey and E. B. Astwood, Endocrinology, 32, 509 (1943).

80. R. E. Cortell, J. Clin. Endocrinol. Metab., 9, 955 (1949).

81. M. V. Mussett and R. Pitt-Rivers, Metabolism, 6, 18 (1957).

82. E. C. Jorgensen and P. Slade, J. Med. Pharm. Chem., 5, 729 (1962).

83. G. S. Boyd and M. F. Oliver, J. Endocrinology, 21, 25 (1960).

84. C. M. Greenberg, C. A. Bocher, J. F. Kerwin, S. M. Greenberg, and T. H. Lin, Amer. J. Physiol., 201, 732 (1961).

85. F. Comite, G. N. Burrow, and E. C. Jorgensen, Clin. Res., 24, 271A (1976).

86. Binding affinities of thyroid hormone analogs to solubilized rat hepatic nuclear protein were determined by M. B. Bolger.

87. The possibility of injection of daily graded doses of a thiouracil solution on a weight basis, as opposed to thiouracil inclusion in the solid feed, might insure a more constant daily thiouracil dose.

88. N. Bailey, "Statistical Methods in Biology", John Wiley & Sons, Inc., New York, 1959, pp. 47-50.

89. J. Ipsen and P. Feigl, "Bancroft's Introduction to Biostatistics", 2nd Ed., Harper & Row, New York, 1970. pp. 55-58.

90. T. H. Wonnacott and R. J. Wonnacott, "Introductory Statistics", 2nd Ed., John Wiley & Sons, Inc., New York, 1972, p. 490.

91. T. C. Bruice, N. Kharasch, and R. J. Winzler, Arch. Biochem. Biophys., 62, 305 (1956).

92. C. Hansch and T. Fujita, J. Amer. Chem. Soc., 86, 1616 (1964).

93. C. Hansch, A. R. Steward, J. Isawa, and E. W. Deutsch, Mol. Pharmacol., 1, 205 (1965).

94. H. Kubinyi and O.-H. Kehrhahn, J. Med. Chem., 19, 578 (1976).

95. These studies.

96. R. W. Rawson, W. L. Money, R. L. Kroc, S. Kumaoka, R. S. Benua, and R. D. Leeper, Amer. J. Medical Sciences, 238, 261 (1959).

97. N. R. Stasilli, R. L. Kroc, and R. I. Meltzer, Endocrinology, 64, 62 (1959).

98. C. M. Greenberg, B. Blank, F. R. Pfeiffer, and J. F. Pauls, Amer. J. Physiol., 205, 821 (1963).

99. M. Wool, V. S. Fang, and H. A. Selenkow, Endocrinology, 78, 29 (1966).

100. J. Roche, R. Michel, M. Wolf, and N. Etling, Compt. rend. Soc. biol., 148, 1738 (1954).

101. H. A. Selenkow, C. A. Plamondon, J. G. Wiswell, and S. P. Asper, Jr., Bull. John Hopkins Hosp., 102, 94 (1958).

102. R. Pitt-Rivers, Metabolism, 66, 628 (1960).

103. K. Tomita and H. A. Lardy, J. Biol. Chem., 219, 595 (1956).

104. C. H. Duncan and M. M. Best, Endocrinology, 63, 169 (1958).

105. C. A. Plamondon, H. A. Selenkow, J. G. Wiswell, and S. P. Asper, Jr., Bull. John Hopkins Hosp., 102, 88 (1958).

106. R. Michel and R. Pitt-Rivers, Biochim. Biophys. Acta, 24, 213 (1957).

107. J. Gross and R. Pitt-Rivers, Lancet, 262, 593 (1952).

108. J. Gross and R. Pitt-Rivers, Biochemical J., 53, 652 (1953).

109. A. E. Heming and D. E. Holtkamp, Proc. Soc. Exptl. Biol. & Med., 83, 875 (1953).

110. K. I. Colville and D. D. Bonnycastle, Lancet, 265, 44 (1953).

111. J. R. Tata, in "The Thyroid Gland", R. Pitt-Rivers and W. R. Trotter, Eds., Vol. I, Buttersworth, Washington, D. C., 1964, p. 163.

112. C. S. Pittman, J. B. Chambers, Jr., and V. H. Read, J. Clin. Invest., 50, 1187 (1971).

113. C. S. Pittman, H. Nakafugi, and V. H. Read, Clin. Res., 18, 75 (1970).

114. L. E. Braverman, S. H. Ingbar, and K. Sterling, J. Clin. Invest., 49, 855 (1970).

115. K. Sterling, M. A. Brenner and E. S. Newman, Science, 169, 1099 (1970).

116. H. L. Schwartz, M. I. Surks, and J. H. Oppenheimer, J. Clin. Invest., 50, 1124 (1971).

117. J. H. Oppenheimer, H. L. Schwartz, and M. I. Surks, J. Clin. Invest., 51, 2493 (1972).

118. M. I. Surks, A. R. Schadlow, J. M. Stock, and J. H. Oppenheimer, J. Clin. Invest., 52, 805 (1973).

119. I. J. Chopra, J. Clin. Invest., 54, 583 (1974).

120. E. V. Flock, J. L. Bollman, and J. H. Grindlay, Proc. Mayo Clin., 35, 75 (1960).

121. E. V. Flock, J. L. Bollman, J. H. Grindlay, and G. H. Stobie, Endocrinology, 69, 626 (1961).

122. E. V. Flock, J. L. Bollman, and J. H. Grindlay, Endocrinology, 67, 419 (1960).

123. See Appendix I for references to analog rat antigoiter bioassay activities.

124. J. R. Tata and C. C. Windell, Biochemical J., 98, 604 (1966).

125. N. Camerman and A. Camerman, Science, 175, 764 (1972).

126. L. B. Kier and J. R. Hoyland, J. Med. Chem., 13, 1182 (1970).

127. J. A. Pople, D. P. Santry, and G. A. Segal, J. Chem. Phys., 43, S129 (1965).

128. J. A. Pople and G. A. Segal, J. Chem. Phys., 43, S136 (1965).

129. The CNDO/2 calculations were carried out with QCPE Program 91.

130. J. A. Pople and M. Gordon, J. Amer. Chem. Soc., 89, 4253 (1967).

131. "Tables of Interatomic Distances and Configuration in Molecules and Ions: Supplement", The Chemical Society of London, Special Publication No. 18, 1965.

132. W. J. Hehre, W. A. Lathan, R. Ditchfield, M. D. Newton, and J. A. Pople, Program No. 236 Quantum Chemistry Exchange, Indiana University, 1973.

133. W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys., 51, 2657 (1969).

134. J. P. Jesson and E. L. Muetterties, "Basic Chemical and Physical Data", Marcel Dekker, N. Y., 1969.

135. E. L. Eliel, "Stereochemistry of Carbon Compounds", McGraw-Hill, N. Y., 1962, p. 134.

136. A. L. McClellan, "Tables of Experimental Dipole Moments", W. W. Freeman and Co., San Francisco, 1963.

137. The following abbreviations are used in this chapter:

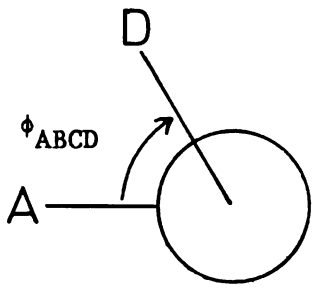
R(A-B) = A-B bond length

R(A - - C) = A - - C internuclear distance

θ_{ABC} = A-B-C angle

ϕ_{ABCD} = the dihedral angle between the C-D bond and the A-B bond. Changes in ϕ_{ABCD} are further defined by a clockwise rotation of the C-D bond relative to the A-B bond, when observed down the C-B bond from C toward B (see next page).

$\Delta E(I \rightarrow II) = E_{II} - E_{I}$.


D
ϕ_{ABCD}
A

138. M. N. Frey, T. F. Koetzle, M. S. Lehmann, and W. C. Hamilton, J. Chem. Phys., 58, 2547 (1973).

139. W. S. Benedict, N. Gailar, and E. K. Plyler, J. Chem. Phys., 24, 1139 (1956).

140. S. A. Arora, R. B. Bates, R. A. Grady,and N. E. Delfel, J. Amer. Chem. Soc., 97, 5752 (1975).

141. L. Farnell, W. G. Richards, and C. R. Ganellin, J. Theor. Biol., 43, 389 (1974).

142. G. H. Loew and D. S. Berkowitz, J. Med. Chem., 18, 656 (1975).

143. K. U. Ingold and D. R. Taylor, Can. J. Chem., 39, 481 (1961).

144. I. Brown, G. Eglinton, and M. Martin-Smith, Spectrochim. Acta, 19, 463 (1963).

145. G. Rossmy, W. Lüttke, and R. Mecke, J. Chem. Phys., 21, 1606 (1953).

146. H. H. Jaffé, J. Amer. Chem. Soc., 79, 2373 (1957).

147. P. A. Kollman and L. C. Allen, Chem. Rev., 72, 283 (1972).

148. A. S. N. Murthy, S. N. Bhat, and C. N. R. Rao, J. Chem. Soc. A, 1251 (1970).

149. P. Schuster, Chem. Phys. Letters, 3, 433 (1969).

150. A. S. N. Murthy, R. E. Davis and C. N. R. Rao, Theor. Chim. Acta, 13, 81 (1969).

151. A. D. Isaacson and K. Morokuma, J. Amer. Chem. Soc., 97, 4453 (1975).

152. M. S. Gordon and R. D. Koob, J. Amer. Chem. Soc., 95, 5863 (1973).

153. G. Karlström, H. Wennerström, B. Jönsson, S. Forsén, J. Almlöf, and B. Roos, J. Amer. Chem. Soc., 97, 4188 (1975).

154. A. Johansson, P. A. Kollman, and S. Rothenberg, Chem. Phys. Letters, 18, 276 (1973).

155. L. Pauling, J. Amer. Chem. Soc., 58, 94 (1936).

156. O. R. Wulf, U. Liddel, and S. B. Hendricks, J. Amer. Chem. Soc., 58, 2287 (1936).

157. G. L. Carlson and W. G. Fateley, J. Chem. Phys., 77, 1157 (1973).

158. A. W. Baker, J. Amer. Chem. Soc., 80, 3598 (1958).

159. A. W. Baker and A. T. Shulgin, Spectrochim. Acta, 19, 1611 (1963).

160. A. W. Baker and A. T. Shulgin, Can. J. Chem., 43, 650 (1965).

161. A. W. Baker and A. T. Shulgin, Nature, 206, 712 (1965).

162. A. W. Baker and A. T. Shulgin, Spectrochim. Acta, 22, 95 (1966).

163. H. Bourassa-Bataille, P. Sauvageau, and C. Sandorfy, Can. J. Chem., 41, 2240 (1963).

164. K. U. Ingold and D. R. Taylor, Can. J. Chem., 39, 471 (1961).

165. I. Brown, G. Eglinton, and M. Martin-Smith, Spectrochim. Acta, 18, 1593 (1962).

166. A. W. Baker and W. W. Kaeding, J. Amer. Chem. Soc., 81, 5904 (1959).

167. G. L. Carlson, W. G. Fateley, A. S. Manocha, and F. F. Bentley, J. Phys. Chem., 76, 1553 (1972).

168. D. Doddrell, E. Wenkert, and P. V. Demarco, J. Mol. Spectrosc., 32, 162 (1969).

169. T. Schaefer, J. Phys. Chem., 79, 1888 (1975).

170. E. A. Allan and L. W. Reeves, J. Phys. Chem., 66, 613 (1962).

171. E. A. Allan and L. W. Reeves, J. Phys. Chem., 67, 591 (1963).

172. R. Bennett, A. Burger, and C. L. Gemmill, J. Med. Pharm. Chem., 2, 493 (1960).

173. E. A. Robinson, H. D. Schreiber, and J. N. Spencer, Spectrochim. Acta, 28A, 397 (1972).

174. W. G. Fateley, G. L. Carlson, and F. F. Bentley, J. Phys. Chem., 79, 199 (1975).

175. J. Del Bene and J. A. Pople [J. Chem. Phys., 52, 4858 (1970); 55, 2296 (1971)] find that $(HF)_2$ and $(H_2O)_2$ have geometries "very near linear". References 176 and 177 find a similar result for $(HCl)_2$ with small deviations from linearity (10°) making a 0.1-0.2 kcal/mole difference in energy.

176. P. A. Kollman and L. C. Allen, J. Amer. Chem. Soc., 93, 4991 (1971).

177. P. A. Kollman, A. Johansson, and S. Rothenberg, Chem. Phys. Letters, 24, 199 (1974).

178. L. Pauling, "The Nature of the Chemical Bond", Cornell University Press, Ithaca, N. Y., 1960.

179. A. Bondi, J. Phys. Chem., 68, 441 (1964).

180. H_2O geometry: θ_{HOH} = 104.52° and R(H-O) = 0.9572 Å; from reference 139.

181. M. L. Josien, Pure Appl. Chem., 4, 33 (1962).

182. R. West, D. L. Powell, L. S. Whatley, M. K. T. Lee, and P. von R. Schleyer, J. Amer. Chem. Soc., 84, 3221 (1962).

183. D. A. K. Jones and J. G. Watkinson, J. Chem. Soc., 2366 (1964).

184. J. E. Del Bene [J. Chem. Phys., 57, 1899 (1972)] gives θ = 57° for ab initio calculations with an STO-3G basis set for the H_2O/H_2O dimer, the H_2O monomer geometry having first been optimized. Hence, our use of the H_2O experimental geometry is not directly comparable with the use of an ab initio/STO-3G optimized H_2O monomer geometry. In addition, our C_6H_5XH/H_2O dimer system is not necesssarily directly comparable with the H_2X/H_2O dimer; various ab initio H_2O/H_2O dimer studies have shown the dimer energy changes very little with θ variation (see reference 147).

Hence, any energy change due to our not using optimized θ values for the C_6H_5XH/H_2O dimers should be negligibly small.

185. N. L. Allinger, J. J. Maul, and M. J. Hickey, J. Org. Chem., 36, 2747 (1971).

186. E. V. Konovalov, Yu. P. Egorov, R. V. Belinskaya, V. N. Boiko,and L. M. Yagupol'skii, Zh. Prikl. Spectrosk., 14, 484 (1971); Chem. Abstr., 75, 42673a (1971).

187. M. Oki and H. Iwamura, Bull. Chem. Soc. Jap., 33, 717 (1960).

188. F. C. Marler, III, and H. P. Hopkins, Jr., J. Phys. Chem., 74, 4164 (1970).

189. N. Mori, Y. Asano, T. Irie, and Y. Tsuzuki, Bull. Chem. Soc. Jap., 42, 482 (1969).

190. But pK_a values may not necessarily correspond to weak hydrogen bond strengths.

191. J. G. David and H. E. Hallam, Spectrochim. Acta, 21, 841 (1965).

192. M-L. Josien, C. Castinel, and P. Saumagne, Bull. Soc. Chim. France, 648 (1957).

193. H_2S geometry: θ_{HSH} = 92.2° and R(H-S) = 1.335 Å; from reference 131.

194. P. Kollman, J. McKelvey, A. Johansson, and S. Rothenberg, J. Amer. Chem. Soc., 97, 955 (1975).

195. P. A. Kollman and L. C. Allen, J. Amer. Chem. Soc., 92, 753 (1970).

196. J. R. Hoyland and L. B. Kier, Theor. Chim. Acta, 15, 1 (1969).

197. R. M. Badger and S. H. Bauer, J. Chem. Phys., 5, 839 (1937).

198. Interestingly, for the phenol/cyclohexyl halide dimers[182] and for the phenol/n-pentyl halide dimers[183] $\Delta\nu_{OH}$ is of the order F < Cl < Br < I while ΔH of intermolecular hydrogen bond

formation is of the order F > Cl > Br > I. This suggests that the lack of correlation between $\Delta\nu_{OH}$ and intramolecular hydrogen bond strength in the ortho-halophenols might in part be due to intrinsic properties of the halogens as proton acceptors and not solely due to factors uniquely present in the ortho-halophenols.

199. L. Random, W. J. Hehre, J. A. Pople, G. L. Carlson, and W. G. Fateley, J. Chem. Soc., Chem. Commun., 308 (1972).

200. W. G. Fateley, F. A. Miller, and R. E. Witkowski, Technical Documentary Report on AFML-TR-66-408, Jan., 1967; see reference 167.

201. R. C. Bingham, M. J. S. Dewar, and D. H. Lo, J. Amer. Chem. Soc., 97, 1285 (1975); and following articles.

202. A minimum energy CNDO/2 search lead to a R(C-O) value in phenoxide of 1.33 Å.

203. D. G. Lister and J. K. Tyler, J. Mol. Struct., 23, 253 (1974).

204. J. H. Oppenheimer, H. L. Schwartz, W. Dillman, and M. I. Surks, Biochem. Biophys. Res. Commun., 55, 544 (1973).

205. Crystallographic structures for NH_4^+ [131] and L-T_3 [11] were used as sources for the following bond lengths and angles R(N-H) = 1.03 Å; R(C_8-N) = 1.50 Å; R(C_8-C_{10}) = 1.51 Å; R(C_{10}-O) = 1.24 Å (avg.); $\theta_{OC_{10}O}$ = 125°; $\theta_{OC_{10}C_8}$ = 117.5° (avg.). $\theta_{C_4O_{12}C_{13}}$ was taken as 117°, the CNDO/2 minimum energy value for θ_{COC} of diphenyl ether.[37] $\phi_4 = \phi_{C_3C_4O_{12}C_{13}}$ was taken as 90° to approximate the diphenyl ether conformation of the conformationally locked thyroid hormone analogs.

206. J. K. Fawcett, N. Camerman, and A. Camerman, J. Amer. Chem. Soc., 98, 587 (1976).

207. A. Camerman and N. Camerman, Acta Cryst., B30, 1832 (1974).

208. V. Cody, W. L. Duax, and H. Hauptman, Int. J. Peptide Protein Res., 5, 297 (1973).

209. T. Fujita, J. Iwasa, and C. Hansch, J. Amer. Chem. Soc., 86, 5175 (1964).

210. H. H. Jaffé, Chem. Rev., 53, 191 (1953).

211. E. Kutter and C. Hansch, J. Med. Chem., 12, 647 (1969).

212. R. W. Taft, Jr., in "Steric Effects in Organic Chemistry", M. S. Newman, Ed., Wiley, N. Y., 1956, p. 556.

213. S. H. Unger and C. Hansch, J. Med. Chem., 16, 745 (1973).

214. R. A. Cuestas, A. Lindall, and R. R. Engel, New England J. Med., 295, 297 (1976).

215. Drawing reproduced courtesy of E. C. Jorgensen.

216. C. Hansch, S. H. Unger, and A. B. Forsythe, J. Med. Chem., 16, 1217 (1973).

217. E. G. Tomich, E. A. Wollett, and M. A. Pratt, J. Endocrinology, 20, 65 (1960).

218. J. Roche, R. Michel and W. Wolf, Bull. soc. chim. France, 462 (1957).

219. E. C. Jorgensen, and K. Tsutsui, Endocrinology, 68, 171 (1961).

220. E. C. Jorgensen and R. A. Wiley, J. Med. Pharm. Chem., 5, 1307 (1962).

221. E. C. Jorgensen and J. Wright, J. Med. Chem., 13, 745 (1970).

222. B. Blank and F. R. Pfeiffer, J. Med. Chem., 10, 653 (1967).

223. E. C. Jorgensen and J. Wright, J. Med. Chem., 13, 367 (1970).

224. T. Matsuura, T. Nagamachi, K. Matsuo, and A. Nishinaga, J. Med. Chem., 11, 899 (1968).

225. E. C. Jorgensen and R. A. Wiley, J. Med. Chem., 6, 459 (1963).

226. E. C. Jorgensen, R. O. Muhlhauser, and R. A. Wiley, J. Med. Chem., 12, 689 (1969).

227. A. Dibbo, J. C. P. Sly, L. Stephenson, T. Walker, W. K. Warburton and K. D. E. Whiting, J. Chem. Soc., 2890 (1961).

228. G. S. Boyd, unpublished results; see reference 227.

229. R. E. Taylor, Jr., T. Tu, S. B. Barker, and E. C. Jorgensen, Endocrinology, 80, 1143 (1967).

230. M. V. Mussett and R. Pitt-Rivers, Lancet, 268, 1212 (1954).

231. C. L. Gemmill, Amer. J. Physiol., 186, 1 (1956).

232. E. C. Jorgensen, E. Frieden, and P. Block, Jr., Proc. West. Pharmacol. Soc., 17, 271 (1974).

233. E. C. Jorgensen, unpublished data.

234. R. W. Doskotch and H. A. Lardy, J. Amer. Chem. Soc., 80, 6230 (1958).

235. E. C. Jorgensen and J. A. W. Reid, Endocrinology, 76, 312 (1965).

236. E. P. Reineke and C. W. Turner, Endocrinology, 36, 200 (1945).

237. E. C. Jorgensen and P. E. Berteau, J. Med. Chem., 14, 1199 (1971).

238. H. Lardy, Endocrinology, 57, 566 (1955).

239. E. C. Jorgensen and R. Cavestri, J. Pharm. Sci., 52, 481 (1963).

240. E. C. Jorgensen and P. Slade, J. Med. Chem., 14, 1023 (1971).

241. E. C. Jorgensen and J. A. W. Reid, J. Med. Chem., 7, 701 (1964).

242. E. Frieden and R. J. Winzler, J. Biol. Chem., 176, 155 (1948).

243. R. J. Winzler and E. Frieden, Federation Proc., 7, 200 (1948).

244. R. Pitt-Rivers, Lancet, 265, 234 (1953).

245. B. Blank, F. R. Pfeiffer, and C. M. Greenberg, J. Med. Chem., 9, 832 (1966).

236. T. Matsuura and J. H. Cahnmann, J. Amer. Chem. Soc., 82, 2055 (1960).

247. C. M. Greenberg, L. F. Mansor, C. A. Bocher, H. L. Saunders, and J. F. Kerwin, Endocrinology, 70, 365 (1962).

248. A. Höfer and H. J. Cahnmann, J. Med. Chem., 7, 326 (1964).

249. E. Frieden and R. J. Winzler, J. Amer. Chem. Soc., 70, 3511 (1948).

250. S. Wawzonek and W. G. Gaffield, J. Med. Chem., 6, 442 (1963).

251. B. Blank, E. G. Rice, F. R. Pfeiffer, and C. M. Greenberg, J. Med. Chem., 9, 10 (1966).

252. R. A. Pages and A. Burger, J. Med. Chem., 10, 435 (1967).

253. K. Tomita and H. A. Lardy, J. Biol. Chem., 235, 3292 (1960).

254. R. O. Muhlhauser and E. C. Jorgensen, J. Pharm. Sci., 57, 151 (1968).

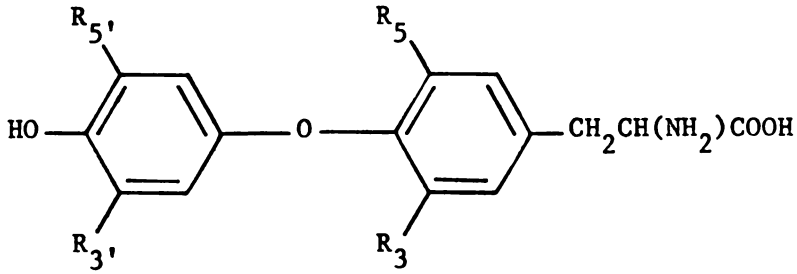
255. C. A. Plamondon, J. G. Wiswell, and S. P. Asper, Jr., Bull. John Hopkins Hosp., 102, 107 (1958).

APPENDIX I: RECALCULATED AND STANDARDIZED RAT ANTIGOITER BIOASSAY ACTIVITIES OF THYROID HORMONE ANALOGS

See text for discussion and explanation of the recalculation and standardization.

Table I-1. Rat Antigoiter Bioassay Activities.

Thyronines: 3, 5, 3', and 5' Substitutions.



No.	DL	R_3	R_5	$R_{3'}$	$R_{5'}$	Relative Molar Activity				
						%	DL-T_3	Ref.	% L-T_3	Ref.
1	DL	H	H	H	H	0		80		
2	DL	H	H	F	H	0		80		
3	L	H	H	I	H				<0.006	217
	DL	H	H	I	H	0		218	0	103,219
4	L	H	H	I	I				<0.008	217
	DL	H	H	I	I	0		218	0	219
5	DL	CH_3	H	H	H				0	220
6	DL	CH_3	H	I	H				0.12	220
7	DL	CH_3	CH_3	H	H				0	220

Table I-1. (Continued)

No.	DL	R_3	R_5	$R_{3'}$	$R_{5'}$	Relative Molar Activity			
						% DL-T_3	Ref.	% L-T_3	Ref.
8	L	CH_3	CH_3	CH_3	H			0.54	9
9	L	CH_3	CH_3	CH_3	CH_3			0.36	9
10	L	CH_3	CH_3	iPr	H			3.60	9,36,95
11	L	CH_3	CH_3	nPr	H			2.36	36,95
12	L	CH_3	CH_3	sBu(±)	H			2.91	36,95
13	L	CH_3	CH_3	I	H			0.90	9
	DL	CH_3	CH_3	I	H			0.65	220,221
14	DL	CH_3	I	H	H			0.054	220
15	DL	CH_3	I	I	H			3.62	220
16	DL	iPr	H	iPr	iPr			0	222
17	DL	iPr	iPr	H	H			0	223
18	DL	iPr	iPr	CH_3	H			0	223
19	DL	iPr	iPr	Br	H			0	223
20	DL	iPr	iPr	I	H			0	223,224
21	DL	sBu(±)	sBu(±)	H	H			0	223
22	DL	sBu(±)	sBu(±)	Br	H			0	223

Table I-1. (Continued)

No.	DL	R_3	R_5	$R_{3'}$	$R_{5'}$	Relative Molar Activity			
						% DL-T_3	Ref.	% L-T_3	Ref.
23	DL	sBu(±)	sBu(±)	I	H			0	223
24	DL	CO_2H	CO_2H	H	H			0	225
25	DL	NH_2	NH_2	H	H			0	225
26	DL	NH_2	NH_2	CH_3	H			0	225
27	L	NO_2	NO_2	H	H			0	101
28	L	NO_2	NO_2	I	I			0	101
29	L	SEt	SEt	H	H			0	226
30	L	SEt	SEt	iPr	H			0	226
31	L	SPh	SPh	H	H			0	226
32	L	SPh	SPh	iPr	H			0	226
33	DL	Cl	Cl	Cl	H	0.091	81,106		
34	DL	Cl	Cl	Cl	OH			0[a]	227,228
35	DL	Cl	Cl	Cl	Cl	>0.014	81,106	<0.20	97
36	L	Cl	I	Cl	I			2.27	102
37	L	Br	Br	iPr	H			30.0	229
38	DL	Br	Br	Br	H	4.63	81,106, 230		

Table I-1. (Continued

No.	DL	R_3	R_5	$R_{3'}$	$R_{5'}$	Relative Molar Activity % DL-T_3	Ref.	% L-T_3	Ref.
39	DL	Br	Br	Br	Br	0.065[b]	81,230		
40	DL	Br	Br	I	H	16.87	81,230		
41	DL	Br	Br	I	I	1.97	81,230		
42	L	Br	I	Br	I			2.83	102
43	DL	Br	I	I	H			41.75	231
44	L	I	H	H	H			<0.006	217
	DL	I	H	H	H			<0.19	97
45	L	I	H	I	H			0.056	97,217
	DL	I	H	I	H			<0.25	97
46	L	I	H	I	I			<0.03	217
	DL	I	H	I	I			<0.125	96,97
47	L	I	I	H	H			0.81	97,101,217
	DL	I	I	H	H			0.90[c]	23
48	L	I	I	CH_3	H			14.47	23,232
	DL	I	I	CH_3	H			8.28	98

Table I-1. (Continued)

No.	DL	R_3	R_5	$R_{3'}$	$R_{5'}$	Relative Molar Activity			
						% DL-T_3	Ref.	% L-T_3	Ref.
49	L	I	I	CH_3	CH_3			9.04	232
50	L	I	I	Et	H			93.5	98
	DL	I	I	Et	H			40.8	98,99
51	D	I	I	iPr	H			7.49	98
	L	I	I	iPr	H			142.1	36,95,98,99
	DL	I	I	iPr	H			47.0	98
52	L	I	I	iPr	I			55.36[d]	99
53	L	I	I	nPr	H			39.5	36,95
54	L	I	I	iBu	H			7.74	99
55	L	I	I	sBu(±)	H			79.9	36,95
56	L	I	I	tBu	H			21.70	54,233
57	DL	I	I	Ph	H			2.03	98
58	L	I	I	NO_2	H			≤0.32	36,95
59	L	I	I	OH	H			0.27	234
60	DL	I	I	F	H	1.12	80,81		

Table I-1. (Continued)

No.	DL	R_3	R_5	$R_{3'}$	$R_{5'}$	Relative Molar Activity % DL-T_3	Ref.	% L-T_3	Ref.
61	DL	I	I	F	F	0.43	80,81		
62	DL	I	I	F	I	$\geq$6.03	80		
63	L	I	I	Cl	H			4.88	235
64	L	I	I	Cl	Cl			3.80	235
65	DL	I	I	Br	H	23.78	81,106, 230		
66	DL	I	I	Br	Br	1.58[b]	81,106		
67	D	I	I	I	H			7.50	84,108
	L	I	I	I	H			100	e
	DL	I	I	I	H	100	e	59	108
68	D	I	I	I	I			3.07[f]	83
	L	I	I	I	I			18.1	36,95,97,98,99, 101,105,108
	DL	I	I	I	I			∿9.04	236

Table I-1. (Continued)

---

[a]"little if any thyroactivity".

