CELL-WALL POLYSACCHARIDES FROM WHITE MUSTARD

by

RAYMOND J. SKERRETT, B.Sc.

A thesis presented for the degree of

Doctor of Philosophy

University of Edinburgh

November 1968

To my mother

PREFACE

I would like to express my thanks to Dr. D. A. Rees for his constant advice and encouragement, and necessary, if often destructive, criticism, during the course of this work. I am also grateful to Professor Sir Edmund Hirst, C.B.E., F.R.S., for the provision of research facilities, and to the Institute of Paper Chemistry, Appleton, Wisconsin, U.S.A., for the provision of a grant under their Pioneering Research Program.

Part of the work in Section III has been published in collaboration with Dr. Rees, and a reprint has been inserted at the end of the thesis.

TABLE OF CONTENTS

GENERAL INTRODUCTION

It is well known that all forms of life, with the exception of viruses, consist of one or more basic structural units called cells. These cells are capable of the fundamental processes of life such as feeding, growing, and reproducing and have, in the case of multicellular organisms, a degree of autonomy from their nearest neighbours. The general structures of cells are similar, whatever the organism of origin, consisting of one or more nuclei, cytoplasm, and a cell wall, or cell membrane.

This thesis is concerned with the cell walls of vascular plants. One of the distinguishing features of plants is their possession of cell walls consisting of certain types of polysaccharides. In the higher plants, cellulose is nearly always the main polysaccharide although, in a few specialised cells, xylans (1a), mannans (2), galactomannans (1b) or pectic substances (1c), may be more abundant. In most fungi, and in some other primitive types, cellulose is replaced by chitin, its 2-acetamido 2-deoxy analogue (1d). The main components of bacterial cell walls are called peptidoglycans or mureins, which differ from chitin in that every other N-acetyl glucosamine residue is replaced by muramic acid, its 3-O-D-lactyl ether; and the polysaccharide chains are cross-linked by oligopeptide chains, attached by peptide links to the muramic acid residues (3).

Although the cell wall may appear to play only a passive role in the life of the cell, it is nonetheless of prime importance, and its properties need to be a balanced compromise between the requirements for different functions. Thus, it needs to be impermeable enough to maintain the integrity of the cell; and yet permeable enough to allow the passage of food, secretions,

oxygen, waste, etc. Similarly, it needs to be flexible enough, at least at some times, to allow growth; but strong and rigid enough to provide some degree of support to the cytoplasm.

It would obviously be of interest to be able to explain these properties of cell walls, and, to do this, it will be necessary to know not only the chemical (i.e.primary) structures of the components, but also their shapes, and their arrangement with respect to neighbouring molecules (i.e. their secondary, tertiary, and quaternary structures), and also the nature and magnitude of interactions within and between them.

The carbohydrates of the cell walls of vascular plants were originally classified on the basis of their solubilities; the hot water soluble fraction was called pectin, the hot alkali soluble fraction was called hemicellulose and the residue was called cellulose. It is now possible to give modified, and more precise, definitions of these fractions on a structural basis.

The pectic fraction consists of (a) polysaccharides with a main-chain consisting partially or wholly of α 1 , 4 linked D-galactopyranuronosyl residues, or their methyl esters and (b) galactans and arabans.

Pectin is regarded as, ideally, a linear chain consisting entirely of galacturonosyl residues, partially or wholly methyl-esterified. This ideal structure has only been found, however, in sunflower seed heads (4). In all other cases, various other sugars, especially D-galactose, L-arabinose, and L-rhamnose, occur as part of the main-chain, or of side-chains. The solubility of pectin decreases with decreasing esterification, and largely unesterified pectins may form insoluble salts with

divalent ions, chiefly calcium.

The galactans and arabans form groups which are biologically associated with pectin, and may be glycosidically linked to it. Lupin galactan (5) is unbranched, with β 1→4 links, but larch galactan (6), which is regarded as a hemicellulose, is branched, with a β 1→3 linked main-chain, and with 1→6 linked side-chains of one or more 1→6 linked galactose residues, or 1→3 linked L-arabinose residues. The commonest general structure for arabans appears to be a 5-O-α-L-arabinofuranosyl 1- main-chain, with some branching at C_3 (1h,8). Arabogalactans are also known.

The gum exudates from species of Acacia and Prunus are highly-branched galactans with complex side-chains. They have main-chains of β 1→3 linked (in Acacia) or β 1→6 linked (in Prunus) D-galactose residues with side chains containing D-glucuronic acid and various neutral sugars (7).

Tragacanthic acid, the acidic component of gum tragacanth, probably has a similar structure, with 1→4 linked D-galacturonosyl residues, and with side-chains of D-xylose and L-fucose (7a).

Another group of polysaccharides, comprising several plant mucilages and gums, is related to pectin, but has a main chain consisting of →4-O-α-D-galactopyranuronosyl 1→2-O-α-L-rhamnopyranosyl 1- residues, with various side-chains. This type of structure is found in Khaya gum (7b) and mucilages from okra, flax, white mustard, cress, quince, Plantago and slippery elm, amongst others (7c).

The hemicelluloses are a group of polysaccharides which usually have a β 1→4 linked main chain of D-glucose, D-mannose, or D-xylopyranose residues. In this, they structurally resemble cellulose, but usually differ in having side-chains. Some

arabans and galactans are also hemicellulosic.

The hemicelluloses with xylan main chain are an important group, particularly in wood, and in the harder parts of monocotyledons, e.g. straw and esparto. Common side-chains to the →4-O-β-D-xylopyranosyl 1- backbone are L-arabino-furanosyl and 4-O-methyl-D-glucopyranuronosyl. D-mannopyranosyl and D-glucopyranuronosyl also occur (9).

Other hemicelluloses are the glucomannans of softwood (9), the galactomannans of the Leguminosae (1g), which have D-galactose linked 1→6 to some of the mannose residues, and the amyloids (1j), which have various sugars, especially xylose, linked 1→6 to a cellulose backbone.

The third fraction of cell wall polysaccharide, cellulose, differs from the others in being a single, apparently well-defined species, $(\rightarrow 4\text{-O-}\beta\text{-D-glucopyranosyl 1-})_n$.

In addition to these polysaccharide components of the cell wall, other components are sometimes important, of which lignin is the chief. This is a condensation product of coniferyl alcohol, 1-hydroxymethyl 2-(m-methoxy-p-hydroxyphenyl) ethylene, or related compounds. Lignin is relatively impermeable to solutions, and hence causes the death of the cell when laid down in quantity in the cell wall. It occurs only in mature cells.

It has been found (1) that lignin can be more easily extracted after hemicelluloses have been removed chemically, and vice-versa; and it has been suggested, from this, that there are covalent links between them, but it could well be that this effect is due to non-bonded interactions.

Other substances which occur occasionally are tannins, and high molecular weight fatty acids and their esters, as cutin, suberin, etc. (1).

Cellulose normally occurs in the cell wall as microfibrils, which are threadlike or ribbon-shaped aggregates of molecules with a width of 50-200Å, and lengths of up to 10μ. These contain crystalline cellulose, as is shown by the X-ray diffraction photographs of cell walls (1k).

Hemicelluloses are intimately associated with the microfibrils, and may at times be included in them, since cellulose is sometimes difficult to obtain free from sugars other than glucose. With the possible exception of ivory-nut mannan, the hemicelluloses are non-crystalline and non-fibrillar in the cell wall (although several can be crystallised after extraction), but some show evidence of orientation.

Pectins occur with the other polysaccharides, and are more abundant in some tissues, such as collenchyma, and parts of fruit and roots, than in others. The highly-branched pectic materials must be amorphous and the less-branched pectins may also be often amorphous, but the birefringence and X-ray diffraction properties of cell walls from which non-pectic materials had been removed, showed that pectin is, at least sometimes, crystalline In this condition, it probably occurs as short microfibrils (10).

When a cell is dividing to produce two daughter cells, the first stage in the production of the intervening walls is the coalescence of many small particles, known as Golgi vesicles, to produce a disc-shaped barrier between the daughter nuclei, known as the cell-plate (11). Cellulose is rapidly formed on either side of the cell-plate and this marks the beginning of the cell walls proper. The cell-plate remains between the two walls as the middle lamella, and contains largely pectic material. The cell wall which is produced first is known as the primary wall, which expands as the cell grows. Although the

composition varies, the primary wall is generally characterised by a relatively low cellulose content, and an appreciable pectin content. When the cell is full-grown, an inner wall is formed, which may consist of several layers and is known as the secondary wall. In this case, too, the composition is variable, but it is generally characterised by a high cellulose, and low pectin content, and by the presence of lignin. The secondary wall itself can be sub-divided into three concentric layers, S_1, S_2, and S_3, the latter innermost. These differ in the orientation of the cellulose microfibrils with respect to the long axis of the cell. In S_1, the microfibrils form two sets of helices, in opposite senses, at a large angle to the axis. In S_2, the microfibrils are aligned at a small angle to the axis. In S_3, there is one set of helices at a large angle to the axis (11).

The interactions between and within the major polysaccharide components of the cell wall are obviously likely to have a determinative effect on the properties of the cell wall. It would be desirable to study these interactions in this system as it occurs <u>in vivo</u>, but the presence of cell contents, and the anatomical structure of the plant, are likely to make chemical and physical investigation more difficult, and more likely to give misleading results. On the other hand, extraction of the components is almost certain to change the physical, if not the chemical, nature of the system. Fortunately, however, there exist a number of mucilages, which are derived from the cell walls of seed-coats, and which can be extracted with cold water (7d). These, on acid hydrolysis, give an insoluble precipitate which proves to be a glucan; whilst examination by electron microscope shows the presence of microfibrils. It, therefore, appears that these mucilages are cellulose microfibrils solubilised, in some

way, by association with hemicellulosic or pectic polysaccharides, and it seems probable that these mucilages resemble, chemically and physically, the situation which occurs in normal cell walls.

These mucilages have been reported from the seeds of flax, white mustard, cress, and quince (7d). This is why this thesis is concerned largely with polysaccharides from the seed coats of white mustard, Sinapis alba.L. Two cellulose-containing mucilages have been reported from the vegetative parts of plants; noriasa mucilage (7d) from the roots of Abelmoschus glutiotextilis, and gum tragacanth (12), from species of Astragalus.

The work described in this thesis aims, first of all, to confirm the identification of the microfibrils as cellulose, and to discover something of the composition of the mucilage. The cellulose is confirmed from its infrared spectrum and its X-ray diffraction pattern, and by enzymolysis. The mucilage is examined for heterogeneity by means of ultracentrifugation and free-solution electrophoresis, and separated into distinct fractions for which the cellulose contents and the monosaccharide components are estimated.

Secondly, the nature of the association between cellulose and the other polysaccharides is investigated. This could be either covalent, as glycosidic or as ester links; or non-covalent, as hydrogen bonding, as hydrophobic bonding, or as chelation or some other form of ionic interaction involving the ions associated with uronic acid residues. It is possible to distinguish between these by their reactions with various reagents. Infrared spectroscopy is also used to investigate the inaccessibility of hydroxyl groups in cellulose and the mucilage to reaction with deuterium oxide vapour, and the orientation of hydrogen bonds in cellulose.

A third line of investigation is a more theoretical one, which involves the calculation of inter-residue non-covalent interactions. The simplest and earliest method of doing this was to build models of the molecule from spheres of appropriate radii, or rods of appropriate lengths, and to find, by adjustment of the parts, the conformation which involves little or no strain due to overlap of the atoms. The outstanding success of this approach was in the elucidation of the double-helical structure of DNA (13); but it has also been used to justify a modification of the Meyer and Misch structure for cellulose (14). The limitation of this method is its representation of van der Waals interactions as equivalent to hard, spherical atoms, which is clearly not strictly true; and its failure to give quantitative estimates of the total interaction. The method which attempts to overcome this limitation is the geometrical construction of given conformations in a molecule, and the calculation of inter-atomic distances, and the corresponding interaction. This is done by means of a computer, which can perform these calculations for every conformation generated by, for example, systematic rotations about covalent bonds. Most of the work described below is concerned with cellulose, but it is intended not only to provide information about cellulose, but also to develop methods of dealing in this way with polysaccharides generally. The use of this approach will, it is hoped, provide information regarding the conformation of polymers and also give some quantitative estimate of the importance of the various types of interaction within, and hence between, polysaccharide molecules.

GENERAL METHODS OF INVESTIGATION

A. Purification and isolation of polysaccharides.

Dialysis. Solutions were dialysed in cellophane tubes against tap-water. Chloroform was added as a bacteriostatic agent.

Concentration of solutions. Solutions were concentrated under reduced pressure, at about 40°C, using a rotary film evaporator.

Isolation of polysaccharides from solution. Polysaccharides were isolated by freeze-drying, or by precipitation from aqueous solution with ethanol or acetone. In the latter case, the precipitate was separated by centrifugation and washed with acetone, and the process was repeated until the precipitate had a feathery texture, when it was dried overnight in a vacuum oven at 50-60°C, in the presence of phosphoric oxide.

B. Methods of hydrolysis.

Monier-Williams degradation (15). The polysaccharide was shaken with 72% sulphuric acid until dissolved, and the solution was diluted to 1-2 N. It was then treated as for the following hydrolysis.

Dilute sulphuric acid hydrolysis. The polysaccharide was dissolved or suspended in 1-2 N sulphuric acid, and heated at 100°C for 4 hours, or overnight, as required. The solution was then diluted, neutralised with calcium carbonate, and filtered. It was then evaporated to the required concentration.

Formic acid hydrolysis. The polysaccharide was dissolved in 90% formic acid, and the solution was diluted with an equal volume of water, and heated at 100°C overnight. It was then evaporated almost to dryness, diluted, and re-evaporated, in

order to hydrolyse any formyl esters. It was then diluted to a suitable concentration.

C. Estimation of sugars.

Phenol/sulphuric acid estimation (for all carbohydrates) (16). Sugar solution (1 ml.) was mixed with 1 ml. of 5% w/v aqueous phenol solution, and AnalaR concentrated sulphuric acid (5 ml.) was added quickly. When cool, the optical density was measured by means of an EEL colorimeter using Filter 623. When necessary, the absolute concentration was obtained from a calibration curve.

Aniline phthalate estimation (for neutral sugars) (17). o-Phthalic acid (1.66 gm.) and aniline (re-distilled from zinc) (0.91 ml.) were dissolved in 48 ml. 1-butanol, 48 ml. diethyl ether, and 4 ml. water.

The unknown sugar mixture was measured on to chromatography paper by means of a micropipette. Standard mixtures were measured on to the same paper, and the chromatogram was developed. The paper was then dipped in the aniline phthalate solution, air-dried, then oven-dried at 105°C for 10 minutes, to develop the spots. The spots were cut out, taking equal areas for the standard and unknown spots of each sugar, and eluted with 0.7N hydrochloric acid in 80% (v/v) ethanol (10 ml.), for one hour. Optical densities were then measured, using 1 cm. cells in a Unicam SP600 spectrometer, at 390 mμ for hexoses and 6-deoxyhexoses, and at 360 mμ for pentoses. By reference to calibration curves, obtained from the optical densities of the standard sugars, the concentrations of the sugars in the mixture can be estimated.

Carbazole estimation (for uronic acids) (18). 12 ml. Aliquots of AnalaR concentrated sulphuric acid were measured

into boiling tubes and cooled to 3°C or less. The solutions (not more than 1 ml.) containing the samples (standard and unknown) were added to these dropwise, and the mixtures were cooled again to 3°C. They were then immersed in boiling water for ten minutes, and allowed to cool to room temperature. 0.15% Ethanolic carbazole solution (1 ml.) was added to each tube, and the tubes were shaken and allowed to stand at room temperature for 25 minutes. The optical densities of the solutions was then measured at 520 mμ by means of a Unicam SP 600 spectrometer, and the concentrations of the unknown solutions were calculated from the calibration curve given by the standard solutions.

D. Paper Chromatography.

Paper chromatograms were developed overnight, on Whatman No. 1 or No. 4 paper. The solvent systems used were:

(a) ethyl acetate/pyridine/water in the ratio 10:4:3;

(b) ethyl acetate/acetic acid/formic acid/water in the ratio 18:3:1:4;

(c) 1-propanol/ethyl acetate/water in the ratio 7:1:2;

(d) 1-butanol/ethanol/water (upper layer) in the ratio 4:1:5.

E. Paper electrophoresis.

Paper electrophoretograms were run on Whatman No. 1 paper using an acetic acid/pyridine buffer with a pH of approximately 6.2, and a potential gradient of 14 volts/cm. Glucose and picric acid were used as standards. In order to detect the sugars, chromatograms and electrophoretograms were sprayed with a 3% solution of p-anisidine hydrochloride in water saturated 1-butanol containing a trace of stannous chloride.

After spraying, the chromatograms were heated at about 110°C for 1 minute.

F. Ultracentrifugation.

Analytical ultracentrifugation. Samples were placed in aluminium sector cells with an optical path of 12 mm., and quartz windows. These were ultracentrifuged at room temperature, and observed by Schlieren optics, in Spinco/Beckman Model E ultracentrifuges.

Preparative ultracentrifugation. Samples were placed in 20 ml. polythene tubes, and ultracentrifuged using the three-bucket angle head of an MSE Superspeed 50, at about 0°C. After ultracentrifugation, the solutions were removed from the tubes, in the required fractions, by piercing a small hole in the bottom of the tube to give a controlled flow.