[b]Reliability questionable.

[c]Only one dose level.

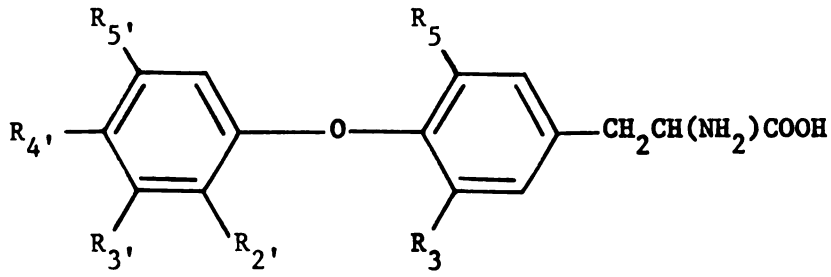
[d]Unsure of analog purity because synthesis was never reported.

[e]Reference compound.

[f]Analog administered orally.

Table I-2. Rat Antigoiter Bioassay Activities.

Thyronines: Sterically Fixed (2' Substituted; Including 4' Substitutions).



No.	DL	R_3=R_5	R_4'	R_2'	R_3'	R_5'	Relative Molar Activity % L-T_3	Ref.
1	DL	NO_2	H	CH_3	H	CH_3	0	21
2	DL	I	H	CH_3	H	H	0	21,23
3	DL	I	H	CH_3	H	CH_3	0	21,23
4	DL	I	H	CH_3	CH_3	H	0.054	21,23
5	DL	I	H	iPr	H	H	0	23
6	DL	I	H	iPr	H	CH_3	0	23
7	L	I	H	OH	H	H	0	237
8[a]	L	I	H	OH	I	I	0.054	237

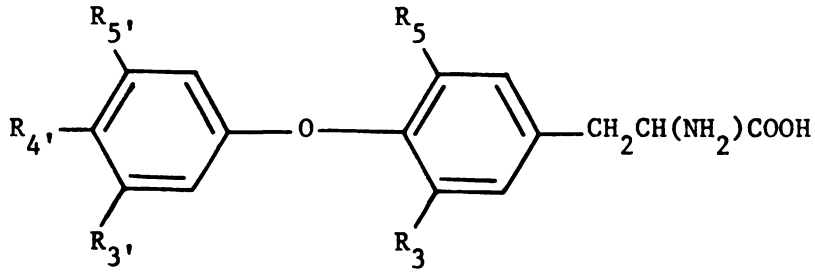
Table I-2. (Continued)

No.	DL	R_3=R_5	$R_{4'}$	$R_{2'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity % L-T_3	Ref.
9	DL	I	CH_3	CH_3	H	H	0	23
10	L	I	CH_3	OH	H	H	0	237
11	DL	I	OH	CH_3	H	H	<0.18	21,23
12	DL	I	OH	CH_3	H	CH_3	<0.18	21,23
13	DL	I	OH	CH_3	H	I	0.36	23
14	L	I	OH	CH_3	CH_3	H	10.85	23
	DL	I	OH	CH_3	CH_3	H	9.04	21,23
17	DL	I	OH	iPr	H	CH_3	0.36	21,23
18	L	I	OH	OH	H	H	0	237
19	DL	I	Cl	CH_3	H	CH_3	0	21,23
20	DL	I	Cl	CH_3	CH_3	H	0	23
21[b]	L	I	I	I	H	OH	0	237

[a]"ortho-Thyroxine".

[b]"meta-Thyroxine".

Table I-3. Rat Antigoiter Bioassay Activities.
Thyronines: 4' Substitutions.



No.	DL	R_3=R_5	$R_{4'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity % L-T_3	Ref.
1	DL	iPr	NH_2	H	H	0	223
2	DL	sBu	NH_2	H	H	0	223
3	L	SEt	OCH_3	iPr	H	0	226
4	DL	I	H	H	H	0.72	23
5	DL	I	H	CH_3	H	2.71	51
6	DL	I	H	CH_3	CH_3	0.054	23
7	DL	I	H	CF_3	H	13.56	51
8	L	I	H	OH	H	0.18	237

Table I-3. (Continued)

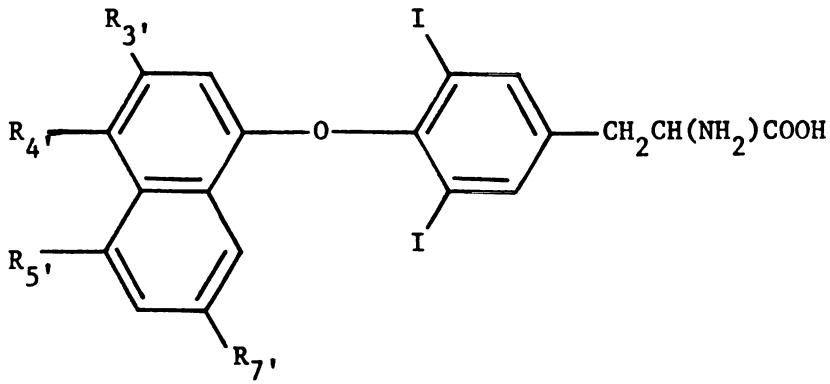
No.	DL	R_3=R_5	$R_{4'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity % L-T_3	Ref.
9	L	I	H	F	H	1.39	36,95
10	L	I	H	Cl	H	7.78	36,95
11	L	I	H	Br	H	18.0	36,95
12	DL	I	H	I	H	>27.12	51
13	DL	I	CH_3	H	H	0	51
14	L	I	CH_3	OH	H	0	237
15	L	I	NH_2	H	H	<0.036	82
16	L	I	NH_2	CH_3	H	<0.036	82
17	L	I	NH_2	CH_3	CH_3	0.13	82
18	L	I	NH_2	I	H	>0.27	82
19	L	I	OCH_3	iPr	H	19	36,95
20	L	I	OCH_3	sBu(±)	H	<21	36,95
21	L	I	OCH_3	tBu	H	≤2.35	54
22	L	I	OCH_3	OCH_3	H	0	234
23	L	I	OCH_3	I	H	11.25	55
24	DL	I	OCH_3	I	I	<0.03	55,103,238

Table I-3.(Continued)

No.	DL	R_3=R_5	R_4'	R_3'	R_5'	Relative Molar Activity % L-T_3	Ref.
25	DL	I	DL-$CH_2CH(NH_2)COOH$	H	H	0	239

Table I-4. Rat Antigoiter Bioassay Activities.

1'-Napthyl Ethers of 3,5-I_2-Tyrosine.



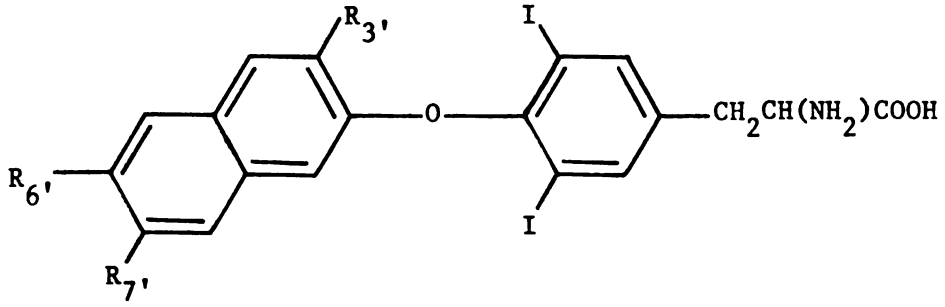
No.	DL	R	$R_{4'}$	$R_{3'}$	$R_{5'}$	$R_{7'}$	Relative Molar Activity % L-T_3	Ref.
1	DL	H	H	H	H	H	0.57	22,23
2	L	H	H	H	H	OH	<0.036	240
3	DL	H	H	H	OH	H	<0.036	240
4	L	H	CH_3	H	H	H	<0.036	240
5	L	H	NH_2	H	H	H	0	240

Table I-4. (Continued)

No.	DL	R	$R_{4'}$	$R_{3'}$	$R_{5'}$	$R_{7'}$	Relative Molar Activity % L-T_3	Ref.
6	L	H	OH	H	H	H	18.08	23,240
	DL	H	OH	H	H	H	$\geq$18.08	23
7	DL	H	OH	Br	H	H	5.24	23
8	DL	H	OCH_3	H	H	H	0.89	22
9	DL	$COCH_3$	OCH_3	H	H	H	0.90	23

Table I-5. Rat Antigoiter Bioassay Activities.

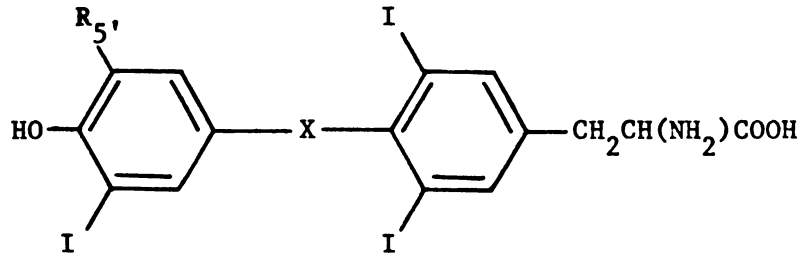
2'-Naphthyl Ethers of 3,5-I_2-Tyrosine.



No.	DL	$R_{3'}$	$R_{6'}$	$R_{7'}$	Relative Molar Activity % L-T_3	Ref.
1	DL	H	H	H	0.036	23
2	L	OH	H	H	<0.036	240
3	L	H	OH	H	0.36	240
4	L	H	H	OH	0	240

Table I-6. Rat Antigoiter Bioassay Activities.

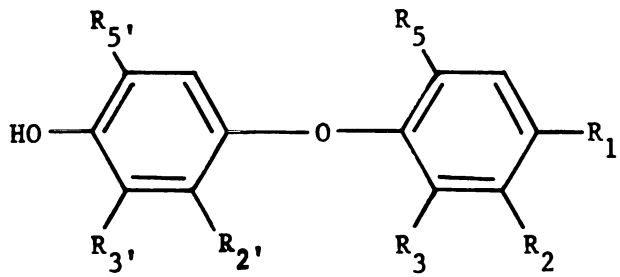
Thyronines: Bridge Variations.



				Relative Molar Activity			
No.	DL	X	$R_{5'}$	% DL-T_3	Ref.	% L-T_3	Ref.
1	L	--	I			0	233
2	DL	S	H	23.82	81,106		
3	DL	S	I	0.21	81		
4	DL	CH_2	H			54.25	9

Table I-7. Rat Antigoiter Bioassay Activities.

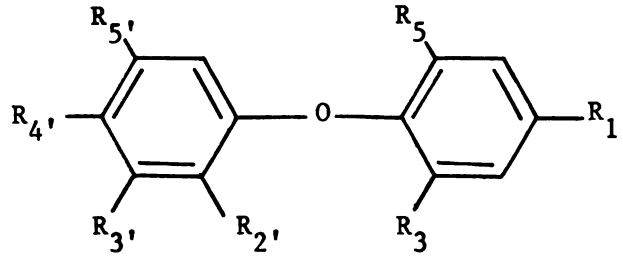
Thyronines: Alanine Side Chain Position Isomers.



									Relative Molar Activity	
No.	DL	R_1	R_2	R_3	R_5	R_2'	R_3'	R_5'	% L-T_3	Ref.
1	DL	H	H	CH_3	CH_3	$CH_2CH(NH_2)COOH$	CH_3	CH_3	0	220,221
2	DL	I	H	$CH_2CH(NH_2)COOH$	I	H	I	H	<0.27	241
3	DL	I	$CH_2CH(NH_2)COOH$	H	I	H	I	H	<0.27	241

Table I-8. Rat Antigoiter Bioassay Activities.

Benzoic, Acetic, Propionic, and Butyric Acid Side Chain Analogs.



								Relative Molar Activity			
No.	R_1	$R_{4'}$	R_3	R_5	$R_{2'}$	$R_{3'}$	$R_{5'}$	% DL-T_3	Ref.	% L-T_3	Ref.
1	COOH	OH	H	H	H	H	H			0	104
2	COOH	OH	I	I	H	I	H			0.081	103,104
3	COOH	OH	I	I	H	I	I	0.027	242,243	0.024	97,101,103,104
4	COOH	OCH_3	NO_2	NO_2	H	H	H			<0.19	97
5	CH_2COOH	OH	I	H	H	H	H			<0.21	97
6	CH_2COOH	OH	I	H	H	I	H			<0.29	97
7	CH_2COOH	OH	I	H	H	I	I			<0.36	97
8	CH_2COOH	OH	I	I	H	H	H			<0.29	97

Table I-8. (Continued)

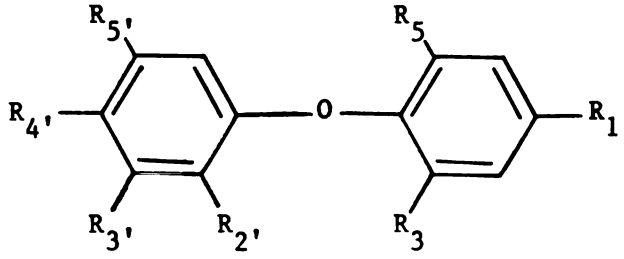
No.	R_1	$R_{4'}$	R_3	R_5	$R_{2'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity			
								% DL-T_3	Ref.	% L-T_3	Ref.
9	CH_2COOH	OH	I	I	H	iPr	H			13.23	98
10	CH_2COOH	OH	I	I	H	I	H	19.1	102	6.43	55,96,97,98,244
11	CH_2COOH	OH	I	I	H	I	I	23.0	102	9.10	96,97,244
12	CH_2COOH	OH	I	I	CH_3	CH_3	H			0.81	245
13	CH_2COOH	OCH_3	NH_2	NH_2	H	H	H			<0.17	97
14	CH_2COOH	OCH_3	NO_2	NO_2	H	H	H			<0.20	97
15	CH_2COOH	OCH_3	I	I	H	H	H			<0.29	97
16	CH_2COOH	OCH_3	I	I	H	I	H			4.58	55,97,98
17	CH_2CH_2COOH	OH	H	H	H	tBu	tBu			<0.05	246
18	CH_2CH_2COOH	OH	NH_2	H	H	H	H			<0.13	97
19	CH_2CH_2COOH	OH	NO_2	NO_2	H	I	I			0	101
20	CH_2CH_2COOH	OH	Br	Br	H	tBu	tBu			<0.05	246
21	CH_2CH_2COOH	OH	I	H	H	H	H			<0.17	97
22	CH_2CH_2COOH	OH	I	H	H	I	H			<0.25	97

Table I-8. (Continued)

No.	R_1	$R_{4'}$	R_3	R_5	$R_{2'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity			
								% DL-T_3	Ref.	% L-T_3	Ref.
23	CH_2CH_2COOH	OH	I	H	H	I	I			<0.31	97
24	CH_2CH_2COOH	OH	I	I	H	H	H			<0.25	97
25	CH_2CH_2COOH	OH	I	I	H	CH_3	CH_3			0.69	97
26	CH_2CH_2COOH	OH	I	I	H	I	H	15.20[a]	100,102	3.58	96,97,103,247
27	CH_2CH_2COOH	OH	I	I	H	I	I	13.35[a]	100,102	2.66	96,97,101,103
28	CH_2CH_2COOH	OCH_3	NH_2	H	H	H	H			<0.14	97
29	CH_2CH_2COOH	OCH_3	I	H	H	H	H			<0.19	97
30	CH_2CH_2COOH	OCH_3	I	I	H	I	H			4.99	247
31	$CH_2CH_2CH_2COOH$	OH	I	I	H	H	H			<0.25	97
32	$CH_2CH_2CH_2COOH$	OH	I	I	H	I	H	2.8	102	0.73	97
33	$CH_2CH_2CH_2COOH$	OH	I	I	H	I	I	3.3	102	0.87	97

[a]Reliability questionable.

Table I-9. Rat Antigoiter Bioassay Activities.
Other Side Chain Variations.



No.	R_1	$R_{4'}$	R_3	R_5	$R_{2'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity % DL-T_3	Ref.	% L-T_3	Ref.
1	$COOCH_3$	OCH_3	NH_2	NH_2	H	H	H			<0.17	97
2	$COOCH_3$	OCH_3	NH_2	NH_2	H	NH_2	H			<0.17	97
3	$COOCH_3$	OCH_3	NO_2	NO_2	H	H	H			<0.20	97
4	$COOCH_3$	OCH_3	I	I	H	H	H			<0.29	97
6	CH_2COOEt	OCH_3	NO_2	NO_2	H	NO_2	H			<0.24	97
7	$CH_2COOCH_2CH_2N(Et)_2 \cdot HCl$	OH	I	I	H	I	H			6.52	98
8	$CH_2COOCH_2CH_2N(Et)_2 \cdot HCl$	OH	I	I	CH_3	CH_3	H			1.01	245

Table I-9. (Continued)

No.	R_1	$R_{4'}$	R_3	R_5	$R_{2'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity			
								% DL-T_3	Ref.	% L-T_3	Ref.
9	$CH_2COOCH_2CH_2N(Et)_2 \cdot HCl$	OCH_3	I	I	CH_3	CH_3	H			0	245
10	CH_2CH_2COOEt	OCH_3	NH_2	H	H	H	H			<0.15	97
11	DL-$CH_2CHCOOH$ (OH)	OH	I	I	H	H	H	<0.40	102		
12	DL-$CH_2CHCOOH$ (OH)	OH	I	I	H	I	H	5.01	102		
13	DL-$CH_2CHCOOH$ (OH)	OH	I	I	H	I	I	3.82	102		
14	$CH_2CH(COOH)_2$	OH	I	I	H	I	H			0.08	248
15	$CH_2CH(COOH)_2$	OH	I	I	H	I	I			0.06	248
16	$CH_2CH(COOEt)_2$	OCH_3	NH_2	NH_2	H	H	H			<0.19	97
17	$CH_2CH(COOEt)_2$	OCH_3	I	I	H	H	H			<0.30	97
18	$CH_2COCOOH$	OH	I	I	H	I	H	0.106	102		
19	$CH_2COCOOH$	OH	I	I	H	I	I	0.031	102		
20	CH=CHCOOH	OH	NO_2	NO_2	H	I	I			0	101

Table I-9. (Continued)

No.	R_1	$R_{4'}$	R_3	R_5	$R_{2'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity			
								% DL-T_3	Ref.	% L-T_3	Ref.
21	CH=CHCOOH	OH	I	I	H	I	I	11.67	102	3.89	97,101
22	CH=CHCOOH	OCH_3	NO_2	H	H	H	H			<0.18	97
23	CH=CHCOOEt	OCH_3	NO_2	H	H	H	H			<0.20	97
24	CH=CHCOOEt	OCH_3	NO_2	NO_2	H	H	H			<0.27	97
25	NH_2	OH	I	I	H	H	H			<0.1	103
26	NH_2	OH	I	I	H	I	H			0.1	103
27	NH_2	OH	I	I	H	I	I			<0.1	103
28	NH_2	OCH_3	I	I	H	H	H			0.1	103
29	$CH_2CH_2NH_2$	OH	I	I	H	H	H			0.1	103
30	$CH_2CH_2NH_2$	OH	I	I	H	I	H			1.1	103
31	$CH_2CH_2NH_2$	OH	I	I	H	I	I			0.1	103
32	DL-$CH(NH_2)COOH$	OH	I	I	H	I	I	0.031[a]	242,243, 249		
33	DL-$CH_2CH(NHCOCH_3)COOH$	OH	I	I	H	I	I	4.29	242		

Table I-9. (Continued)

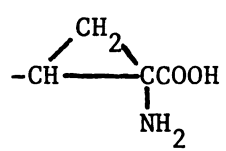
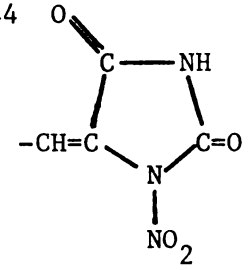
No.	R_1	$R_{4'}$	R_3	R_5	$R_{2'}$	$R_{3'}$	$R_{5'}$	Relative Molar Activity: % DL-T_3	Ref.	% L-T_3	Ref.
34	L-$CH_2CH(NHCOCH_3)COOH$	OCH_3	I	I	H	OCH_3	H			0	234
35	L-$CH_2CH(NHCOCH_3)COOEt$	OCH_3	I	I	H	OCH_3	H			0	234
36	L-$CH_2CH(NHCOCH_3)COOEt$	OCH_3	I	I	H	I	H			12.14	96
37	D-$CH_2CH(NMe_2)COOH$	OH	I	I	H	H	H			0[b]	227,228
38	DL/DL-$CH(HO)CH(NH_2)COOH$	OH	I	I	H	I	I			0[c]	250
39	DL-$CH_2C(NH_2)(CH_3)COOH$	OH	I	I	H	I	I	2.04	251		
40		OH	I	I	H	Br	Br			0[d]	252

Table I-9. (Continued)

No.	R_1	R_4'	R_3	R_5	R_2'	R_3'	R_5'	Relative Molar Activity: % DL-T_3	Ref.	% L-T_3	Ref.
41	CH_2CH_2OH	OH	I	I	H	H	H	0	253		
42	CH_2CH_2OH	OH	I	I	H	I	H	5.42	253		
43	$CONHCH_2COOH$	OH	I	I	H	I	I	<0.12	80		
44		OCH_3	NO_2	NO_2	H	H	H			<0.25	97

[a]Mouse antigoiter assay.

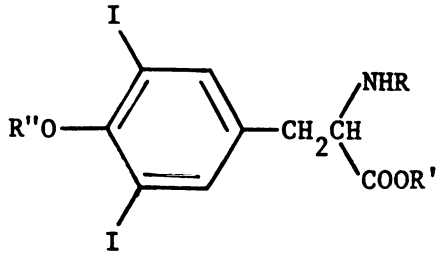
[b]"little or no thyroactivity".

[c]Possibly very low activity.

[d]"weak biological activity".

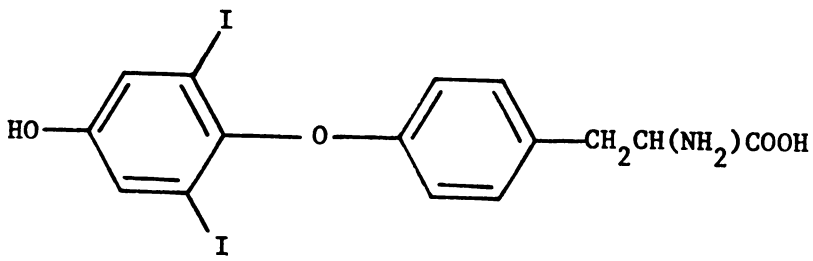
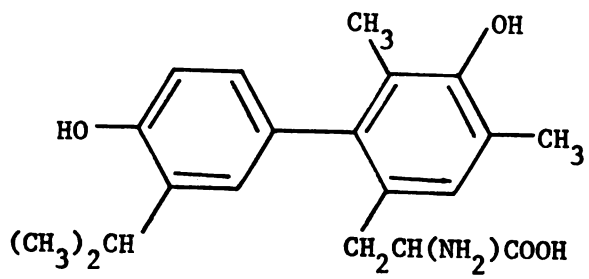
Table I-10. Rat Antigoiter Bioassay Activities.

Aliphatic and Alicyclic Ethers of 3,5-I_2-Tyrosine.



No.	DL	R	R'	R"	Relative Molar Activity % L-T_3	Ref.
1	DL	H	H	$CH_2CH_2CH_2CH_3$	0	23
2	DL	H	H	$CH_2CH_2CH_2CH_2COOH$	0	23
3	DL	H	H	$CH_2CH_2CH_2N(CH_3)_2$	0	23
4	L	H	H	4-pyridyl	0	254
5	DL	$COCH_3$	H	$CH_2CH{=}CH_2$	0	23
6	DL	$COCH_3$	H	cyclohexyl	0	23
7	DL	$COCH_3$	H	3-cyclohexenyl	0	23
8	DL	$COCH_3$	H	6-OH-3-cyclohexenyl	0	23
9	L	$COCH_3$	C_2H_5	4-pyridyl	0	254

Table I-11. Rat Antigoiter Bioassay Activities.
Odds and Ends.

No.	DL	Analog	Relative Molar Activity			
			% DL-T_3	Ref.	% L-T_3	Ref.
1	DL		0	80	0	255
2	DL				0	9,221

APPENDIX II: COMPUTER PROGRAMS

This appendix contains listings of the computer programs written and used in the quantitative structure-activity studies. PROGRAM QSAR47 was used to perform the actual QSAR calculations and requires three on-line magnetic disk direct access files:

1. FTABLES: 510 formatted (7F9.2) records each with a length of 63 bytes.
   : Contains the critical F statistic values needed by PROGRAM QSAR47.
   : Created by PROGRAM FILLFTBL.
2. WCFILE: 346 formatted (60I1, 612) records each with a length of 72 bytes.
   : Contains the weight card values needed by PROGRAM QSAR47 when the "ALL" option is used.
   : Created by PROGRAM FILLWCF from cards punched by PROGRAM PUNCHWC2.
3. SFILE: A temporary (scratch) disk space.
   : Used by PROGRAM QSAR47 to save a summary of the title cards and regression equations for listing at the end of the regressions.
   : Fixed-format, blocked records each with a length of 120 bytes.

PROGRAM QSAR47

```
C*************************************************************************
C
C                              PROGRAM QSAR47
C
C*************************************************************************
C
C
C     QSAR = QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS
C
C     47 = THAT MAGICAL NUMBER FROM WHICH ALL GOOD THINGS ARE DERIVED
C
C
C*************************************************************************
C
C
C PROGRAM QSAR47 IS A MULTIPLE LEAST SQUARES LINEAR REGRESSION ANALYSIS
C PROGRAM ESPECIALLY DESIGNED FOR (BUT NOT LIMITED TO APPLICATIONS WITH)
C QSAR STUDIES IN MEDICINAL CHEMISTRY
C
C QSAR47 WAS PROGRAMMED AND DEBUGGED ON AN IBM 370 MODEL 145
C COMPUTER BY:      STEVE DIETRICH
C                   DEPARTMENT OF PHARMACEUTICAL CHEMISTRY
C                   SCHOOL OF PHARMACY
C                   UNIVERSITY OF CALIFORNIA
C                   SAN FRANCISCO, CALIFORNIA  94143
C                   APRIL, 1976
C
C THE PROGRAM WAS DEBUGGED IN WATFIV AND STORED AS AN ON-LINE DISK.
C PRIVATE PROGRAM IN FORTRAN H.
C
C
C*************************************************************************
C
C
C GENERAL REFERENCES:  C. DANIEL AND F. S. WOOD
C                      'FITTING EQUATIONS TO DATA'
C                      WILEY INTERSCIENCE, N. Y.  (1971)
C
C                    : G. A. PALL
C                      'INTRODUCTION TO SCIENTIFIC COMPUTING'
C                      APPLETON-CENTURY-CROFTS, N. Y. (1971)
C
C                    : I. M. WONNACOTT AND R. J. WONNACOTT
C                      'INTRODUCTORY STATISTICS'
C                      SECOND EDITION
C                      JOHN WILEY & SONS, INC., NY (1972)
C
C
C*************************************************************************
C
C
C THE FEATURES OF THIS PROGRAM ARE:
C        1. ALL CALCULATIONS ARE DONE IN DOUBLE PRECISION (REAL*8).
C        2. ANY NUMBER OF TITLE CARDS MAY BE USED.
C        3. VALUES FROM 1 TO 15 INDEPENDENT VARIABLES MAY BE SUPPLIED BY
C           THE USER.
```

```
C        4. THE PROGRAM WILL (AT THE USER'S OPTION) GENERATE AND ALLOW
C           REGRESSION WITH ANY NUMBER OF 1 TO 10 ADDITIONAL
C           INDEPENDENT VARIABLES, WHICH ARE GENERATED FROM THE
C           SUPPLIED INDEPENDENT VARIABLES: SEE BELOW.
C        5. DATA FROM 2 TO 99 (BUT SEE BELOW) DATA POINTS MAY BE USED.
C        6. AT THE USER'S OPTION, SPECIFIC DATA POINTS WILL NOT BE
C           INCLUDED IN CALCULATING THE REGRESSION EQUATIONS.
C           DEPENDENT VARIABLE VALUES DO NOT NEED TO BE ENTERED FOR
C           THESE DATA POINTS.
C        7. ANY SINGLE REGRESSION MAY USE FROM 1 TO 15 OF THE
C           INDEPENDENT VARIABLES.
C        8. THE REGRESSIONS THEMSELVES ARE SPECIFIED BY WEIGHT CARDS IN
C           ONE OR BOTH OF 2 METHODS: SINGLE REGRESSIONS WITH SPECIFIC
C                                     INDEPENDENT VARIABLES MAY BE
C                                     INDICATED BY A SINGLE WEIGHT CARD.
C                                   : WITH THE 'ALL' OPTION, ALL
C                                     POSSIBLE COMBINATIONS OF ANY 1 TO
C                                     10 INDEPENDENT VARIABLES
C                                     SPECIFIED ON THE WEIGHT CARD WILL
C                                     BE RUN.
C        9. PROGRAM OUTPUT CONSISTS OF: LISTINGS OF ALL DEPENDENT
C                                       AND INDEPENDENT VARIABLE VALUES,
C                                       MEANS, STANDARD DEVIATIONS, AND
C                                       '**' BY THE NUMBERS OF THE DATA
C                                       POINTS NOT USED IN THE
C                                       REGRESSION CALCULATIONS.
C                                     : FOR EACH REGRESSION:
C                                       :THE REGRESSION EQUATION WITH
C                                        95% CONFIDENCE INTERVALS FOR
C                                        THE INTERCEPT AND INDEPENDENT
C                                        VARIABLE COEFFICIENTS.
C                                       :ESTIMATED VALUES (AND
C                                        DEVIATIONS FROM THE ACTUAL
C                                        VALUES) OF THE DEPENDENT
C                                        VARIABLE, BASED ON THE
C                                        REGRESSION EQUATION.
C                                       :ALL RELEVANT STATISTICAL DATA
C                                        FOR THE REGRESSION IN AN
C                                        ANOVA (ANALYSIS OF VARIANCE)
C                                        TABLE.
C                                       :'IDEAL' VALUES FOR EACH
C                                        INDEPENDENT VARIABLE X FOR
C                                        EQUATIONS CONTAINING THE
C                                        INDEPENDENT VARIABLE X ONLY AS
C                                        X AND X**2 TERMS.
C                                       :ALL THE CALCULATED F TEST
C                                        VALUES AND THE INTERPOLATED
C                                        PERCENTAGE CONFIDENCE LEVELS
C                                        FOR COMPARISON OF EACH
C                                        EQUATION WITH ALL OTHER
C                                        EQUATIONS WHICH ARE SUBSETS OF
C                                        IT.  SEE SUBROUTINE FTESTS.
C                                       :THE INDEPENDENT VARIABLE CROSS
C                                        CORRELATION AND SQUARED CROSS
C                                        CORRELATION MATRICES.
C       10. MAXPTS, THE MAXIMUM NUMBER OF DATA POINTS ALLOWED, MAY BE
C           CHANGED BY ALTERING: ARRAY SIZES IN EACH SUBROUTINE AND IN
C                                THE MAINPROGRAM.
C                              : VALUE OF MAXPTS IN THE FIRST EXECUTABLE
C                                STATEMENT OF THE MAINPROGRAM.
C       11. A PERMANENT ON-LINE DISK DIRECT ACCESS FILE FTABLES
C           (CREATED BY THE PROGRAM FILLFTBL) CONTAINS THE CRITICAL F
```

```
C           STATISTIC VALUES NEEDED BY THIS PROGRAM.
C       12. A PERMANENT ON-LINE DISK DIRECT ACCESS FILE WCFILE
C           (CREATED BY THE PROGRAM FILLWCF FROM CARDS PUNCHED BY THE
C           PROGRAM PURCHWC2) CONTAINS THE WEIGHT CARD VALUES NEEDED BY
C           THIS PROGRAM WHEN THE 'ALL' OPTION IS USED.
C       13. THE TEMPORARY (SCRATCH) DISK SPACE SFILE IS USED BY THE
C           PROGRAM TO SAVE A SUMMARY OF THE TITLE CARDS AND
C           REGRESSION EQUATIONS FOR LISTING AT THE END OF THE
C           REGRESSIONS.
C       14. THE MATRIX INVERSION INVOLVED WITH EACH REGRESSION EMPLOYS
C           LARGEST PIVOT ELEMENT SEARCHES IN ORDER TO MINIMIZE
C           ROUND-OFF ERRORS DUE TO SMALL DIVISORS.
C       15. AT THE USER'S OPTION, THE PROGRAM WILL INTERNALLY CALCULATE
C           FOR EACH DATA POINT:
C           DEPENDENT VARIABLE VALUE
C              = LOG(VALUE OF ONE OF THE INDEPENDENT VARIABLES)
C                              OR
C              = LOG(1./VALUE OF ONE OF THE INDEPENDENT VARIABLES).
C       16. AT THE USER'S OPTION, THE PROGRAM WILL INTERNALLY CALCULATE
C           FOR EACH DATA POINT:
C           VALUE OF ONE OF THE INDEPENDENT VARIABLES
C                     = LOG(VALUE OF ONE OF THE OTHER INDEPENDENT
C                       VARIABLES).
C           THIS CAN BE DONE FOR FROM 1 TO 3 OF THE INDEPENDENT
C           VARIABLES.
C       17. FOR ANY SINGLE SET OF DATA POINTS, THE MAXIMUM NUMBER OF
C           ALLOWED REGRESSIONS (1099 = 76 MORE THAN THE NUMBER OF
C           REGRESSIONS RUN WITH THE 'ALL' OPTION AND 10 INDEPENDENT
C           VARIABLES) MAY BE CHANGED BY ALTERING ARRAY SIZES IN EACH
C           SUBROUTINE AND IN THE MAIN PROGRAM.
C
C
C*****************************************************************************
C
C
C THE DATA IS ENTERED IN THE FOLLOWING STEPS:
C       1. ANY NUMBER OF TITLE CARDS.
C          FORMAT (A76).
C       2. ENTER THE VALUE 1 ON THE NEXT CARD.
C          FORMAT (79X, I1).
C       3. ON THE NEXT CARD ENTER: DVAR = THE DEPENDENT VARIABLE NAME
C                                : LOGY = 0 = DEFAULT
C                                       = USER WILL ENTER DEPENDENT
C                                         VARIABLE VALUES
C                                  > 0 AND < 16 = OPTION
C                                               = PROGRAM WILL
C                                                 INTERNALLY
C                                                 GENERATE FOR
C                                                 EACH DATA POINT
C                                                 THE DEPENDENT
C                                                 VARIABLE VALUE =
C                                                 LOG(VALUE OF THE
C                                                 (LOGY)TH
C                                                 INDEPENDENT
C                                                 VARIABLE)
C                                  > -16 AND < 0 = OPTION
C                                                = PROGRAM WILL
C                                                  INTERNALLY
C                                                  GENERATE FOR
C                                                  EACH DATA POINT
C                                                  THE DEPENDENT
C                                                  VARIABLE VALUE
```

```
C                                                  = -LOG(VALUE OF
C                                                  THE (-LOGY)TH
C                                                  INDEPENDENT
C                                                  VARIABLE).
C                                 : NI = THE NUMBER OF USER SUPPLIED
C                                        INDEPENDENT VARIABLES.
C                                        AN INDEPENDENT VARIABLE
C                                        INTERNALLY GENERATED ACCORDING
C                                        TO THE LOGX OPTION (SEE BELOW)
C                                        IS CONSIDERED TO BE A USER
C                                        SUPPLIED INDEPENDENT VARIABLE.
C                                 : OPTION(K), K=1,10
C                                     = 0 = DEFAULT
C                                     = 1 = OPTION
C                                   WHERE: THE PROGRAM WILL GENERATE AND
C                                          ALLOW REGRESSION WITH THE
C                                          (K+15)TH INDEPENDENT VARIABLE
C                                   IF X(I,J) = THE VALUE OF THE JTH
C                                   INDEPENDENT VARIABLE FOR THE ITH
C                                   DATA POINT:
C                                     X(I,16) = X(I,1)**2
C                                     X(I,17) = X(I,2)**2
C                                     X(I,18) = X(I,3)**2
C                                     X(I,19) = X(I,4)**2
C                                     X(I,20) = X(I,1)**3
C                                     X(I,21) = X(I,2)**3
C                                     X(I,22) = X(I,1)*X(I,2)
C                                     X(I,23) = X(I,1)*X(I,3)
C                                     X(I,24) = X(I,1)*X(I,2)**2
C                                     X(I,25) = X(I,1)*X(I,3)**2
C                                 : LOGX(K,1), K=1,2
C                                   LOGX(K,2), K=1,2
C                                   LOGX(K,3), K=1,2
C                                      LOGX(1,L) = 0
C                                                = DEFAULT
C                                      LOGX(1,L) > 0 AND < 16
C                                                = OPTION
C                                                = PROGRAM WILL
C                                                  INTERNALLY
C                                                  GENERATE FOR EACH
C                                                  DATA POINT.
C                                                  VALUE OF THE
C                                                  LOGX(1,L)TH
C                                                  INDEPENDENT
C                                                  VARIABLE =
C                                                  LOG(VALUE OF THE
C                                                  LOGX(2,L)TH
C                                                  INDEPENDENT
C                                                  VARIABLE).
C        FORMAT (A8, I3, 1X, I2, 1X, 10I1, 5X, 6I2)
C     4. NEXT ENTER INDVAR(I), I=1,IVTOT
C        WHERE: INDVAR(I) = THE NAME OF THE ITH INDEPENDENT
C                           VARIABLE
C             : IVTOT = NI + NUMBER OF PROGRAM-GENERATED INDEPENDENT
C                       VARIABLES
C        ENTER THE NAMES OF THE SUPPLIED AND PROGRAM-GENERATED
C        INDEPENDENT VARAIABLES IN ORDER: EG. FOR I=1,2,3,16,18,
C        ON ONE CARD WITH FORMAT (5(A8, 2X)) THE FIVE INDEPENDENT
C        VARIABLE NAMES WOULD BE ENTERED.
C        FORMAT (8(A8, 2X)).
C     5. FOR EACH DATA POINT ENTER: ICALC = 0 IF THE DATA POINT WILL
C                                           BE USED IN THE
```

```
C                                          REGRESSION CALCULATIONS
C                                        = 1 IF THE DATA POINT WILL
C                                          NOT BE USED IN THE
C                                          REGRESSION CALCULATIONS
C                                          BUT A VALUE FOR THE
C                                          DEPENDENT VARIABLE WILL
C                                          BE ENTERED.
C                                        = 2 IF THE DATA POINT WILL
C                                          NOT BE USED IN THE
C                                          REGRESSION CALCULATIONS
C                                          AND A VALUE FOR THE
C                                          DEPENDENT VARIABLE
C                                          WILL NOT BE ENTERED.
C                                  : XNAME = THE NAME OF THE DATA
C                                            POINT.
C                                  : Y = THE DEPENDENT VARIABLE VALUE
C                                        FOR THE DATA POINT.
C                                      = IF LOGY OPTION IS USED,
C                                        VALUES OF 0. SHOULD BE
C                                        ENTERED FOR EACH DATA POINT.
C                                  : X(I,J), J=1,NI = THE INDEPENDENT
C                                                     VARIABLE VALUES
C                                                     OF THE ITH DATA
C                                                     POINT FOR THE NI
C                                                     USER SUPPLIED
C                                                     INDEPENDENT
C                                                     VARIABLES
C                                                   = VALUES OF 0.
C                                                     SHOULD BE
C                                                     ENTERED FOR
C                                                     THOSE
C                                                     INDEPENDENT
C                                                     VARIABLES WHICH
C                                                     WILL BE
C                                                     GENERATED BY THE
C                                                     LOGX OPTION.
C        FORMAT (I1, 1X, A16, 2X, (6F10.5)).
C     6. ENTER THE VALUE 1 ON THE NEXT CARD.
C        FORMAT (1X, I1).
C        ALSO ENTER 0, 1, OR 2 BLANK CARDS SO THAT THE TOTAL NUMBER
C        OF CARDS ENTERED IN THIS STEP = NUMBER OF CARDS ENTERED FOR
C        EACH DATA POINT IN STEP 5.
C     7. ENTER THE WEIGHT CARDS, USING ONE AND/OR BOTH OF THE TWO
C        METHODS BELOW:
C              A. FOR A SINGLE REGRESSION, PUT A 1 IN EACH COLUMN
C                 CORRESPONDING TO THE NUMBERS OF THE INDEPENDENT
C                 VARIABLES DESIRED IN THE REGRESSION EQUATION.
C                 FORMAT (2511).
C              B. THE 'ALL' OPTION.  PUT A 1 IN EACH COLUMN
C                 CORRESPONDING TO THE NUMBERS OF THE INDEPENDENT
C                 VARIABLES OF WHICH ALL POSSIBLE COMBINATIONS WILL
C                 BE RUN AS REGRESSIONS.  ALSO PUT A 1 IN COLUMN 79.
C                 FORMAT (2511, 53X, I1).
C     8. ENTER THE VALUE 1 ON THE NEXT CARD.
C        FORMAT (79X, I1).
C     9. IF ANOTHER REGRESSION ANALYSIS IS TO BE DONE, RETURN TO
C        STEP 1.  OTHERWISE, THAT IS ALL.
C
C
C*****************************************************************
C
C
```

```
C*************************************************************************
C                                                                        *
C MAINPROGRAM: CONTROLS THE CALLING OF THE VARIOUS SUBROUTINES           *
C                                                                        *
C*************************************************************************
C
C
C*************************************************************************
C                                                                        *
C VARIABLE DEFINITIONS:                                                  *
C              DVAR = NAME OF THE DEPENDENT VARIABLE                     *
C              I = CONVENTIONAL USE AS DO LOOP VARIABLES, SUBSCRIPTS,    *
C                  COUNTERS, ETC.                                        *
C              IBEGIN = COUNTING VARIABLE WHICH IS USED TO CONTROL       *
C                       NUMBER AND WHICH OF OTHER VARIABLES ARE          *
C                       PRINTED, CALCULATED WITH, ETC.                   *
C              ICALC(I) = 1 = PROGRAM WILL NOT USE ITH DATA POINT IN     *
C                             CALCULATING THE REGRESSION EQUATIONS       *
C                       = 0 = PROGRAM WILL USE THE ITH DATA POINT IN     *
C                             CALCULATING THE REGRESSION EQUATIONS       *
C              IEND = COUNTING VARIABLE WHICH IS USED TO CONTROL         *
C                     NUMBER AND WHICH OF OTHER VARIABLES ARE PRINTED,   *
C                     CALCULATED WITH, ETC.                              *
C              IF   = DATA SET REFERENCE NUMBER FOR PERMANENT ON-LINE    *
C                     MAGNETIC DISK DIRECT ACCESS FILE OF CRITICAL F     *
C                     STATISTIC VALUE TABLES                             *
C              IMDONE = PROGRAM TERMINATION FLAG                         *
C              INDVAR(J) = NAME OF THE JTH INDEPENDENT VARIABLE          *
C              IR = DATA SET REFERENCE NUMBER FOR CARD READER            *
C              IV(J) = EXTERNAL NUMBER OF THE JTH INTERNAL               *
C                      INDEPENDENT VARIABLE: J=1,IVTOT: SEE              *
C                      SUBROUTINE READ1 FOR MORE DETAILS                 *
C              IVTOT = TOTAL NUMBER OF INDEPENDENT VARIABLES             *
C                    = NI + NOPT                                         *
C              IW = DATA SET REFERENCE NUMBER FOR PRINTER                *
C              J = CONVENTIONAL USE AS DO LOOP VARIABLES, SUBSCRIPTS,    *
C                  COUNTERS, ETC.                                        *
C              K = CONVENTIONAL USE AS DO LOOP VARIABLES, SUBSCRIPTS,    *
C                  COUNTERS, ETC.                                        *
C              L = CONVENTIONAL USE AS DO LOOP VARIABLES, SUBSCRIPTS,    *
C                  COUNTERS, ETC.                                        *
C              LINES = NUMBER OF LINES THAT HAVE BEEN PRINTED ON A       *
C                      PAGE: USED FOR PRINTING CONTROL                   *
C              M = CONVENTIONAL USE AS DO LOOP VARIABLES, SUBSCRIPTS,    *
C                  COUNTERS, ETC.                                        *
C              MATXX(J,K) = SUMMING OVER ALL N DATA POINTS               *
C                           ((X(I,J) - XAVG(J))*(X(I,K) - XAVG(K)))      *
C              MAXPTS = MAXIMUM NUMBER OF DATA POINTS ALLOWED            *
C              N = NUMBER OF DATA POINTS WHICH WILL BE USED IN           *
C                  CALCULATING THE REGRESSION EQUATIONS                  *
C              NI = NUMBER OF SUPPLIED INDEPENDENT VARIABLES             *
C              NIV(J) = THE NUMBER OF INDEPENDENT VARIABLES INDICATED    *
C                       FOR REGRESSION BY THE JTH REGRESSION WEIGHT      *
C                       CARD                                             *
C              NLINES = MAXIMUM NUMBER (EITHER SYSTEM DEFAULT OR USER    *
C                       SUPPLIED VALUE) OF LINES PRINTED PER PAGE        *
C              NOPT = NUMBER OF INDEPENDENT VARIABLES WHICH WILL BE      *
C                     INTERNALLY GENERATED BY THE PROGRAM FROM THE       *
C                     SUPPLIED INDEPENDENT VARIABLES                     *
C              NOREG = NUMBER OF DATA POINTS WHICH WILL NOT BE USED      *
C                      IN CALCULATING THE REGRESSION EQUATIONS           *
C              NTOT = TOTAL NUMBER OF DATA POINTS                        *
```

```
C                 = N + NOREG                                           *
C              NWC = TOTAL NUMBER OF REGRESSION WEIGHT CARDS READ       *
C                    (FROM CARD READER AND FROM WCFILE), EACH OF WHICH *
C                    WILL CAUSE A SINGLE REGRESSION TO BE PERFORMED    *
C              RR(WCN) = THE SQUARED (MULTIPLE) REGRESSION COEFFICIENT *
C                        FOR THE (WCN)TH REGRESSION                    *
C                      = R**2                                          *
C                      = VREG/TOTVAR                                   *
C              SDX(J) = STANDARD DEVIATION OF THE JTH INDEPENDENT       *
C                       VARIABLE                                       *
C              SDY = STANDARD DEVIATION OF THE DEPENDENT VARIABLE       *
C              SFILE = DATA SET REFERENCE NUMBER FOR TEMPORARY          *
C                      (SCRATCH) ON-LINE MAGNETIC DISK SPACE FOR       *
C                      STORAGE OF SUMMARY OF TITLE CARDS AND           *
C                      REGRESSION EQUATIONS WITH ASSOCIATED            *
C                      STATISTICS (TO BE LISTED AT END OF RUN)         *
C              SUMXY(J) = SUMMING OVER ALL N DATA POINTS               *
C                         ((X(I,J) - XAVG(J))*(Y(I) - YAVG))           *
C              TOTVAR = TOTAL VARIATION OF THE DEPENDENT VARIABLE       *
C              WC(J,K) = FOR THE JTH REGRESSION, THE K = NIV(J)         *
C                        INTERNAL NUMBERS OF THE K INDEPENDENT          *
C                        VARIABLES SPECIFIED FOR THE REGRESSION         *
C              WCFILE = DATA SET REFERENCE NUMBER FOR PERMANENT         *
C                       ON-LINE MAGNETIC DISK DIRECT ACCESS FILE OF     *
C                       REGRESSION WEIGHT CARDS                        *
C              X(I,J) = VALUE OF JTH INDEPENDENT VARIABLE FOR THE ITH   *
C                       DATA POINT                                     *
C              XAVG(J) = MEAN OF THE JTH INDEPENDENT VARIABLE           *
C              XNAME(I,K) = NAME OF THE ITH DATA POINT: K=1,2           *
C              Y(I) = VALUE OF DEPENDENT VARIABLE FOR ITH DATA POINT    *
C              YAVG = MEAN OF THE DEPENDENT VARIABLE                    *
C                                                                       *
C***********************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15)
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25)
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
C
C INITIALIZE MAXPTS AND NLINES
      MAXPTS = 99
      NLINES = 56
C
C INITIALIZE DATA SET REFERENCE NUMBER VARIABLES
      IR = 5
      IW = 6
      IF = 12
      WCFILE = 13
      SFILE = 14
C
C CALL SUBROUTINE READ1 TO READ IN DATA FOR THE REGRESSIONS
   10 CALL READ1
C
C IF IMDONE = 1 UPON RETURN FROM SUBROUTINE READ1, NO MORE REGRESSION
```

```
C ANALYSES ARE TO BE DONE AND EXIT FROM PROGRAM OCCURS.
      IF (IMDONE .EQ. 1) GO TO 99999
C
C IF THE PROGRAM IS TO INTERNALLY GENERATE ANY OF THE INDEPENDENT
C VARIABLES (FROM THE SUPPLIED INDEPENDENT VARIABLES). SUBROUTINE CALC1
C IS CALLED TO CALCULATE THEM.
      IF (NOPT .NE. 0) CALL CALC1
C
C CALL SUBROUTINE CALC2 TO CALCULATE: MEAN VALUES AND STANDARD
C                                     DEVIATIONS OF THE DEPENDENT AND
C                                     INDEPENDENT VARIABLES
C                                   : VARIATION OF THE DEPENDENT
C                                     VARIABLE
      CALL CALC2