G. Infrared spectrometry.

General. Infrared spectra were produced by a Perkin-Elmer 237 double-beam grating instrument. Non-aqueous samples were examined between sodium chloride plates; others were examined between barium fluoride plates. General values of group absorption frequencies were obtained from A. D. Cross, Introduction to Practical Infrared Spectroscopy (2nd ed.) London, 1964.

Deuteration studies. Samples were enclosed in fairly airtight cells, through which a stream of D_2O-saturated nitrogen was passed. This was produced by bubbling dry nitrogen from a cylinder or, more usually, from a Dewar flask of liquid nitrogen, through heavy water. The cells, at first, were made of glass tubing and microscope cover-slips sealed, where necessary, with rubber washers and Apiezon grease. All glass in the optical path was placed in dilute hydrofluoric acid for about five minutes,

before use (19, 20). In later work, barium fluoride windows were used in the optical path.

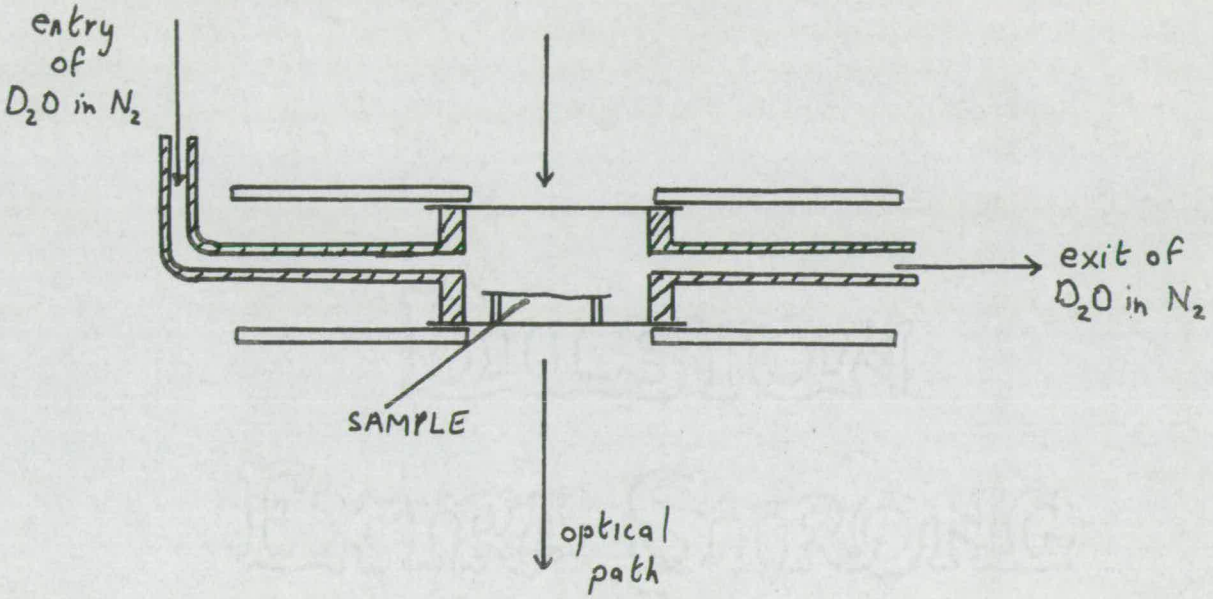


Figure 1

Samples were prepared by evaporating solutions or suspensions of polysaccharides in a vacuum desiccator; either on mercury, thus producing a free film; or on a treated cover-slip; or, in later work, on a slip of transparent silver chloride.

Polarization studies. A gold-wire polarizer was incorporated into the spectrometer for this purpose. Preparation of samples for these studies will be described in the main part of the thesis.

H. Computing

Computer programs were written in Atlas Autocode, (Edinburgh University ISO version), on eight-hole paper tape, and run on the English Electric KDF 9 at the Edinburgh Regional Computing Centre.

Some simpler calculations were performed on a Facit desk Calculator.

S E C T I O N I

Introduction

This section describes work done on the extraction, fractionation and analysis of the seed-coat mucilage of Sinapis alba. It also covers investigation of the nature of the linkage between cellulose and the associated polysaccharide, and the conclusive identification of the cellulose as such. Finally, it reports some comparative studies on other cellulose-containing mucilages.

The precipitate from these mucilages has generally been assumed to be cellulose on the basis of (a) the glucose content, and (b) the insolubility and fibrous appearance. The only rigorous characterisation of this type of glucan was performed on the precipitate formed by the action of dilute acid on noriasa (Abelmoschus glutiotextilis) mucilage (21). This was shown by means of a Monier-Williams degradation, to contain only glucose. Acetolysis gave a disaccharide identified as cellobiose octa-acetate, and the X-ray diffraction pattern was almost identical to that of cotton cellulose.

Indirect indication of the presence in these mucilages of cellulose has been provided by Muhlethaler (12). He has shown the presence of cellulose-like microfibrils in several mucilages by means of electron microscopy. In one species, Cobaea scandens, larger macrofibrils are found, up to 5 cm. long, which are visible under the light microscope.

Cellulose can occur in a number of crystalline forms, which are distinguishable by their infrared spectra after deuteration and by their X-ray diffraction patterns. Cellulose I is the name which has been given to most natural celluloses. There are slight variations in the structure of cellulose I, particularly when it is derived from algae (22), and Valonia cellulose is

triclinic, whereas most cellulose I is monoclinic. Bacterial cellulose, produced extracellularly by Acetobacter xylinum, and Valonia cellulose, have been distinguished (23) as cellulose IA on the basis of a small peak at 3242 $cm.^{-1}$ in the infrared spectrum, which does not occur in that of plant cellulose (cellulose IB). Cellulose II has been reported in the alga Halicystis (24), but is commonest as regenerated cellulose, e.g. rayon, etc. Other artificial types have been reported, but these are the commonest forms, and mucilage celluloses are almost certain to belong to one or other type.

These forms can be recognised by their infrared spectra (20); the main hydroxyl peak in cellulose I is at 3350 $cm.^{-1}$ whilst that in cellulose II is at 3450 $cm.^{-1}$ In addition, the pattern of hydroxyl peaks is quite characteristic. They may also be distinguished by their X-ray powder diffraction patterns; cellulose I has the 101 peak at 12° (25), whilst in cellulose II it is at 15° (26).

There are three possible types of covalent linkage between cellulose and the other polysaccharides, viz. ether, ester and glycosidic links. Ether links can be discounted immediately since these would not be acid-labile. The other two are both possible on the present experimental evidence. One method of distinguishing between them would be by means of the infrared spectrum; esters have a strong peak at 1735-1750 $cm.^{-1}$, due to carbonyl stretching, which is not given by a glycoside.

Non-covalent links between the polysaccharides are also possible. These could be cationic interactions, hydrogen bonding or hydrophobic bonding. Divalent cations, by interacting with a pair of uronate anions, or by chelation with hydroxyl groups,

might hold a pectic polysaccharide, in some way, around a cellulose molecule or microfibril, thus solubilizing it. Such an interaction could easily be identified by means of a standard ion-exchange method. Substitution of hydrogen (which would form largely non-ionic uronic acid), or a monovalent cation for the divalent cation would cause the structure to collapse, and hence the cellulose would be precipitated.

Hydrogen-bonding is a unique type of interaction between two polar atoms, through a proton covalently linked to one of them. It has been shown (27) to include electrostatic, delocalisation, repulsion and dispersion interactions. It is difficult to detect chemically, although removal of the proton by ionization or substitution should destroy it. Infrared spectroscopy can provide clear proof of its existence in a system, since hydrogen-bonding lowers the hydroxyl stretch frequency from its non-hydrogen-bonded value of 3580-3670 $cm.^{-1}$ Hydrogen-bonding, however, is unlikely to be the sole interaction between cellulose and the polysaccharide, since the strength of a water-alcohol bond is very similar to that of an alcohol-bond, and hence, dissolving the mucilage would probably be sufficient to dissociate the polysaccharides, if this were the case.

Hydrophobic interactions occur when non-polar molecules, or non-polar moieties within molecules, are dissolved in water. The non-polar groups (typically hydrocarbon in biological systems) tend to avoid the solvent by remaining in contact with one another. This is because the water molecules tend to assume an ordered structure around such groups, so that, when such groups are in contact, they present a smaller total surface to the water, and hence more of the water can exist in a disordered

form. Thus, the driving-force in this type of interaction is not decreased enthalpy, but increased entropy. In polysaccharides, the C-H groups are non-polar, and might exhibit hydrophobic bonding, although the presence of hydroxyl groups might counteract this by maintaining disorder amongst the neighbouring water molecules, and by their own hydrophilicity. Indications that hydrophobic bonding may occur in polysaccharides is given by the helices of V amylose (28) and κ-carrageenan (29), in which the C-H groups are on the internal surfaces of the helix. V amylose is obtained by the formation of complexes, such as amylose-butanol, amylose-cyclohexane, or amylose-iodine. In these complexes the amylose is coiled around the other molecule and, in the case of butanol and similar predominantly hydrophobic molecules, the interaction may well be hydrophobic bonding.

Hydrophobic interactions may be disrupted by three methods:

(a) Competitive hydrophobic interaction. Another hydrophobic species is introduced into the system, and disrupts the former hydrophobic interactions by forming new ones. Thus, if a hydrophobic species C is introduced into a solution of a hydrophobically-bound complex A:B the result is

$$A{:}B \xrightarrow{C} A{:}C, B{:}C$$

Hence, if cellulose and the other polysaccharides are hydrophobically bound, addition of a competitive reagent, such as a detergent, should dissociate them.

(b) Disruption of the ordered structure of water. Certain chemical species such as urea, can only fit, geometrically, into disordered regions of water molecules. This lowers the chemical

potential of the disordered water, and hence discourages the formation of ordered regions. Because of this, there is no longer any entropy advantage in the association of hydrophobic groups, and dissociation may occur. This seems to be the mode of action of protein denaturants such as urea (30), though in this case, competitive hydrogen-bonding can also occur.
(c) Production of a superior repulsion between the hydrophobic groups. For instance, two sugar residues might be parted by partial ionization of the hydroxyl groups.

The only previous structural work which has been performed on the seed-coat mucilage of Sinapis alba was done by Bailey et al. (31, 32) over thirty years ago. The mucilage was extracted from the seeds with cold water, and could be fractionated into two portions by the addition of saturated barium hydroxide solution. Material which remained in solution contained 7% pentosan, 10% 6-deoxyhexosan, 24% uronic acid, 45% hexosan, and 5% methoxyl, and had an optical rotation $[\alpha]_D$ in water $+10^{\circ}$. L-arabinose, D-galactose, and an aldobiouronic acid composed of L-rhamnose and D-galacturonic acid, were detected in the hydrolysate.

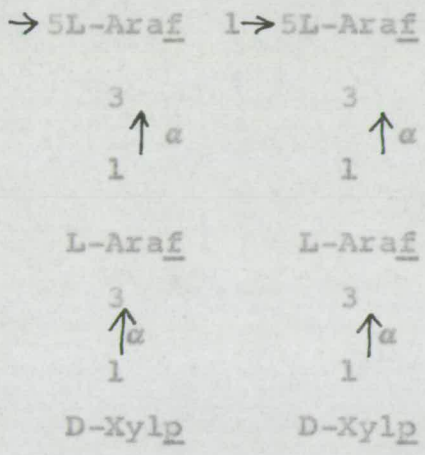
The gel produced by treatment with barium hydroxide was shown to contain cellulose, and also L-arabinose, D-galactose, D-galacturonic acid, and D-glucuronic acid. Quantitative analysis of this fraction showed the presence of 9% pentosan, 15% uronic acid, 72% hexosan and 3.4% methoxyl. Acid hydrolysis of the whole mucilage produced a cellulose precipitate which amounted to 43% of the original. The acid hydrolysate was a mixture of aldobiouronic acids, which probably contained a uronic acid-hexose, and a uronic acid-6-deoxyhexose, in the ratio 2:1.

Of the other cellulose-containing mucilages, that form the seeds of cress (Lepidium sativum) is likely to be most similar to mustard-seed mucilage, since the two species are related. This has recently been investigated (33, 34), much more thoroughly than has mustard-seed mucilage.

Fractional precipitation (33) of cress-seed mucilage with ethanol gave two fractions which differed in the amount of cellulose which they contained, but were otherwise similar. The first fraction to be precipitated contained 9% of cellulose, whilst the second was virtually cellulose-free. Both were shown by acid hydrolysis and paper chromatography, to contain galacturonic acid and 4-O-methyl-glucuronic acid (total uronic acid contents 28.4% and 30.9% respectively); galactose, arabinose, xylose and rhamnose in the ratios 10:13:7:7 and 10:12:8:8 respectively; and traces of glucose.

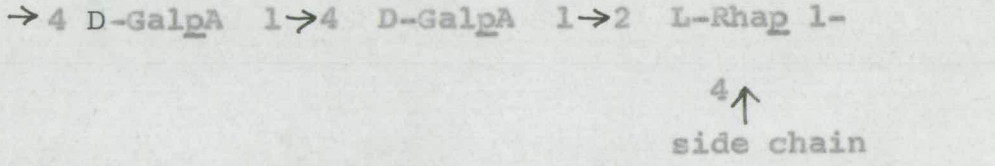
When the cellulose-free fraction was methylated, it could be further fractionated by extracting the neutralised mixture from Haworth methylation with chloroform. The chloroform layer was shown to contain the methyl ether of a xyloaraban contaminated by an acidic component, which remained in solution when the xyloaraban was partially precipitated with petroleum ether. The aqueous layer contained the methyl ether of an acidic polysaccharide.

The structure of the xyloaraban was elucidated further by hydrolysis of the fully-methylated polysaccharide, and fractionation and identification of the products; and by characterisation of oligosaccharides produced by autohydrolysis of the original material. From the results, structure I was suggested, although the main-chain could also contain 1,4-linked arabinopyranosyl, or 1,3-linked arabinofuranosyl units.

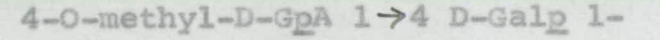


I. Probable structure of cress-seed xyloaraban

The aqueous solution from the Haworth methylation, after the extraction with chloroform, contained an acidic polysaccharide. Hydrolyses were performed (34) on this polysaccharide when fully methylated, and on the polysaccharide obtained by autohydrolysis of the whole mucilage, and methylation of the residue. These showed that the structure of the acidic polysaccharide was essentially represented by IIa, with IIb as the commonest side-chain.



IIa Approximate structure of the acidic polysaccharide.



IIb Commonest side chain

About one third of the rhamnose residues are also branched at position 3, and a few unbranched terminal rhamnose residues also occur. A few galactose residues are terminal or are branched at position 3, or at positions 2 and 3, and could occur in the main-chain. α-D-Xylp 1→4 D-Galp 1- sometines occurs as a side chain instead of IIb.

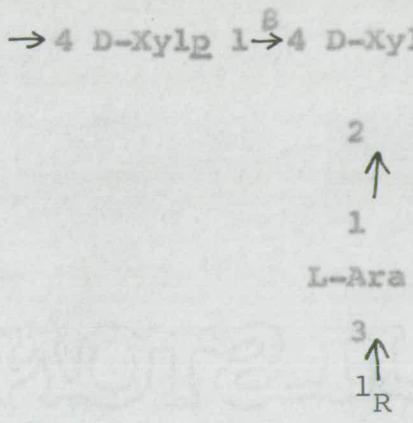
When the supposed cellulose, obtained by autohydrolysis, was subjected to dilute acid hydrolysis, the first hydrolysate contained galactose, glucose, mannose, arabinose, rhamnose, galacturonic acid, and, mainly xylose. A further dilute acid hydrolysis on the residue gave only mannose, glucose and xylose. Graded Monier-Williams degradations on the residue gave similar results, with glucose increasing in intensity, and xylose decreasing to a trace.

The mucilage from the seeds of flax (Linum usitatissimum) also contains cellulose although, in this case, it amounts to only 0.5% (32).

This mucilage has been fractionated by cupric acetate and ethanol (35) or by cetyl-trimethylammonium bromide ("Cetavlon") (36). In both cases a neutral pentosan fraction, and an acidic fraction were obtained.

Analysis of the pentosan fraction (35) showed the presence of xylose and arabinose in the approximate molar ratio of 3:1, together with traces of galactose and glucose. Methylation of the pentosan acetate, followed by hydrolysis, indicated 1,4-linked xylose, branched at position 2, and perhaps also at position 3, with 3-linked arabinopyranosyl and arabinofuranosyl units, and a trace of D-galactose end groups. Partial hydrolysis of the original pentosan caused preferential loss of

arabinose, and gave oligosaccharides characterised as β-D-Xylp 1→4 D-Xyl, α-D-Xylp 1→5 L-Ara, and (36) α-D-Xylp 1→3 D-Xyl. These results suggest a structure of the type illustrated in III.

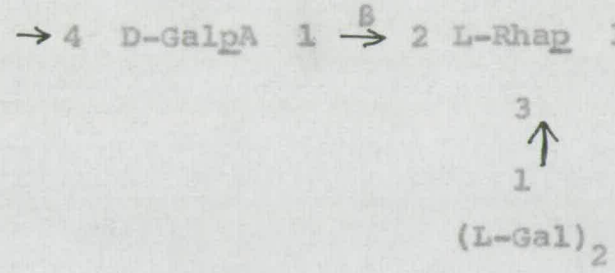


R=H, $(L\text{-Ara})_n$, D-Gal or D-Xylp (linked α 1,5).