C
C CALL SUBROUTINE WRITE1 TO PRINT:
C               : FOR EACH DATA POINT: NUMBER, NAME, AND DEPENDENT AND
C                                      INDEPENDENT VARIABLE VALUES,
C                                      USED IN THE REGRESSION
C                                      CALCULATIONS
C               : MEANS AND STANDARD DEVIATIONS OF THE DEPENDENT AND
C                 INDEPENDENT VARIABLES
      CALL WRITE1
C
C CALL SUBROUTINE MAT1 TO CREATE THE MASTER MATRICES MATXX AND SUMXY
      CALL MAT1
C
C CALL SUBROUTINE WTCARD TO READ AND STORE VALUES FROM REGRESSION WEIGHT
C CARDS FROM CARD READER AND/OR, IF SPECIFIED, FROM THE DIRECT ACCESS
C FILE WCFILE.
      CALL WTCARD
C
C CALL SUBROUTINE REGRES  TO PERFORM EACH OF THE REGRESSIONS, AS
C SPECIFIED BY THE REGRESSION WEIGHT CARDS, AND TO PRINT OUT THE
C RESULTS.
      CALL REGRES
C
C CALL SUBROUTINE FTESTS TO CALCULATE AND PRINT ALL OF THE F TEST
C COMPARISONS OF THE REGRESSION EQUATIONS.
      CALL FTESTS
C
C CALL SUBROUTINE SUMUP TO: PRINT (AS STORED ON THE TEMPORARY (SCRATCH)
C                           FILE SFILE) TITLE CARDS AND A SUMMARY OF THE
C                           REGRESSION EQUATIONS AND ASSOCIATED
C                           STATISTICAL DATA.
C                         : CALCULATE AND PRINT INDEPENDENT VARIABLE
C                           CROSS CORRELATION COEFFICIENT AND SQUARED
C                           CROSS CORRELATION COEFFICIENT MATRICES.
      CALL SUMUP
C
C RETURN TO BEGINNING OF MAINPROGRAM TO SEE IF ANOTHER REGRESSION SERIES
C IS TO BE RUN.
      GO TO 10
C
C TRANSFER IS TO STATEMENT 99999 (AFTER RETURN FROM SUBROUTINE READ1)
C IF NO MORE REGRESSION ANALYSES ARE TO BE RUN.
99999 CONTINUE
      STOP
      END
      SUBROUTINE READ1
C
C
```

```
C*************************************************************************
C                                                                        *
C SUBROUTINE READ1: PRINTS HEADING FOR REGRESSION                        *
C                 : READS TITLE CARDS, CHECKING TO SEE IF THERE ARE NO   *
C                   MORE REGRESSION RUNS TO BE RUN                       *
C                 : PRINTS TITLE CARDS                                   *
C                 : WRITES TITLE CARDS ON SCRATCH FILE SFILE FOR         *
C                   SUMMARY AT END OF REGRESSIONS                        *
C                 : READS AND PRINTS DEPENDENT AND INDEPENDENT           *
C                   VARIABLE NAMES                                       *
C                 : READS NUMBER OF INDEPENDENT VARIABLES TO BE          *
C                   SUPPLIED AND WHICH (IF ANY) INDEPENDENT VARIABLES    *
C                   WILL BE INTERNALLY GENERATED BY THE PROGRAM          *
C                 : READS (FOR EACH DATA POINT) DEPENDENT AND            *
C                   INDEPENDENT VARIABLE VALUES AND WHETHER THE DATA     *
C                   POINT IS TO BE USED IN CALCULATING THE REGRESSION    *
C                   EQUATIONS, CHECKING FOR END OF DATA POINT CARDS      *
C                   FLAG                                                 *
C                 : IF SPECIFIED, CALCULATE FOR EACH DATA POINT          *
C                     DEPENDENT VARIABLE VALUE = LOG(VALUE OF ONE OF     *
C                                                THE INDEPENDENT         *
C                                                VARIABLES)              *
C                                          OR = LOG(1./VALUE OF ONE      *
C                                                OF THE INDEPENDENT      *
C                                                VARIABLES)              *
C                       VALUE OF ONE OF THE INDEPENDENT VARIABLES        *
C                            = LOG(VALUE OF ONE OF THE OTHER             *
C                              INDEPENDENT VARIABLES)                    *
C                            = THIS CAN BE DONE FOR FROM 1 TO 3          *
C                              DIFFERENT INDEPENDENT VARIABLES.          *
C                                                                        *
C*************************************************************************
C
C
C*************************************************************************
C                                                                        *
C VARIABLES NOT DEFINED IN MAINPROGRAM:                                  *
C              LAST1 = END OF TITLE CARDS FLAG                           *
C              LAST2 = END OF DATA POINTS FLAG                           *
C              LOGX(K,L) = AN OPTION USED TO CALCULATE THE VALUES OF     *
C                          FROM 1 TO 3 INDEPENDENT VARIABLES AS          *
C                          LOG(VALUES OF 1 TO 3 OTHER INDEPENDENT        *
C                          VARIABLES)                                    *
C                        : K=1,2: L=1,3                                  *
C                        : IF LOGX(1,L) > 0 AND < 16, THEN PROGRAM       *
C                          WILL CALCULATE FOR EACH ITH DATA POINT:       *
C                              X(I,LOGX(1,L)) = LOG(X(I,LOGX(2,L)))      *
C              LOGY = 0 = DEFAULT                                        *
C                   > 0 AND < 16 = OPTION                                *
C                                = PROGRAM WILL GENERATE INTERNALLY      *
C                                  FOR EACH ITH DATA POINT THE VALUE     *
C                                  OF THE DEPENDENT VARIABLE Y(I) AS     *
C                                  Y(I) = LOG(X(I,LOGY))                 *
C                   > -16 AND < 0 = OPTION                               *
C                                 = PROGRAM WILL GENERATE INTERNALLY     *
C                                   FOR EACH ITH DATA POINT THE VALUE    *
C                                   OF THE DEPENDENT VARIABLE Y(I) AS    *
C                                   Y(I) = -LOG(X(I,-LOGY))              *
C                                        = LOG(1./X(I,-LOGY))            *
C              OPTION(K)  = 0 = DEFAULT                                  *
C                         = 1 = OPTION                                   *
C                             = PROGRAM WILL GENERATE INTERNALLY.        *
```

```
C                                LIST. AND ALLOW REGRESSION WITH THE     *
C                                (K+15)TH INDEPENDENT VARIABLE: K=1,10   *
C               TITLES = TEMPORARY STORAGE VARIABLE FOR LITERAL          *
C                        CHARACTERS OF DATA CARDS                        *
C                                                                        *
C************************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15),  INVMAT(15,15),
     X TITLES(19), DLOG10
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25),
     X LOGX(2,3)
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
      EQUIVALENCE (LAST1, LAST2), (XNAME, TITLES)
C
C INITIALIZE IMDONE = PROGRAM TERMINATION FLAG
      IMDONE = 0
C
C PRINT HEADING FOR REGRESSION
      WRITE (IW,10)
   10 FORMAT (1H1, 120(1H*)// 41X, 40HLEAST SQUARES LINEAR REGRESSION AN
     XALYSIS/ 1H0, 120(1H*)//)
C
C READ (CARD READER) AND WRITE (PRINTER AND SCRATCH FILE SFILE) TITLE
C CARDS, CHECKING FOR END OF FILE (TRANSFER IS THEN TO STATEMENT 180)
C AND IF THE PREVIOUS TITLE CARD READ IN WAS THE LAST TITLE CARD (IE, IF
C LAST1 =1)
      REWIND SFILE
      LINES = 7
   20 READ (IR,30,END=180) TITLES, LAST1
   30 FORMAT (19A4, 3X, I1)
      WRITE (IW,40) TITLES
   40 FORMAT (1X,19A4)
      WRITE (SFILE,30) TITLES
      LINES = LINES + 1
      IF (LINES .GT. NLINES) LINES = 1
      IF (LAST1 .EQ. 0) GO TO 20
      IF (LINES .GE. NLINES-1) GO TO 60
      WRITE (IW,50)
   50 FORMAT (1H , 120(1H*))
      LINES = LINES + 1
C
C READ VALUES FOR DVAR, LOGY, NI, OPTION, AND LOGX
   60 READ (IR,70) DVAR, LOGY, NI, OPTION, LOGX
   70 FORMAT (A8, I3, 1X, I2, 1X, 10I1, 5X, 6I2)
C
C OPTION VALUES ARE INSPECTED AND USED TO CALCULATE NOPT AND IVTOT.
C IN ADDITION, AND OF SPECIAL IMPORTANCE FOR COMPREHENSION OF THE REST
C OF THE PROGRAM, THE ARRAY IV(IVTOT) IS GENERATED.
C IV(IVTOT) ALLOWS FURTHER CALCULATIONS INVOLVING THE INDEPENDENT
C VARIABLES TO BE DONE SUCH THAT INTERNALLY WITHIN THIS PROGRAM THE
C PROGRAM-GENERATED INDEPENDENT VARIABLES ARE SHIFTED TO THE UNFILLED
C SUPPLIED INDEPENDENT VARIABLE POSITIONS.
C FOR EXAMPLE, IF NI = 3 AND ONLY INDEPENDENT VARIABLE 16 IS INTERNALLY
```

```
C GENERATED, ALL FURTHER CALCULATIONS WILL OCCUR INTERNALLY WITHIN THE
C PROGRAM AS IF INDEPENDENT VARIABLE 16 WERE REALLY INDEPENDENT VARIABLE
C NUMBER 4.
C THIS IS DONE FOR PROGRAMMING EFFICIENCY BUT IS OF NO IMPORTANCE TO THE
C PROGRAM USER, TO WHOM IT WILL APPEAR THAT THE PROGRAM-GENERATED
C INDEPENDENT VARIABLES STILL HAVE THEIR ORIGINALLY ASSIGNED NUMBERS.
C
      NOPT = 0
      DO 80 J=1,10
      IF (OPTION(J) .EQ. 0) GO TO 80
      NOPT = NOPT + 1
      IV(NI+NOPT) = J + 15
   80 CONTINUE
      IVTOT = NI + NOPT
      DO 90 J=1,NI
      IV(J) = J
   90 CONTINUE
C
C READ NAMES OF SUPPLIED AND PROGRAM-GENERATED INDEPENDENT VARIABLES
      READ (IR,100) (INDVAR(I), I=1,IVTOT)
  100 FORMAT (8(A8,2X))
C
C WRITE DEPENDENT AND INDEPENDENT VARIABLE NAMES AND NUMBERS
      IF (LINES .GT. NLINES-9-2*IVTOT) WRITE (IW,110)
  110 FORMAT (1H1)
      WRITE (IW,120) DVAR, (IV(I), INDVAR(I), I=1,IVTOT)
  120 FORMAT (22H0DEPENDENT VARIABLE = , A8/ 18H0INDEPT VARIABLES:, 3X,
     X12, 1H., 3X, A8/ (21X, I2, 1H., 3X, A8))
C
C READ FOR EACH ITH DATA POINT: XNAME(I), Y(I), (X(I,J), J=1,NI), AND
C                               ICALC(I)
C IF SPECIFIED BY LOGY, FOR EACH ITH DATA POINT CALCULATE:
C      Y(I) = LOG(X(I,LOGY)),  IF LOGY > 0
C             OR
C      Y(I) = -LOG(X(I,-LOGY)), IF LOGY < 0
C IF SPECIFIED BY THE LOGX VALUES, CALCULATE FOR EACH ITH DATA POINT:
C      X(I,LOGX(1,J)) = LOG(X(I,LOGX(2,J))),  J=1,3
C ALSO CHECK WHETHER THE PREVIOUS DATA POINT READ IN WAS THE LAST DATA
C POINT (LAST2 = 1) OR NOT (LAST2 = 0)
C ALSO CALCULATE: N, NOREG, NTOT
      N = 0
      NTOT = 0
      K = MAXPTS + 1
      DO 160 I=1,K
      READ (IR,130) ICALC(I), LAST2, XNAME(I,1), XNAME(I,2), Y(I),
     X (X(I,J), J=1,NI)
  130 FORMAT (2I1, 2A8, 2X, (6F10.5))
      IF (LAST2 .EQ. 1) GO TO 170
      IF (ICALC(I) .EQ. 2) GO TO 140
      IF (LOGY .GT. 0) Y(I) = DLOG10(X(I,LOGY))
      IF (LOGY .LT. 0) Y(I) = -DLOG10(X(I,-LOGY))
  140 DO 150 J=1,3
      IF (LOGX(1,J) .NE. 0) X(I,LOGX(1,J)) = DLOG10(X(I,LOGX(2,J)))
  150 CONTINUE
      NTOT = NTOT + 1
      IF (ICALC(I) .NE. 0) GO TO 160
      N = N + 1
  160 CONTINUE
  170 NOREG = NTOT - N
      GO TO 190
C
C TRANSFER IS TO THIS POINT WHEN AN END OF FILE IS ENCOUNTERED WHILE
```

```
C ATTEMPTING TO READ TITLE CARDS.
C IF A VALUE OF 1 FOR IMDONE IS RETURNED TO THE MAINPROGRAM, NO MORE
C REGRESSION ANALYSES ARE TO BE CONDUCTED AND PROGRAM EXITS.
C IF A VALUE OF 0 FOR IMDONE IS RETURNED TO THE MAINPROGRAM, REGULAR
C EXECUTION OF THE REGRESSION ANALYSIS IS CONTINUED.
  180 IMDONE = 1
C
  190 RETURN
      END
      SUBROUTINE CALC1
C
C
C*****************************************************************************
C                                                                             *
C SUBROUTINE CALC1: CALCULATES THE VALUES OF THOSE INDEPENDENT                *
C                   VARIABLES WHICH WERE SPECIFIED TO BE INTERNALLY           *
C                   GENERATED BY THE PROGRAM FROM THE SUPPLIED                *
C                   INDEPENDENT VARIABLES.                                    *
C                 : THIS SUBROUTINE IS CALLED IF AND ONLY IF NOPT DOES        *
C                   NOT EQUAL 0.                                              *
C                                                                             *
C*****************************************************************************
C
C
C*****************************************************************************
C                                                                             *
C ALL VARIABLES ARE DEFINED IN THE MAINPROGRAM.                               *
C                                                                             *
C IN PARTICULAR: X(I,16) = X(I,1)**2                                          *
C                X(I,17) = X(I,2)**2                                          *
C                X(I,18) = X(I,3)**2                                          *
C                X(I,19) = X(I,4)**2                                          *
C                X(I,20) = X(I,1)**3                                          *
C                X(I,21) = X(I,2)**3                                          *
C                X(I,22) = X(I,1)*X(I,2)                                      *
C                X(I,23) = X(I,1)*X(I,3)                                      *
C                X(I,24) = X(I,1)*X(I,2)**2                                   *
C                X(I,25) = X(I,1)*X(I,3)**2                                   *
C                                                                             *
C*****************************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15)
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25)
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
C
      DO 90 J=1,NOPT
      K = IV(NI+J) - 15
      GO TO (10,10,10,10,30,30,50,50,70,70), K
   10 DO 20 I=1,NTOT
      X(I,NI+J) = X(I,K)**2
   20 CONTINUE
      GO TO 90
   30 DO 40 I=1,NTOT
```

```
      X(I,NI+J) = X(I,K-4)**3
   40 CONTINUE
      GO TO 90
   50 DO 60 I=1,NTOT
      X(I,NI+J) = X(I,1)*X(I,K-5)
   60 CONTINUE
      GO TO 90
   70 DO 80 I=1,NTOT
      X(I,NI+J) = X(I,1)*X(I,K-7)**2
   80 CONTINUE
   90 CONTINUE
      RETURN
      END
      SUBROUTINE CALC2
C
C
C*****************************************************************************
C                                                                             *
C SUBROUTINE CALC2: CALCULATES MEAN, VARIATION, AND STANDARD                  *
C                   DEVIATION OF THE DEPENDENT VARIABLE                       *
C                 : CALCULATES MEANS AND STANDARD DEVIATIONS OF THE           *
C                   SUPPLIED AND PROGRAM-GENERATED (IF ANY)                   *
C                   INDEPENDENT VARIABLES                                     *
C                                                                             *
C*****************************************************************************
C
C
C*****************************************************************************
C                                                                             *
C ALL VARIABLES ARE DEFINED IN THE MAINPROGRAM.                               *
C                                                                             *
C*****************************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15),
     X DSQRT
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25)
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
C
C INITIALIZE VARIABLES
      YAVG = 0.
      TOTVAR = 0.
      DO 10 J=1,IVTOT
      XAVG(J) = 0.
      SDX(J) = 0.
   10 CONTINUE
C
C CALCULATE YAVG, TOTVAR, AND SDY
      DO 20 I=1,NTOT
      IF (ICALC(I) .EQ. 0) YAVG = YAVG + Y(I)
   20 CONTINUE
      YAVG = YAVG/N
      DO 30 I=1,NTOT
      IF (ICALC(I) .EQ. 0) TOTVAR = TOTVAR + (Y(I) - YAVG)**2
```

```
   30 CONTINUE
      SDY = DSQRT(TOTVAR/(N-1))
C
C CALCULATE MEANS AND STANDARD DEVIATIONS OF INDEPENDENT VARIABLES
      DO 60 J=1,IVTOT
      DO 40 I=1,NTOT
      IF (ICALC(I) .EQ. 0) XAVG(J) = XAVG(J) + X(I,J)
   40 CONTINUE
      XAVG(J) = XAVG(J)/N
      DO 50 I=1,NTOT
      IF (ICALC(I) .EQ. 0) SDX(J) = SDX(J) + (X(I,J) - XAVG(J))**2
   50 CONTINUE
      SDX(J) = DSQRT(SDX(J)/(N-1))
   60 CONTINUE
C
      RETURN
      END
      SUBROUTINE WRITE1
C
C
C*************************************************************************
C                                                                        *
C SUBROUTINE WRITE1: PRINTS VALUES FOR: N                                *
C                                     : NOREG                            *
C                                     : NTOT                             *
C                                     : NI                               *
C                                     : NOPT                             *
C                                     : IVTOT                            *
C                  : PRINTS FOR EACH                                     *
C                         DATA POINT: NUMBER                             *
C                                   : NAME                               *
C                                   : DEPENDENT VARIABLE VALUE           *
C                                   : SUPPLIED AND PROGRAM-GENERATED     *
C                                     INDEPENDENT VARIABLE VALUES        *
C                                   : '*' BY DATA POINT NUMBER IF THE    *
C                                     DATA POINT WILL NOT BE USED IN     *
C                                     THE REGRESSION CALCULATIONS        *
C                  : PRINTS MEAN AND STANDARD DEVIATION FOR DEPENDENT    *
C                    AND EACH INDEPENDENT VARIABLE                       *
C                                                                        *
C*************************************************************************
C
C
C*************************************************************************
C                                                                        *
C VARIABLES NOT DEFINED IN MAINPROGRAM:                                  *
C              : ULINE = THE LITERAL VARIABLE '________'                 *
C                                                                        *
C*************************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15),
     X ULINE/8H________/
      INTEGER*2  IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FLEWC(10,6), IV(25)
      COMMON /REAL/  DVAR, INDVAR, XNAME, Y, X, YAVG, TOTVAR,
     X XAVG, SDX, SDY, MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE, NOPT, NLINES, LINES,          NI, OPTION,  IV,
     X IVTOT, ICALC, N, NOREG, NTOT, IBEGIN, IEND, NWC, NIV,
```

```
     X I, J, K, L, M, WC, FILEWC
C
C PRINT VALUES FOR N, NOREG, NTOT, NI, NOPT, AND IVTOT
      IF (LINES .GT. NLINES-12) WRITE (IW,10)
   10 FORMAT (1H1)
      WRITE (IW,20) N, NOREG, NTOT, NI, NOPT, IVTOT
   20 FORMAT (1H0,I3, 46H = NUMBER OF DATA PTS USED IN REGRESSION CALCS/
     X 1H0, I3, 50H = NUMBER OF DATA PTS NOT USED IN REGRESSION CALCS/
     X 1H0, I3, 24H = TOTAL NUMBER DATA PTS/
     X 1H0, I3, 38H = NUMBER OF SUPPLIED INDEPT VARIABLES/
     X 1H0, I3, 43H = NUMBER OF PGM-GENERATED INDEPT VARIABLES/
     X 1H0, I3, 35H = TOTAL NUMBER OF INDEPT VARIABLES)
C
C INITIALIZE IBEGIN AND IEND
      IBEGIN = 1
   30 IEND = IBEGIN + 6
      IF (IVTOT .LT. IEND) IEND = IVTOT
C
C PRINT TABLE CONTAINING: DATA POINT NUMBERS, NAMES, DEPENDENT AND
C                         INDEPENDENT VARIABLE VALUES, AND (IF
C                         INDICATED) '*' BY DATA POINT NUMBER
C                       : DEPENDENT AND INDEPENDENT VARIABLE MEANS AND
C                         STANDARD DEVIATIONS
      DO 160 I=1,NTOT
      IF (I .EQ. 1) GO TO 40
      IF (LINES .LE. NLINES-3) GO TO 90
   40 WRITE (IW,10)
      LINES = 1
C
C PRINT OUT TABLE HEADINGS
      IF (IBEGIN .NE. 1) GO TO 60
      WRITE (IW,50) DVAR
   50 FORMAT (1H+, 27X, A8/ 30X, 4HOBSD/ 1H+, 27X, 8(1H_))
      LINES = LINES + 1
   60 WRITE (IW,70) (INDVAR(J), J=IBEGIN,IEND)
   70 FORMAT (1H+, 6X, 10HDATA POINT, 19X, 7(3X,A8))
      WRITE (IW,80) (ULINE, J=IBEGIN,IEND)
   80 FORMAT (1H+, 24(1H_), 11X, 7(3X, A8))
C
C FILLING TABLE
   90 WRITE (IW,100) I, (XNAME(I,J), J=1,2), (X(I,J), J=IBEGIN,IEND)
  100 FORMAT (2X, I3, 1H., 3X, 2A8, 11X, 7(3X, F8.4))
      IF (IBEGIN .NE. 1) GO TO 120
C
C DEPENDENT VARIABLE VALUE LISTED ONLY IF PROVIDED
      IF (ICALC(I) .LE. 1) WRITE (IW,110) Y(I)
  110 FORMAT (1H+, 27X, F8.4)
C
C PUT A '*' BY THE DATA POINTS WHICH WILL NOT BE USED IN THE REGRESSION
C CALCULATIONS
  120 IF (ICALC(I) .GE. 1) WRITE (IW,130)
  130 FORMAT (2H+*)
      LINES = LINES + 1
      IF (LINES .EQ. NLINES-3) GO TO 140
      IF (I .LT. NTOT) GO TO 160
C
C PAGE IS FILLED AND/OR THIS PART OF THE TABLE IS FINISHED.
C BLOCK OFF BOTTOM OF TABLE.
  140 WRITE (IW,80) (ULINE, J=IBEGIN,IEND)
      IF (IBEGIN .EQ. 1) WRITE (IW,150) ULINE
  150  FORMAT (1H+, 27X, A8)
  160 CONTINUE
```

```
C
C ADD MEAN AND STANDARD DEVIATION VALUES AT BOTTOM OF TABLE
      WRITE (IW,170) (XAVG(J), J=IBEGIN,IEND)
  170 FORMAT (5H MEAN, 31X, 7(3X, F8.4))
      IF (IBEGIN .EQ. 1) WRITE (IW,110) YAVG
      WRITE (IW,180)  (SDX(J), J=IBEGIN,IEND)
  180 FORMAT (3H SD, 33X, 7(3X, F8.4))
      IF (IBEGIN .EQ. 1) WRITE (IW,110) SDY
C
C IF ALL THE INDEPENDENT VARIABLES HAVE BEEN LISTED, THEN EXIT FROM
C SUBROUTINE.  OTHERWISE CONTINUE WITH THE TABLE.
      IF (IEND .EQ. IVTOT) GO TO 190
      IBEGIN = IEND + 1
      GO TO 30
  190 CONTINUE
      RETURN
      END
      SUBROUTINE MAT1
C
C
C*****************************************************************************
C                                                                             *
C SUBROUTINE MAT1: CALCULATES THE MASTER MATRICES SUMXY (A 1 X IVTOT           *
C                  MATRIX) AND MATXX (A IVTOT X IVTOT MATRIX)                 *
C                                                                             *
C*****************************************************************************
C
C
C*****************************************************************************
C                                                                             *
C ALL VARIABLES ARE DEFINED IN THE MAINPROGRAM.                               *
C                                                                             *
C*****************************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15),
     X SUMXY(25)
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25)
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
      EQUIVALENCE (SDX, SUMXY)
C
C INITIALIZE SUMXY
      DO 10 J=1,IVTOT
      SUMXY(J) = 0.
   10 CONTINUE
C
C FILL SUMXY
      DO 30 J=1,IVTOT
      DO 20 I=1,NTOT
      IF (ICALC(I) .EQ. 0) SUMXY(J) = SUMXY(J) + (X(I,J) - XAVG(J))*(Y(I
     X) - YAVG)
   20 CONTINUE
   30 CONTINUE
C
```

```
C INITIALIZE MATXX (ACTUALLY ONLY NEED TO INITIALIZE UPPER RIGHT
C TRIANGULAR HALF OF MATXX)
      DO 50 J=1,IVTOT
      DO 40 K=J,IVTOT
      MATXX(J,K) = 0.
   40 CONTINUE
   50 CONTINUE
C
C FILL MATXX, TAKING ADVANTAGE OF THE FACT THAT MATXX(J,K) = MATXX(K,J)
      DO 80 J=1,IVTOT
      DO 70 K=J,IVTOT
      DO 60 I=1,NTOT
      IF (ICALC(I) .EQ. 1) GO TO 60
      MATXX(J,K) = MATXX(J,K) + (X(I,J) - XAVG(J))*(X(I,K) - XAVG(K))
   60 CONTINUE
      MATXX(K,J) = MATXX(J,K)
   70 CONTINUE
   80 CONTINUE
      RETURN
      END
      SUBROUTINE WTCARD
C
C
C*****************************************************************************
C                                                                             *
C SUBROUTINE  WTCARD: READS THE REGRESSION WEIGHT CARDS WHICH TELL            *
C                     WHAT VARIABLES ARE TO BE INCLUDED IN EACH OF THE        *
C                     REGRESSIONS                                             *
C                   : WHEN THE OPTION 'ALL' IS INVOKED TO CAUSE THE           *
C                     REGRESSION OF ALL POSSIBLE COMBINATIONS OF A            *
C                     SELECTED SET OF 1 TO 10 INDEPENDENT VARIABLES,          *
C                     THE APPROPRIATE REGRESSION WEIGHT CARDS ARE READ        *
C                     FROM THE DIRECT ACCESS FILE WCFILE                      *
C                                                                             *
C*****************************************************************************
C
C
C*****************************************************************************
C                                                                             *
C VARIABLES NOT DEFINED IN THE MAINPROGRAM:                                   *
C               ALL = THE 'ALL' OPTION FLAG                                   *
C                   = 0 = DEFAULT                                             *
C                   = 1 = OPTION                                              *
C                       = THE REGRESSION OF ALL POSSIBLE COMBINATIONS         *
C                         OF THE 1 TO 10 INDEPENDENT VARIABLES                *
C                         SPECIFIED ON THE WEIGHT CARD WILL BE                *
C                         PERFORMED                                           *
C               ALLVAR(J) = THE INTERNAL NUMBER OF THE JTH (J=1,NALL)         *
C                           INDEPENDENT VARIABLE SPECIFIED ALONG WITH         *
C                           THE 'ALL' OPTION                                  *
C               FILEWC(I,J) = TEMPORARY STORAGE OF THE I = 1 TO 10            *
C                             WEIGHT CARD VALUES FOR EACH OF J = 6            *
C                             WEIGHT CARDS OF A SINGLE RECORD OF              *
C                             WCFILE                                          *
C               LAST = END OF REGRESSION WEIGHT CARDS FLAG                    *
C               NALL = THE NUMBER OF INDEPENDENT VARIABLES OF WHICH           *
C                      ALL POSSIBLE REGRESSION COMBINATIONS WILL BE           *
C                      PERFORMED                                              *
C               NIVFWC(6) = TEMPORARY STORAGE OF THE NUMBER OF                *
C                           INDEPENDENT VARIABLES VALUES FOR EACH OF          *
C                           J = 6 WEIGHT CARDS OF A SINGLE RECORD OF          *
C                           WCFILE                                            *
```

```
C               READWC(25) = TEMPORARY STORAGE OF THE 25 WEIGHT CARD     *
C                            VALUES FOR THE POSSIBLE INDEPENDENT         *
C                            VARIABLES OF A REGRESSION, AS READ FROM A   *
C                            SINGLE REGRESSION WEIGHT CARD               *
C               RECORD = THE ASSOCIATED INTEGER VARIABLE FOR THE DIRECT  *
C                        ACCESS FILE WCFILE                              *
C                      = INDICATES THE RELATIVE POSITION OF A RECORD     *
C                        WITHIN WCFILE                                   *
C               WCREC(J) = THE RECORD NUMBER OF THE DIRECT ACCESS FILE   *
C                          WCFILE AT WHICH THE READING OF RECORDS        *
C                          COMMENCES WHEN THE 'ALL' OPTION HAS BEEN      *
C                          INVOKED FOR J = NALL INDEPENDENT VARIABLES    *
C                                                                        *
C************************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15)
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25),
     X READWC(25), ALL, ALLVAR(10), RECORD, NIVREC(6),
     X WCREC(10)/1, 2, 3, 5, 8, 14, 25, 47, 90, 176/
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
      EQUIVALENCE  (OPTION, ALLVAR)
      DEFINE FILE 13 (346,72,E,RECORD)
C
C INITIALIZE NWC
      NWC = 0
C
C READ VALUES FOR READWC(25), ALL, AND LAST FROM REGRESSION WEIGHT CARD
   10  READ (IR,20) READWC, ALL, LAST
   20 FORMAT (25I1, 53X, 2I1)
C
C IF LAST = 1, THE PREVIOUS REGRESSION WEIGHT CARD WAS THE LAST AND THE
C SUBROUTINE RETURNS TO THE MAINPROGRAM
      IF (LAST .EQ. 1) GO TO 110
C
C CHECK TO SEE IF THE 'ALL' OPTION HAS BEEN INDICATED
      IF (ALL .EQ. 1) GO TO 40
C
      NWC = NWC + 1
      NIV(NWC) = 0
C
C STORE THE INTERNAL NUMBERS OF THOSE K OF THE INDEPENDENT VARIABLES
C WHICH WERE INDICATED FOR REGRESSION BY THE REGRESSION WEIGHT CARD IN
C WC(NWC,J), J=1,K
      DO 30 J=1,IVTOT
      IF (READWC(IV(J)) .EQ. 0) GO TO 30
      NIV(NWC) = NIV(NWC) + 1
      WC(NWC,NIV(NWC)) = J
   30 CONTINUE
C
C RETURN TO STATEMENT 10 AND READ ANOTHER REGRESSION WEIGHT CARD
      GO TO 10
C
C TRANSFER IS TO THIS POINT WHEN THE 'ALL' OPTION IS INDICATED (IE, WHEN
```

```
C ALL = 1).
C
C DETERMINE THE NUMBER = NALL AND WHICH (ALLVAR(J), J=1,NALL) OF THE
C INDEPENDENT VARIABLES OF WHICH ALL POSSIBLE REGRESSION COMBINATIONS
C WILL BE PERFORMED
   40 NALL = 0
      DO 50 J=1,IVTOT
      IF (READWC(IV(J)) .EQ. 0) GO TO 50
      NALL = NALL + 1
      ALLVAR(NALL) = J
   50 CONTINUE
C
C THE APPROPRIATE WEIGHT CARD VALUES FOR ALL POSSIBLE COMBINATIONS OF
C THE NALL INDEPENDENT VARIABLES, AS WELL AS NIVFWC, ARE READ FROM THE
C DIRECT ACCESS FILE WCFILE.
C READING STARTS AT RECORD NUMBER WCREC(NALL) OF WCFILE
C AS EACH RECORD IS READ, NIVFWC AND THE INTERNAL NUMBERS OF THE
C INDEPENDENT VARIABLES INDICATED FOR REGRESSION ARE APPROPRIATELY
C STORED. FOR THE 6 WEIGHT CARDS CONTAINED IN THE RECORD, IN
C NIV(NWC) AND WC(NWC,K), K=1,NALL
      RECORD = WCREC(NALL)
      DO 90 I=1,200
      READ (WCFILE'RECORD,60) FILEWC, NIVFWC
   60 FORMAT (60I1, 6I2)
      IF (NIVFWC(6) .NE. 0) FIND (WCFILE'RECORD)
      DO 80 L=1,6
C
C WHEN NIVFWC(K+1) = 0, THE KTH WEIGHT CARD OF THIS RECORD OF WCFILE
C COMPLETES THE SET OF REGRESSION WEIGHT CARDS INVOKED BY THE 'ALL'
C OPTION
      IF (NIVFWC(L) .EQ. 0) GO TO 100
      NWC = NWC + 1
      NIV(NWC) = NIVFWC(L)
      K = 0
      DO 70 J=1,NALL
      IF (FILEWC(J,L) .EQ. 0) GO TO 70
      K = K + 1
      WC(NWC,K) = ALLVAR(J)
   70 CONTINUE
   80 CONTINUE
   90 CONTINUE
C
C RETURN TO STATEMENT 10 AND READ ANOTHER REGRESSION WEIGHT CARD
  100 GO TO 10
C
C TRANSFER IS TO THIS POINT WHEN THE LAST WEIGHT CARD HAS BEEN READ.
C SUBROUTINE NOW RETURNS TO THE MAINPROGRAM.
  110 CONTINUE
      RETURN
      END
      SUBROUTINE REGRES
C
C
C*****************************************************************************
C                                                                             *
C SUBROUTINE REGRES: IS THE GUTS OF THIS PROGRAM AND DOES THE                 *
C                    FOLLOWING FOR EACH REGRESSION:                           *
C                       : CREATES THE REGRESSION MATRIX MAT, AS               *
C                         SPECIFIED BY THE (WC(N)TH REGRESSION WEIGHT         *
C                         CARD, FROM THE MASTER MATRIX MATXX                  *
C                       : CALCULATES INVMAT, THE INVERSE MATRIX OF MAT        *
C                       : FROM THE ELEMENTS OF THE MATRICES INVMAT AND        *
```

```
C                       SUMXY. CALCULATES AND PRINTS FOR THE          *
C                       REGRESSION EQUATION:                          *
C                          : THE INTERCEPT AND ITS 95% CONFIDENCE     *
C                            INTERVAL                                 *
C                          : THE INDEPENDENT VARIABLE REGRESSION      *
C                            COEFFICIENTS AND THEIR 95% CONFIDENCE    *
C                            INTERVALS                                *
C                          : ESTIMATES OF THE DEPENDENT VARIABLE      *
C                            VALUE BASED ON THE CALCULATED            *
C                            REGRESSION EQUATION. FOR EACH DATA       *
C                            POINT                                    *
C                          : DEVIATION OF THE ESTIMATED FROM THE      *
C                            ACTUAL DEPENDENT VARIABLE VALUE FOR      *
C                            EACH DATA POINT                          *
C                          : STANDARD ERROR AND MULTIPLE REGRESSION   *
C                            COEFFICIENT                              *
C                          : ANALYSIS OF VARIANCE (ANOVA) TABLE       *
C                          : THE 'IDEAL' INDEPENDENT VARIABLE         *
C                            VALUES. IF APPLICABLE (SEE MAINPROGRAM)  *
C                       : ALSO WRITES THE REGRESSION EQUATION AND A    *
C                         SUMMARY OF THE ASSOCIATED STATISTICS ON THE  *
C                         SCRATCH FILE SFILE                           *
C                                                                      *
C**********************************************************************
C
C
C**********************************************************************
C                                                                      *
C VARIABLES NOT DEFINED IN THE MAINPROGRAM:                            *
C              CICOEF(I) = 95% CONFIDENCE INTERVAL FOR THE REGRESSION  *
C                          COEFFICIENT OF THE ITH INDEPENDENT VARIABLE *
C              CIINT = 95% CONFIDENCE INTERVAL FOR THE REGRESSION      *
C                      INTERCEPT                                       *
C              DF1 = 1./(NUMBER OF DEGREES OF FREEDOM)                 *
C                  = USED FOR INTERPOLATION OF STUDENTS T VALUES WITH  *
C                    THE STATEMENT FUNCTION TINTRP; DF1 IS ASSOCIATED  *
C                    WITH THE STUDENTS T VALUE T1                      *
C              DF2 = 1./(NUMBER OF DEGREES OF FREEDOM)                 *
C                  = USED FOR INTERPOLATION OF STUDENTS T VALUES WITH  *
C                    THE STATEMENT FUNCTION TINTRP; DF2 IS ASSOCIATED  *
C                    WITH THE STUDENTS T VALUE T2                      *
C              IFLAG(I) = FLAG FOR A PARABOLIC MAXIMUM OR MINIMUM FOR  *
C                         THE ITH (I = 1, 2, 3, OR 4) INDEPENDENT      *
C                         VARIABLE                                     *
C              INTER = THE REGRESSION INTERCEPT                        *
C              INVMAT = ORIGINALLY A NIV(WCN) X NIV(WCN) IDENTITY      *
C                       MATRIX. IT IS TRANSFORMED INTO THE INVERSE OF  *
C                       THE REGRESSION MATRIX MAT FOR THE (WCN)TH      *
C                       REGRESSION IN THE COURSE OF THE MATRIX         *
C                       INVERSION                                      *
C              IPORM = 1 IF INTER > OR = 0.0                           *
C                    = 2 IF INTER < 0.0                                *
C              IVCOEF(I) = THE REGRESSION COEFFICIENT FOR THE ITH      *
C                          INDEPENDENT VARIABLE                        *
C              MAT(I,J) = FOR THE (WCN)TH REGRESSION, MAT IS THE       *
C                         NIV(WCN) X NIV(WCN) REGRESSION MATRIX        *
C                         CREATED FROM MATXX (THE MASTER MATRIX)       *
C                       = MATXX(WC(WCN,I),WC(WCN,J))                   *
C              MINMAX(I) = THE 'IDEAL' VALUE OF THE ITH (I = 1, 2, 3,  *
C                          OR 4) INDEPENDENT VARIABLE FOR MAXIMIZING   *
C                          OR MINIMIZING THE DEPENDENT VARIABLE VALUE  *
C              NDFRES = NUMBER OF RESIDUAL DEGREES OF FREEDOM FOR THE  *
```

```
C                (WCN)TH REGRESSION                                    *
C              = N - NIV(WCN) - 1                                      *
C      NDFTOT = TOTAL NUMBER DEGRESS OF FREEDOM FOR THE DATA           *
C              = N - 1                                                 *
C      NMINUS = NUMBER OF DATA FOR WHICH YDEV < 0.0                    *
C      NPLUS = NUMBER OF DATA POINTS FOR WHICH YDEV > OR = 0.0          *
C      PARENS(1) = THE LITERAL VARIABLE '('                            *
C      PARENS(2) = THE LITERAL VARIABLE ')'                            *
C      PFLAG = THE PIVOT FLAG, WHICH TELLS WHEN EACH FIRST             *
C              PIVOT IS FOUND                                          *
C      PIVCOL(I) = THE PIVOT ELEMENT FLAG FOR THE ITH COLUMN           *
C                  OF MAT                                              *
C                = TELLS NOT ONLY IF THE ITH COLUMN HAS                *
C                  CONTAINED A PIVOT (NOT = 0) OR NOT (= 0),           *
C                  BUT ALSO TELLS WHAT ROW CONTAINED THE PIVOT         *
C                  (VALUE OF PIVCOL(I)).  THIS IS USED IN              *
C                  PLACE OF ACTUALLY EXCHANGING ROWS (COLUMNS          *
C                  NEED NOT BE EXCHANGED AT ALL).                      *
C      PIVOTC = THE COLUMN NUMBER OF THE CURRENT PIVOT ELEMENT         *
C      PIVOTR = THE ROW NUMBER OF THE CURRENT PIVOT ELEMENT            *
C      PIVROW(I) = THE PIVOT ELEMENT FLAG FOR THE ITH ROW OF           *
C                  MAT                                                 *
C                = TELLS WHETHER THE ITH ROW HAS CONTAINED A           *
C                  PIVOT (NOT = 0) OR NOT (=0)                         *
C      PORM(I) = 1 IF IVCOEF(I) > OR = 0.0                             *
C              = 2 IF IVCOEF(I) < 0.0                                  *
C      R = THE (MULTIPLE) REGRESSION COEFFICIENT                       *
C      S = THE REGRESSION STANDARD DEVIATION                           *
C      SIGN(1) = THE LITERAL VARIABLE '+'                              *
C      SIGN(2) = THE LITERAL VARIABLE '-'                              *
C      SIGN(3) = THE LITERAL VARIABLE '_'                              *
C      SS = THE RESIDUAL (OR UNEXPLAINED) VARIANCE OF THE DATA         *
C           POINTS                                                     *
C         = THE REGRESSION VARIANCE                                    *
C         = S**2                                                       *
C         = VRES/NDFRES                                                *
C      STORE = TEMPORARY STORAGE OF VALUES DURING THE                  *
C              CALCULATIONS INVOLVED WITH THE INVERSION OF MAT         *
C      T = STUDENTS T VALUE WHICH WILL BE USED WITH THE                *
C          REGRESSION FOR CALCULATING THE 95% CONFIDENCE               *
C          INTERVALS FOR THE REGRESSION INTERCEPT AND                  *
C          INDEPENDENT VARIABLE COEFFICIENTS                           *
C      T1 = ONE OF THE STUDENTS T VALUES USED FOR                      *
C           INTERPOLATION WITH THE STATEMENT FUNCTION TINTRP;          *
C           T1 IS ASSOCIATED WITH NUMBER OF DEGREES OF FREEDOM         *
C           = 1/DF1                                                    *
C      T2 = ONE OF THE STUDENTS T VALUES USED FOR                      *
C           INTERPOLATION WITH THE STATEMENT FUNCTION TINTRP;          *
C           T2 IS ASSOCIATED WITH NUMBER OF DEGREES OF FREEDOM         *
C           = 1/DF2                                                    *
C      TINTRP = THE STATEMENT FUNCTION TINTRP WHICH IS USED            *
C               TO INTERPOLATE STUDENTS T VALUES FOR THE               *
C               NUMBER OF DEGREES OF FREEDOM = NDFRES > 30             *
C      TOTALV = TOTAL VARIANCE OF THE DATA                             *
C             = TOTVAR/NDFTOT                                          *
C      TVAL(I) = STUDENTS T VALUE FOR I DEGREES OF FREEDOM             *
C                (I = 1 TO 30)                                         *
C      VREG = VARIATION OF THE DATA POINTS EXPLAINED BY THE            *
C             REGRESSION                                               *
C      VRES = THE RESIDUAL (OR UNEXPLAINED) VARIATION OF THE           *
C             DATA POINTS                                              *
C      VVREG = VARIANCE OF THE DATA POINTS EXPLAINED BY THE            *
```

```
C                       REGRESSION                                      *
C                     = VREG/NIV(WCN)                                   *
C               WCN = AN INTEGER VARIABLE THAT KEEPS TRACK OF THE       *
C                     NUMBER OF THE REGRESSION EQUATION CURRENTLY BEING *
C                     RUN                                               *
C               YDEV(I) = Y(I) - YEST(I)                                *
C               YEST(I) = CALCULATED VALUE OF Y(I), BASED ON THE        *
C                         REGRESSION EQUATION                           *
C                                                                       *
C***********************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15),
     X SUMXY(25),   DABS, INTER, IVCOEF(15), MINMAX(4), YEST(100),
     X YDEV(100), CICOEF(15), TINTRP, DSQRT,
     X   TVAL(30)/ 12.706, 4.303, 3.182, 2.776, 2.571, 2.447, 2.365,
     X 2.306, 2.262, 2.228, 2.201, 2.179, 2.160, 2.145, 2.131, 2.120,
     X 2.110, 2.101, 2.093, 2.086, 2.080, 2.074, 2.069, 2.064, 2.060,
     X 2.056, 2.052, 2.048, 2.045, 2.042/
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25),
     X IFLAG(4), SIGN(3)/1H+, 1H-, 1H_/, PARENS(2)/1H(, 1H)/, WCN,
     X PIVROW(15), PIVCOL(15), PFLAG, PIVOTR, PIVOTC, PORM(15)
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
      EQUIVALENCE (MAT(1), IVCOEF), (MAT(16), YEST), (MAT(116), YDEV),
     X (MAT(216), MINMAX), (SDX,SUMXY), (FILEWC(1), PIVROW),
     X (FILEWC(16), PIVCOL), (FILEWC(31), PORM)
C
C DEFINE THE STATEMENT FUNCTION TINTRP
      TINTRP(T1, T2, DF1, DF2) = T1 + (T2 - T1)*
     X ((1./NDFRES - DF1)/(DF2 - DF1))
C
C CALCULATE TOTALV AND NDFTOT
      NDFTOT = N - 1
      TOTALV = TOTVAR/NDFTOT
C
C AS SPECIFIED BY (WC(WCN,J), J=1,NIV(WCN)) FOR THE (WCN)TH REGRESSION
C WEIGHT CARD, CREATE THE REGRESSION MATRIX MAT (A NIV(WCN) X NIV(WCN)
C MATRIX) FROM THE MASTER MATRIX MATXX, TAKING ADVANTAGE OF THE FACT
C MATXX(J,K) = MATXX(K,J) AND THAT MAT(J,K) = MAT(K,J)
      WCN = 1
   10 IEND = NIV(WCN)
      DO 30 I=1,IEND
      DO 20 J=I,IEND
      MAT(I,J) = MATXX(WC(WCN,I),WC(WCN,J))
      IF (I .NE. J) MAT(J,I) = MAT(I,J)
   20 CONTINUE
   30 CONTINUE
C
C INITIALIZE INVMAT AS AN IDENTITY MATRIX OF ORDER NIV(WCN) X NIV(WCN)
C AND INITIALIZE PIVROW AND PIVCOL, THE PIVOT ELEMENT FLAGS
      DO 50 I=1,IEND
      PIVROW(I) = 0
      PIVCOL(I) = 0
      DO 40 J=1,IEND
```

```
      INVMAT(I,J) = 0.
      IF (I .EQ. J) INVMAT(I,J) = 1.
   40 CONTINUE
   50 CONTINUE
C
C NOW PERFORM THE MATRIX INVERSION, CONVERTING MAT INTO AN IDENTITY
C MATRIX OF ORDER NIV(WCN) AND INVMAT INTO THE INVERSE MATRIX OF THE
C ORIGINAL MAT MATRIX.
C ROUND OFF ERRORS DUE TO SMALL DIVISORS IS MINIMIZED BY UTILIZATION OF
C A SEARCH FOR THE LARGEST PIVOT ELEMENT.
C
C FIRST DO THE PIVOT SEARCH
      DO 120 I=1,IEND
      PFLAG = 0
      DO 80 J=1,IEND
      IF (PIVROW(J) .NE. 0) GO TO 80
      DO 70 K=1,IEND
      IF (PIVCOL(K) .NE. 0) GO TO 70
      IF (PFLAG .EQ. 1) GO TO 60
      PIVOTR = J
      PIVOTC = K
      PFLAG = 1
      GO TO 70
   60 IF (DABS(MAT(PIVOTR,PIVOTC)) .GE. DABS(MAT(J,K))) GO TO 70
      PIVOTR = J
      PIVOTC = K
   70 CONTINUE
   80 CONTINUE
      PIVROW(PIVOTR) = 1
      PIVCOL(PIVOTC) = PIVOTR
C
C IF THE PIVOT ELEMENT IS EQUAL TO 0.0, THE REGRESSION MATRIX IS
C SINGULAR AND SO HAS NO UNIQUE SOLUTION: CONTROL IS THEN PASSED TO
C STATEMENT 640.
      IF (MAT(PIVOTR,PIVOTC) .EQ. 0.) GO TO 640
C
C NOW DO THE MATRIX INVERSION
      STORE = MAT(PIVOTR,PIVOTC)
      DO 90 K=1,IEND
      MAT(PIVOTR,K) = MAT(PIVOTR,K)/STORE
      INVMAT(PIVOTR,K) = INVMAT(PIVOTR,K)/STORE
   90 CONTINUE
      IF (NIV(WCN) .EQ. 1) GO TO 120
      DO 110 J=1,IEND
      IF (PIVOTR .EQ. J) GO TO 110
      STORE = MAT(J,PIVOTC)
      DO 100 K=1,IEND
      MAT(J,K) = MAT(J,K) - STORE*MAT(PIVOTR,K)
      INVMAT(J,K) = INVMAT(J,K) - STORE*INVMAT(PIVOTR,K)
  100 CONTINUE
  110 CONTINUE
  120 CONTINUE
C
C USING INVMAT, CALCULATE THE INTERCEPT AND INDEPENDNT VARIABLE
C REGRESSION COEFFICIENTS FOR THE REGRESSION EQUATION, NOTING WHETHER
C EACH IS GREATER THAN OR EQUAL TO 0.0 (IPORM OR PORM = 1) OR LESS THAN
C 0.0 (IPORM OR PORM = 2).
      INTER = YAVG
      DO 140 I=1,IEND
      IVCOEF(I) = 0.
      DO 130 J=1,IEND
      IVCOEF(I) = IVCOEF(I) + INVMAT(PIVCOL(I),J)*SUMXY(WC(WCN,J))
```

```
  130 CONTINUE
      PORM(I) = 1
      IF (IVCOEF(I) .LT. 0.) PORM(I) = 2
      INTER = INTER - XAVG(WC(WCN,I))*IVCOEF(I)
  140 CONTINUE
      IPORM = 1
      IF (INTER .LT. 0.) IPORM = 2
C
C IF NOPT > 0, THEN THE VARIABLES OF THE REGRESSION ARE CHECKED TO SEE
C IF THE EQUATION CONTAINS THE X AND X**2 TERMS AS THE ONLY TERMS
C INVOLVING X, WHERE X = THE 1ST, 2ND, 3RD, OR 4TH INDEPENDENT VARIABLE.
      IF (NOPT .EQ. 0) GO TO 250
      DO 240 I=1,4
      IF (I .EQ. NIV(WCN)) GO TO 250
      IFLAG(I) = 0
      J = IV(WC(WCN,I))
      IF (J .GT. 4) GO TO 240
      K = I + 1
      DO 210 L=K,IEND
      M = IV(WC(WCN,L)) - 14 - J
      GO TO (150,160,170,180),J
  150 GO TO (190,210,210,210,200,210,200,200,200,200),M
      GO TO 210
  160 GO TO (190,210,210,210,200,200,210,200),M
      GO TO 210
  170 GO TO (190,210,210,210,210,200,210,200),M
      GO TO 210
  180 IF (M .EQ. 1) GO TO 190
      GO TO 210
  190 IFLAG(I) = L
      GO TO 210
  200 IFLAG(I) = 0
  210 CONTINUE
C
C IF APPROPRIATE, THE 'IDEAL' VALUE FOR THE ITH (I = 1, 2, 3, OR 4)
C INDEPENDENT VARIABLE IS CALCULATED.
  220 IF (IFLAG(I) .EQ. 0) GO TO 240
      MINMAX(I) = -1.*IVCOEF(I)/(2.*IVCOEF(IFLAG(I)))
C
C IFLAG(I) IS NOW USED TO STORE THE SIGN OF THE COEFFICIENT OF THE
C SQUARED INDEPENDENT VARIABLE TERM AS WELL AS THE MAGNITUDE OF THE
C INTERNAL NUMBER OF THE CORRESPONDING UNSQUARED INDEPENDENT VARIABLE
      IF (IVCOEF(IFLAG(I)) .LT. 0.) GO TO 230
      IFLAG(I) = WC(WCN,I)
      GO TO 240
  230 IFLAG(I) = -WC(WCN,I)
  240 CONTINUE
  250 CONTINUE
C
C CALCULATE ESTIMATES OF THE DEPENDENT VARIABLE VALUES (AND DEVIATIONS
C OF THESE FROM THE ACTUAL VALUES), BASED ON THE CALCULATED REGRESSION
C EQUATION (IE, BASED ON INTER AND THE IVCOEF VALUES) FOR EACH DATA
C POINT.
      NPLUS = 0
      NMINUS = 0
      DO 270 I=1,NTOT
      YEST(I) = INTER
      DO 260 J=1,IEND
      YEST(I) = YEST(I) + X(I,WC(WCN,J))*IVCOEF(J)
  260 CONTINUE
      IF (ICALC(I) .LE. 1) YDEV(I) = Y(I) - YEST(I)
      IF (ICALC(I) .GE. 1) GO TO 270
```

```
      IF (YDEV(I) .LT. 0.) NMINUS = NMINUS + 1
      IF (YDEV(I) .GT. 0.) NPLUS = NPLUS + 1
  270 CONTINUE
C
C CALCULATE ANOVA TABLE ENTRIES: VREG, VVREG, VRES, NDFRES, S, SS, R,
C                                AND RR
      NDFRES = N - NIV(WCN) - 1
      VRES = 0.
      DO 280 I=1,NTOT
      IF (ICALC(I) .NE. 0) GO TO 280
      VRES = VRES + YDEV(I)**2
  280 CONTINUE
      VREG = TOTVAR - VRES
      VVREG = VREG/NIV(WCN)
      SS = VRES/NDFRES
      S = DSQRT(SS)
      RR(WCN) = VREG/TOTVAR
      R = DSQRT(RR(WCN))
C
C CALCULATE STUDENT'S T VALUE WHICH WILL BE USED WITH THIS REGRESSION
C FOR CALCULATING THE 95% CONFIDENCE INTERVALS OF THE REGRESSION
C INTERCEPT (CIINT) AND OF THE INDEPENDENT VARIABLE REGRESSION
C COEFFICIENTS (CICOEF).
      IF (NDFRES .GT. 30) GO TO 290
      T = TVAL(NDFRES)
      GO TO 300
  290 IF (NDFRES .LE. 40) T = TINTRP(2.042,2.021,1./30.,1./40.)
      IF (NDFRES .GT. 40 .AND. NDFRES .LE. 60) T = TINTRP(2.021,2.000,
     X 1./40.,1./60.)
      IF (NDFRES .GT. 60 .AND. NDFRES .LE. 120) T = TINTRP(2.000,1.980,
     X 1./60.,1./120.)
      IF (NDFRES .GT. 120) T = TINTRP(1.980,1.970,1./120.,1./240.)
  300 CONTINUE
C
C CALCULATE 95% CONFIDENCE INTERVALS OF THE REGRESSION INTERCEPT
C (CIINT) AND OF THE INDEPENDENT VARIABLE REGRESSION COEFFICIENTS (CICOEF).
      CIINT = 1./N
      DO 320 J=1,IEND
      CICOEF(J) = T*DSQRT(INVMAT(PIVCOL(J),J)*SS)
      DO 310 K=1,IEND
      CIINT = CIINT + XAVG(WC(WCN,J))*XAVG(WC(WCN,K))*INVMAT(PIVCOL(J),K
     X )
  310 CONTINUE
  320 CONTINUE
      CIINT = T*DSQRT(CIINT*SS)
C
C BECAUSE THE SIGNS OF THE INTERCEPT AND THE INDEPENDENT VARIABLE
C REGRESSION COEFFICIENTS ARE CONTAINED IN IPORM AND PORM, FOR OUTPUT
C PURPOSES THE INTERCEPT AND INDEPENDENT VARIABLE REGRESSION
C COEFFICIENTS ARE SET EQUAL TO THEIR ABSOLUTE VALUES.
      INTER = DABS(INTER)
      DO 330 J=1,IEND
      IVCOEF(J) = DABS(IVCOEF(J))
  330 CONTINUE
C
C NOW, FINALLY, ITS TIME TO PRINT OUT THE REGRESSION EQUATION AND
C ASSOCIATED STATISTICS.
C AT THE SAME TIME, ALSO WRITE THE EQUATION AND A SUMMARY OF THE
C ASSOCIATED STATISTICS ON THE SCRATCH FILE SFILE.
      LINES = NLINES
      DO 400 I=1,NTOT
      IF (LINES .LE. NLINES-4) GO TO 360
```

```
      IF (I .NE. 1) WRITE (IW,340)
  340 FORMAT (1H+, 24(1H_), 3(5X, 8(1H_)))
C
C PRINT HEADINGS FOR TABLE TO CONTAIN Y(I), YEST(I), AND YDEV(I) VALUES
      WRITE (IW,350) WCN, DVAR, DVAR
  350 FORMAT (1H1, 21(1H*)/
     X 2H *, 19X, 1H*/
     X 2H *, 2X, 10HREGRESSION, I5, 2X, 1H*/
     X 2H *, 19X, 1H*/
     X 1X, 21(1H*)/
     X 1H0, 24X, 2(5X,A8)/
     X 8X, 10HDATA POINT, 13X, 6H(OBSD), 7X, 6H(CALC), 8X,4HDEV./
     X 1H+, 24(1H_), 3(5X, 8(1H_)))
      LINES = 8
C
C FOR EACH DATA POINT, PRINT DATA POINT NUMBER, NAME, AND DEPENDENT
C VARIABLE VALUE
  360 WRITE (IW,370) I, (XNAME(I,J), J=1,2), YEST(I)
  370 FORMAT (2X, I3, 1H., 3X, 2A8, 18X, F8.4)
      LINES = LINES + 1
C
C FOR EACH DATA POINT, PRINT VALUES FOR Y(I) AND YDEV(I) IF Y(I) WAS
C PROVIDED
      IF (ICALC(I) .LE. 1) WRITE (IW,380) Y(I), YDEV(I)
  380 FORMAT(1H+, 11X, 2(18X, F8.4))
C
C PUT A '*' BY THE DATA POINT NUMBER IF THE DATA POINT WAS NOT USED IN
C THE REGRESSION CALCULATIONS
      IF (ICALC(I) .GE. 1) WRITE (IW,390)
  390 FORMAT (2H+*)
  400 CONTINUE
C
C PRINT, AT THE BOTTOM OF THE TABLE, VALUES FOR YAVG, SDY, NPLUS, AND
C NMINUS
      WRITE (IW,410) YAVG, SDY, NPLUS, NMINUS
  410 FORMAT (1H+, 24(1H_), 3(5X, 8(1H_))/
     X 5H MEAN, 25X, F8.4/
     X 3H SD, 27X, F8.4/
     X 54X, 7HDEV+ = , I3/
     X 54X, 7HDEV- = , I3)
      LINES = LINES + 4
      IF (LINES .LE. NLINES-9) GO TO 430
      WRITE (IW,420)
  420 FORMAT (1H1)
      LINES = 1
C
C PRINT THE ANOVA TABLE