III. Possible structure of linseed pentosan.

The precipitate from cetavlon fractionation could be further fractionated, with 7% cupric acetate, into copper-insoluble and copper-soluble fractions (36). The copper-insoluble fraction contained L-rhamnose, L-galactose, and D-galacturonic acid, in the approximate molar ratios 2:1:2. Reduction, methylation and hydrolysis, showed that the D-galacturonic acid was linked 1,4, whilst the L-rhamnose was linked 1,2, with some branching at position 3. L-Galactose derivatives were absent, suggesting that this occurred as side-chains which were lost during methylation. Autohydrolysis of the polysaccharide gave L-galactose first, followed by D-galacturonic acid, and finally L-rhamnose. These results suggest that this polysaccharide is probably a main chain of the common galacturonorhamnan

type, with side-chains on every fourth rhamnose residue containing, on average, two L-galactose residues (IV). The possibility of other arrangements cannot be excluded, however.



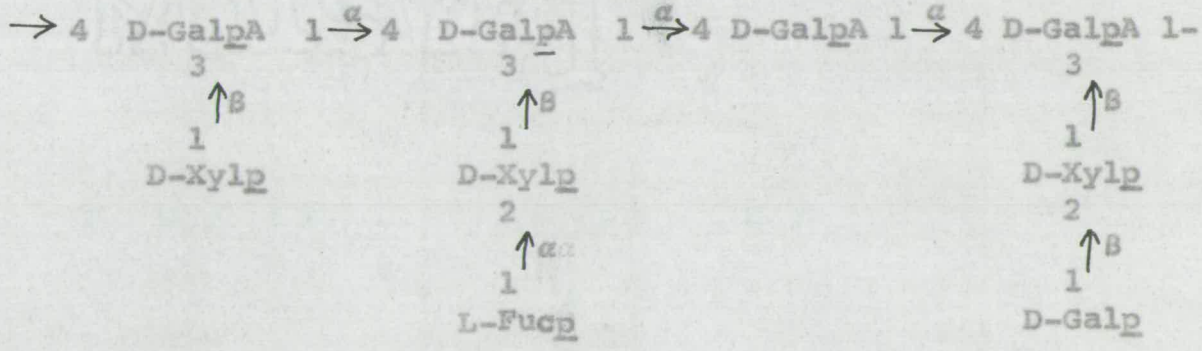
IV. Probable structure of linseed acidic polysaccharide

The copper-soluble fraction contained L-rhamnose, L-fucose, L-galactose, and D-galacturonic acid, in the molar ratios 4:1:2:2. Reduction, methylation and hydrolysis showed that the L-fucose and L-galactose occurred largely as end-groups, whilst the D-galacturonic acid was linked 1,4 and the L-rhamnose was linked at positions 2 and 3, or positions 2, 3 and 4. This suggests a structure similar to, but more branched than IV. Periodate oxidation results, however, suggest that some of the L-galactose may occur in the main-chain.

Gum tragacanth is not a seed-coat mucilage, but a gum exudate from species of Astragalus. It is obviously heterogeneous, since it is only partially soluble in water. The methylated gum can also be separated into an arabinogalactan and an acidic component (tragacanthic acid) by dissolving the sodium salt of the latter in water (37). Later workers (38) prepared tragacanthic acid by addition of ethylene oxide to the water-soluble fraction, fractional precipitation with acetone, and saponification

of the ester as required; and the arabinogalactan by extraction from the crude gum with ethanol/water (7:3), after removal of glycosides with hot ethanol.

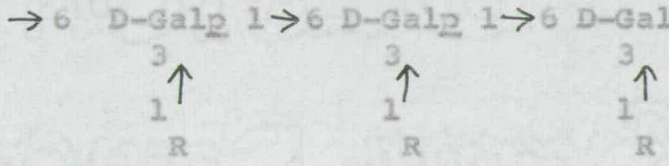
Tragacanthic acid, after purification on a diethylaminoethyl cellulose column, contained uronic acid, xylose, fucose and galactose, in the approximate ratios 8:8:2:1; and a trace of arabinose. Acetolysis of tragacanthic acid, followed by deacetylation, gave two disaccharides, recognised as α-L-Fucp 1→2 D-Xyl and β-D-Galp 1→2 D-Xyl. Partial hydrolysis of tragacanthis acid with 0.05N sulphuric acid gave a degraded product which, with commercial hemicellulase and pectinase, gave a disaccharide recognised as β-D-Xylp 1→3 GalA. More rigorous partial hydrolysis with 0.5N sulphuric acid gave another degraded product which, with these enzymes, gave α 1,4-linked GalA oligomers. These results, confirmed by methylation and hydrolysis of the reduced polysaccharide, show that tragacanthic acid can be represented by V.



V. Structure of tragacanthic acid.

Partial hydrolysis of the arabinogalactan (39) gave a preferential loss of arabinose, and a residual branched galactan. Further partial hydrolysis gave two disaccharides identified as

1,6-galactosylgalactose, and a small amount of 1,3-galactosyl-galactose. Methylation and hydrolysis showed that the main-chain consists mainly of 1,6-linked galactopyranosyl residues, with occasional 1,3 links, and some branching at position 3, and perhaps also position 4. Methylation and hydrolysis of the whole arabinogalactan showed that the arabinose occurred as highly-branched arabinofuranosyl side-chains linked to position 3 of galactose residues (see structure VI). Small amounts of L-rhamnose and D-galacturonic acid were also present in the side-chains.



R=L-Araf or $(\text{L-Araf})_n$ branched 2 and/or 3 and/or 5.

VI. Structure of gum tragacanth arabinogalactan.

The water-insoluble fraction of gum-tragacanth can be partially dissolved in alkali, and it is found (38) that the alkali-soluble fraction resembles the tragacanthic acid/arabinogalactan mixture in composition. The residue is shown, by dilute acid hydrolysis and Monier-Williams degradation to be a glucan contaminated by tragacanthic acid and arabinogalactan. A similar residue is formed when 2-hydroxyethyl tragacanthate is allowed to stand. The alkali-insoluble residue constitutes 21% of the whole, and the residue apparently solubilised by tragacanthic acid constitutes another 7.3%.

The mucilage from the seeds of quince (Cydonia vulgaris) has not been studied in detail, but it has been shown (7d) to

contain partially methyl-etherified uronic acid, L-arabinose, probably in the furanose form, and xylose. Acid hydrolysis gives a mixture of sugars, including one which appears to be a mono-O-methyl-hexosyluronic acid-xylose. This mucilage contains about 33% cellulose.

EXPERIMENTAL

Extraction and initial fractionation of mucilage

Experiment 1.1.

Crushed mustard-seeds were sieved to give a residue consisting very largely of seed coats. This residue (201.7 g.) was stirred with cold water (2.5 l.) for 4½ hours, after which it was removed by straining through muslin, and re-extracted with cold water (2 l.) for 18 hours. The process was repeated thrice, each time with 2 l. of water. The solutions were combined and concentrated, and insoluble material was removed by centrifugation. The solution was further concentrated to a final volume of 1.53 l. After the addition of ethanol (1.98 l.; methylated spirit), and dilution with water (0.6 l.) precipitation had still not occurred. Part of the solution (200 ml. was removed, and the mucilage was totally precipitated by the addition of excess ethanol. A 1:1 ethanol/water mixture (7.5 l.) was added to the remainder to give a first precipitate, fraction A. Subsequent fractions were produced by the addition of further quantities of ethanol.

A sample of each fraction was hydrolysed overnight with 2N H_2SO_4. The precipitated cellulose was filtered off, washed, dried and weighed and after neutralisation the supernatant solutions were examined chromatographically (solvent a).

Results:

Final volume of water in main fraction = 5.87 l.

Fraction:	A	B	C	D	E	Whole
Weight (gm.)	2.36	4.24	9.69	2.82	0.91	1.25
Total vol. of ethanol (l.)	5.7o	6.45	7.45	7.95	8.95	excess
% cellulose	27.5	26.9	24.9	33.6*	21.6	27.6

*This was later shown to be erroneous. The true percentage was intermediate between those of C and E.

Total weight of mucilage = 21.27 g. (10.54% of seed-coats).

Sugars:

unknown	+	+	+	+	+	+	+
GalA	++	++	++	++	++	++	++
Gal	+++	+++	+++	+++	+++	+++	+++
G	++	++	++	++	++	++	++
Man	+	+	+	+	+	+	+
Ara	++	++	++	++	++	++	++
Xyl	+	+	+	+	+	+	+
Rha	+	+	+	+	+	+	+

The Nature of the cellulose'polysaccharide linkage

Experiment 1.2 Infrared spectroscopy of the mucilage

A thin film of mucilage, fraction D, cast on a mercury surface (see General Methods), was examined between sodium chloride plates by infrared spectroscopy. The hydroxyl stretch peak was broad and flattened, extending from 3580 $cm.^{-1}$ to 3120 $cm.^{-1}$ indicating hydrogen bonding. The carbonyl stretch peak was centred at 1600 $cm.^{-1}$ which is characteristic of the carboxylate anion (see Fig. 1.1). A solution (1%) of fraction C was dialysed 3 times, for 1 day each, against 0.1N HCl, and then 3 times against distilled water. A film of this mucilage was cast on a mercury surface, and examined as before. The only major difference in the spectrum was a peak at 1730 $cm.^{-1}$ due to un-ionized uronic acid; and a smaller peak at 1650 $cm.^{-1}$; due to uronate anion, showing that cation-exchange to give the H^{+} form, was probably incomplete.

Experiment 1.3 Effects of reagents on mucilage

Various reagents were added to solutions of mucilage (approximately 0.25%) and the visible effects, before and after boiling, were noted. The reagents which were used were: water-structure breaking agents; urea, guanidine hydrochloride and ethyl glycol;

competitive hydrophobic bond-formers: sodium lauryl sulphate (SLS);
DNA denaturants (40): lithium thiocyanate and trichloroacetic acid;
cation chelating agents: sodium diaminoethane-tetra-acetate (EDTA) and sodium hexametaphosphate (HMP);
acids: hydrochloric and formic acids;
bases: sodium tetraborate, ammonium and sodium hydroxides.
The effects of some combinations of reagents, and of ultrasonic vibrations, were also tested. Several reagents produced a sharp increase in viscosity, which is listed as 'gel' in the table.

Reagent	Conditions	Effect, cold	Effect, hot
$C(NH_2)_3^+ Cl^-$	1M	Cloudy gel	Clear gel
$CO(NH_2)_2$	7-8M	Cloudy solution	Clear gel
$(CH_2OH)_2$	10-30%	Thicker solution	No change
SLS	0.05-0.2M	" " "	Some fibrous ppt.
LiCNS	50% w/v	Thick gel	Clearer gel
CCl_3COOH	5-50%	Thicker solution	Gelatinous ppt.
EDTA	2-5% w/v; pH 4.9 or 7.5	Thick solution	Gel or gelatinous ppt.
HMP	2%; pH 4.0	" " "	" " " "
HCl	6N	Cloudy gel	Clear gel
HCOOH	10-90%	Thicker solution	Fibrous ppt.
$Na_2B_4O_7$	0.1M	Thicker solution	Thick bubbly gel
NH_4OH	6N	Cloudy solution	No change
NaOH	0·1N	Clear gel	" " "

Reagent	Conditions	Effect, cold	Effect, hot
NaOH	1N	Slow formation of gelatinous ppt.	Fibrous ppt.
" "	2N	Immediate formation of fibrous ppt.	No change
$(CH_2OH)_2$ + $C(NH_2)_3^+Cl^-$	10-30% 8M	Thicker solution	No change
EDTA+ SLS	2% 0.2M; pH 5.0	Thicker solution	Fibrous ppt.
ultrasound		No change	

Experiment 1.4 Quantitative examination of alkali precipitation

Portions of mucilage, fractions B and C (250 mg.), were dissolved in water (100 ml.) and an equal volume of 4N NaOH was added. The solutions were diluted to 1 l. and filtered by suction, through glass paper. The cellulose was washed, dried and weighed. B. contained 33.4% cellulose, and C contained 32.3%. This shows that alkali-precipitation gives yields substantially higher than acid hydrolysis.

Experiment 1.5 Quantitative examination of the effects of EDTA and SLS.

0.5% solutions of fraction C were made up, containing 2% EDTA, 0.2M SLS and 2% EDTA + 0.2M SLS, respectively. The solutions were adjusted to an approximate pH of 5 and boiled. The precipitates were removed by centrifugation, dried and weighed.

Results:	2% EDTA	0.2M SLS	both
Percentage cellulose	30.92	36.38	38.20

Thus, SLS and SLS+EDTA are even more efficient than NaOH in precipitating cellulose. The difference is probably due

to the degradation of cellulose to soluble products by the NaOH.

Experiment 1.6 Re-examination of the effect of urea

0.1% solutions of mucilage in urea solutions of various concentrations were boiled and centrifuged. Visual comparisons were made between the various supernatants and precipitates and the latter were further compared by Monier-Williams degradation and paper chromatography (solvent a).

Results:

Conc. of urea	ppt.	supernatant	Hydrolysis products of ppt. GalA	Gal	G	Ara	Xyl	Rha
0%	slimy	cloudy	-	++	+	+	trace	trace
10%	"	"	-	++	+	trace	-	-
20%	"	"	-	++	+	" "	-	-
30%	"	clearer	-	++	++	" "	-	-
40%	fibrous	clear	trace	++	++	+	-	trac
50%	" "	"	-	++	++	trace	-	-

Experiment 1.7 Examination of cellulose-free mucilage.

The supernatant from dissociation of the mucilage with NaOH was dialysed and concentrated, and hydrolysed by dilute acid. The solution was filtered, and solid material was obtained which amounted to 0.85% of the original cellulose-containing mucilage. This was dark in colour, and appeared to be debris rather than cellulose. The hydrolysate was examined by paper chromatograph (solvent a) and shown to contain unknown (++), GalA (++), Gal (+++), G (+), Man (trace), Ara (++), Xyl (+), and Rha (+).

Samples of cellulose-free mucilage produced by NaOH dissociation and by EDTA/SLS dissociation respectively (1% in 0.15M NaCl) were examined by analytical ultracentrifugation, the former at 50, 74 0 rpm, (see Fig. 1.5) and the latter at 15,220 rpm. (10,090 x g) for one hour, then at 39,460 rpm

(67,870 x g). Both gave a single large, broad peak moving rather more rapidly than component 5 of the whole mucilage (see Expt. 1.16).

Experiment 1.8. Attempted recombination of mucilage and cellulose.

A solution of cellulose in Schweizer's reagent was added, dropwise, to solutions of whole mucilage (1%) and of cellulose-free mucilage (1%; prepared by alkaline precipitation of cellulose) made approximately 0.1N with respect to H_2SO_4. Cellulose was precipitated at the first drop; the presence of mucilage therefore did not inhibit the precipitation of cellulose.

A solution of whole mucilage (1%) was made 2N with respect to NaOH to precipitate the cellulose. It was then dialysed, but the cellulose did not re-dissolve. A sample was homogenized, but the cellulose re-flocculated on standing.

A sample of this cellulose/pectic mixture was examined by infrared spectroscopy during deuteration (see Expt. 1.12). It was deuterated more rapidly and more completely than whole mucilage, thus indicating that inaccessible regions occur in the pectic polysaccharide, as well as in the cellulose.

Identification of cellulose

Experiment 1.9 Monier-Williams degradation

Samples of cellulose, precipitated by various methods, were subjected to Monier-Williams degradations, and examined chromatographically (solvent a).

Results:

Cellulose	GalA	unknown	Gal	G	Man	Ara	Xyl	Rha
Acid-pptd.[a]	-	-	-	+++	trace	-	trace	-
Alkali-pptd.	trace	trace	+	+++	-	trace	trace?	trace?
EDTA/SLS pptd[b]	++	"	++	++++	+	+	+	-

[a] Fraction A; acid-precipitated celluloses from other fractions were not significantly different.

[b] Celluloses precipitated by EDTA alone, or SLS alone, were not significantly different.

The precipitates produced by urea have already been described (Expt. 1.6).

In order to test whether mannose (which would be obscured by arabinose in solvent a) occurred in the hydrolysates, they were re-examined in solvent b. Mannose was not detected.

The hydrolysate from the alkali-precipitated cellulose was examined by paper electrophoresis, and was shown to contain two minor components. The slower one has the same mobility as glucuronic acid ($R_{picric} = 1.03$) and is probably galacturonic acid. The faster one has an R_{picric} value of 1.34.

To investigate how firmly the non-cellulosic material is attached to the cellulose in alkali-precipitated material, cellulose was precipitated in the usual way from a solution of fraction C (500 mg.). It was filtered off and washed with water (100 ml.) 3 times. A third was removed (sample a) and the remainder was washed a further 17 times with water (100 ml.) then divided into two. One part was kept (sample b), and the other was washed with 2N NaOH (100 ml.), and a further 5 times with water (100 ml.) to give sample c. A fourth sample

(d) of about the same size, was produced by washing regenerated cellulose (from filter paper) with 2N NaOH (100 ml.). The samples were degraded and examined as before.