  430 WRITE (IW,440) VREG, NIV(WCN), VVREG, VRES, NDFRES, SS, TOTVAR,
     X NDFTOT, TOTALV
  440 FORMAT (1H0, 24X, 11HANOVA TABLE/
     X 1H , 60(1H-)/
     X 7H SOURCE, 13X, 9HVARIATION, 12X, 2HDF, 10X, 8HVARIANCE/
     X 1H+, 10(1H_), 9X, 9(1H_), 11X, 3H___, 10X, 8(1H_)/
     X 11H REGRESSION, 10X, F8.4, 11X, I3, 10X, F8.4/
     X 9H RESIDUAL, 12X, F8.4, 11X, I3, 10X, F8.4/
     X 1H+, 10(1H_), 9X, 9(1H_), 11X, 3H___, 10X, 8(1H_)/
     X 6H TOTAL, 15X, F8.4, 11X, I3, 10X, F8.4///)
      LINES = LINES + 10
      IF (LINES .LE. NLINES-3) GO TO 450
      WRITE (IW,420)
      LINES = 1
C
```

```
C PRINT VALUES FOR N, R, RR(WCN), S, AND SS
  450 WRITE (IW,460) N, R, RR(WCN), S, SS
  460 FORMAT (4H N =, I4, 10X, 4HR = , F6.4, 10X, 7HR**2 = , F6.4, 10X,
     X 4HS = , F8.4, 10X, 7HS**2 = , F8.4//)
      LINES = LINES + 3
      IF (LINES .LE. NLINES-9) GO TO 470
      WRITE (IW,420)
      LINES = 1
C
C NOW PRINT THE REGRESSION EQUATION (WITH 95% CONFIDENCE INTERVALS FOR
C THE INTERCEPT AND THE INDEPENDENT VARIABLE REGRESSION COEFFICIENTS),
C PLACING THE ENTIRE EQUATION IN A BOX OF '*'S TO HELP IT STAND OUT FROM
C THE REST OF THE PRINTOUT.
C THE REGRESSION EQUATION IS ALSO WRITTEN ON THE SCRATCH FILE SFILE FOR
C LISTING AT THE END OF THE REGRESSIONS.
  470 WRITE (IW,480) WCN
  480 FORMAT (1X, 120(1H*)/ 2H *, 118X, 1H*/ 12H *  EQUATION, I5, 103X,
     X 1H*/ 2H *, 118X, 1H*)
      LINES = LINES + 3
      IBEGIN = 1
  490 IEND = IBEGIN + 3
      IF (IEND .GT. NIV(WCN)) IEND = NIV(WCN)
      IF (LINES .LE. NLINES-6) GO TO 500
      WRITE (IW,420)
      LINES = 1
  500 IF (IBEGIN .EQ. 1) WRITE (SFILE,510) WCN, DVAR, SIGN(IPORM),
     X INTER, (SIGN(PORM(J)), IVCOEF(J), INDVAR(WC(WCN,J)), J=IBEGIN,
     X IEND)
  510 FORMAT (///3HEQN, I5, 1H., 5X, A8, 2H =, 3X, A1, F8.4,
     X 4(3X, A1, F8.4, 1H*, A8))
      IF (IBEGIN .NE. 1) WRITE (SFILE,520) (SIGN(PORM(J)), IVCOEF(J),
     X INDVAR(WC(WCN,J)), J=IBEGIN,IEND)
  520 FORMAT (///36X, 4(3X, A1, F8.4, 1H*, A8))
      WRITE (IW,530) (SIGN(PORM(J)), IVCOEF(J), INDVAR(WC(WCN,J)),
     X J=IBEGIN,IEND)
  530 FORMAT (2H *,118X, 1H*/ 2H *,118X,1H*/ 120X,1H*/2H**, 27X, 4(5X,
     X A1, F8.4, 1H*, A8))
      IF (IBEGIN .EQ. 1) WRITE (IW,540) DVAR, SIGN(IPORM), INTER
  540 FORMAT (1H*, 3X, A8, 2H =, 5X, A1, F8.4)
      WRITE (IW,550) (PARENS(1), SIGN(1), CICOEF(J), PARENS(2),
     X J=IBEGIN,IEND)
  550 FORMAT (21X, 4(12X, 2A1, F8.4, A1))
      WRITE(IW,560) (SIGN(3), J=IBEGIN,IEND)
  560 FORMAT (2H**, 118X, 1H*/
     X 1H*, 11X, 4(22X, A1))
      IF (IBEGIN .EQ. 1) WRITE (IW,570) CIINT
  570 FORMAT (1H*, 17X, 2H(*, F8.4, 1H)/ 1H*, 18X, 1H_)
      LINES = LINES + 4
      IF (IEND .EQ. NIV(WCN)) GO TO 580
      IBEGIN = IEND + 1
      GO TO 490
  580 WRITE (IW,590)
  590 FORMAT (2H *, 118X,1H*/1X, 120(1H*))
      WRITE (SFILE,600) N, R, RR(WCN), S
  600 FORMAT (/14X, 3HN =, I4, 10X, 4HR = , F6.4, 10X, 7HR**2 = , F6.4,
     X 10X, 4HS = , F8.4)
C
C NOW, IF INDICATED BY THE IFLAG(I), I=1,4 VALUES, PRINT THE 'IDEAL'
C INDEPENDENT VARIABLE VALUES FOR MAXIMIZED AND/OR MINIMIZED DEPENDENT
C VARIABLES.
      IF (NOPT .EQ. 0) GO TO 680
      DO 630 I=1,4
```

```
      IF (I .EQ. NIV(WCN)) GO TO 680
      IF (IFLAG(I) .EQ. 0) GO TO 630
      IF (IFLAG(I) .GT. 0) WRITE (IW,610) DVAR, INDVAR(IFLAG(I)),
     X MINMAX(I)
  610 FORMAT (1H0, A8,21H MINIMIZED FOR IDEAL , A8, 3H = , F8.4)
      IF (IFLAG(I) .LT. 0) WRITE (IW,620) DVAR, INDVAR(-1*IFLAG(I)),
     X MINMAX(I)
  620 FORMAT (1H0, A8,21H MAXIMIZED FOR IDEAL , A8, 3H = , F8.4)
  630 CONTINUE
      GO TO 680
C
C TRANSFER IS TO THIS POINT WHEN A PIVOT ELEMENT IS FOUND TO BE ZERO
C (IE, WHEN THE REGRESSION MATRIX IS SINGULAR AND SO HAS NO UNIQUE
C SOLUTION).  APPROPRIATE COMMENTS TO THIS EFFECT ARE PRINTED AND
C WRITTEN ON THE SCRATCH FILE SFILE.
  640 IEND = NIV(WCN)
C
C NIV(WCN) IS SET TO 99 AS A FLAG TO INDICATE (UPON ANY FURTHER
C REFERENCES TO THIS REGRESSION) THAT THIS REGRESSION HAS NO UNIQUE
C SOLUTION.
      NIV(WCN) = 99
      WRITE (IW,650) WCN, (INDVAR(WC(WCN,J)), J=1,IEND)
  650 FORMAT (1H1, 120(1H*)/
     X 2H *, 118X, 1H*/
     X 14H *  REGRESSION, I5, 101X, 1H*/
     X 2H *, 118X, 1H*/
     X 22H *  MATRIX IS SINGULAR, 98X, 1H*/
     X 2H *, 118X, 1H*/
     X 22H *  INDEPT VARIABLES: , A8, 90X, 1H*/
     X 2H *, 118X, 1H*/
     X (2H *, 19X, A8, 90X, 1H*))
      WRITE (IW,660)
  660 FORMAT (1H+, 119X, 1H*/  1X, 120(1H*))
      WRITE (SFILE,670) WCN, (INDVAR(WC(WCN,J)), J=1,IEND)
  670 FORMAT (/8HEQUATION, I5, 2X, 10(1H*), 2X, 15HMATRIX SINGULAR//
     X (5X, 8(A8, 5X)))
C
C CHECK AND SEE IF ALL THE REGRESSIONS SPECIFIED BY THE WEIGHT CARDS
C HAVE BEEN COMPLETED
  680 IF (WCN .EQ. NWC) GO TO 690
      WCN = WCN + 1
C
C RETURN TO STATEMENT 10 AND DO ANOTHER REGRESSION BASED ON THE NEXT
C REGRESSION WEIGHT CARD
      GO TO 10
C
C TRANSFER IS TO THIS POINT WHEN ALL THE REGRESSIONS HAVE BEEN DONE
  690 CONTINUE
      END FILE SFILE
      RETURN
      END
      SUBROUTINE FTESTS
C
C
C*****************************************************************************
C                                                                             *
C SUBROUTINE FTESTS: 1. FROM THE DIRECT ACCESS FILE FTABLES, CREATES          *
C                       FMAT, A MATRIX CONTAINING THE CRITICAL F              *
C                       STATISTIC VALUES (FOR 7 DIFFERENT                     *
C                       PROBABILITIES) FOR COMPARISON WITH CALCULATED         *
C                       F STATISTIC VALUES                                    *
C                    2. CALCULATES F STATISTIC VALUES (AND ASSOCIATED         *
```

```
C                       DEGREES OF FREEDOM IN THE NUMERATOR AND          *
C                       DENOMINATOR) FOR COMPARISON OF EACH REGRESSION   *
C                       EQUATION WITH EVERY OTHER REGRESSION EQUATION    *
C                       WHICH IS A SUBSET OF IT                          *
C                    3. DETERMINES BETWEEN WHICH CRITICAL F STATISTIC    *
C                       VALUES THE CALCULATED F STATISTIC VALUES LIE     *
C                    4. CALCULATES THE ESTIMATED PERCENTAGE CONFIDENCE   *
C                       LEVELS FOR THE CALCULATED F VALUES               *
C                    5. PRINTS: 2. AND 4. ABOVE                          *
C                             : WHICH INDEPENDENT VARIABLES THE          *
C                               COMPARED EQUATIONS CONTAIN               *
C                                                                        *
C************************************************************************
C
C
C************************************************************************
C                                                                        *
C VARIABLES NOT DEFINED IN THE MAINPROGRAM:                              *
C               BLANK = THE LITERAL VARIABLE ' '                         *
C               DFD = POINTER WHICH TELLS THAT THE CRITICAL F            *
C                     STATISTIC VALUES NEEDED FOR COMPARISON WITH THE    *
C                     CALCULATED F VALUES ARE TO BE FOUND IN             *
C                     FMAT(DFD,NNUM,K)T K=1,7                            *
C               F = THE CALCULATED F TEST VALUE                          *
C               FINTRP = TEMPORARY STORAGE OF F VALUES WHICH ARE READ    *
C                        FROM FTABLES AND ARE THEN USED TO CALCULATE     *
C                        BY INTERPOLATION THE NEEDED CRITICAL F          *
C                        STATISTIC VALUES                                *
C               FMAT(I,J,K) = A MATRIX WHICH IS FILLED ( FROM THE        *
C                             DIRECT ACCESS FILE FTABLES) WITH THE       *
C                             CRITICAL F STATISTIC VALUES NEEDED FOR     *
C                             COMPARISON WITH THE F VALUES TO BE         *
C                             CALCULATED                                 *
C                           = SEE BELOW                                  *
C               FPROB = THE INTERPOLATED PERCENTAGE CONFIDENCE LEVEL     *
C                       OF THE CALCULATED F TEST VALUE                   *
C               FVAR1(I) = BLANK, IF THE EQUATION WCN1 CONTAINS THE      *
C                          INDEPENDENT VARIABLE WITH THE INTERNAL        *
C                          NUMBER I (EXTERNAL NUMBER IV(I))              *
C                        = XX, IF THE EQUATION WCN1 DOES NOT CONTAIN     *
C                          THE INDEPENDENT VARIABLE WITH THE INTERNAL    *
C                          NUMBER I                                      *
C               FVAR2(I) = BLANK, IF THE EQUATION WCN2 CONTAINS THE      *
C                          INDEPENDENT VARIABLE WITH THE INTERNAL        *
C                          NUMBER I (EXTERNAL NUMBER IV(I))              *
C                        = XX, IF THE EQUATION WCN2 DOES NOT CONTAIN     *
C                          THE INDEPENDENT VARIABLE WITH THE INTERNAL    *
C                          NUMBER I                                      *
C               ICOUNT = ASSOCIATED VARIABLE FOR THE DIRECT ACCESS FILE  *
C                        FTABLES                                         *
C                      = SEE BELOW                                       *
C               IEND1 = TEMPORARY STORAGE OF NIV(WCN1)                   *
C               IEND2 = TEMPORARY STORAGE OF NIV(WCN2)                   *
C               IFLAG = FLAG TO INDICATE IF THE F VALUES NEEDED FOR      *
C                       INTERPOLATION HAVE BEEN READ FROM THE DIRECT     *
C                       ACCESS FILE FTABLES (= 1) OR NOT (= 0)           *
C               INTRP = DEGREES OF FREEDOM (OR THEIR RECIPRICALS) WHICH  *
C                       ARE USED IN INTERPOLATION BETWEEN CRITICAL F     *
C                       STATISTIC VALUES IN ORDER TO CALCULATE FPROB     *
C               JEND = AN INDICATOR FOR WHICH VALUE OF J FMAT(I,J,K) IS  *
C                      BEING FILLED                                      *
C               JJ = WHEN VALUES ARE READ FROM FTABLES, JJ IS SET TO     *
```

```
C                    THE VALUE OF JEND SUCH THAT THE MINIMUM NUMBER      *
C                    OF READS FROM FTABLES WILL OCCUR                    *
C               NDENOM = NUMBER OF DEGREES OF FREEDOM IN THE             *
C                      DENOMINATOR OF THE CALCULATED F VALUE FOR         *
C                      EQUATION WCN1                                     *
C                    = N - NIV(WCN1) - 1                                 *
C               NNUM = NUMBER OF DEGREES OF FREEDOM IN THE NUMERATOR     *
C                      OF THE CALCULATED F VALUE FOR COMPARING           *
C                      EQUATION WCN2 TO EQUATION WCN1                    *
C                    = NIV(WCN1) - NIV(WCN2)                             *
C                      WHERE NIV(WCN2) = 0 WHEN COMPARING THE MEAN WITH  *
C                      WCN1                                              *
C               PPOB = PERCENTAGE VALUES USED IN INTERPOLATING BETWEEN   *
C                      CRITICAL F STATISTIC VALUES IN ORDER TO           *
C                      CALCULATE FPROB                                   *
C               WCN1 = THE NUMBER OF THE REGRESSION EQUATION WITH WHICH  *
C                      ALL EQUATIONS WHICH ARE A SUBSET OF IT ARE BEING  *
C                      COMPARED BY MEANS OF THE F TEST                   *
C               WCN2 = THE NUMBER OF THE REGRESSION EQUATION WHICH IS A  *
C                      SUBSET OF WCN1 AND WHICH IS BEING COMPARED WITH   *
C                      WCN1 BY MEANS OF THE F TEST                       *
C               XX = THE LITERAL VARIABLE 'X'                            *
C                                                                        *
C************************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15),
     X FMAT(15,15,7), FINTRP(5,15,7), DLOG10,
     X   INTRP(5)/ 30., 40., 60., 120., 0./,
     X   PROB(7)/ 75.0, 90.0, 95.0, 97.5, 99.0, 99.5, 99.9/
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25),
     X WCN1, WCN2, FVAR1(25), FVAR2(25), BLANK/1H /, XX/1HX/, DFD
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
      EQUIVALENCE  (X(1), FINTRP), (X(526), FMAT), (ICALC(1), FVAR1),
     X (ICALC(26), FVAR2)
C
      DEFINE FILE 12 (510,63,E,ICOUNT)
C
C THE DIRECT ACCESS FILE FTABLES CONTAINS THE CRITICAL F STATISTIC
C VALUES FOR: DFN = 1 TO 15
C           : DFD = 1 TO 30, 40, 60, 120, INFINITY
C
C ICOUNT = SEQUENTIAL RECORD NUMBER OF THE DIRECT ACCESS FILE FTABLES
C        = 15*(I-1) + J
C               WHERE: DFN = J
C                    : DFD = I FOR I = 1 TO 30
C                          = 40 FOR I = 31
C                          = 60 FOR I = 32
C                          = 120 FOR I = 33
C                          = INFINITY FOR I = 34
C
C EACH RECORD OF FTABLES CONTAINS CRITICAL F STATISTIC VALUES FOR 7
C DIFFERENT PERCENTAGE CONFIDENCE LEVELS: 75.0%, 90.0%, 95.0%, 97.5%,
C 99.0%, 99.5%, AND 99.9%.
```

```
C
C READING RECORDS FROM THE DIRECT ACCESS FILE FTABLES, FILL FMAT(I,J,K)
C WITH THE CRITICAL F STATISTIC VALUES NEEDED FOR COMPARISON WITH THE F
C VALUES TO BE CALCULATED: I = N - 1 - IVTOT (OR N - 1 - 15: WHICHEVER
C                              IS LARGER) TO N - 2
C                            = DFD
C                        : J = 1 TO N - 1 - I
C                            = DFN
C                        : K = 1 TO 7
C                            = THE 7 DIFFERENT PERCENTAGE CONFIDENCE
C                              LEVELS 75.0% TO 99.9%
      IFLAG = 0
      IBEGIN = N - 1 - IVTOT
      IF (IVTOT .GT. 15) IBEGIN = N - 1 - 15
      IF (IBEGIN .LE. 0) IBEGIN = 1
      IEND = N - 2
      I = 15 - (N - 1 - IBEGIN)
      DO 200 L=IBEGIN,IEND
      I = I + 1
      JEND = 16 - I
      IF (L .GT. 30) GO TO 30
      ICOUNT = 15*(L-1) + 1
      FIND (IF'ICOUNT)
      DO 20 J=1,JEND
      READ (IF'ICOUNT,10) (FMAT(I,J,K), K=1,7)
   10 FORMAT(7F9.2)
   20 CONTINUE
      GO TO 200
C
C FOR I > 30, FMAT(I,J,K) VALUES ARE CALCULATED BY INTERPOLATION.
C ACTUALLY, ONLY THE UPPER LEFT TRIANGULAR HALF OF FMAT(I,J,K) NEEDS TO
C BE (AND IS) FILLED.
C
C IF IFLAG = 1, THE CRITICAL F STATISTIC VALUES NEEDED FOR THE
C INTERPOLATIONS HAVE ALREADY BEEN READ FROM FTABLES
   30 IF (IFLAG .EQ. 1) GO TO 130
      DO 40 M=1,4
      INTRP(M) = 1./INTRP(M)
   40 CONTINUE
      IFLAG = 1
C
C FIRST, SET UP THE VARIABLES WHICH WILL DEFINE WHICH CRITICAL F
C STATISTIC VALUES WILL BE READ FROM FTABLES FOR THE INTERPOLATIONS
      JJ = JEND
      DO 120 M=1,5
      ICOUNT = 421 + 15*M
      FIND (IF'ICOUNT)
      GO TO (50,60,70,80,90),M
   50 IF (L .GE. 40) GO TO 120
      GO TO 100
   60 IF (L .GE. 60) GO TO 120
      GO TO 100
   70 IF (IEND .LE. 40 .OR. L .GE. 120) GO TO 120
      IF (IEND .LE. 53) JJ = JEND - 40
      GO TO 100
   80 IF (IEND .LE. 60) GO TO 120
      IF (IEND .LE. 73) JJ = JEND - 60
      GO TO 100
   90 IF(IEND .LE. 120) GO TO 120
      IF (IEND .LE. 133) JJ = JEND - 120
C
C READ THE CRITICAL F STATISTIC VALUES NEEDED FOR THE INTERPOLATIONS
```

```
  100 DO 110 J=1,JJ
      READ (IF'ICOUNT,10) (FINTRP(M,J,K), K=1,7)
  110 CONTINUE
  120 CONTINUE
  130 IF (L .GT. 40) GO TO 140
      M = 1
      GO TO 170
  140 IF (L .GT. 60) GO TO 150
      M = 2
      GO TO 170
  150 IF (L .GT. 120) GO TO 160
      M = 3
      GO TO 170
  160 M = 4
C
C NOW DO THE ACTUAL INTERPOLATIONS
  170 DO 190 J=1,JEND
      DO 180 K=1,7
      FMAT(I,J,K) = FINTRP(M,J,K) + (FINTRP(M+1,J,K) - FINTRP(M,J,K))*
     X (1./L - INTRP(M))/(INTRP(M+1) - INTRP(M))
  180 CONTINUE
  190 CONTINUE
  200 CONTINUE
C
C NOW DO THE F TESTS FOR COMPARING EACH EQUATION (WCN1) WITH EACH OTHER
C EQUATION (WCN2) WHICH IS A SUBSET OF IT.
C
C AN EQUATION IS A SUBSET OF ANOTHER EQUATION IF:
C         1. THE NUMBER OF INDEPENDENT VARIABLES OF THE FIRST IS LESS
C            THAN THAT OF THE SECOND.
C         2. ALL THE INDEPENDENT VARIABLES OF THE FIRST EQUATION ARE CONTAINED
C            IN THE SECOND EQUATION.
C         3. THE MEAN (OF THE DEPENDENT VARIABLE) IS A SUBSET OF EVERY
C            EQUATION; IE. THE MEAN IS EQUIVALENT TO AN EQUATION
C            CONTAINING NO INDEPENDENT VARIABLES.
C
      LINES = 0
      DO 440 WCN1 = 1,NWC
C
C IF NIV(WCN1) = 99, THE REGRESSION MATRIX WAS SINGULAR AND NO F TESTS
C CAN BE DONE FOR THIS EQUATION
      IF (NIV(WCN1) .EQ. 99) GO TO 440
      IEND1 = NIV(WCN1)
      DO 210 L=1,IVTOT
      FVAR1(L) = BLANK
      FVAR2(L) = BLANK
  210 CONTINUE
      DO 220 M=1,IEND1
      FVAR1(WC(WCN1,M)) = XX
  220 CONTINUE
      NDENOM = N - NIV(WCN1) - 1
      DFD = 16 - NIV(WCN1)
      K = NWC + 1
      DO 430 J=1,K
      WCN2 = J - 1
      IF (WCN2 .EQ. 0) GO TO 260
C
C FIRST DETERMINE IF EQUATION WCN2 IS A SUBSET OF EQUATION WCN1
      IF (NIV(WCN2) .GE. NIV(WCN1)) GO TO 430
      IEND2 = NIV(WCN2)
      DO 230 L=1,IVTOT
      FVAR2(L) = BLANK
```

```
  230 CONTINUE
      DO 250 L=1,IEND2
      DO 240 M=L,IEND1
      IF (WC(WCN2,L) .NE. WC(WCN1,M)) GO TO 240
      FVAR2(WC(WCN2,L)) = XX
      GO TO 250
  240 CONTINUE
      GO TO 430
  250 CONTINUE
C
C CALCULATE THE F VALUE FOR COMPARING EQUATIONS WCN1 AND WCN2
      NNUM = NIV(WCN1) - NIV(WCN2)
      F = ((RR(WCN1) - RR(WCN2))/NNUM)/((1. - RR(WCN1))/NDENOM)
      GO TO 270
  260 NNUM = NIV(WCN1)
      F = (RR(WCN1)/NNUM)/((1. - RR(WCN1))/NDENO )
  270 IF (LINES .EQ. 0) GO TO 290
      IF (LINES .LE. NLINES-2) GO TO 310
C
C PRINT THE F TESTS TABLE HEADINGS, INSERTING THE EXTERNAL INDEPENDENT
C VARIABLE NUMBERS
      WRITE (IW,280)
  280 FORMAT (1H+, 11(1H_), 2X, 74(1H_), 2X, 3(1H_), 1X, 5(1H_), 2X,
     X 19(1H_))
  290 WRITE (IW,300) (IV(L), L=1,IVTOT)
  300 FORMAT (1H1, 120(1H*)/ 8HOF TESTS/ 1H0, 120(1H*)//
     X 1H0, 40X, 20HEQN INDEPT VARIABLES, 25X, 2(4X, 2HDF)/
     X 14X, 74(1H-), 2X, 2HIN, 4X, 2HIN/
     X 2X, 9HEQUATIONS, 79X, 3HNUM, 1X, 5HDENOM, 6X, 11HF VALUE (%)/
     X 1H+, 12X, 25(1X, I2))
      WRITE (IW,280)
      LINES = 10
C
C PRINT THE F TEST VALUE AND ASSOCIATED DATA
  310 WRITE (IW,320) WCN1, NNUM, NDENOM, F, (FVAR1(I), FVAR2(I),
     X I=1,IVTOT)
  320 FORMAT (1H0, I4, 2H &, 81X, 2(2X, I3), 3X, F9.2/
     X 1H+, 12X, 25(1X, 2A1))
      IF (WCN2 .EQ. 0) GO TO 340
      WRITE (IW,330) WCN2
  330 FORMAT (1H+, 7X, I4)
      GO TO 360
  340 WRITE(IW,350)
  350 FORMAT (1H+, 7X, 4HMEAN)
  360 IF (F .GT. FMAT(DFD,NNUM,1)) GO TO 380
      WRITE (IW,370) PROB(1)
  370 FORMAT (1H+, 110X, 3H(< , F4.1, 2H%))
      GO TO 420
C
C CALCULATE THE % CONFIDENCE LEVEL OF THE F TEST VALUE
  380 DO 400 L=2,7
      IF (F .GT. FMAT(DFD,NNUM,L)) GO TO 400
      PROB1 = PROB(L-1)
      PROB2 = PROB(L)
      FPROB =        10.**(DLOG10(PROB1) + (DLOG10(PROB2) -
     X DLOG10(PROB1)  )*(F - FMAT(DFD,NNUM,L-1))/(FMAT(DFD,NNUM,L) -
     X FMAT(DFD,NNUM,L-1)))
C
C PRINT THE % CONFIDENCE LEVEL OF THE F TEST VALUE
      WRITE (IW,390) FPROB
  390 FORMAT (1H+, 110X, 1H(, F4.1, 2H%))
      GO TO 420
```

```
  400 CONTINUE
      WRITE (IW,410) PROB(7)
  410 FORMAT (1H+, 110X, 3H(> , F4.1, 2H%))
  420 LINES = LINES + 2
  430 CONTINUE
  440 CONTINUE
      WRITE (IW,280)
      RETURN
      END
      SUBROUTINE SUMUP
C
C
C*****************************************************************************
C                                                                             *
C SUBROUTINE SUMUP: SEQUENTIALLY READING FROM THE SCRATCH FILE SFILE,         *
C                   PRINTS AS A SUMMARY OF THE REGRESSION ANALYSES:           *
C                            : THE TITLE CARDS                                *
C                            : THE REGRESSION EQUATIONS AND ASSOCIATED        *
C                              STATISTICAL DATA                               *
C                 : CALCULATES AND PRINTS THE CROSS CORRELATION AND           *
C                   SQUARED CROSS CORRELATION MATRICES FOR THE                *
C                   INDEPENDENT VARIABLES                                     *
C                                                                             *
C*****************************************************************************
C
C
C*****************************************************************************
C                                                                             *
C VARIABLES NOT DEFINED IN THE MAINPROGRAM:                                   *
C              CCMAT = TEMPORARY STORAGE, DURING CALCULATIONS AND             *
C                      BEFORE PRINTING, OF THE INDEPENDENT VARIABLE           *
C                      CROSS CORRELATION MATRIX AND LATER OF THE              *
C                      INDEPENDENT VARIABLE SQUARED CROSS CORRELATION         *
C                      MATRIX                                                 *
C              IFLAG = 1 WHEN THE INDEPENDENT VARIABLE CROSS                  *
C                        CORRELATION MATRIX IS BEING CALCULATED AND           *
C                        PRINTED                                              *
C                    = 2 WHEN THE INDEPENDENT VARIABLE CROSS                  *
C                        CORRELATION MATRIX IS OR HAS BEEN CALCULATED         *
C                        OR PRINTED                                           *
C              SUMING = TEMPORARY STORAGE OF THE RECORDS READ FROM THE        *
C                       SCRATCH FILE SFILE JUST BEFORE THEY ARE               *
C                       PRINTED                                               *
C              ULINE = THE LITERAL VARIABLE '--------'                        *
C                                                                             *
C*****************************************************************************
C
C
      IMPLICIT REAL*8 (A-H,O-Z), INTEGER*2 (I-N)
      REAL*8  INDVAR(25),  XNAME(100,2),  Y(100),  X(100,25),  XAVG(25),
     X SDX(25),  MATXX(25,25),  RR(1100),  MAT(15,15), INVMAT(15,15),
     X ULINE/8H--------/, SUMING(30), CCMAT(25,25), DSQRT
      INTEGER*2   IF*4, WCFILE*4, SFILE*4, IR*4, IW*4, OPTION(10),
     X ICALC(100),  NIV(1100),  WC(1100,15), FILEWC(10,6), IV(25)
      COMMON /REAL/   DVAR, INDVAR,  XNAME,  Y,  X,  YAVG,  TOTVAR,
     X XAVG,  SDX,  SDY,  MATXX, RR, MAT, INVMAT
      COMMON /INTGER/ IF, IR, IW, WCFILE, SFILE, MAXPTS,
     X IMDONE,  NOPT,  NLINES,  LINES,          NI,  OPTION,   IV,
     X IVTOT,  ICALC,  N,  NOREG,  NTOT,  IBEGIN, IEND, NWC,  NIV,
     X I, J, K, L, M, WC, FILEWC
      EQUIVALENCE (XNAME, SUMING), (MATXX, CCMAT)
C
```

```
C REWIND THE SCRATCH FILE SFILE TO THE FIRST RECORD
      REWIND SFILE
C
C SEQUENTIALLY READ AND PRINT RECORDS OF THE SCRATCH FILE SFILE UNTIL AN
C END OF FILE IS REACHED. TRANSFER AT THAT TIME IS TO STATEMENT 50.
      WRITE (IW,10)
   10 FORMAT (1H1, 120(1H*)/ 8H0SUMMARY/
     X 1H0, 120(1H*)/)
   20 READ (SFILE,30,END=50) SUMING
   30 FORMAT (30A4)
      WRITE (IW,40) SUMING
   40 FORMAT (1X, 30A4)
      GO TO 20
   50 CONTINUE
C
C FROM THE MASTER MATRIX MATXX CREATE THE THE SYMMETRICAL CROSS
C CORRELATION MATRIX CCMAT, TAKING ADVANTAGE OF THE FACT THAT CCMAT(I,J)
C = CCMAT(J,I) AND THAT CCMAT(I,I) = 1.0. CCMAT IS EQUIVALENCED WITH
C MATXX.
      IF (IVTOT .EQ. 1) GO TO 80
      IEND = IVTOT - 1
      DO 70 I=1,IEND
      K = I + 1
      DO 60 J=K,IVTOT
      CCMAT(I,J) = MATXX(I,J)/DSQRT(MATXX(I,I)*MATXX(J,J))
      CCMAT(J,I) = CCMAT(I,J)
   60 CONTINUE
   70 CONTINUE
   80 DO 90 I=1,IVTOT
      CCMAT(I,I) = 1.
   90 CONTINUE
C
C PRINT THE INDEPENDENT VARIABLE CROSS CORRELATION MATRIX
      IFLAG = 1
      IBEGIN = 1
  100 WRITE (IW,110)
  110 FORMAT (1H1,       40(1H*), 40HINDEPT VARIABLE CROSS CORRELATION M
     XATRIX, 40(1H*)/)
      GO TO 140
C
C PRINT THE INDEPENDENT VARIABLE SQUARED CROSS CORRELATION MATRIX
  120 WRITE (IW,130)
  130 FORMAT (1H1,       36(1H*), 48HINDEPT VARIABLE SQUARED CROSS CORRE
     XLATION MATRIX, 36(1H*)/)
      IFLAG = 2
  140 IEND = IBEGIN + 9
      IF (IEND .GT. IVTOT) IEND = IVTOT
      WRITE (IW,150) (INDVAR(J), J=IBEGIN,IEND)
  150 FORMAT (1H0, 10X, 10(3X, A8))
      WRITE (IW,160) (ULINE, J=IBEGIN,IEND)
  160 FORMAT (1X, 8(1H-), 2X, 10(3X, A8))
      DO 180 I=1,IVTOT
      WRITE (IW,170) INDVAR(I), (CCMAT(I,J), J=IBEGIN,IEND)
  170 FORMAT (1H0, A8, 1X, 10(5X, F6.3))
  180 CONTINUE
      WRITE (IW,160) (ULINE, J=IBEGIN,IEND)
      IF (IEND .EQ. IVTOT) GO TO 190
      IBEGIN = IEND + 1
      GO TO 100
  190 CONTINUE
      IF (IFLAG .EQ. 2) GO TO 220
C
```

```
C NOW CALCULATE THE INDEPENDENT VARIABLE SQUARED CROSS CORRELATION
C MATRIX BY SQUARING EACH ELEMENT OF CCMAT.
      DO 210 I=1,IVTOT
      DO 200 J=1,IVTOT
      CCMAT(I,J) = CCMAT(I,J)**2
  200 CONTINUE
  210 CONTINUE
      GO TO 120
  220 CONTINUE
      RETURN
      END
```

Supporting Programs for PROGRAM QSAR47

PROGRAM FILLFTBL.

```
C
C
C PROGRAM FILLFTBL
C
C
C A SHORT PROGRAM TO READ (FROM DATA CARDS) F CRITICAL POINT VALUES
C AND THEN TO WRITE THEM ON THE ON-LINE DISK DIRECT ACCESS
C FILE FTABLES
C
C PROGRAM WRITTEN BY: STEVE DIETRICH
C                     DEPARTMENT OF PHARMACEUTICAL CHEMISTRY
C                     SCHOOL OF PHARMACY
C                     UNIVERSITY OF CALIFORNIA
C                     SAN FRANCISCO, CALIFORNIA 94143
C                     APRIL, 1976
C
C PROGRAM WRITTEN, DEBUGGED, AND USED IN WATFIV ON AN IBM MODEL 370
C 145 COMPUTER AT UCSF
C
C F CRITICAL POINT VALUES FROM:
C     BIOMETRICA TABLES FOR STATISTICIANS
C     VOLUME I
C     EDITED BY E. S. PEARSON AND H. O. HARTLEY
C     CAMBRIDGE AT THE UNIVERSITY PRESS
C     FOURTH EDITION (1956)
C
C F CRITICAL POINT VALUES ARE READ FOR:
C     PROBABILITY (P) = 75.0%, 90.0%, 95.0%, 97.5%, 99%, 99.5%,
C                       AND 99.9%.
C     DEGREES OF FREEDOM IN THE DENOMINATOR (DFD) = 1-30, 40, 60,
C                                                   120, AND INFINITY.
C     DEGREES OF FREEDOM IN THE NUMERATOR (DFN) = 1-10, 12, AND 15.
C
C F CRITICAL POINT VALUES FOR DFN = 11, 13, AND 14 ARE OBTAINED
C BY INTERPOLATION
C
C THE ON-LINE DISK DIRECT ACCESS FILE FTABLES CONTAINS 510 FORMATTED
C (7F9.2) RECORDS EACH WITH A LENGTH OF 63 BYTES.
C
C
      REAL*4 INTERP
      DIMENSION F(34,15,7)
      DEFINE FILE 12 (510,63,E,ICOUNT)
      INTERP(A,C,IA,IB,IC) = A + (C-A)*(((1./IB)-(1./IA))/((1./IC)-(1./I
     XA)))
      IF = 12
      IR = 5
      DO 50 K=1,7
      DO 50 I=1,34
      READ (IR,150) (F(I,J,K),J=1,10), F(I,12,K), F(I,15,K)
  150 FORMAT (6(F9.2,1X))
      F(I,11,K) = INTERP(F(I,10,K),F(I,12,K),10,11,12)
      F(I,13,K) = INTERP(F(I,12,K),F(I,15,K),12,13,15)
      F(I,14,K) = INTERP(F(I,12,K),F(I,15,K),12,14,15)
   50 CONTINUE
      ICOUNT = 1
```

```
C NOW CALCULATE THE INDEPENDENT VARIABLE SQUARED CROSS CORRELATION
C MATRIX BY SQUARING EACH ELEMENT OF CCMAT.
      DO 210 I=1,IVTOT
      DO 200 J=1,IVTOT
      CCMAT(I,J) = CCMAT(I,J)**2
  200 CONTINUE
  210 CONTINUE
      GO TO 120
  220 CONTINUE
      RETURN
      END
```

Data Cards for PROGRAM FILLFTBL.

5.83	7.50	8.20	8.58	8.82	8.98	75.0% DFD= 1 #1
9.10	9.19	9.26	9.32	9.41	9.49	75.0% DFD= 1 #2
2.57	3.00	3.15	3.23	3.28	3.31	75.0% DFD= 2 #1
3.34	3.35	3.37	3.38	3.39	3.41	75.0% DFD= 2 #2
2.02	2.28	2.36	2.39	2.41	2.42	75.0% DFD= 3 #1
2.43	2.44	2.44	2.44	2.45	2.46	75.0% DFD= 3 #2
1.81	2.00	2.05	2.06	2.07	2.08	75.0% DFD= 4 #1
2.08	2.08	2.08	2.08	2.08	2.08	75.0% DFD= 4 #2
1.69	1.85	1.88	1.89	1.89	1.89	75.0% DFD= 5 #1
1.89	1.89	1.89	1.89	1.89	1.89	75.0% DFD= 5 #2
1.62	1.76	1.78	1.79	1.79	1.78	75.0% DFD= 6 #1
1.78	1.78	1.77	1.77	1.77	1.76	75.0% DFD= 6 #2
1.57	1.70	1.72	1.72	1.71	1.71	75.0% DFD= 7 #1
1.70	1.70	1.69	1.69	1.68	1.68	75.0% DFD= 7 #2
1.54	1.66	1.67	1.66	1.66	1.65	75.0% DFD= 8 #1
1.64	1.64	1.63	1.63	1.62	1.62	75.0% DFD= 8 #2
1.51	1.62	1.63	1.63	1.62	1.61	75.0% DFD= 9 #1
1.60	1.60	1.59	1.59	1.58	1.57	75.0% DFD= 9 #2
1.49	1.60	1.60	1.59	1.59	1.58	75.0% DFD= 10 #1
1.57	1.56	1.56	1.55	1.54	1.53	75.0% DFD= 10 #2
1.47	1.58	1.58	1.57	1.56	1.55	75.0% DFD= 11 #1
1.54	1.53	1.53	1.52	1.51	1.50	75.0% DFD= 11 #2
1.46	1.56	1.56	1.55	1.54	1.53	75.0% DFD= 12 #1
1.52	1.51	1.51	1.50	1.49	1.48	75.0% DFD= 12 #2
1.45	1.55	1.55	1.53	1.52	1.51	75.0% DFD= 13 #1
1.50	1.49	1.49	1.48	1.47	1.46	75.0% DFD= 13 #2
1.44	1.53	1.53	1.52	1.51	1.50	75.0% DFD= 14 #1
1.49	1.48	1.47	1.46	1.45	1.44	75.0% DFD= 14 #2
1.43	1.52	1.52	1.51	1.49	1.48	75.0% DFD= 15 #1
1.47	1.46	1.46	1.45	1.44	1.43	75.0% DFD= 15 #2
1.42	1.51	1.51	1.50	1.48	1.47	75.0% DFD= 16 #1
1.46	1.45	1.44	1.44	1.43	1.41	75.0% DFD= 16 #2
1.42	1.51	1.50	1.49	1.47	1.46	75.0% DFD= 17 #1
1.45	1.44	1.43	1.43	1.41	1.40	75.0% DFD= 17 #2
1.41	1.50	1.49	1.48	1.46	1.45	75.0% DFD= 18 #1
1.44	1.43	1.42	1.42	1.40	1.39	75.0% DFD= 18 #2
1.41	1.49	1.49	1.47	1.46	1.44	75.0% DFD= 19 #1
1.43	1.42	1.41	1.41	1.40	1.38	75.0% DFD= 19 #2
1.40	1.49	1.48	1.47	1.45	1.44	75.0% DFD= 20 #1
1.43	1.42	1.41	1.40	1.39	1.37	75.0% DFD= 20 #2
1.40	1.48	1.48	1.46	1.44	1.43	75.0% DFD= 21 #1
1.42	1.41	1.40	1.39	1.38	1.37	75.0% DFD= 21 #2
1.40	1.48	1.47	1.45	1.44	1.42	75.0% DFD= 22 #1
1.41	1.40	1.39	1.39	1.37	1.36	75.0% DFD= 22 #2
1.39	1.47	1.47	1.45	1.43	1.42	75.0% DFD= 23 #1
1.41	1.40	1.39	1.38	1.37	1.35	75.0% DFD= 23 #2
1.39	1.47	1.46	1.44	1.43	1.41	75.0% DFD= 24 #1
1.40	1.39	1.38	1.38	1.36	1.35	75.0% DFD= 24 #2
1.39	1.47	1.46	1.44	1.42	1.41	75.0% DFD= 25 #1
1.40	1.39	1.38	1.37	1.36	1.34	75.0% DFD= 25 #2
1.38	1.46	1.45	1.44	1.42	1.41	75.0% DFD= 26 #1
1.39	1.38	1.37	1.37	1.35	1.34	75.0% DFD= 26 #2
1.38	1.46	1.45	1.43	1.42	1.40	75.0% DFD= 27 #1
1.39	1.38	1.37	1.36	1.35	1.33	75.0% DFD= 27 #2
1.38	1.46	1.45	1.43	1.41	1.40	75.0% DFD= 28 #1
1.39	1.38	1.37	1.36	1.34	1.33	75.0% DFD= 28 #2

1.38	1.45	1.45	1.43	1.41	1.40	75.0% DFD= 29 #1
1.38	1.37	1.36	1.35	1.34	1.32	75.0% DFD= 29 #2
1.38	1.45	1.44	1.42	1.41	1.39	75.0% DFD= 30 #1
1.38	1.37	1.36	1.35	1.34	1.32	75.0% DFD= 30 #2
1.36	1.44	1.42	1.40	1.39	1.37	75.0% DFD= 40 #1
1.36	1.35	1.34	1.33	1.31	1.30	75.0% DFD= 40 #2
1.35	1.42	1.41	1.38	1.37	1.35	75.0% DFD= 60 #1
1.33	1.32	1.31	1.30	1.29	1.27	75.0% DFD= 60 #2
1.34	1.40	1.39	1.37	1.35	1.33	75.0% DFD=120 #1
1.31	1.30	1.29	1.28	1.26	1.24	75.0% DFD=120 #2
1.32	1.39	1.37	1.35	1.33	1.31	75.0% DFD=INF #1
1.29	1.28	1.27	1.25	1.24	1.22	75.0% DFD=INF #2
39.86	49.50	53.59	55.83	57.24	58.20	90.0% DFD= 1 #1
58.91	59.44	59.86	60.19	60.71	61.22	90.0% DFD= 1 #2
8.53	9.00	9.16	9.24	9.29	9.33	90.0% DFD= 2 #1
9.35	9.37	9.38	9.39	9.41	9.42	90.0% DFD= 2 #2
5.54	5.46	5.39	5.34	5.31	5.28	90.0% DFD= 3 #1
5.27	5.25	5.24	5.23	5.22	5.20	90.0% DFD= 3 #2
4.54	4.32	4.19	4.11	4.05	4.01	90.0% DFD= 4 #1
3.98	3.95	3.94	3.92	3.90	3.87	90.0% DFD= 4 #2
4.06	3.78	3.62	3.52	3.45	3.40	90.0% DFD= 5 #1
3.37	3.34	3.32	3.30	3.27	3.24	90.0% DFD= 5 #2
3.78	3.46	3.29	3.18	3.11	3.05	90.0% DFD= 6 #1
3.01	2.98	2.96	2.94	2.90	2.87	90.0% DFD= 6 #2
3.59	3.26	3.07	2.96	2.88	2.83	90.0% DFD= 7 #1
2.78	2.75	2.72	2.70	2.67	2.63	90.0% DFD= 7 #2
3.46	3.11	2.92	2.81	2.73	2.67	90.0% DFD= 8 #1
2.62	2.59	2.56	2.54	2.50	2.46	90.0% DFD= 8 #2
3.36	3.01	2.81	2.69	2.61	2.55	90.0% DFD= 9 #1
2.51	2.47	2.44	2.42	2.38	2.34	90.0% DFD= 9 #2
3.29	2.92	2.73	2.61	2.52	2.46	90.0% DFD= 10 #1
2.41	2.38	2.35	2.32	2.28	2.24	90.0% DFD= 10 #2
3.23	2.86	2.66	2.54	2.45	2.39	90.0% DFD= 11 #1
2.34	2.30	2.27	2.25	2.21	2.17	90.0% DFD= 11 #2
3.18	2.81	2.61	2.48	2.39	2.33	90.0% DFD= 12 #1
2.28	2.24	2.21	2.19	2.15	2.10	90.0% DFD= 12 #2
3.14	2.76	2.56	2.43	2.35	2.28	90.0% DFD= 13 #1
2.23	2.20	2.16	2.14	2.10	2.05	90.0% DFD= 13 #2
3.10	2.73	2.52	2.39	2.31	2.24	90.0% DFD= 14 #1
2.19	2.15	2.12	2.10	2.05	2.01	90.0% DFD= 14 #2
3.07	2.70	2.49	2.36	2.27	2.21	90.0% DFD= 15 #1
2.16	2.12	2.09	2.06	2.02	1.97	90.0% DFD= 15 #2
3.05	2.67	2.46	2.33	2.24	2.18	90.0% DFD= 16 #1
2.13	2.09	2.06	2.03	1.99	1.94	90.0% DFD= 16 #2
3.03	2.64	2.44	2.31	2.22	2.15	90.0% DFD= 17 #1
2.10	2.06	2.03	2.00	1.96	1.91	90.0% DFD= 17 #2
3.01	2.62	2.42	2.29	2.20	2.13	90.0% DFD= 18 #1
2.08	2.04	2.00	1.98	1.93	1.89	90.0% DFD= 18 #2
2.99	2.61	2.40	2.27	2.18	2.11	90.0% DFD= 19 #1
2.06	2.02	1.98	1.96	1.91	1.86	90.0% DFD= 19 #2
2.97	2.59	2.38	2.25	2.16	2.09	90.0% DFD= 20 #1
2.04	2.00	1.96	1.94	1.89	1.84	90.0% DFD= 20 #2
2.96	2.57	2.36	2.23	2.14	2.08	90.0% DFD= 21 #1
2.02	1.98	1.95	1.92	1.87	1.83	90.0% DFD= 21 #2
2.95	2.56	2.35	2.22	2.13	2.06	90.0% DFD= 22 #1
2.01	1.97	1.93	1.90	1.86	1.81	90.0% DFD= 22 #2
2.94	2.55	2.34	2.21	2.11	2.05	90.0% DFD= 23 #1
1.99	1.95	1.92	1.89	1.84	1.80	90.0% DFD= 23 #2
2.93	2.54	2.33	2.19	2.10	2.04	90.0% DFD= 24 #1
1.98	1.94	1.91	1.88	1.83	1.78	90.0% DFD= 24 #2
2.92	2.53	2.32	2.18	2.09	2.02	90.0% DFD= 25 #1
1.97	1.93	1.89	1.87	1.82	1.77	90.0% DFD= 25 #2

2.91	2.52	2.31	2.17	2.08	2.01	90.0% DFD= 26 #1
1.96	1.92	1.88	1.86	1.81	1.76	90.0% DFD= 26 #2
2.90	2.51	2.30	2.17	2.07	2.00	90.0% DFD= 27 #1
1.95	1.91	1.87	1.85	1.80	1.75	90.0% DFD= 27 #2
2.89	2.50	2.29	2.16	2.06	2.00	90.0% DFD= 28 #1
1.94	1.90	1.87	1.84	1.79	1.74	90.0% DFD= 28 #2
2.89	2.50	2.28	2.15	2.06	1.99	90.0% DFD= 29 #1
1.93	1.89	1.86	1.83	1.78	1.73	90.0% DFD= 29 #2
2.88	2.49	2.28	2.14	2.05	1.98	90.0% DFD= 30 #1
1.93	1.88	1.85	1.82	1.77	1.72	90.0% DFD= 30 #2
2.84	2.44	2.23	2.09	2.00	1.93	90.0% DFD= 40 #1
1.87	1.83	1.79	1.76	1.71	1.66	90.0% DFD= 40 #2
2.79	2.39	2.18	2.04	1.95	1.87	90.0% DFD= 60 #1
1.82	1.77	1.74	1.71	1.66	1.60	90.0% DFD= 60 #2
2.75	2.35	2.13	1.99	1.90	1.82	90.0% DFD=120 #1
1.77	1.72	1.68	1.65	1.60	1.55	90.0% DFD=120 #2
2.71	2.30	2.08	1.94	1.85	1.77	90.0% DFD=INF #1
1.72	1.67	1.63	1.60	1.55	1.49	90.0% DFD=INF #2
161.4	199.5	215.7	224.6	230.2	234.0	95.0% DFD= 1 #1
236.8	238.9	240.5	241.9	243.9	245.9	95.0% DFD= 1 #2
18.51	19.00	19.16	19.25	19.30	19.33	95.0% DFD= 2 #1
19.35	19.37	19.38	19.40	19.41	19.43	95.0% DFD= 2 #2
10.13	9.55	9.28	9.12	9.01	8.94	95.0% DFD= 3 #1
8.89	8.85	8.81	8.79	8.74	8.70	95.0% DFD= 3 #2
7.71	6.94	6.59	6.39	6.26	6.16	95.0% DFD= 4 #1
6.09	6.04	6.00	5.96	5.91	5.86	95.0% DFD= 4 #2
6.61	5.79	5.41	5.19	5.05	4.95	95.0% DFD= 5 #1
4.88	4.82	4.77	4.74	4.68	4.62	95.0% DFD= 5 #2
5.99	5.14	4.76	4.53	4.39	4.28	95.0% DFD= 6 #1
4.21	4.15	4.10	4.06	4.00	3.94	95.0% DFD= 6 #2
5.59	4.74	4.35	4.12	3.97	3.87	95.0% DFD= 7 #1
3.79	3.73	3.68	3.64	3.57	3.51	95.0% DFD= 7 #2
5.32	4.46	4.07	3.84	3.69	3.58	95.0% DFD= 8 #1
3.50	3.44	3.39	3.35	3.28	3.22	95.0% DFD= 8 #2
5.12	4.26	3.86	3.63	3.48	3.37	95.0% DFD= 9 #1
3.29	3.23	3.18	3.14	3.07	3.01	95.0% DFD= 9 #2
4.96	4.10	3.71	3.48	3.33	3.22	95.0% DFD= 10 #1
3.14	3.07	3.02	2.98	2.91	2.85	95.0% DFD= 10 #2
4.84	3.98	3.59	3.36	3.20	3.09	95.0% DFD= 11 #1
3.01	2.95	2.90	2.85	2.79	2.72	95.0% DFD= 11 #2
4.75	3.89	3.49	3.26	3.11	3.00	95.0% DFD= 12 #1
2.91	2.85	2.80	2.75	2.69	2.62	95.0% DFD= 12 #2
4.67	3.81	3.41	3.18	3.03	2.92	95.0% DFD= 13 #1
2.83	2.77	2.71	2.67	2.60	2.53	95.0% DFD= 13 #2
4.60	3.74	3.34	3.11	2.96	2.85	95.0% DFD= 14 #1
2.76	2.70	2.65	2.60	2.53	2.46	95.0% DFD= 14 #2
4.54	3.68	3.29	3.06	2.90	2.79	95.0% DFD= 15 #1
2.71	2.64	2.59	2.54	2.48	2.40	95.0% DFD= 15 #2
4.49	3.63	3.24	3.01	2.85	2.74	95.0% DFD= 16 #1
2.66	2.59	2.54	2.49	2.42	2.35	95.0% DFD= 16 #2
4.45	3.59	3.20	2.96	2.81	2.70	95.0% DFD= 17 #1
2.61	2.55	2.49	2.45	2.38	2.31	95.0% DFD= 17 #2
4.41	3.55	3.16	2.93	2.77	2.66	95.0% DFD= 18 #1
2.58	2.51	2.46	2.41	2.34	2.27	95.0% DFD= 18 #2
4.38	3.52	3.13	2.90	2.74	2.63	95.0% DFD= 19 #1
2.54	2.48	2.42	2.38	2.31	2.23	95.0% DFD= 19 #2
4.35	3.49	3.10	2.87	2.71	2.60	95.0% DFD= 20 #1
2.51	2.45	2.39	2.35	2.28	2.20	95.0% DFD= 20 #2
4.32	3.47	3.07	2.84	2.68	2.57	95.0% DFD= 21 #1
2.49	2.42	2.37	2.32	2.25	2.18	95.0% DFD= 21 #2
4.30	3.44	3.05	2.82	2.66	2.55	95.0% DFD= 22 #1
2.46	2.40	2.34	2.30	2.23	2.15	95.0% DFD= 22 #2

4.28	3.42	3.03	2.80	2.64	2.53	95.0% DFD= 23 #1
2.44	2.37	2.32	2.27	2.20	2.13	95.0% DFD= 23 #2
4.26	3.40	3.01	2.78	2.62	2.51	95.0% DFD= 24 #1
2.42	2.36	2.30	2.25	2.18	2.11	95.0% DFD= 24 #2
4.24	3.39	2.99	2.76	2.60	2.49	95.0% DFD= 25 #1
2.40	2.34	2.28	2.24	2.16	2.09	95.0% DFD= 25 #2
4.23	3.37	2.98	2.74	2.59	2.47	95.0% DFD= 26 #1
2.39	2.32	2.27	2.22	2.15	2.07	95.0% DFD= 26 #2
4.21	3.35	2.96	2.73	2.57	2.46	95.0% DFD= 27 #1
2.37	2.31	2.25	2.20	2.13	2.06	95.0% DFD= 27 #2
4.20	3.34	2.95	2.71	2.56	2.45	95.0% DFD= 28 #1
2.36	2.29	2.24	2.19	2.12	2.04	95.0% DFD= 28 #2
4.18	3.33	2.93	2.70	2.55	2.43	95.0% DFD= 29 #1
2.35	2.28	2.22	2.18	2.10	2.03	95.0% DFD= 29 #2
4.17	3.32	2.92	2.69	2.53	2.42	95.0% DFD= 30 #1
2.33	2.27	2.21	2.16	2.09	2.01	95.0% DFD= 30 #2
4.08	3.23	2.84	2.61	2.45	2.34	95.0% DFD= 40 #1
2.25	2.18	2.12	2.08	2.00	1.92	95.0% DFD= 40 #2
4.00	3.15	2.76	2.53	2.37	2.25	95.0% DFD= 60 #1
2.17	2.10	2.04	1.99	1.92	1.84	95.0% DFD= 60 #2
3.92	3.07	2.68	2.45	2.29	2.17	95.0% DFD=120 #1
2.09	2.02	1.96	1.91	1.83	1.75	95.0% DFD=120 #2
3.84	3.00	2.60	2.37	2.21	2.10	95.0% DFD=INF #1
2.01	1.94	1.88	1.83	1.75	1.67	95.0% DFD=INF #2
647.8	799.5	864.2	899.6	921.8	937.1	97.5% DFD= 1 #1
948.2	956.7	963.3	968.6	976.7	984.9	97.5% DFD= 1 #2
38.51	39.00	39.17	39.25	39.30	39.33	97.5% DFD= 2 #1
39.36	39.37	39.39	39.40	39.41	39.43	97.5% DFD= 2 #2
17.44	16.04	15.44	15.10	14.88	14.73	97.5% DFD= 3 #1
14.62	14.54	14.47	14.42	14.34	14.25	97.5% DFD= 3 #2
12.22	10.65	9.98	9.60	9.36	9.20	97.5% DFD= 4 #1
9.07	8.98	8.90	8.84	8.75	8.66	97.5% DFD= 4 #2
10.01	8.43	7.76	7.39	7.15	6.98	97.5% DFD= 5 #1
6.85	6.76	6.68	6.62	6.52	6.43	97.5% DFD= 5 #2
8.81	7.26	6.60	6.23	5.99	5.82	97.5% DFD= 6 #1
5.70	5.60	5.52	5.46	5.37	5.27	97.5% DFD= 6 #2
8.07	6.54	5.89	5.52	5.29	5.12	97.5% DFD= 7 #1
4.99	4.90	4.82	4.76	4.67	4.57	97.5% DFD= 7 #2
7.57	6.06	5.42	5.05	4.82	4.65	97.5% DFD= 8 #1
4.53	4.43	4.36	4.30	4.20	4.10	97.5% DFD= 8 #2
7.21	5.71	5.08	4.72	4.48	4.32	97.5% DFD= 9 #1
4.20	4.10	4.03	3.96	3.87	3.77	97.5% DFD= 9 #2
6.94	5.46	4.83	4.47	4.24	4.07	97.5% DFD= 10 #1
3.95	3.85	3.78	3.72	3.62	3.52	97.5% DFD= 10 #2
6.72	5.26	4.63	4.28	4.04	3.88	97.5% DFD= 11 #1
3.76	3.66	3.59	3.53	3.43	3.33	97.5% DFD= 11 #2
6.55	5.10	4.47	4.12	3.89	3.73	97.5% DFD= 12 #1
3.61	3.51	3.44	3.37	3.28	3.18	97.5% DFD= 12 #2
6.41	4.97	4.35	4.00	3.77	3.60	97.5% DFD= 13 #1
3.48	3.39	3.31	3.25	3.15	3.05	97.5% DFD= 13 #2
6.30	4.86	4.24	3.89	3.66	3.50	97.5% DFD= 14 #1
3.38	3.29	3.21	3.15	3.05	2.95	97.5% DFD= 14 #2
6.20	4.77	4.15	3.80	3.58	3.41	97.5% DFD= 15 #1
3.29	3.20	3.12	3.06	2.96	2.86	97.5% DFD= 15 #2
6.12	4.69	4.08	3.73	3.50	3.34	97.5% DFD= 16 #1
3.22	3.12	3.05	2.99	2.89	2.79	97.5% DFD= 16 #2
6.04	4.62	4.01	3.66	3.44	3.28	97.5% DFD= 17 #1
3.16	3.06	2.98	2.92	2.82	2.72	97.5% DFD= 17 #2
5.98	4.56	3.95	3.61	3.38	3.22	97.5% DFD= 18 #1
3.10	3.01	2.93	2.87	2.77	2.67	97.5% DFD= 18 #2
5.92	4.51	3.90	3.56	3.33	3.17	97.5% DFD= 19 #1
3.05	2.96	2.88	2.82	2.72	2.62	97.5% DFD= 19 #2