Results:

Sample	GalA	unknown	Gal	G	Man	Ara
a	trace	trace	+	+++	-	+
b	"	-	+	+++	-	+
c	"	trace	+	+++	trace?	+
d	-	+	-	++++	-	-

The hydrolysates of an alkali-precipitated sample, and of alkali-treated cellulose II, were examined by paper electrophoresis. The minor highly mobile, component, noticed previously, occurs in both samples. Hence it must be an artefact, probably a saccharinic acid-type artefact, caused by the action of NaOH on cellulose.

Experiment 1.10. Enzymolysis

Samples of regenerated cellulose, acid-precipitated cellulose, and alkali-precipitated cellulose (5 mg.) and mucilage fraction D (20 mg.) were each suspended in phthalic acid/sodium hydroxide buffer (0.5 ml.), with a phthalic acid concentration of 0.05M, and a pH of 4.3; and a solution of Myrothecium verrucaria cellulase (kindly supplied by Dr. G. Halliwell of Strathclyde University) (0.5 ml.) was added to each. Two control samples were also prepared (i) mucilage fraction D (15 mg.) + buffer (0.5 ml.) + water (0.5 ml.), and (ii) buffer (0.5 ml.) + enzyme (0.5 ml.). The samples were covered with toluene, as a bacteriostatic agent, and incubated at 35°C for one week. The mixtures were then heated in boiling water, and polysaccharide

material was removed by the addition of ethanol (4 ml.) followed by filtration. The filtrates were concentrated to about 0.5 ml., and examined chromatographically (solvent c).

Results:

Sample	unknown	unknown	cellobiose	unknown	G	Ara
regenerated	-	-	trace	-	++	-
acid-pptd.	-	-	trace	-	++	-
alkali-pptd.	-	-	trace	-	+++	trace
mucilage + enzyme	+	+	trace	trace	+	+

The controls gave blank chromatograms.

These results indicated that the cellulase preparation might also contain an arabanase. This possibility was tested using, as substrates, pectin from the cotyledons of ungerminated mustard seeds (UP), mucilage fraction D, and the cellulose-free polysaccharide obtained from the mucilage by means of alkali (AP).

Results:

Sample	unknown	unknown	cellobiose	G	Ara	unknown	unknown
UP+enzyme	trace?	+	-	trace	+++	++	+
UP-enzyme	trace	-	-	trace	trace	-	-
mucilage	++	-	trace	++	++	-	-
AP+enzyme	trace	-	trace?	+	trace	-	-
AP-enzyme	trace?	-	-	-	-	-	-

Experiment 1.11 X-ray powder diffraction patterns

X-ray powder diffraction patterns of acid-precipitated cellulose (a mixture of the precipitates from every fraction), alkali-precipitated cellulose (from fractionC), and mucilage ('whole' fraction) were produced, using a Pye PW1051 powder

diffractometer, with a filament current of 20 mA, accelerating potential 40 kV, a copper target, and a nickel filter. These were compared with patterns from authentic cellulose I (dispersed filter-paper) and cellulose II [cellulose regenerated from tetramminocupric hydroxide (Schweizer's reagent)] (see Fig. 1.2). Positions and relative intensities (giving the peak at 20-23° an intensity of 100) of the peaks were:

Cellulose I	Cellulose II	Acid-pptd.	Alkali-pptd.	Mucilage
2.2 (at 9.5°)	-	0.85 (9°)	-	-
2.2 (12°)	10.6 (12.5°)	0.85 (11°)	-	-
52.7 (15°)	-	32.5 (15°)	-	29.7 (16.5°)
-	100 (20.3°)	-	-	-
100 (22.8°)	-	100 (21.3°)	100 (21°)	100 (21.5°)
-	2.4 (28.3°)	-	-	-
11 (34.5°)	10.2 (35.8°)	4.3 (35°)	7.1 (36.5°)	5.4 (35°)
-	10.6 (40.8°)	-	-	-
1.1 (46.5°)	-	3.4 (44°)	-	-

Experiment 1.12 Infrared spectroscopy

Thin films were prepared on glass cover-slips, ofmucilage, and of celluloses prepared by acid, alkali- and EDTA/SLS precipitation. The samples were then deuterated (see General Methods). All showed resistance to complete deuteration, indicating crystallinity, and gave, after the initial rapid deuteration, a hydroxyl stretch peak at 3350 $cm.^{-1}$, with shoulders at 3410 $cm.^{-1}$ and 3300 $cm.^{-1}$ If high local concentrations of alkali were allowed during precipitation, however, it gave no such clear peaks. This pattern is characteristic of the cellulose I type B spectrum of Marrinan and Mann (23) (see Fig.1.3).

Samples of acid-precipitated, and of alkali-precipitated,

celluloses, were dissolved in Schweizer's reagent, and re-precipitated with dilute acetic acid. The precipitated material gave the very distinctive cellulose II spectrum, with hydroxyl peaks at 3500 $cm.^{-1}$ and 3450 $cm.^{-1}$ and with shoulders at 3300 $cm.^{-1}$ and 3200 $cm.^{-1}$

A sample of cellulose-free polysaccharide isolated after precipitation of cellulose by alkali, was also deuterated. This appeared to be completely accessible to deuteration, and gave no sign of containing cellulose I.

Experiment 1.13 Electron Microscopy

Droplets of mustard-seed mucilage, cress-seed mucilage, and gum tragacanth solutions (0.03%), containing 0.005-0.05% bovine serum albumin (to give good spreading conditions), were placed on a collodion membrane on a copper grid of the type used for electron microscopy, and allowed to stand for 2-3 minutes (41). After this time, superfluous solution was removed by means of filter-paper, and a drop of potassium phosphotungstate solution (BDH; 2%; pH 7, prepared immediately before use by combining solutions of pH 6 and 8) is added. After a few seconds, the grid is blotted again, and allowed to dry. The grids are then examined by means of an AEI EM6 electron microscope with the help of Messrs. Tulett and Wilson. Despite the use of bovine serum albunin, later specimens did not wet the collodion well. In all cases, however, cellulose-like microfibrils could be seen. The diameter of mustard-seed mucilage were measured from the photograph, and found to be about 37 Å; those in cress-seed mucilage, and in gum tragacanth, were, at least quantitatively, similar (see Fig. 1.4).

Fractionation of the mucilage

Experiment 1.14 Gel filtration

A column of agar gel was made for chromatography (42), consisting of equal proportions of 2%, 4%, 6% and 8% agar-gel particles which would pass through a 40-mesh screen. Commercial 'Ionagar No. 2' was used to produce the gels. The column was equilibrated with 0.15M NaCl solution, and a 1% solution of mucilage (25 ml.) in 0.15M NaCl solution was layered on to the top of the column, and eluted with 0.15M NaCl solution. Measurement of the concentrations of eluate fractions by the phenol sulphuric acid method, and comparison of two of the fractions by analytical ultracentrifugation, suggested that little, if any, fractionation had occurred.

Experiment 1.15. Tiselius electrophoresis

Solutions of fractions A, C and E (0.7%) in 0.05M borate buffer (pH 9.2), were examined by means of a Spinco Model H Electrophoresis-Diffusion instrument, using a field-strength of 6.5 volts/cm. Each fraction had a major component with high electrophoretic mobility, and three minor slow components. Repetition in phosphate buffer (M/30 salt; pH 6.05) gave similar, but less well-defined results.

Experiment 1.16. Analytical ultracentrifugation

1% solutions of fractions A, C and E in 0.15M NaCl solution were ultracentrifuged at 50,740 rpm (112,000 x g), and photographed at intervals up to an hour after commencement. These showed that an initial component was sedimented within the first 2 minutes. Accordingly, a 0.5% solution of A, in the same solvent, was ultracentrifuged at 20,410 rpm (18,150 x g). From the photographs of these runs, (see Fig. 1.5) it was possible to distinguish the following components:

Component 1: The major peak; very sharp; very rapidly sedimented at 50,740 rpm.

Component 2: Minor peak; broad; almost sedimented after 15 minutes at 50,740 rpm.

Component 3: Minor peak, very sharp; just slower than 2.

Component 4: Minor peak; broad; about 1/3 of the distance down the cell after one hour at 50,740 rpm.

Component 5: Minor peak; broad; just behind 4.

Experiment 1.17 High-speed centrifugation

A solution (1%) of fraction B (375 ml.) was centrifuged for 1 hour at 31,440 x g. The supernatant and pellet were treated with ethanol, and the precipitated polysaccharide was dried and weighed. Allowing for the loss of a small portion of supernatant, the proportions of the two fractions were: supernatant 34.2%; gel, 65.8% .

Samples of each fraction were treated with 2N NaOH and the precipitated cellulose was removed, dried and weighed. Cellulose contents were: supernatant 15.2%; gel 51.2%.

Experiment 1.18 Preparative ultracentrifugation

Attempts were made to fractionate mucilage solution (1%, 0.5 ml.) by layering it on to a urea solution decreasing linearly in concentration from 40% at the base to 10% at the surface (19 ml.), in 20 ml. tubes. These were ultracentrifuged for 30 minutes at 12,000 rpm (15,750 x g), or for 1 hour at 25,000 rpm (69,000 xg). The tubes were pierced at the bottom, and the contents were separated into 1 ml. fractions. In both cases, phenol sulphuric acid estimations showed no distinct peaks. Indeed there was an almost linear increase in polysaccharide concentration from bottom to top. The

following more crude approach seemed more successful:

A solution of the whole mucilage (8 ml.; 1% in 10% urea solution) was layered on to a 40% urea solution (12ml.) in 20 ml. tubes, and ultracentrifuged for three hours at 20,000 rpm. (44,000 x g). The samples were then divided into 'light' (top 10 ml. in tube), 'heavy' (bottom 10 ml.), and 'pellet' fractions. The solutions were dialysed for two da ys. The pellet consisted of granular, undissolved material, and a clear genatinous upper layer. The latter was scraped off, re-dissolved in 40% urea solution, and centrifuged to remove any remaining debris. It was then dialysed, to provide a third fraction ('very heavy'). The three fractions were freeze-dried and weighed.

Results:

Weight of starting material 2.21 g.

light fraction	o.52 g.	23.5%
heavy fraction	0.40 g.	18.1%
very heavy fraction	0.07 g.	2.95%
Total yield	0.99 g.	(44.6%)

Small samples of each were hydrolysed by formic acid, and the hydrolysates were examined by chromatography (solvent a).

Results:

Sample	unknown	GalA	Gal	G	Man	Ara	Xyl	Rha
v. heavy	-	trace	+	trace	-	+	trace	trace
heavy	-	+	++	++	trace	++	+	" "
light	++	++	+++	++	+	++	+	++

In view of the similarity of the 'heavy' and 'very heavy' fractions, the latter was included in the former.

An analytical ultracentrifuge examination of solutions of 'heavy' (0.3%) and 'light' (1%) in 10% urea, at 24,630 rmp (26,430 x g), showed no apparent 'light' contaminant in the 'heavy' fraction, although there was still 'heavy' material in the 'light' fraction.

Samples of the two fractions were analysed quantitatively. Cellulose was precipitated by boiling a 1% solution with 40% urea, filtered off, and weighed. Uronic acid was estimated by the carbazole method (see General Methods).

Results:

Fraction	Cellulose content	Uronic acid content
light	25.23%	22.73%
heavy	43.17%	27.46%

The quantities of neutral sugars in the 'heavy' fraction, and in whole mucilage, were compared by quantitative paper chromatography (see aniline phthalate estimations in General Methods)

Results:

Fraction	Approximate molar ratios			
	Gal+G	Ara+Man	Xyl	Rha
whole	17	6	1	2
heavy	18	8	1	2

N.B. The rhamnose figure is a minimum one, since some rhamnose will be present as galacturonosyl 1-2 rhamnose, or corresponding oligosaccharides.

Comparisons with cress-seed mucilage and gum tragacanth.

The following experiments were done with the help of Mr. D. Wright. Mucilages were obtained from the seeds of cress (Lepidium sativum) by the same method as was used for mustard-seed mucilage. Two varieties of cress were used; these were 'plain' and 'curled' cress.

Gum tragacanth was the same sample as used in the classical structural work of Aspinall and Baillie (38,39).

Comparison of the seed-coat mucilages of mustard and curled cress by analytical ultracentrifugation at 20,140 rpm (0.5% solutions in 10% urea) showed that the heavy component of cress mucilage sedimented approximately twice as rapidly as that of mustard mucilage (see Fig. 1.5). Hence, this component was isolated by preparative ultracentrifugation as in Expt. 1.18, except that it was ultracentrifuged for only 1¼ hours, and a 0.75% solution was used, instead of a 1% solution, because of the greater viscosity.

Gum tragacanth, as was mentioned in the Introduction to this Section, is only partially soluble. The whole gum was dispersed in water, and the precipitate, later found to be cellulose-rich was obtained by centrifugation.

Experiment 1.20. Chromatographic comparisons of mucilages.

Samples of mustard-seed mucilage (whole and heavy fraction), plain cress-seed mucilage (whole), curled cress-seed mucilage (whole and heavy fraction), and gum tragacanth (supernatant, precipitate, and whole) were hydrolysed by formic acid (see General Methods), and examined by chromatography (solvent a).

Results:

Sample	GalA	Gal	G	Ara	Fuc	Xyl	Rha
mustard (whole)	+++	+++	+	++	-	+	+
" " (heavy)	++	+++	+	++	-	++	+
plain cress	+++	++	trace	+++	-	+++	++
curled cress (whole)	+++	++	"	+++	-	+++	++
" "" (heavy)	+	+	+	+++	-	+++	trace
tragacanth (whole	++	+	+	+++	++	++	-
" " (supernatant)	++	+	-	+++	++	++	-
" " (ppt.)	++	+	+	+++	++	++	-

Fibrous precipitates formed during hydrolysis of the heavy fraction of cress-seed mucilage suggested that this was also the cellulose-rich fraction.

Experiment 1.21 Determination of cellulose contents

Solutions of mustard-seed, plain and curled cress-seed mucilages, and gum tragacanth (all whole; 1%), were mixed with equal volumes of 4N NaOH, and the cellulose was filtered off, dried and weighed.

Results:

	Cellulose contents
mustard	37.1%
cress (plain)	11.6%
cress (curled)	14.2%
gum tragacanth	3.1%

Experiment 1.22 Comparison of plain and curled cress mucilages

Since the two cress-seed mucilages appeared to be slightly different in cellulose content, they were compared by analytical ultracentrifugation (0.3% solutions in 10% urea), at 20,410 rpm, increasing to 50,740 rpm. This showed that the heaviest peak in 'curled' mucilage was rather larger and faster than that in 'plain'mucilage. Apart from the heavy peak, both samples contained only one light component, which had sedimented only one-tenth of the length of the cell after 1 hour at 20,410 rpm, followed by 45 minutes at 50,740 rpm.

Experiment 1.23 Infrared spectroscopy of mucilages

Thin films of curled mucilage (cellulose-rich fraction) and gum tragacanth (whole), dried on to silver chloride strips, were observed by means of infrared spectroscopy during deuteration. Cress-seed mucilage was partially resistant to deuteration, and gave the typical cellulose IB spectrum. Gum tragacanth, however,

appeared to be completely accessible to deuteration, and gave no cellulose spectrum.

Samples of cellulose, prepared from gum tragacanth by precipitation with 2N NaOH and 40% urea respectively, were also examined. These were partially resistant to deuteration, and gave an ill-defined peak at 3350 $cm.^{-1}$ thus resembling cellulose I.

Experiment 1.24 Effects of reagents on mucilage

Cress-seed mucilage and gum tragacanth were tested with the same reagents as mustard-seed mucilage (see Expt. 1.3) and gave comparable results. They were also tested with anions of the so-called chaotropic series (40). Solutions of curled cress-seed mucilage (0.3%) and gum tragacanth (0.5%) were made 4M with respect to: sodium trichloroacetate, potassium thiocyanate, sodium iodide, sodium acetate, potassium bromide and sodium chloride and 1M with respect to sodium perchlorate. The effectiveness of these anions (which with DNA, follows the order Cl^-, Br^-, CH_3COO^- I^- ClO_4^- CNS^- CCl_3COO^-) was observed, firstly in the cold, and then after boiling for 30 minutes.

Results:

	Cress		Gum tragacanth	
	cold	hot	cold	hot
CCl_3COONa	slight precipitate	precipitate	gel separates	aggregation of suspended matter
KCNS	precipitate	" "	" "	" "
$NaClO_4$	slight ppt.	some ppt	" "	" "
NaI	slight thickening	thickening	" "	slight aggregation
CH_3COONa	-	slight thickening	- -	- -
KBr	-	-	- -	- -
NaCl	-	-	- -	- -

DISCUSSION

The experiments described in this section have confirmed, first of all, that the insoluble glucan found in these mucilages is, indeed, cellulose. It has also been shown by its infrared spectrum (Expt. 1.12), and its X-ray diffraction pattern (Expt. 1.11) that it is present in the crystalline form that is normal in higher plants, that is, cellulose IB; and occurs in microfibrils similar to those found in the plant cell-wall. There is no conclusive proof that anomalous linkages or branch-points do not occur, but it is safe to conclude, from the enzymolysis (Expt. 1.10) and electron microscopy (Expt. 1.13), that such abnormalities are uncommon, if they occur at all. In the former case, the result of anomalies would be the presence of di- and oligosaccharides other than cellobiose in the products of enzymolysis; in the latter case, it is unlikely that regular microfibrils would be produced. One anomaly that has been suggested previously (33) on the basis of hydrolysis products, is the presence of a small amount of mannose in an otherwise typical cellulose. Mannose has been recorded in acid-precipitated cellulose (Expt. 1.9); however, this could be due to a mannose-containing hemicellulose cortex rather than to a mannose-containing "cellulose".