5.87	4.46	3.86	3.51	3.29	3.13	97.5% DFD= 20 #1
3.01	2.91	2.84	2.77	2.68	2.57	97.5% DFD= 20 #2
5.83	4.42	3.82	3.48	3.25	3.09	97.5% DFD= 21 #1
2.97	2.87	2.80	2.73	2.64	2.53	97.5% DFD= 21 #2
5.79	4.38	3.78	3.44	3.22	3.05	97.5% DFD= 22 #1
2.93	2.84	2.76	2.70	2.60	2.50	97.5% DFD= 22 #2
5.75	4.35	3.75	3.41	3.18	3.02	97.5% DFD= 23 #1
2.90	2.81	2.73	2.67	2.57	2.47	97.5% DFD= 23 #2
5.72	4.32	3.72	3.38	3.15	2.99	97.5% DFD= 24 #1
2.87	2.78	2.70	2.64	2.54	2.44	97.5% DFD= 24 #2
5.69	4.29	3.69	3.35	3.13	2.97	97.5% DFD= 25 #1
2.85	2.75	2.68	2.61	2.51	2.41	97.5% DFD= 25 #2
5.66	4.27	3.67	3.33	3.10	2.94	97.5% DFD= 26 #1
2.82	2.73	2.65	2.59	2.49	2.39	97.5% DFD= 26 #2
5.63	4.24	3.65	3.31	3.08	2.92	97.5% DFD= 27 #1
2.80	2.71	2.63	2.57	2.47	2.36	97.5% DFD= 27 #2
5.61	4.22	3.63	3.29	3.06	2.90	97.5% DFD= 28 #1
2.78	2.69	2.61	2.55	2.45	2.34	97.5% DFD= 28 #2
5.59	4.20	3.61	3.27	3.04	2.88	97.5% DFD= 29 #1
2.76	2.67	2.59	2.53	2.43	2.32	97.5% DFD= 29 #2
5.57	4.18	3.59	3.25	3.03	2.87	97.5% DFD= 30 #1
2.75	2.65	2.57	2.51	2.41	2.31	97.5% DFD= 30 #2
5.42	4.05	3.46	3.13	2.90	2.74	97.5% DFD= 40 #1
2.62	2.53	2.45	2.39	2.29	2.18	97.5% DFD= 40 #2
5.29	3.93	3.34	3.01	2.79	2.63	97.5% DFD= 60 #1
2.51	2.41	2.33	2.27	2.17	2.06	97.5% DFD= 60 #2
5.15	3.80	3.23	2.89	2.67	2.52	97.5% DFD=120 #1
2.39	2.30	2.22	2.16	2.05	1.94	97.5% DFD=120 #2
5.02	3.69	3.12	2.79	2.57	2.41	97.5% DFD=INF #1
2.29	2.19	2.11	2.05	1.94	1.83	97.5% DFD=INF #2
4052.	4999.5	5403.	5625.	5764.	5859.	99.0% DFD= 1 #1
5928.	5982.	6022.	6056.	6106.	6157.	99.0% DFD= 1 #2
98.50	99.00	99.17	99.25	99.30	99.33	99.0% DFD= 2 #1
99.36	99.37	99.39	99.40	99.42	99.43	99.0% DFD= 2 #2
34.12	30.82	29.46	28.71	28.24	27.91	99.0% DFD= 3 #1
27.67	27.49	27.35	27.23	27.05	26.87	99.0% DFD= 3 #2
21.20	18.00	16.69	15.98	15.52	15.21	99.0% DFD= 4 #1
14.98	14.80	14.66	14.55	14.37	14.20	99.0% DFD= 4 #2
16.26	13.27	12.06	11.39	10.97	10.67	99.0% DFD= 5 #1
10.46	10.29	10.16	10.05	9.89	9.72	99.0% DFD= 5 #2
13.75	10.92	9.78	9.15	8.75	8.47	99.0% DFD= 6 #1
8.26	8.10	7.98	7.87	7.72	7.56	99.0% DFD= 6 #2
12.25	9.55	8.45	7.85	7.46	7.19	99.0% DFD= 7 #1
6.99	6.84	6.72	6.62	6.47	6.31	99.0% DFD= 7 #2
11.26	8.65	7.59	7.01	6.63	6.37	99.0% DFD= 8 #1
6.18	6.03	5.91	5.81	5.67	5.52	99.0% DFD= 8 #2
10.56	8.02	6.99	6.42	6.06	5.80	99.0% DFD= 9 #1
5.61	5.47	5.35	5.26	5.11	4.96	99.0% DFD= 9 #2
10.04	7.56	6.55	5.99	5.64	5.39	99.0% DFD= 10 #1
5.20	5.06	4.94	4.85	4.71	4.56	99.0% DFD= 10 #2
9.65	7.21	6.22	5.67	5.32	5.07	99.0% DFD= 11 #1
4.89	4.74	4.63	4.54	4.40	4.25	99.0% DFD= 11 #2
9.33	6.93	5.95	5.41	5.06	4.82	99.0% DFD= 12 #1
4.64	4.50	4.39	4.30	4.16	4.01	99.0% DFD= 12 #2
9.07	6.70	5.74	5.21	4.86	4.62	99.0% DFD= 13 #1
4.44	4.30	4.19	4.10	3.96	3.82	99.0% DFD= 13 #2
8.86	6.51	5.56	5.04	4.69	4.46	99.0% DFD= 14 #1
4.28	4.14	4.03	3.94	3.80	3.66	99.0% DFD= 14 #2
8.68	6.36	5.42	4.89	4.56	4.32	99.0% DFD= 15 #1
4.14	4.00	3.89	3.80	3.67	3.52	99.0% DFD= 15 #2
8.53	6.23	5.29	4.77	4.44	4.20	99.0% DFD= 16 #1
4.03	3.89	3.78	3.69	3.55	3.41	99.0% DFD= 16 #2

8.40	6.11	5.18	4.67	4.34	4.10	99.0% DFD= 17 #1
3.93	3.79	3.68	3.59	3.46	3.31	99.0% DFD= 17 #2
8.29	6.01	5.09	4.58	4.25	4.01	99.0% DFD= 18 #1
3.84	3.71	3.60	3.51	3.37	3.23	99.0% DFD= 18 #2
8.18	5.93	5.01	4.50	4.17	3.94	99.0% DFD= 19 #1
3.77	3.63	3.52	3.43	3.30	3.15	99.0% DFD= 19 #2
8.10	5.85	4.94	4.43	4.10	3.87	99.0% DFD= 20 #1
3.70	3.56	3.46	3.37	3.23	3.09	99.0% DFD= 20 #2
8.02	5.78	4.87	4.37	4.04	3.81	99.0% DFD= 21 #1
3.64	3.51	3.40	3.31	3.17	3.03	99.0% DFD= 21 #2
7.95	5.72	4.82	4.31	3.99	3.76	99.0% DFD= 22 #1
3.59	3.45	3.35	3.26	3.12	2.98	99.0% DFD= 22 #2
7.88	5.66	4.76	4.26	3.94	3.71	99.0% DFD= 23 #1
3.54	3.41	3.30	3.21	3.07	2.93	99.0% DFD= 23 #2
7.82	5.61	4.72	4.22	3.90	3.67	99.0% DFD= 24 #1
3.50	3.36	3.26	3.17	3.03	2.89	99.0% DFD= 24 #2
7.77	5.57	4.68	4.18	3.85	3.63	99.0% DFD= 25 #1
3.46	3.32	3.22	3.13	2.99	2.85	99.0% DFD= 25 #2
7.72	5.53	4.64	4.14	3.82	3.59	99.0% DFD= 26 #1
3.42	3.29	3.18	3.09	2.96	2.81	99.0% DFD= 26 #2
7.68	5.49	4.60	4.11	3.78	3.56	99.0% DFD= 27 #1
3.39	3.26	3.15	3.06	2.93	2.78	99.0% DFD= 27 #2
7.64	5.45	4.57	4.07	3.75	3.53	99.0% DFD= 28 #1
3.36	3.23	3.12	3.03	2.90	2.75	99.0% DFD= 28 #2
7.60	5.42	4.54	4.04	3.73	3.50	99.0% DFD= 29 #1
3.33	3.20	3.09	3.00	2.87	2.73	99.0% DFD= 29 #2
7.56	5.39	4.51	4.02	3.70	3.47	99.0% DFD= 30 #1
3.30	3.17	3.07	2.98	2.84	2.70	99.0% DFD= 30 #2
7.31	5.18	4.31	3.83	3.51	3.29	99.0% DFD= 40 #1
3.12	2.99	2.89	2.80	2.66	2.52	99.0% DFD= 40 #2
7.08	4.98	4.13	3.65	3.34	3.12	99.0% DFD= 60 #1
2.95	2.82	2.72	2.63	2.50	2.35	99.0% DFD= 60 #2
6.85	4.79	3.95	3.48	3.17	2.96	99.0% DFD=120 #1
2.79	2.66	2.56	2.47	2.34	2.19	99.0% DFD=120 #2
6.63	4.61	3.78	3.32	3.02	2.80	99.0% DFD=INF #1
2.64	2.51	2.41	2.32	2.18	2.04	99.0% DFD=INF #2
16211.	20000.	21615.	22500.	23056.	23437.	99.5% DFD= 1 #1
23715.	23925.	24091.	24224.	24426.	24630.	99.5% DFD= 1 #2
198.5	199.0	199.2	199.2	199.3	199.3	99.5% DFD= 2 #1
199.4	199.4	199.4	199.4	199.4	199.4	99.5% DFD= 2 #2
55.55	49.80	47.47	46.19	45.39	44.84	99.5% DFD= 3 #1
44.43	44.13	43.88	43.69	43.39	43.08	99.5% DFD= 3 #2
31.33	26.28	24.26	23.15	22.46	21.97	99.5% DFD= 4 #1
21.62	21.35	21.14	20.97	20.70	20.44	99.5% DFD= 4 #2
22.78	18.31	16.53	15.56	14.94	14.51	99.5% DFD= 5 #1
14.20	13.96	13.77	13.62	13.38	13.15	99.5% DFD= 5 #2
18.63	14.54	12.92	12.03	11.46	11.07	99.5% DFD= 6 #1
10.79	10.57	10.39	10.25	10.03	9.81	99.5% DFD= 6 #2
16.24	12.40	10.88	10.05	9.52	9.16	99.5% DFD= 7 #1
8.89	8.68	8.51	8.38	8.18	7.97	99.5% DFD= 7 #2
14.69	11.04	9.60	8.81	8.30	7.95	99.5% DFD= 8 #1
7.69	7.50	7.34	7.21	7.01	6.81	99.5% DFD= 8 #2
13.61	10.11	8.72	7.96	7.47	7.13	99.5% DFD= 9 #1
6.88	6.69	6.54	6.42	6.23	6.03	99.5% DFD= 9 #2
12.83	9.43	8.08	7.34	6.87	6.54	99.5% DFD= 10 #1
6.30	6.12	5.97	5.85	5.66	5.47	99.5% DFD= 10 #2
12.23	8.91	7.60	6.88	6.42	6.10	99.5% DFD= 11 #1
5.86	5.68	5.54	5.42	5.24	5.05	99.5% DFD= 11 #2
11.75	8.51	7.23	6.52	6.07	5.76	99.5% DFD= 12 #1
5.52	5.35	5.20	5.09	4.91	4.72	99.5% DFD= 12 #2
11.37	8.19	6.93	6.23	5.79	5.48	99.5% DFD= 13 #1
5.25	5.08	4.94	4.82	4.64	4.46	99.5% DFD= 13 #2

11.06	7.92	6.68	6.00	5.56	5.26	99.5% DFD= 14 #1
5.03	4.86	4.72	4.60	4.43	4.25	99.5% DFD= 14 #2
10.80	7.70	6.48	5.80	5.37	5.07	99.5% DFD= 15 #1
4.85	4.67	4.54	4.42	4.25	4.07	99.5% DFD= 15 #2
10.58	7.51	6.30	5.64	5.21	4.91	99.5% DFD= 16 #1
4.69	4.52	4.38	4.27	4.10	3.92	99.5% DFD= 16 #2
10.38	7.35	6.16	5.50	5.07	4.78	99.5% DFD= 17 #1
4.56	4.39	4.25	4.14	3.97	3.79	99.5% DFD= 17 #2
10.22	7.21	6.03	5.37	4.96	4.66	99.5% DFD= 18 #1
4.44	4.28	4.14	4.03	3.86	3.68	99.5% DFD= 18 #2
10.07	7.09	5.92	5.27	4.85	4.56	99.5% DFD= 19 #1
4.34	4.18	4.04	3.93	3.76	3.59	99.5% DFD= 19 #2
9.94	6.99	5.82	5.17	4.76	4.47	99.5% DFD= 20 #1
4.26	4.09	3.96	3.85	3.68	3.50	99.5% DFD= 20 #2
9.83	6.89	5.73	5.09	4.68	4.39	99.5% DFD= 21 #1
4.18	4.01	3.88	3.77	3.60	3.43	99.5% DFD= 21 #2
9.73	6.81	5.65	5.02	4.61	4.32	99.5% DFD= 22 #1
4.11	3.94	3.81	3.70	3.54	3.36	99.5% DFD= 22 #2
9.63	6.73	5.58	4.95	4.54	4.26	99.5% DFD= 23 #1
4.05	3.88	3.75	3.64	3.47	3.30	99.5% DFD= 23 #2
9.55	6.66	5.52	4.89	4.49	4.20	99.5% DFD= 24 #1
3.99	3.83	3.69	3.59	3.42	3.25	99.5% DFD= 24 #2
9.48	6.60	5.46	4.84	4.43	4.15	99.5% DFD= 25 #1
3.94	3.78	3.64	3.54	3.37	3.20	99.5% DFD= 25 #2
9.41	6.54	5.41	4.79	4.38	4.10	99.5% DFD= 26 #1
3.89	3.73	3.60	3.49	3.33	3.15	99.5% DFD= 26 #2
9.34	6.49	5.36	4.74	4.34	4.06	99.5% DFD= 27 #1
3.85	3.69	3.56	3.45	3.28	3.11	99.5% DFD= 27 #2
9.28	6.44	5.32	4.70	4.30	4.02	99.5% DFD= 28 #1
3.81	3.65	3.52	3.41	3.25	3.07	99.5% DFD= 28 #2
9.23	6.40	5.28	4.66	4.26	3.98	99.5% DFD= 29 #1
3.77	3.61	3.48	3.38	3.21	3.04	99.5% DFD= 29 #2
9.18	6.35	5.24	4.62	4.23	3.95	99.5% DFD= 30 #1
3.74	3.58	3.45	3.34	3.18	3.01	99.5% DFD= 30 #2
8.83	6.07	4.98	4.37	3.99	3.71	99.5% DFD= 40 #1
3.51	3.35	3.22	3.12	2.95	2.78	99.5% DFD= 40 #2
8.49	5.79	4.73	4.14	3.76	3.49	99.5% DFD= 60 #1
3.29	3.13	3.01	2.90	2.74	2.57	99.5% DFD= 60 #2
8.18	5.54	4.56	3.92	3.55	3.28	99.5% DFD=120 #1
3.09	2.93	2.81	2.71	2.54	2.37	99.5% DFD=120 #2
7.88	5.30	4.28	3.72	3.35	3.09	99.5% DFD=INF #1
2.90	2.74	2.62	2.52	2.36	2.19	99.5% DFD=INF #2
405300.	500000.	540400.	562500.	576400.	585900.	99.9% DFD= 1 #1
592900.	598100.	602300.	605600.	610700.	615800.	99.9% DFD= 1 #2
998.5	999.0	999.2	999.2	999.3	999.3	99.9% DFD= 2 #1
999.4	999.4	999.4	999.4	999.4	999.4	99.9% DFD= 2 #2
167.0	148.5	141.1	137.1	134.6	132.8	99.9% DFD= 3 #1
131.6	130.6	129.9	129.2	128.3	127.4	99.9% DFD= 3 #2
74.14	61.25	56.18	53.44	51.71	50.53	99.9% DFD= 4 #1
49.66	49.00	48.47	48.05	47.41	46.76	99.9% DFD= 4 #2
47.18	37.12	33.20	31.09	29.75	28.84	99.9% DFD= 5 #1
28.16	27.64	27.24	26.92	26.42	25.91	99.9% DFD= 5 #2
35.51	27.00	23.70	21.92	20.81	20.03	99.9% DFD= 6 #1
19.46	19.03	18.69	18.41	17.99	17.56	99.9% DFD= 6 #2
29.25	21.69	18.77	17.19	16.21	15.52	99.9% DFD= 7 #1
15.02	14.63	14.33	14.08	13.71	13.32	99.9% DFD= 7 #2
25.42	18.49	15.83	14.39	13.49	12.86	99.9% DFD= 8 #1
12.40	12.04	11.77	11.54	11.19	10.84	99.9% DFD= 8 #2
22.86	16.39	13.90	12.56	11.71	11.13	99.9% DFD= 9 #1
10.70	10.37	10.11	9.89	9.57	9.24	99.9% DFD= 9 #2
21.04	14.91	12.55	11.28	10.48	9.92	99.9% DFD= 10 #1
9.52	9.20	8.96	8.75	8.45	8.13	99.9% DFD= 10 #2

19.69	13.81	11.56	10.35	9.58	9.05	99.9% DFD= 11 #1
8.66	8.35	8.12	7.92	7.63	7.32	99.9% DFD= 11 #2
18.64	12.97	10.80	9.63	8.89	8.38	99.9% DFD= 12 #1
8.00	7.71	7.48	7.29	7.00	6.71	99.9% DFD= 12 #2
17.81	12.31	10.21	9.07	8.35	7.86	99.9% DFD= 13 #1
7.49	7.21	6.98	6.80	6.52	6.23	99.9% DFD= 13 #2
17.14	11.78	9.73	8.62	7.92	7.43	99.9% DFD= 14 #1
7.08	6.80	6.58	6.40	6.13	5.85	99.9% DFD= 14 #2
16.59	11.34	9.34	8.25	7.57	7.09	99.9% DFD= 15 #1
6.74	6.47	6.26	6.08	5.81	5.54	99.9% DFD= 15 #2
16.12	10.97	9.00	7.94	7.27	6.81	99.9% DFD= 16 #1
6.46	6.19	5.98	5.81	5.55	5.27	99.9% DFD= 16 #2
15.72	10.66	8.73	7.68	7.02	6.56	99.9% DFD= 17 #1
6.22	5.96	5.75	5.58	5.32	5.05	99.9% DFD= 17 #2
15.38	10.39	8.49	7.46	6.81	6.35	99.9% DFD= 18 #1
6.02	5.76	5.56	5.39	5.13	4.87	99.9% DFD= 18 #2
15.08	10.16	8.28	7.26	6.62	6.18	99.9% DFD= 19 #1
5.85	5.59	5.39	5.22	4.97	4.70	99.9% DFD= 19 #2
14.82	9.95	8.10	7.10	6.46	6.02	99.9% DFD= 20 #1
5.69	5.44	5.24	5.08	4.82	4.56	99.9% DFD= 20 #2
14.59	9.77	7.94	6.95	6.32	5.88	99.9% DFD= 21 #1
5.56	5.31	5.11	4.95	4.70	4.44	99.9% DFD= 21 #2
14.38	9.61	7.80	6.81	6.19	5.76	99.9% DFD= 22 #1
5.44	5.19	4.99	4.83	4.58	4.33	99.9% DFD= 22 #2
14.19	9.47	7.67	6.69	6.08	5.65	99.9% DFD= 23 #1
5.33	5.09	4.89	4.73	4.48	4.23	99.9% DFD= 23 #2
14.03	9.34	7.55	6.59	5.98	5.55	99.9% DFD= 24 #1
5.23	4.99	4.80	4.64	4.39	4.14	99.9% DFD= 24 #2
13.88	9.22	7.45	6.49	5.88	5.46	99.9% DFD= 25 #1
5.15	4.91	4.71	4.56	4.31	4.06	99.9% DFD= 25 #2
13.74	9.12	7.36	6.41	5.80	5.38	99.9% DFD= 26 #1
5.07	4.83	4.64	4.48	4.24	3.99	99.9% DFD= 26 #2
13.61	9.02	7.27	6.33	5.73	5.31	99.9% DFD= 27 #1
5.00	4.76	4.57	4.41	4.17	3.92	99.9% DFD= 27 #2
13.50	8.93	7.19	6.25	5.66	5.24	99.9% DFD= 28 #1
4.93	4.69	4.50	4.35	4.11	3.86	99.9% DFD= 28 #2
13.39	8.85	7.12	6.19	5.59	5.18	99.9% DFD= 29 #1
4.87	4.64	4.45	4.29	4.05	3.80	99.9% DFD= 29 #2
13.29	8.77	7.05	6.12	5.53	5.12	99.9% DFD= 30 #1
4.82	4.58	4.39	4.24	4.00	3.75	99.9% DFD= 30 #2
12.61	8.25	6.60	5.70	5.13	4.73	99.9% DFD= 40 #1
4.44	4.21	4.02	3.87	3.64	3.40	99.9% DFD= 40 #2
11.97	7.76	6.17	5.31	4.76	4.37	99.9% DFD= 60 #1
4.09	3.87	3.69	3.54	3.31	3.08	99.9% DFD= 60 #2
11.38	7.32	5.79	4.95	4.42	4.04	99.9% DFD=120 #1
3.77	3.55	3.38	3.24	3.02	2.78	99.9% DFD=120 #2
10.83	6.91	5.42	4.62	4.10	3.74	99.9% DFD=INF #1
3.47	3.27	3.10	2.96	2.74	2.51	99.9% DFD=INF #2

PROGRAM PUNCHWC2.

```
C
C
C PROGRAM PUNCHWC2
C
C
C A SHORT PROGRAM TO PUNCH THE REGRESSION WEIGHT CARDS FOR LEAST SQUARES
C LINEAR REGRESSION ANALYSIS FOR ALL POSSIBLE COMBINATIONS OF 1 TO 10
C INDEPENDENT VARIABLES.
C PROGRAM ALSO LISTS WHAT IS PUNCHED.
C
C PROGRAM WRITTEN BY: STEVE DIETRICH
C                     DEPARTMENT OF PHARMACEUTICAL CHEMISTRY
C                     SCHOOL OF PHARMACY
C                     UNIVERSITY OF CALIFORNIA
C                     SAN FRANCISCO, CALIFORNIA 94143
C                     APRIL, 1976
C
C PROGRAM WRITTEN, DEBUGGED, AND USED IN WATFIV ON AN IBM MODEL 370
C 145 COMPUTER AT UCSF
C
C PROGRAM FIRST PUNCHES THE ONE POSSIBLE COMBINATION FOR 1 VARIABLE;
C THEN ALL POSSIBLE COMBINATIONS OF FROM 1 TO 2 VARIABLES;
C THEN ALL POSSIBLE COMBINATIONS OF FROM 1 TO 3 VARIABLES;
C AND SO ON UP TO ALL POSSIBLE COMBINATIONS OF FROM 1 TO 10 VARIABLES.
C
C FOR ALL POSSIBLE COMBINATIONS OF FROM 1 TO X VARIABLES, THE PROGRAM
C FIRST PUNCHES ALL POSSIBLE 1 VARIABLE COMBINATIONS OF THE X VARIABLES;
C THEN ALL POSSIBLE 2 VARIABLE COMBINATIONS OF THE X VARIABLES;
C AND SO ON UP TO THE ONE POSSIBLE X VARIABLE COMBINATION OF THE X
C VARIABLES
C
C
      DIMENSION I(10)
   91 FORMAT (1X, 10I1, 63X, I2, 1H-, I4)
   92 FORMAT (    1011, 63X, I2, 1H-, I4)
      DO 10 II =1,10
   10 I(II) = 0
      DO 200 N=1,10
      NCARDS = 0
      DO 111 NN=1,N
      DO 110 J1=1,N
      IF (J1 .GT. N+1-NN) GO TO 111
      DO 31 II=1,N
   31 I(II)=0
      I(J1 )=1
      IF (NN .NE. 1) GO TO 11
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 110
   11 K2 = J1 + 1
      DO 109 J2=K2,N
      IF (J2 .GT. N+2-NN) GO TO 110
      DO 32 II=K2,N
   32 I(II)=0
      I(J2 )=1
```

```
      IF (NN .NE. 2) GO TO 12
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 109
   12 K3 = J2 + 1
      DO 108 J3=K3,N
      IF (J3 .GT. N+3-NN) GO TO 109
      DO 33 II=K3,N
   33 I(II)=0
      I(J3 )=1
      IF (NN .NE. 3) GO TO 13
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 108
   13 K4 = J3 + 1
      DO 107 J4=K4,N
      IF (J4 .GT. N+4-NN) GO TO 108
      DO 34 II=K4,N
   34 I(II)=0
      I(J4 )=1
      IF (NN .NE. 4) GO TO 14
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 107
   14 K5 = J4 + 1
      DO 106 J5=K5,N
      IF (J5 .GT. N+5-NN) GO TO 107
      DO 35 II=K5,N
   35 I(II)=0
      I(J5 )=1
      IF (NN .NE. 5) GO TO 15
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 106
   15 K6 = J5 + 1
      DO 105 J6=K6,N
      IF (J6 .GT. N+6-NN) GO TO 106
      DO 36 II=K6,N
   36 I(II)=0
      I(J6 )=1
      IF (NN .NE. 6) GO TO 16
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 105
   16 K7 = J6 + 1
      DO 104 J7=K7,N
      IF (J7 .GT. N+7-NN) GO TO 105
      DO 37 II=K7,N
   37 I(II)=0
      I(J7 )=1
      IF (NN .NE. 7) GO TO 17
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 104
   17 K8 = J7 + 1
      DO 103 J8=K8,N
```

```
      IF (J8 .GT. N+8-NN) GO TO 104
      DO 38 II=K8,N
   38 I(II)=0
      I(J8 )=1
      IF (NN .NE. 8) GO TO 18
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 103
   18 K9 = J8 + 1
      DO 102 J9=K9,N
      IF (J9 .GT. N+9-NN) GO TO 103
      DO 39 II=K9,N
   39 I(II)=0
      I(J9 )=1
      IF (NN .NE. 9) GO TO 19
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
      GO TO 102
   19 I(10) = 1
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
      WRITE (7,92) I, N, NCARDS
  102 CONTINUE
  103 CONTINUE
  104 CONTINUE
  105 CONTINUE
  106 CONTINUE
  107 CONTINUE
  108 CONTINUE
  109 CONTINUE
  110 CONTINUE
  111 CONTINUE
      DO 140 M=1,10
  140 I(M) = 0
      L = 3
      GO TO (150,160,150,160,150,160,150,160,150,160), N
  150 L = 5
  160 DO 170 LL = 1,L
      NCARDS = NCARDS + 1
      WRITE (6,91) I, N, NCARDS
  170 WRITE (7,92) I, N, NCARDS
  200 CONTINUE
      STOP
      END
```

PROGRAM FILLWCF.

```
C
C
C PROGRAM FILLWCF
C
C
C PROGRAM FILLWCF READS REGRESSION WEIGHT CARDS FOR LEAST SQUARES
C LINEAR REGRESSION ANALYSES FOR ANY COMBINATION OF 1 TO 10 INDEPENDENT
C VARIABLES. THE WEIGHT CARDS HAVING BEEN PREVIOUSLY PUNCHED BY
C THE PROGRAM PUNCHWC2.
C
C PROGRAM CALCULATES THE NUMBER OF INDEPENDENT VARIABLES SPECIFIED BY
C EACH WEIGHT CARD FOR REGRESSION (NIVWC).
C
C VALUES OF 6 WEIGHT CARDS (FILEWC(10,J), J=1,6) AND THE NUMBER OF
C INDEPENDENT VARIABLES INDICATED FOR REGRESSION BY EACH (NIVWC(J),
C J=1,6) ARE WRITTEN INTO EACH RECORD OF THE ON-LINE DISK DIRECT ACCESS
C FILE WCFILE.
C
C THE ON-LINE DISK DIRECT ACCESS FILE WCFILE CONTAINS 346 FORMATTED
C (6011, 612) RECORDS EACH WITH A LENGTH OF 72 BYTES.
C
C PROGRAM WRITTEN BY: STEVE DIETRICH
C                     DEPARTMENT OF PHARMACEUTICAL CHEMISTRY
C                     SCHOOL OF PHARMACY
C                     UNIVERSITY OF CALIFORNIA
C                     SAN FRANCISCO, CALIFORNIA 94143
C                     APRIL, 1976
C
C PROGRAM WRITTEN, DEBUGGED, AND USED IN WATFIV ON AN IBM MODEL 370
C 145 COMPUTER AT UCSF
C
C
      INTEGER*2 FILEWC(10,6), NIVWC(6)
      DEFINE FILE 13 (346,72,E,JCOUNT)
      IWCF = 13
      IR = 5
      JCOUNT = 1
   50 READ (IR,100,END=900) FILEWC
  100 FORMAT (1011)
      DO 120 J=1,6
      NIVWC(J) = 0
      DO 110 K=1,10
      IF (FILEWC(K,J) .EQ. 1) NIVWC(J) = NIVWC(J) + 1
  110 CONTINUE
  120 CONTINUE
      WRITE (IWCF'JCOUNT,150) FILEWC, NIVWC
  150 FORMAT (6011, 612)
      GO TO 50
  900 CONTINUE
      STOP
      END
```