Structural studies of cellulose-containing mucilages (see Introduction to Section) have established that these are a mixture of two or more polysaccharides in addition to cellulose. These fall into two groups, both of which are pectic. Firstly, there are the branched acidic polysaccharides of the galacturonan or galacturonorhamnan type; secondly, there are the branched neutral polysaccharides, frequently pentosans. The analyses reported in this section are not detailed enough to prove that

this is also true of mustard-seed mucilage, but they are certainly amenable to this interpretation. There is, however, from these experiments, no unmistakable evidence of chemical heterogeneity in mustard-seed mucilage, other than in cellulose content. The hydrolyses reported in Expt. 1.18 indicated that the cellulose-rich fraction contained less galacturonic acid and rhamnose, but this seems to be disproved by the sugar estimations reported in the same experiment. The work of Bailey in 1932 (31,32) is relevant to this question. The analytical figures given for his cellulose-rich fraction, obtained by barium hydroxide precipitation, are very similar to those reported, for a corresponding fraction, in this section, if the cellulose is included in the hexosan figure (Table I).

	cellulose-rich fraction, present work	cellulose-rich fraction, Bailey	cellulose-poor fraction, Bailey
uronic anhydride	15.6	15	24
hexosan	70.8	72	45
pentosan	10.8*	9	7
6-deoxyhexosan	2.5	-	10

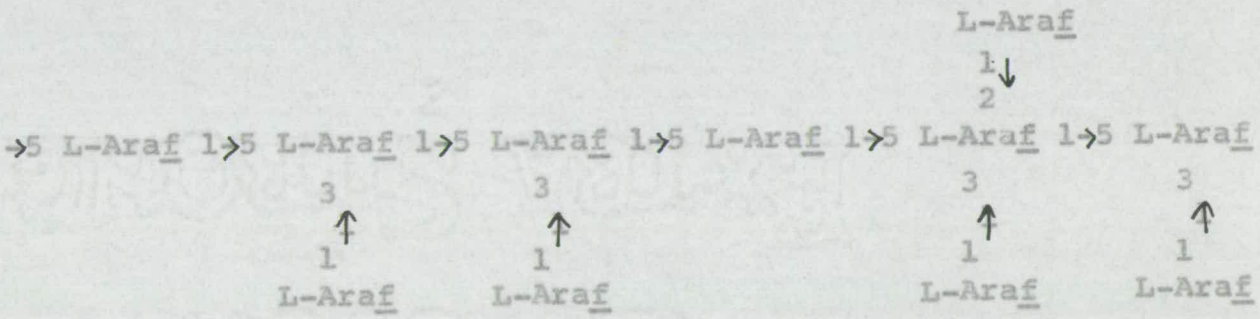
*This will include traces of mannose

Table I. Percentage composition of mucilage fractions.

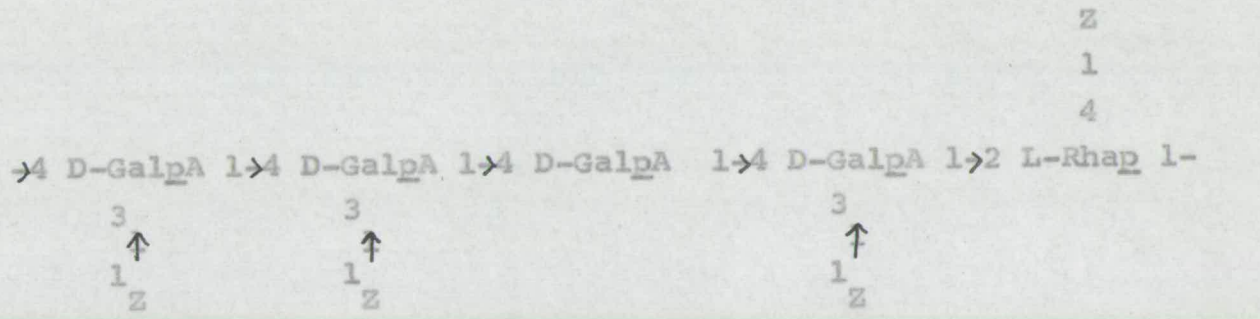
His cellulose-poor fraction was richer in uronic acid and 6-deoxyhexosan and thus supports the qualitative findings of Expt. 1.18. It should be noted, however, that Bailey recorded glucuronic acid in his samples. There has been no sign of this in the present work, and it could be that Bailey's mucilage

was contaminated, or in some other way distinct, from the mucilage used by us. Further work is needed on this, but it does seem possible that there are chemically distinct pectic components in mustard-seed mucilage. Analytical ultracentrifugation and Tiselius electrophoresis certainly showed that physically, the mucilage was markedly heterogeneous, but this could be explained by different cellulose contents.

It is of interest to see, at this stage, how far these results show a relationship to the normal cell-wall pectins from mustard. A study of the pectic polysaccharides in mustard cotyledons (43) showed that they are of two types, represented by structures VII and VIII.



VII. Possible structure for mustard araban.



Z = Xylp or a highly-branched araban chain, probably like VII

VIII. Representative structure for mustard acidic polysaccharide.

It can be seen that these are similar to those found in cress-seed mucilage (see Introduction to Section) and would fit the analytical data for mustard-seed mucilage fairly well, except that they contain no galactose. Further support for this type of structure is given by Bailey's identification (32) of a galacturonosyl-rhamnose in the hydrolysates of mustard mucilage. The slow 'unknown', recorded in the chromatograms of several of the hydrolysates reported in this Section, has a mobility similar to that of galacturonosyl 1→2 rhamnose.

Cellulose appears to be no more intimately associated with one component than another, in the mucilages which are definitely heterogeneous. In gum tragacanth (38), both components occur in association with cellulose. There are signs, however, that it shows a slight preference for a neutral sugar-rich component in cress (33) and mustard (ref. 32 and Expt. 1.18). The results of Monier-Williams degradation of cress-seed cellulose (33) definitely show that there is a minor component, a xylan or xyloglucan, closely associated with the cellulose. Traces of mannose (ref. 33 and Expt. 1.9) may be similarly explained, and perhaps also the galactose (Expt. 1.6) which persists with the cellulose after urea dissociation. This fits in with an earlier suggestion (22) that the cellulose microfibril in plant cell-walls consist of a crystalline glucan core, surrounded by a paracrystalline cortex of probably hemicellulosic materials. In higher plants, this may contain sugars such as mannose, xylose and galactose.

Investigation of the nature of the cellulose-pectic linkage (Expt. 1.2 to 1.8) demonstrated clearly that the linkage was not

covalent, since reagents such as urea etc. would have no effect on possible covalent bonds. There is apparently some indication that hydrophobic interactions play a dominant part in this linkage because of the effects of urea etc. Hydrogen-bonding is present (Expts. 1.2 and 1.12), and the size of the D_2O-resistant peak (Expt. 1.8) suggests that some inaccessible areas occur in the polysaccharides other than cellulose. On the other hand, there is no sign of regular hydrogen-bonding other than in the cellulose, and it is unlikely, as suggested in the Introduction, that hydrogen-bonding makes a major contribution to the linkage. The main remaining uncertainty is whether cations have any role in maintaining the structure of the mucilage. EDTA and HMP, chelating agents, do, it is true, cause cellulose precipitation. This, however, might be due purely to the "chaotropic" effect (40) of the anions. The fact that no precipitate was noticed in the preparation of the H^+ form of the mucilage (Expt. 1.2), or in the presence of reagents with exchangable cations, such as NH_4OH or $Na_2B_4O_7$, suggests that electrostatic interactions with cations are not of major importance.

The electron micrographs (Expt. 1.13) show the pectic polysaccharides tending to aggregate around the cellulose microfibrils. The possibility that this also occurs in solution is suggested by the results of the enzymolysis (Expt. 1.10). Although the cellulose was present in a highly dispersed form, its degradation was significantly retarded; and this despite the presence of a mucilage-degrading impurity in the enzyme.

In summary: cellulose-containing mucilages and gums contain cellulose IB, in microfibrils surrounded, at least in some cases, by a hemicellulosic sheath, to which are aggregated, more loosely,

the pectic polysaccharides of the mucilage, particularly, perhaps, the neutral components. The forces which hold these polysaccharide associations together might include hydrophobic interactions, and, to some extent, hydrogen-bonding and chelation of cations.

SECTION II

Introduction

The work described in this section aims to investigate the structures of polysaccharides by means of infrared spectroscopy. It attempts to find the spectral features characteristic of different types of polysaccharides and, in particular, it explores the uses of deuteration and beam polarisation in the investigation of crystallinity and group orientation.

Detailed analyses of monosaccharide and glycoside spectra and earlier work on polysaccharides, have recently been reviewed by Spedding (44). These have established differences in detail, due to anomeric configuration, and to the configurations of the secondary hydroxyl groups. Substituent groups also give such differences. It is hoped that the different types of polysaccharide may show comparable differences in the shape and positions of spectral peaks, either when deuterated or when undeuterated.

The use of deuteration to give spectra of the hydroxyl groups in crystalline regions was described in the previous chapter (Expt. 1.12). It has also been used (20) to estimate the crystallinity of cellulose II from the infrared spectrum. This method is based on the fact that the amorphous regions of cellulose are completely deuterated before the crystalline regions are significantly affected; this being shown by the virtual absence of crystalline OD peaks when the cellulose is back-exchanged after brief exposure to D_2O. Since this is so, a comparison of the sizes of the OH and OD peaks will give an estimate of the crystallinity, or more correctly, the accessibility

to deuteration. The Lambert-Beer law expresses the optical density $\log_{10}(I_o/I)$, in terms of concentration, c, and thickness of sample d, by the expression $\log_{10}(I_o/I)=kcd$, where k is a constant, the extinction coefficient, for a given chromophore. Hence, for the OH and OD peaks,

$$\frac{\log_{10}(I_o/I)OD}{\log_{10}(I_o/I)OH} = \frac{k_{OD}\cdot c_{OD}}{k_{OH}\cdot c_{OH}}$$

and the crystallinity is given by $\frac{COH \times 100\%}{c_{OH}+c_{OD}}$

The beam intensities are measured at wavelengths where they are independent of any accidental polarisation. In the particular case of cellulose II, these are at 3360 $cm.^{-1}$ for OH, and at 2530 $cm.^{-1}$ for OD. The ratio k_{OD}/k_{OH} is found to be 1·11. For other polysaccharides, the wavelengths and ratio will probably be different.

Prolonged exposure of cellulose II to D_2O deuterates a small fraction of the crystalline material; this is shown by the presence of crystalline peaks in the OD region after back-exchanging (20). It is also found that crystallinities measured by infrared spectroscopy are smaller than those measured by X-ray diffraction. Mann (45) explained these features of cellulose II by postulating the existence of four types of material, viz:

(a) Material crystalline to X-ray and infrared, and inaccessible to deuteration;

(b) Material crystalline to X-ray and infrared, but accessible to deuteration;

(c) Material crystalline to X-ray but amorphous to infrared, and accessible to deuteration;

(d) Material amorphous to both X-ray and infrared, and accessible to deuteration.

These are not regarded as distinct species, but as definable points on a continum between (a) and (d).

It seems, therefore, that the infrared measurements are of 'accessibility' rather than 'crystallinity' and that there is no sharp boundary between accessible and inaccessible regions. Nonetheless, it is a fairly close approximation to assume that such a boundary exists, since the rate constants of the deuterations of 'amorphous' and 'crystalline' regions are very different, and intermediate regions seem to be of minor importance. In view of this, it is possible to use a different method for calculating crystallinity in polysaccharides by means of infrared spectroscopy.

The interaction between polysaccharides and D_2O can be represented by

$$ROH + D_2O = ROD + DOH$$

The rate of disappearance of ROH is given by the equation

$$-d[ROH]/dt = k[ROH][D_2O] \qquad (i)$$

In the cells used for infrared spectroscopy, the D_2O vapour is continuously being replenished, so that $[D_2O]$ is constant. Hence, putting [ROH] = c, we can write (i) as $-dc/dt = Kc$ ... (ii)
If $c=c_o$ when t=0, this equation becomes $c=c_o.\exp(-Kt)$ ..... (iii)
If the fractional crystallinity of the sample is n, and the crystalline and amorphous parameters are denoted by subscripts x and a respectively, then the total concentration of ROH, $C=c_o(n.\exp(-K_x t) + (1-n).\exp(-K_a t))$ ..... (iv).
If we assume that radiation attenuation due to scattering and chromophore absorption are additive, then the optical density

is given by $\log_{10}(I_o/I) = (S + A.C.)d.$, where S is the scattering factor, and d is the sample thickness.

If we further assume that S is not altered by a small change of wavelength, we can measure its magnitude in a non-chromophoric region, and $\log_{10}(I_o/I_s)_s = Sd.$

$$\log_{10}(I_o/I) - \log_{10}(I_o/I_s)_s = A.C.d = A' (n.\exp(-K_x t) + (1-n).\exp(-K_a t)). \quad .....(v)$$

Equation (v) contains four unknown quantities; thus four sets of corresponding values of I and t are needed, in addition to the values of I_o and $(I_o/I_s)_s$, in order to calculate these quantities. Direct solution of (v) is impossible, due to its form; in practice, an iterative process is used, first calculating approximate values for A', n, and K_x from experimental values measured during deuteration of the crystalline region, then using these to calculate approximate amorphous parameters, using the latter to improve the crystalline parameters, and so on, by successive approximations.

There is also a simpler method for testing whether or not a sample is crystalline. Since $c = c_o.\exp(-Kt)$ (eqn. iii), $\ln(c_o/c) = Kt$. The time, $T_{1/4}$ required to reduce the size of the OH peak by one quarter is obtained from $\ln(c_o/\frac{3}{4}c_o) = KT_{1/4}$. Thus, the ratios $T_{1/4}:T_{1/2}:T_{3/4} = \log(4/3):\log(2):\log(4) = 1:2.41:4.82$ if the sample is homogeneous. If there are large relatively inaccessible regions in the sample, then the ratios of $T_{3/4}$ and $T_{1/2}$ to $T_{1/4}$ will be much larger. On the other hand, minor inaccessible regions (say 5%) would probably not affect these ratios, so that the choice of fractional times would need to be modified.

Polarised infrared spectroscopy has also been used to provide information about the structure of polysaccharides, notably in celluloses (46,47,48) and β-1,4 xylans (49). In this method of investigation, oriented samples, films or fibres, are used. An infrared chromophore absorbs radiation most strongly when the electric vector of the radiation is parallel to the transition moment of the vibration. Thus, it is possible, by means of polarised radiation, to find the orientation of the transition moment causing a given peak, relative to the fibre, or macromolecule, axis. When the origin of the peak is known, it is then possible to assign an orientation to the chromophore with respect to the fibre axis. Conversely, if the general configuration of the molecule is assumed, peaks can be assigned to chromophores on the basis of their orientations. In practice, both methods are generally used. Above about 1500 $cm.^{-1}$ the chromophores are generally widely spaced and easily assigned, so that this region is mostly used to find chromophore orientation. Below this wavelength, however, the peaks tend to be crowded, and the main use of polarised radiation in this region is the assignment of peaks to chromophores.

Although polarised infrared spectroscopy can thus be useful over the whole range of the spectrophotometer, its chief use is in the investigation of the orientations of hydrogen bonds in the polymers. These may be inferred from the X-ray and infrared data for homologous oligomers, but they can be directly investigated only by the use of polarised infrared spectroscopy.

Four distinct methods have been used to prepare oriented samples. These, with the instances relevant to the present

study are:

(a) Pressing aligned fibres into a sheet; Ramie fibres (cellulose IB) (46).

(b) Stretching a wet film of polysaccharide; bacterial cellulose (cellulose IA) (47).

(c) Spreading a viscous solution of polysaccharide on a smooth surface and allowing it to dry; cellulose II (47) and xylans (49).

(d) Spreading a suspension of insoluble crystallites, and allowing it to dry; Ramie crystallites (48).

In addition, the oriented film produced in the latter cases may be modified by reagents. Thus, cellulose II is produced by saponification of cellulose acetate film, produced as in (c) (47), and cellulose III_I is produced by the action of liquid ammonia or ethylamine on oriented bacterial cellulose (47). It is generally necessary, too, to deuterate the sample before investigation of the hydroxyl peaks.

The ratio of the optical intensities ($I_{\parallel}, I_{\perp}$) of a peak when the electric vector of the radiation is parallel and perpendicular respectively to the molecular axis, is termed the dichroic ratio R. This has been related (50) to the angle, α, between the transition moment and the axis, and to the disorder, by considering the sample as containing a perfectly ordered fraction, f, and a perfectly disordered fraction, 1-f. The relationship is

$$R = \frac{f.\cos^2\alpha + (1-f)}{\frac{1}{2}f.\sin^2\alpha + (1-f)} \qquad \text{(vi)}$$

It is instructive to see how R varies with α, over the whole range of order from f=0 (perfect disorder) to f=1 (perfect order).

$\alpha = 0^\circ$ $\quad R = \frac{1+2f}{1-f}$ Range 1 to ∞

$\alpha = 45^\circ$ $\quad R = \frac{2(2+f)}{4-f}$ Range 1 to 2

$\alpha = 54^\circ 44' = \tan^{-1}(\sqrt{2})$; $R = 1$

$\alpha = 90^\circ$ $\quad R = \frac{2(1-f)}{2+f}$ Range 1 to 0

Hence, a chromophore with a transition moment parallel, or nearly parallel, to the axis, will have a large value of I_l and a small value of I_r; if the transition moment is approximately perpendicular to the axis, the converse will be true; and if it is at an angle of about 55°, I_l and I_r will be equal. Thus, for a given sample (so that f is constant), some relative estimate of the azimuthal angles of chromophores can be made.

The relationship between the optical density, I, and the angle θ between the axis and the electric vector is sinusoidal. Using subscripts l, r to represent parallel and perpendicular parameters respectively, the electric vector at any angle θ is given by

$$E^2 = E_l^2.\sin^2\theta + E_r^2.\cos^2\theta$$

Since $I \propto E^2$. this becomes

$$I = I_l.\sin^2\theta + I_r.\cos^2\theta$$

This can be rearranged, to give the relationship

$$I = \tfrac{1}{2}(I_l + I_r) + \tfrac{1}{2}(I_l - I_r)\cos 2\theta \qquad \text{(vii)}$$

The extrema of this curve occur when $dI/d\theta = 0$, i.e. at $\theta = 0^\circ$, 90°, 180° etc Hence, the positions of the extrema are independent of I_l and I_r, but the sizes and natures of the extrema are not.

EXPERIMENTAL

Experiment 2.1 General characteristics of spectra

The spectra of several polysaccharides, examined as films on glass slips, were recorded during deuteration (see General Methods). The polysaccharides which were used were:

(i) Cellulose II. This was presented by the acidification of a solution of filter paper in Schweizer's reagent.

(ii) Alginic acids. These were prepared by acidification of sodium alginate solutions. A guluronic acid-rich sample from Laminaria hyperborea and a mannuronic acid-rich sample from L. digitata, were provided by Alginate Industries Ltd.

(iii) 2-Hydroxyethyl alginate. This was prepared by the action of ethylene oxide on mannuronic acid-rich alginic acid.

(iv) Gum tragacanth. The same material was used as previously (Expt. 1.19, et seq.)

(v) Citrus pectin. This was kindly supplied by Mr. Ian Cottrell.

(vi) Pectic (acidic) polysaccharide from mustard-seed cotyledons. This was extracted from the cotyledons of (a) ungerminated, and (b) germinated mustard seeds with EDTA solution (51), after removal of lipids with aqueous ethanol and acetone. After dialysis, protein was removed from the crude extract with phenol, and the polysaccharide was isolated by freeze-drying.

(vii) Mustard seed-coat mucilage Fraction B (Expt. 1.1.)

Results:

Polysaccharide	O-H stretch	C-H stretch	O-D stretch
(i) (see Fig. 2.1)	(a) undeuterated; fairly broad peak centred on sharp peak at 3440 cm^{-1}.	Sharp peak centred at 2900 cm^{-1}	deuterated peak relatively narrow, centred at 2530 cm^{-1}
	(b) deuterated; peak still large, fine features apparent.		

Results (contd.)

Polysaccharide	O-H stretch	C-H stretch	O-D stretch
(ii), (vi a) (see Fig. 2.1)	(a) undeuterated; peak broad or very broad centred at 3360-3480 $cm.^{-1}$ (b) deuterated; peak still present; no fine features in (ii); sharper peak at 3300 $cm.^{-1}$ in (vi a).	peak sharp to almost swamped by O-H peak; centred at 2920-2930 $cm.^{-1}$	(a) undeuterated; unassigned peak centred at 2500-2630 $cm.^{-1}$ not always visible, but shown by initial decrease around 2500-2600 $cm.^{-1}$ during deuteration. (b) deuterated; peak not always clearly visible centred at 2500-2540 $cm.^{-1}$
(iii), (iv), (v) (vi b) (see Fig. 2.1.)	(a) undeuterated; fairly broad peak centred at 3380-3440 $cm.^{-1}$ (b) deuterated; peak absent.	Fairly sharp peak centred at 2930 $cm.^{-1}$	deuterated peak fairly broad centred at 2500-2530 $cm.^{-1}$
(vii) (see Fig. 1.3)	(a) undeuterated; rather narrow peak centred at 3345 $cm.^{-1}$ (b) deuterated; peak still large, features like cellulose I.	Fairly sharp peak, truncated or slightly split, apices at 2915, 2895 $cm.^{-1}$	Deuterated peak rather narrow centred at 2510 $cm.^{-1}$

Experiment 2.2. Investigation of the pectic peak at 3300 $cm.^{-1}$

The peak at 3300 $cm.^{-1}$ in germinated cotyledon pectin was suspected to be due to the N-H stretch of protein contaminants, since the N-H peak of proteins, like the pectin peak (a) occurs

at 3300 $cm.^{-1}$; (b) is narrower than amorphous O-H peaks; and (c) is less accessible to proton exchange. That it was, in fact, due to this, was confirmed by the following tests:
(i) Crude mustard-seed protein, isolated from phenol extract, was examined by infrared. It gave a similar peak at 3300 $cm.^{-1}$
(ii) A sample of polysaccharide, on a silver chloride slip, was wetted with methanolic HCl (0.5%). After being dried, it was examined by infrared during deuteration. This showed the 3300 $cm.^{-1}$ peak almost disappeared, a new peak at 3080 $cm.^{-1}$ and a narrow peak appearing, after some time, at 2470 $cm.^{-1}$ It is suggested that these changes are due to the following reactions:

$$N\text{-}H + HCl \rightarrow NH_2^+$$

(peak at 3300 cm^{-1})

$$NH_2^+ + D_2O \rightarrow NHD^+$$

(3080 cm^{-1})

$$NHD^+ - N\text{-}D$$

(2470 cm^{-1})

Experiment 2.3 Investigation of the pectic peak at 2500 $cm.^{-1}$

This peak was suspected to be due to chelated water or secondary hydroxyl groups, which would give a peak in this region. Further evidence for this was obtained by the following tests:
(i) A sample with a large peak at 2500 $cm.^{-1}$ was dried overnight, in a high vacuum at about 40°C, over phosphoric oxide. When the sample was re-examined, the peak was absent.
(ii) Samples of calcium methyl β-D-glucopyranosiduronate (film; see Fig. 2.2.) and $BaCl_2.2H_2O$ (Nujol mull) were examined by infrared. These had peaks at 2680 $cm.^{-1}$ and at 2720 and 2660 $cm.^{-1}$ respectively.

Experiment 2.4. Test for crystallinity

Series of spectra, run at noted times during deuteration, were produced for the samples used in Expt. 2.1. From these, graphs of the decrease with time of the hydroxyl groups were produced, which gave the $T_{\frac{1}{4}}$, $T_{\frac{1}{2}}$ and $T_{\frac{3}{4}}$ (see Introduction to Section).

Results:

Polysaccharide	$T_{\frac{1}{4}}$, $T_{\frac{1}{2}}$, and $T_{\frac{3}{4}}$ minutes and ratio (theoretical amorphous ratio = 1:2.41:4.82)	Conclusion
(i)	0.67, 5.0, very large =1:7.5: -	crystalline
(ii) (mannuronic)	10,240, very large = 1:24: -	crystalline
(ii) (guluronic)	5,15, very large =1:3: -	crystalline
(iii)	2,4,9.5 =1:2:4.75	Non-crystalline
(iv)	2.25, 5.0, 10.0 =1:2.22:4.45	Non-crystalline
(v)	0.5, 1.25, 2.5 =1:2.5:5.0	Non-crystalline
(vi)(a)	0.67, 1.5, 3.5 =1:2.25:5.25	Non-crystalline
(vi)(b)	0.5, 1.0, 4.5 =1:2:9	Crystalline*
(vii)	7.5, 100 very large =1:13.3: -	Crystalline

*See Expt. 2.2

Experiment 2.5 Calculations of rate constants and crystallinity

The rate constants and crystallinities were calculated, from equation (v) (see Introduction to Section), by the method outlined in the introduction for the following polysaccharides:

(i) Cellulose II

(ii) "Cloustonii" alginic acid

(iii) Gum tragacanth

(iv) Mustard seed mucilage, fraction B.

See Expt. 2.1 for details of samples (i)-(iv).

(v) Ultracentrifuge precipitate from the mucilage

(vi) Ultracentrifuge supernatant from the mucilage

(v) and (vi) were prepared by ultracentrifugation of a mucilage solution for one hour at 20,410 rpm (see Expt. 1.16).

(vii) Acid-precipitated cellulose I (see Expt. 1.1).

(viii) Very heavy fraction isolated with the preparative ultracentrifuge

(ix) Heavy fraction isolated as (viii)

(x) Light fraction isolated as (viii).

For samples (viii)-(x) see Expt. 1.18.

The scattering correction was calculated from the transmittances of the cell, with and without the sample at 4000 $cm.^{-1}$ The hydroxyl transmittances were all measured at 3360 $cm.^{-1}$ The calculations were made by computer, using five iterative cycles to compute accurate values. In the case of cellulose II, the values of the parameters were printed after each cycle, and showed that, as might be expected, the values are virtually unaltered, so that iteration is likely to be unnecessary.

Results:

Sample	K_x ($hr.^{-1}$)	K_a ($hr.^{-1}$)	n (%)
i	0.2283	11.3599	19.67
ii	0.0231	5.2495	23.11
iii	1.0737	19.3186	15.98
iv	0.1507	7.8446	57.21
v	0.0529	43.9579	60.71
vi	0.2414	17.2416	18.15
vii	0.4178	27.9473	47.39
viii	0.1506	176.4587	51.91
ix	0.1512	34.8739	65.65
x	0.1739	36.1719	58.90

Experiment 2.6 Polarised spectroscopy of Fortisan fibres.

The method used for the examination of fibres was an adaptation of that used by Cumberbirch and Spedding (52). The fibres were wound around a small metal plate with bevelled edges (to avoid cutting them), and teased into a continuous thin layer. They were held in place with a piece of sellotape on one side, and wetted, and covered with a single layer of aluminium foil. The sample was then placed in a KBr-disc press, and left overnight under a pressure of about 25 tons/sq.in. This produced an oriented sample generally stuck to the foil, from which it could be removed when dry. The sample was deuterated for about 24 hours, then immersed in carbon tetrachloride or hexachlorobutadiene to reduce the scattering. If carbon tetrachloride was used, it needed to be replenished frequently during infrared examination, to replace evaporation losses. The spectrometer was held at a

given wavelength, and the transmittance at that wavelength was recorded for a series of angles of polarisation. This enabled a graph to be drawn up of the variation of transmittance with polarisation angle. This method was used, rather than single measurements of maximum and minimum transmittances in order to give more accurate values of the maximum and minimum.

Results (see also Fig.2.2):

Peak	Assignment	Dichroic ratio	Orientation	ref. 47
3500	OH (s)	1.15	//	//
3450	OH (s)	1.05	//	//
3300	OH (s)	0.80	⊥r	⊥r
3200	OH (s)	0.62	⊥r	⊥r
2930	CH_2 (as)	0.62	⊥r	-
2900	CH (s)	0.61	⊥r	-
2880	CH_2 (ss)	0.59	⊥r	-
2500	OD (s)	0.84	⊥r	-
1380	COH (b)	1.03	//	-

s = stretch; as = antisymmetric stretch; ss = symmetric stretch; b = in-plane bend

Experiment 2.7 Polarised spectroscopy of the mustard-seed mucilage

A thick solution (2 or 3%) of mucilage was placed between barium fluoride plates. These were then slid apart under acetone, and left there for 30 seconds. This process gives two oriented films, on plates which can be placed, with the oriented films parallel and facing inwards, to observe the polarised spectrum of the mucilage during deuteration. The transmittances for various angles of polarisation were recorded as in the previous experiment.

Results:

Peak	Assignment	Dichroic ratio	Orientation
3350	OH (s)	1.086	//
1600	COO^- (as)	0.902	$\perp^r$

DISCUSSION

Experiments 2.1 to 2.3 have shown that infrared spectroscopy yields some useful information regarding the structures of polysaccharides. The shape and position of the hydroxyl peak appears to be quite unrelated to the nature of the polysaccharide, but the changes in it which occur during deuteration give some information regarding its state of aggregation. It seems legitimate to distinguish three types of aggregation on this basis, viz:

(a) Amorphous (polysaccharides iii, iv, v, vib). The hydroxyl groups are completely and readily accessible to deuteration, and no fine features appear in the spectrum at any stage. This type of aggregation occurs in well-branched polysaccharides, and also in 2-hydroxyethyl alginate. In branched polysaccharides, the packing of monosaccharide residues must be fairly random, and there is no chance of extensive areas of ordered packing, with strong hydrogen-bonds, such as occur in cellulose. 2-Hydroxyethyl alginate is a linear ester, but it presumably occurs in a disordered state, due to (i) the presence of bulky 2-hydroxyethyl groups, and (ii) the attachment of these at two distinct configurations, depending on whether the uronic acid residue is mannuronic or guluronic.

(b) Occluded (polysaccharide ii). The hydroxyl groups are not completely accessible to deuteration, but no fine features are visible in the hydroxyl peak. This type of aggregation is shown by the alginic acids, which are linear 1,4-linked polymers of D-mannuronic and L-guluronic acids. Because of their linearity, these are capable of closer and more orderly packing than are branched polysaccharides, and are probably not truly amorphous. Nonetheless, they do not appear to be

crystalline, at least in the samples used in Expt. 1.2. This is presumably due to the disorder within the chain caused by the presence of two types of residue.

(c) Crystalline (polysaccharide i). The hydroxyl groups are not completely accessible, and fine features occur in the hydroxyl peak, at least after partial deuteration. This type of aggregation is shown by cellulose. The fine features are caused by the presence of large numbers of hydroxyl groups having the same strength. Since this strength is both distance and angle-dependent, it follows that there must be a considerable degree of three-dimensional order, i.e. crystallinity, in order for them to occur.

These states of aggregation will be definable points on a continuum rather than three distinct species.

Another feature which is of interest is the hydroxyl peak at about 2500 $cm.^{-1}$ in the spectrum of germinated cotyledon pectin. This is due to the presence of hydrated calcium ions, or to chelated hydroxyl groups or water molecules. The fact that the peak disappeared after rigorous dehydration [Expt. 2.3 (i)] suggests that it is due to water. This evidence is not conclusive, however; removal of water might merely have caused the collapse of a hydroxyl-chelated structure. Whatever the exact cause, it is likely to be biologically significant. Certainly it is sensitive to minor changes in structure for, although the compositions of ungerminated and germinated pectins are similar (51), only the latter gives this peak. Further investigation shows that germinated pectin contains a component which is insoluble in water, but soluble in EDTA. It is this fraction which is responsible for the peak at 2500 $cm.^{-1}$

The cause of the hydroxyl absorption as low as 2500 $cm.^{-1}$ in alginic acids is not clear; it could be due to the presence of residual hydrated or chelated metal ions. More probably, however, it is caused by weak hydrogen bonds, which are rapidly deuterated, or to water, which becomes deuterated, or vapourised by the heat of the infrared beam.

It appears, from the results of Expt. 2.1 that cellulose has its main C-H absorption at 2900 $cm.^{-1}$, whereas the acidic polysaccharides have theirs at 2920-2930 $cm.^{-1}$, but its relation to structure is not known. It is impossible to say whether this is truly related to their structures, or is merely accidental.

Of the quantitative methods of analysis of deuteration rates, the simpler method (Expt. 2.4) based on the ratios of the quarter-, half-, and three quarter-times of deuteration, seems to be very successful. It clearly differentiates between amorphous polysaccharides and occluded or crystalline ones. The second method, however, Expt. 2.5, the calculation of rate constants and crystallinities, does not appear to have been very successful. Samples (iv) to (vi) and (viii) to (x), for example, are all mucilage fractions, and should all have closely similar crystalline rate constants. They are, in fact, fairly similar (mean = 0.1565, standard deviation = 0.0559 (35.76%), but still show appreciable variation. The crystallinity values, too, are often unrealistic. It is to be expected that the amorphous rate constant will be variable, since it is likely to be affected by minor changes in the preparation of the sample. On the other hand, the crystalline rate constant should be fairly constant in a given type of

polysaccharide. There are two possible reasons for this failure. Firstly, the assumption that $[D_2O]$ is constant, made in the deviation of $c = c_o.\exp(-Kt)$ (eqn. iii) may be faulty. It will be faulty if the rate of flow of D_2O vapour is too slow, so that it is not always in excess. It is also likely to be faulty if two different deuterations are compared, since the D_2O readily back-exchanges with atmospheric H_2O, and hence may be effectively diluted in some runs. The second possible reason for this failure lies in the assumptions made in the deviation of eqn. (v) in order to allow for scattering.

The orientations of transition moments of chromophores in Fortisan (cellulose II; Expt. 2.6) show a general resemblance to those of corresponding chromophores in Ramie (cellulose I; refs. 46,47). One interesting feature, which does not appear to have been noted previously, is the orientation of the 'amorphous' OD groups, perpendicular to the fibre. This is, however, a reasonable result. The 'amorphous' OD groups which are rapidly formed during deuteration, will be weakly hydrogen bonded, and would therefore, be expected to be long-range intermolecular bonds, since intramolecular bonds would tend to be shorter, and therefore stronger. It is noticeable that the dichroic ratio (0.84) is higher (i.e. there is probably less orientation) than for any other perpendicular chromophore.

Only two peaks, viz. those due to hydroxyl and carboxyl stretches, were investigated in the mucilage (Expt. 2.7). The 3350 $cm.^{-1}$ peak, due to cellulose I, was parallel to the orientation, a finding which agrees with that of Tsuboi (46). The carboxylate ion is found to be perpendicularly orientated. It follows from this result that the acidic component of the mucilage is oriented by the experimental method used, though it

is impossible to say whether or not this is due to close association with the aligned microfibrils.

SECTION III

Introduction

X-ray crystallography has shown (53) that pyranoses and their derivatives normally occur in the C1 conformation, and it can be assumed that this is the case in polysaccharides. Theconformations, i.e. the relative spatial positions of adjacent residues in a polysaccharide chain are, therefore, dependent only on the various types of non-bonded interaction between the residues, and on similar interactions with neighbouring molecules. If the sizes of these interactions are known, then the most stable conformation can easily be predicted.

The simplest method of representing one type of interaction, the van der Waals, is by assigning each atom a definite radius, the van der Waals radius. If two atoms approach to within a distance less than the sum of their van der Waals radii, the interatomic interactions cause strong repulsion. Thus, this method represents atoms as hard spheres of definite radii. This approach can be applied most simply by the construction of molecular models, and the measurement of interatomic distances. Permissible conformations will have no non-bonded atoms closer than the sum of their van der Waals radii. Hermans (14) used this method to predict the 'bent-chain' structure for cellulose as long ago as 1943. It was later used by Pauling and his co-workers (54) to predict the α-helix of proteins and polypeptides, and was important (13) in the elucidation of the structure of DNA. It is obvious. therefore, that the estimation of non-bonded interactions, even by this simplest method, is capable of yielding valuable information about the conformation of macromolecules.

One disadvantage of this method is that commercially available models have standard and invariable bond-lengths and

angles. Pauling (54) surmounted the first difficulty by using models based on the structures of peptides as obtained from X-ray diffraction studies. An alternative solution is to represent the atoms by Cartesian co-ordinates, which can be obtained from X-ray work, or constructed from the standard bond-lengths and angles. The conformation can then be altered by standard mathematical operations. The advantages of this method are the increased accuracy of the interatomic measurements, and the ease with which bond-lengths and bond-angles can be changed. It is also more amenable to systematic investigation of all possible conformations. This method has been used to investigate peptides (55), cellulose (55,56), β-1,4 xylan (57), and carrageenan (29).

The main limitation of the hard sphere method is that it does not give a quantitative estimate of the interactions; it suggests either an infinite repulsion, or no interaction at all. In practice, it is made semi-quantitative by using two radii instead of one to represent repulsion. Two atoms separated by a distance greater than, or equal to, the sum of their outer radii do not repel at all. There is slight repulsion if the distance lies between the sums of the outer and inner radii; and at distances less than the sum of the inner radii there is strong repulsion. The outer radius is generally termed the 'fully-allowed' radius; the inner radius is the 'marginally-allowed' or 'outer limit' radius.

Quantitative estimates of the total interaction between different residues in a polymer chain can be made if the potential equations are known. The earliest work in this direction was performed on synthetic hydrocarbon polymers and peptides by Brant and Flory (58) and Liquori and co-workers (59). These

calculations require the use of an electronic computer, and have been mostly concerned with synthetic polymers and polypeptides, though studies of amylose (60) and its triacetate (61) have been made.

Attention has centred mainly on van der Waals interactions in these studies, largely because of the large repulsion forces which they cause at small interatomic distances. Unfortunately the equations representing van der Waals interactions remain semi-empirical, and most of the groups of workers in this field have produced their own versions. The preferred form of equation is the Buckingham type $V=-A/r^6 + B.\exp.(-\mu r)$, though some workers use the Lennard-Jones type, $V=-A/r^6 + B/r^{12}$, or some other form. The maximum attractive van der Waals potential between two atoms, at about 3 to 4 Ångstroms, is a fraction of a kilocalorie per mole, but repulsive potentials can be thousands of kcal./mole.

Next to van der Waals interactions, dipolar interactions and hydrogen bonding are most important. Dipolar interactions can be treated accurately, since there are theoretically derived functions to represent them. In practice, however, most workers have adopted the simplifying expedient of regarding each atom as a monopole, and calculating the potential between point charges. The dipolar potentials between two atoms a few Å apart will be up to a few kcal./mole, and will be attractive or repulsive, depending on the signs of the atomic monopoles.

No equation, soundly based on theory, exists which represents the potential energy of hydrogen bonding. Several semi-empirical equations are known, however. The maximum attractive potential due to hydrogen bonding is about five or six kcal./mole.

At small interatomic distances, the potential is large and repulsive.

The least important of the non-bonded interactions is the torsional potential. Unlike the other non-bonded interactions torsional potential cannot cause large repulsive forces. It ranges in size from zero to about three kcal./mole, and is always repulsive.

Because of the approximations involved, and the semi-empirical nature of the potential functions, this method is not yet as accurate as experiment for macromolecules, although it is claimed to be so for small hydrocarbons (62). For macromolecules, a difference of no more than about 10° between the experimental and theoretical conformations is regarded as fair agreement.

The work described in this section is aimed at applying the methods of macromolecular conformational analysis to polysaccharides. The purpose of this is, firstly, to discover the relative importance of the various types of non-bonded interaction in determining the properties of polysaccharides; secondly, to provide theoretical confirmation of the known conformations of some well-known polysaccharides; and thirdly, to develop potential functions, mathematical methods, and computer programs, with which to predict the conformations of other polysaccharides. The work is concerned very largely with cellulose and cellobiose. There are several reasons for this. Firstly, cellulose is chemically well-characterised and simple; it is a linear molecule with only one type of linkage. Secondly, it is an extended molecule. It does not curve back on itself so that remote residues come into contact. This means that it is sufficient to consider interactions between neighbouring molecules. Thirdly, cellobiose was the

only disaccharide obtained from a polysaccharide, whose structure had been accurately determined by X-ray crystallography, at the commencement of this work. Fourthly, some preliminary results of conformational analysis already exist for cellulose (55,56), which are useful as checks. Fifthly, it is, of course, of great biological and industrial importance.

The first part of the work is concerned with the derivation of atomic co-ordinates and conformations from experimental data. The hard-sphere approach, and van der Waals interactions are then considered, followed by dipolar and torsional potentials and hydrogen-bonding. A new van der Waals function for oxygen interactions is derived, and the total potential, resulting from the sum of all types of non-bonded interaction, is calculated. Finally, a beginning is made on the treatment of polysaccharides which, unlike cellulose, can curve back on themselves.

CHAPTER I

Introduction

In the earliest conformational studies of polysaccharides, atomic co-ordinates were derived by assuming standard bond-lengths and tetrahedral bond-angles. Settineri (57) calculated the co-ordinates mathematically; Jones (56) and Anderson (29) constructed models, and obtained the atomic co-ordinates from the shadows cast by a parallel light source. It is now known, however, that some bond-lengths and angles differ appreciably from standard values (53,63), so that these treatments are inevitably approximate. Ramachandran et al.(55), realizing this, compared the co-ordinates of various pyranose structures which had been accurately determined by X-ray crystallography. The survey showed that bond-lengths and angles are fairly constant within this group. A standard residue was therefore constructed by averaging the co-ordinates of the various structures. This standard residue included co-ordinates from sugars such as L-arabinopyranose. In order to include these co-ordinates, it is necessary to (a) alter the chirality of the co-ordinate axes; and (b) modify the lengths of bonds C_5-H_5 (eq.) (to correspond to C_5-C_6), and C_4-O_4 and C_4-H_4 (to invert the configuration at C_4). The latter operation is likely to introduce some minor errors into the co-ordinates. Thus, Ramachandran's method still contains approximations, although it is a great improvement on the earlier studies.

The glucose residues in cellodextrins are likely to be a good approximation to those in cellulose, and it might be possible to obtain atomic co-ordinates from these. So far, no accurately determined structures have been published, and a

preliminary report on the structure of cellotetraose (64) suggests that it may be difficult, if not impossible, to obtain sufficiently accurate structures. The best compromise structure, therefore, would appear to be that of cellobiose, which has been accurately determined (65,66,67).

In the present work four sets of co-ordinates have been used:

(i) the Ramachandran co-ordinates (55), which include co-ordinates from Brown's cellobiose work (66);

(ii) Brown's cellobiose co-ordinates (66);

(iii) Chu and Jeffrey's cellobiose co-ordinates (67);

(iv) a set of co-ordinates, derived from Chu and Jeffrey's (67), which are intended to represent the situation in cellulose more accurately.

Ramachandran did not give co-ordinates for the hydrogens, and these were calculated as in his subsequent work on amylose (60), using the bond-angles given there.

The studies of Jones and Settineri used variable bridge angles. Anderson gave the bridge angle the tetrahedral value. Ramachandran, on the basis of X-ray determinations of disaccharide structure, used a bridge angle of 117.5°.

Different workers have also based their approaches on different principles. Jones and Settineri were interested only in conformations compatible with the experimental data; their rotations were about the line O_1-O_4, and were thus expressed in terms of only one variable angle (see Fig. 3.1). Ramachandran expressed conformations in terms of two angles, representing rotations of the residues about their bonds to the bridge oxygen, as did Anderson. Their axes, however, were defined differently. Anderson defined the origin as the bridge oxygen, with the x-axis

lying along $O\text{-}C'_4$ for the reducing residue, and along $C_1\text{-}O$ for the non reducing residue, and H_1 and H'_4 lying in the xz planes at the zero conformation. For the co-ordinates which he published, Ramachandran defined the origin as C_3, the x-axis as along $C_3 \ldots C_5$, and C_1 in the xy plane, thus giving left-handed axes (Ramachandran says that they are right-handed, but has since pointed out that, on this basis, the co-ordinates refer to L-glucose). It is not clear whether Ramachandran used these axes in his calculations, but, in the work on amylose (60), he used axes very similar to those of Anderson. Ramachandran's zero conformation is taken to be that in which $C_4C_1OC'_4C'_1$ are coplanar; very similar, though not identical, to that of Anderson. The Anderson axes require the simplest rotation matrices possible, and can themselves be related by a simple rotation matrix. Accordingly, the present work also uses these axes. Previous definitions of zero conformation are ambiguous, in that they describe either $\phi=\psi=0^\circ$ or $\phi=\psi=180^\circ$, where ϕ and ψ are the rotations about the bonds to the non reducing and reducing residues respectively. In view of this, it is worthwhile to define unambiguously the axes and rotations used in the present work on cellobiose.

Origin at the bridge oxygen for both residues. Both sets of axes are right-handed.

Reducing residue x-axis along $O\text{-}C'_4$, in that sense. H'_4 in the xz plane, and pointing in the same sense as the apex of the angle $C_1\text{-}O\text{-}C'_4$ when in the zero conformation. A positive rotation, ψ, about the x-axis, is seen as an anticlockwise rotation of the atoms of the reducing residue, when viewed along $C'_4\text{-}O$.

Non reducing residue. x-axis the extrapolation of C_1-O, in that sense. H_1 in the xz plane, and pointing in the opposite sense to the apex of the angle C_1-O-C_4' when in the zero conformation. A positive rotation, ϕ, about the x-axis, is seen as a clockwise rotation of the atoms of the non reducing residue, when viewed along C_1-O (see Fig. 3.1).

In order to relate the results of conformational analysis to X-ray diffraction data for cellulose and xylan, it is necessary to obtain relationships between ϕ and ψ and the repeat distance and screw axis. Ramachandran appears to have used an approximation, namely that C_1, O_1, C_4, and O_4 are coplanar, in order to derive the relationships. Miyazawa (68) derives more general relationships for polymers with one to six distinct atoms in the repeating unit, all of which are joined by bonds which allow free rotation. Cellulose and xylan, on this basis, are best regarded as polymers of the type $(-M_1-M_2-M_3-)_n$, where M_1=O_1, M_2=C_1 and M_3=C_4.

X-ray studies of celluloses (28) have established that the repeat distance is 10.30-10.34 Å. They have also been generally interpreted as evidence of a 2-fold screw-axis, but it has been pointed out (69) that this is not necessarily so. Hence, to comply with the experimental evidence, preferred conformations should (a) give a repeat distance of 10.30-10.34 Å and (b) possibly have a 2-fold screw axis.

The xylan study (57) shows that this has a repeat distance of 14.84 Å, and a 3-fold or 6-fold screw-axis. Conformational analysis by the same author suggests that the screw-axis is left-handed, and 3-fold.

In the analyses of Jones and Settineri, only two variables were studied at any one time, namely the rotation about O_1-O_4, and one other, e.g. interatomic distance or bridge angle. The results could therefore be represented as a two-dimensional graph.

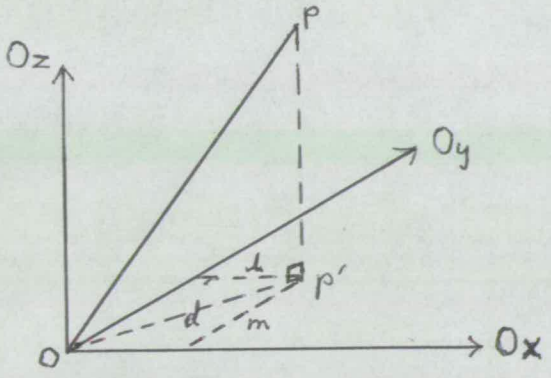
In the studies by Ramachandran and Anderson, however, there were three variables, namely, ϕ, ψ and interatomic distance. Hence, these results were not represented by a graph but by a map, the third variable being plotted as a contour.

The Ramachandran work on cellulose had the disadvantage of not considering all relevant interatomic interactions. In particular, H_1 and H_4' were omitted. Jones considered the $H_1 \ldots H_4'$ distance, and Settineri considered both this and $O_2 \ldots H_4'$, but their studies were restricted to the limits imposed by the experimental data, whilst that of Ramachandran looked at the whole area around the most probable conformation.

The work described in this chapter aims to provide accurate atomic co-ordinates from available crystallographic data, and also accurate bridge angle and ϕ and ψ values from the cellobiose data. Approximate and accurate maps of the variation of repeat distance and screw-axis are obtained, and the Ramachandran work is repeated, on axes as defined above, in order to link the two studies. Finally, a complete hard-sphere map for cellulose is obtained.

Derivation of a rotation matrix

A rotation matrix which is often used in the following work is one which rotates the axes so that the x-axis becomes parallel to a line originally having direction cosines l,m,n. This matrix can be derived as follows:



$$d = \sqrt{l^2 + m^2}$$

(a) rotate about Oz so that Ox coincides with OP';

(b) rotate about Oy so that Ox coincides with OP.

These rotations are given (70) by

$$(a)\quad \begin{bmatrix} l/d & m/d & n/d \\ -m/d & l/d & . \\ . & . & 1 \end{bmatrix}, \qquad (b)\begin{bmatrix} d & . & n \\ . & 1 & . \\ -n & . & d \end{bmatrix}$$

Hence the total rotation is represented (71) by

$$R=\begin{bmatrix} d & . & n \\ . & 1 & . \\ -n & . & d \end{bmatrix}\cdot\begin{bmatrix} l/d & m/d & . \\ -m/d & l/d & . \\ . & . & 1 \end{bmatrix}=\begin{bmatrix} l & m & n \\ -m/d & l/d & . \\ -ln/d & -nm/d & d \end{bmatrix} \qquad \text{(viii)}$$

The reciprocal matrix is also useful. It is given by (71) $R^{-1}=adj(R)/det(R)$ where adj(R) is the determinant of R.

This becomes

$$R^{-1} = \begin{bmatrix} l & -m/d & -ln/d \\ m & l/d & -mn/d \\ n & . & d \end{bmatrix} \qquad \text{(ix)}$$

EXPERIMENTAL

Calculation 3.1. Modification of Ramachandran co-ordinates

A set of atomic co-ordinates was derived, using the Ramchandran figures (55) except that only the figures relating to β-D-glucose residues (5 sets) were used. This was done in the following stages:

(i) averaging of the 5 sets of co-ordinates;

(ii) transference of the origin to the bridge oxygen. If the co-ordinates of the bridge oxygen are given by $\underline{x}_o$ the new co-ordinates are given by $\underline{x}' = \underline{x} - \underline{x}_o$ ($\underline{x}$ etc. underlined denotes the column vector $\{x\ y\ z\}$);

(iii) rotation of the Ox axis, using the R matrix (eqn. viii), to coincide with one of the O-C bonds.

Since these co-ordinates did not include hydrogen atoms, these were calculated, using a C-H bond length of 1.1 Å, and the angles given by Ramachandran in his study of amylose (60). The direction cosines of the other three bonds from a carbon atom can be calculated from the atomic co-ordinates. Then, if l,m,n are the direction cosines of the C-H bond, l_i, m_i, n_i (i=1 to 3) are the direction cosines of the other bonds, and c_i is the cosine of the angle between C-H and bond i, we have three equations of the type $ll_i + mm_i + nn_i = c_i$, or $L\underline{l} = \underline{c}$, where L is the 3 x 3 matrix of the other direction cosines. Therefore, $\underline{l} = L^{-1}\underline{c}$. The co-ordinates of each hydrogen, $\underline{x}$, were then calculated from $\underline{x} = \underline{x}_o + l.d$, where $\underline{x}_o$ is the position of the carbon atom, and d is the bond length.

Finally, the sets of co-ordinates were rotated about Ox so that the appropriate hydrogen lay in the xz plane. If y,z are co-ordinates of the hydrogen, the rotation is an angle of $\tan^{-1}(y/z)$.

For results (set A co-ordinates) see Table 3.1.

Table 3.1. Set "A" Co-ordinates

	Non-reducing			Reducing		
	x	y	z	x	y	z
C_1	-1.390	0.000	0.000	3.820	0.147	-1.401
C_2	-1.890	-1.251	0.687	3.470	-1.147	-0.759
C_3	-3.400	-1.229	0.756	1.980	-1.246	-0.598
C_4	-3.910	0.067	1.269	1.420	0.000	0.000
C_5	-3.260	1.289	0.633	1.930	1.285	-0.592
C_6	-3.610	2.608	1.265	1.530	2.561	0.096
O_1	0.000	0.000	0.000	5.200	0.256	-1.492
O_2	-1.440	-2.418	0.090	3.960	-2.263	-1.429
O_3	-3.850	-2.349	1.459	1.660	-2.410	0.053
O_4	-5.320	0.174	1.165	0.000	0.000	0.000
O_5	-1.850	1.139	0.637	3.330	1.243	-0.702
H_1	-1.740	0.000	-1.030	3.400	0.157	-2.422
H_2	-1.500	-1.241	1.717	3.920	-1.148	-0.231
H_3	-3.760	-1.288	-0.273	1.560	-1.285	-1.618
H_4	-3.650	0.098	2.348	1.750	0.000	1.070
H_5	-3.580	1.309	-0.408	1.520	1.335	-1.623

Calculation 3.2. Derivation of co-ordinates from the Brown and Jeffrey studies.

Atomic co-ordinates were derived from the crystal data of Brown (66) and Chu and Jeffrey (67) in the following stages:
(i) transformation from crystal co-ordinates to rectangular co-ordinates on an Ångstrom scale. In cellobiose, β is fractionally over 90°. If a,b,c are the unit cell dimensions, the new co-ordinates are given by

$$x' = a.x - c.z\sin(\beta-90)$$

$$y' = b.y$$

$$z' = c.z.\cos(\beta-90)$$

(ii) transference of the origin to the bridge oxygen, and rotation of Ox to coincide with one of the C-O bonds, as in stages (ii) and (iii) of Calc. 3.1.;
(iii) rotation about Ox so that the appropriate hydrogen lies in the xz plane, as in Calc. 3.1.

This procedure was then repeated, except that in stage (ii), Ox was made to coincide with the other O-C bond.

For results (sets B and C co-ordinates) see Table 3.2.

Calculation 3.3. Derivation of standard co-ordinates

The co-ordinates of atoms close to the glycosidic link in cellobiose are likely to be similar to those in cellulose. At the reducing and non-reducing extremities of the molecule, however, they may well be different. Accordingly, a "standard residue" was constructed from the two residues of the Jeffrey determination as follows: The origins of both sets of co-ordinates were transferred to C_3. Oy was rotated to lie along $C_3 \ldots C_5$, and C_2 was made to lie in the xy plane. Co-ordinates were then combined as follows (see Fig. 3.1):

Table 3.2. Set "B" Co-ordinates

	Non-reducing			Reducing		
	x	y	z	x	y	z
C_1	-1.373	0.000	0.000	3.718	0.150	-1.748
C_2	-1.894	-1.177	0.782	3.476	-1.162	-0.988
C_3	-3.370	-1.161	0.796	1.967	-1.140	-0.694
C_4	-3.867	0.130	1.407	1.467	0.000	0.000
C_5	-3.187	1.359	0.748	1.903	1.315	-0.700
C_6	-3.463	2.663	-1.566	1.652	2.530	0.126
O_1	0.000	0.000	0.000	5.086	0.278	-1.960
O_2	-1.375	-2.372	0.159	3.731	-2.262	-1.880
O_3	-3.968	-2.283	1.439	1.760	-2.435	0.037
O_4	-5.267	0.252	1.252	0.000	0.000	0.000
O_5	-1.791	1.195	0.721	3.314	1.237	-0.874
H_1	-1.760	0.000	-1.009	3.173	0.171	-2.680
H_2	-1.523	-1.129	1.795	4.069	-1.224	-0.087
H_3	-3.698	-1.172	-0.234	1.397	-1.331	-1.622
H_4	-3.633	0.129	2.462	1.826	0.000	1.019
H_5	-3.557	1.478	-0.260	1.414	1.408	-1.658

Table 3.2 Set "C" Co-ordinates

	Non-reducing			Reducing		
	x	y	z	x	y	z
C_1	-1.397	0.000	0.000	3.701	0.287	-1.763
C_2	-1.894	-1.193	0.811	3.422	-1.070	-1.153
C_3	-3.414	-1.179	0.796	1.946	-1.196	-0.815
C_4	-3.919	0.157	1.356	1.446	0.000	0.000
C_5	-3.254	1.338	0.642	1.877	1.335	-0.606
C_6	-3.532	2.641	1.372	1.618	2.511	0.289
O_1	0.000	0.000	0.000	5.055	0.448	-1.970
O_2	-1.332	-2.391	0.308	3.710	-2.106	-2.085
O_3	-3.966	-2.286	1.507	1.777	-2.419	-0.136
O_4	-5.323	0.216	1.157	0.000	0.000	0.000
O_5	-1.825	1.193	0.651	3.293	1.312	-0.844
H_1	-1.776	0.000	-0.934	-3.297	0.290	-2.799
H_2	-1.511	-1.146	1.787	3.980	-1.185	-0.291
H_3	-3.578	-1.234	-0.157	1.321	-1.264	-1.673
H_4	-3.729	0.127	2.388	1.773	0.000	0.940
H_5	-3.567	1.397	-0.314	1.315	1.390	-1.503

Kept: C_1, H_1, C_4', H_4', O (bridge)

Averaged: $C_2, C_3, C_5, C_6, H_2, H_3, H_5, O_2, O_3, O_5$ from both residues

Rejected: $C_4, H_4, O_4, C_1', H_1', O_1'$

Finally, the new set of standard co-ordinates was transferred back to the former axes, as in Calc. 3.1.

Obviously, only the co-ordinates of relatively rigidly-held atoms could be treated in this way, since the positions of those which are free to rotate would be different in the two residues; but in any case, these are not yet needed.

For results (set D co-ordinates) see Table 3.3.

Calculation 3.4. Calculation of some important parameters for cellobiose and lactose.

Cellobiose parameters which are of interest in the present work are: the O_3'-O_5 distance, across which an intramolecular hydrogen-bond can form; the bridge angle, ß and Ø and Ø'. The O_3'-O_5 distance and ß are given in all the published work, but Ø and Ø' need to be calculated. In Calc. 3.2., atomic co-ordinates were calculated for the entire molecule for each set of axes. Ø and Ø' could easily be calculated from these since Ø, for instance, is the dihedral angle between C_1-H_1 and O-C_4'. On the "Ø" axes, C_1-H_1 lies in the xz plane, so that the dihedral angle is simple $\tan^{-1}$ (y/z), where y,z are co-ordinates of C_4'.

Ø and Ø' were also calculated from the data of Lipscomb et al. (65), although a complete set of co-ordinates was not calculated.

C. A. Beevers and H. N. Hansen have obtained preliminary crystallographic data for α-lactose (4-O-ß-D-galactopyranosyl D-glucose) monohydrate, which has the same configuration as

Table 3.3. Set "D" Co-ordinates

	Non-reducing			Reducing		
	x	y	z	x	y	z
C_1	-1.397	0.000	0.000	3.619	0.263	-1.860
C_2	-1.894	-1.193	0.810	3.421	-1.067	-1.138
C_3	-3.413	-1.168	0.795	1.945	-1.200	-0.800
C_4	-3.955	0.168	1.296	1.446	0.000	0.000
C_5	-3.254	1.356	0.640	1.879	1.327	-0.621
C_6	-3.557	2.663	1.325	1.656	2.501	0.297
O_1	0.000	0.000	0.000	4.970	0.409	-2.182
O_2	-1.332	-2.392	0.308	3.930	-2.128	-1.925
O_3	-3.916	-2.273	1.539	1.720	-2.435	-0.129
O_4	-5.353	0.273	0.946	0.000	0.000	0.000
O_5	-1.826	1.192	0.651	3.274	1.314	-0.963
H_1	-1.776	0.000	-0.934	3.025	0.338	-2.672
H_2	-1.511	-1.146	1.786	4.024	-1.100	-0.281
H_3	-3.704	-1.277	-0.165	1.439	-1.219	-1.673
H_4	-3.853	0.290	2.280	1.773	0.000	0.940
H_5	-3.577	1.360	-0.341	1.328	1.417	-1.490

cellobiose at all asymmetric centres except C_4 and C_1'. Hydrogen atoms have not yet been fixed, but a rough estimate of ϕ and ψ can be made by a comparison of the dihedral angles of O_1-C_4' and C_1-C_2 (for ϕ) and C_1-O_1 and C_4'-C_5' (for ψ) in cellobiose and lactose. This comparison was kindly made by Dr. D. A. Rees.

Results:

	$O_3'\ldots O_5$ (Å)	β	ϕ	ψ
cellobiose (65)	2.80	117.5	-56.88	168.25
" " (66)	2.772	116.7	-44.44	168.18
" " (67)	2.767	116.1	-42.33	162.13
lactose	2.84	118.5	-31	155

Calculation 3.5. Derivation of the Ramachandran map.

A hard-sphere map was produced, using set A co-ordinates, and measuring the same interatomic distances as did Ramachandran, namely $C_1\ldots C_3'$, $C_1\ldots C_5'$, $C_4'\ldots O_5$, and $C_4'\ldots C_2$. Van der Waals radii were as Calc. 3.8, and the bridge angle was 117.5°.

ψ was given an initial value, and ϕ was varied over the whole 360° range, at 10° intervals. At each value of ϕ, the interatomic distances were measured. When ϕ had covered the whole range, ψ was increased by 10°, and the process was repeated. Thus, a comprehensive map was produced, covering a 360° range for both ϕ and ψ.

The atomic co-ordinates in the reducing residue, with respect to the "ψ" axes, at a rotation of ψ, are given by $\underline{x}' = S\underline{x}$, where $\underline{x}$ is the position at $\psi = 0$, and $S = \begin{bmatrix} 1 & . & . \\ . & \cos\psi & \sin\psi \\ . & -\sin\psi & \cos\psi \end{bmatrix}$

Similarly, the co-ordinates in the non-reducing residue, with respect to the "ϕ" axes, are given by $\underline{x}' = P\underline{x}$, where

$$P = \begin{bmatrix} 1 & . & . \\ . & \cos\phi & -\sin\phi \\ . & \sin\phi & \cos\phi \end{bmatrix}$$

To make the axes equivalent, a further rotation must be applied to allow for the bridge angle. If α is the supplement of the bridge angle, the "ϕ" axes must be rotated clockwise through an angle α. Thus, the atomic co-ordinates in the non-reducing residue become $\underline{x}' = AP\underline{x}$, where $A = \begin{bmatrix} \cos\alpha & . & -\sin\alpha \\ . & 1 & . \\ \sin\alpha & . & \cos\alpha \end{bmatrix}$

Now that the axes are equivalent, the distance between any two atoms can be calculated by the three-dimensional Pythagoras formula. This was done for all the required contacts. If any interatomic distance was less than the sum of the outer-limit van der Waals radii, the conformation was disallowed. If any interatomic distance was greater than this, but still less than the sum of the fully allowed radii, the conformation was marginally allowed. If all interatomic distances were greater than this, the conformation was fully allowed.

A computer program was written to give a conformational map, by printing a different symbol according to whether the conformation was disallowed, marginally, or fully, allowed (see Fig. 3.2). The resultant map (see Fig.3.2) was similar to Ramachandran's (55) except that an additional marginally allowed area was found, outside of the area that he investigated, and the signs of our ϕ and Ramachandran's ϕ' were opposite, due to the fact that he used L-glucose co-ordinates (see Introduction to Chapter).

Calculation 3.6. Qualitative test for hydrogen bonding.

In order to form strong hydrogen-bonds, two oxygen atoms need to be between 2.5 and 2.8 Å apart. At distances greater than this, weak hydrogen bonds may still form. Intramolecular hydrogen bonding in cellulose, between O_5 and O_3', can be qualitatively assessed by measuring the distance at various conformations.

A conformational map was produced by the same principles, as those used in Calc. 3.5, except that the $O_5 \ldots O_3'$ distance was calculated at each conformation and printed. The 2.5 and 2.8 Å contours were added to the hard-sphere map, as in Ramachandran's work, with which the results were closely comparable. See Fig. 3.2.

Calculation 3.7. Approximate repeat distances and screw axes, and the experimental conformation.

Approximate expressions for repeat-distance and screw-axis multiplicity can be obtained by assuming that C_1-O_1 and O_4-C_4 areparallel. The rotations ϕ, ψ and α can be replaced by a single rotation, θ, of the non reducing residue, about a new x-axis. Suppose that the new axis has direction cosines, l, m, n with respect to the old axes, and that the new origin is at X,Y,Z. Five stages are necessary in order to represent on the old axes a single rotation of θ about the new axis, viz:

(i) displacement of the origin to X,Y,Z;
(ii) rotation of Ox to lie along l,m,n;
(iii) anticlockwise rotation θ;
(iv) reciprocal rotation of (ii);
(v) displacement of origin to old origin.

Considering steps (i) and (v) first, and representing the

operations (ii) to (iv) by a matrix M step (i) is given by $\underline{x}'=\underline{x}-\underline{X}$; steps (ii) to (iv) are given by $\underline{x}'' = M\underline{x} = M(\underline{x}-\underline{X})$; and step (v) is given by $\underline{x}'' = M(\underline{x}-\underline{X}) - \underline{X}'$, where $\underline{X}'$ is the position of the old origin. But this was $-\underline{X}$ after step (i), and $M(-\underline{X})$ after steps (ii)-(iv). Therefore, $\underline{x}'' = M(\underline{x}-\underline{X})+M\underline{X}=M\underline{x}$. Thus the double translation of origin has no effect on the result, and the operation can be regarded as purely rotational, steps (ii) to (iv). Step (ii) is represented by matrix R (eqn. viii), and step (iv) is represented by R^{-1}. Therefore, the entire operation is represented by $R^{-1}TR$ where $T = \begin{bmatrix} 1 & . & . \\ . & \cos\theta & \sin\theta \\ . & -\sin\theta & \cos\theta \end{bmatrix}$

If we put $\cos\theta = C$, $\sin\theta = S$, $R^{-1}TR$ becomes

$$\begin{bmatrix} l^2(1-C)+C & lm(1-C)-nS & ln(1-C)+mS \\ lm(1-C)+nS & m^2(1-C)+C & mn(1-C)-lS \\ ln(1-C)-mS & mn(1-C)+lS & n^2(1-C)+C \end{bmatrix}$$

If we now put $a = \cos\tfrac{1}{2}\theta$, $b = \sin\tfrac{1}{2}\theta$, $c = m.\sin\tfrac{1}{2}\theta$, $d = n.\sin\tfrac{1}{2}\theta$, the matrix becomes

$$\begin{bmatrix} a^2+b^2-c^2-d^2 & 2(bc-ad) & 2(bd+ac) \\ 2(bc+ad) & a^2-b^2+c^2-d^2 & 2(cd-ab) \\ 2(bd-ac) & 2(cd+ab) & a^2-b^2-c^2+d^2 \end{bmatrix} \qquad (x)$$

This is similar to the matrix derived by Ramakrishnan (72), but differs in that Ramakrishnan worked on the basis of rotating axes, rather than of rotating atoms.

This matrix, then, represents a rotation θ of the non reducing residue, whilst the reducing residue remains stationary. In terms of ϕ, ψ and α, this is $S^{-1}AP$, since we are now holding the reducing residue stationary, i.e., instead of rotating the

reducing residue an angle Ψ about C_4-O_4, we are rotating the non reducing residue an angle $-\Psi$ about that bond. This matrix is

$$\begin{bmatrix} \cos\alpha & -\sin\alpha.\sin\phi & -\sin\alpha.\cos\phi \\ \sin\Psi.\sin\alpha & \cos\phi.\cos\Psi + \cos\alpha.\sin\phi.\sin\Psi & -\sin\phi.\cos\Psi + \cos\alpha.\cos\phi.\sin\Psi \\ \cos\Psi.\sin\alpha & -\cos\phi.\sin\Psi + \cos\alpha.\sin\phi.\cos\Psi & \sin\phi.\sin\Psi + \cos\alpha.\cos\phi.\cos\Psi \end{bmatrix} \quad \ldots. \text{(xi)}$$

Since (x) and (xi) are equivalent, the corresponding elements of each are equivalent, and hence θ can be related to ϕ, Ψ and α by equating suitable elements.

An n-fold screw-axis is defined as one along which successive corresponding entities (molecules or parts of molecules) are related by a rotation of $360/n^{\circ}$ about that axis. Thus, in the present case, $\theta = 360/n$. Also $a = \cos\frac{1}{2}\theta = \cos 180/n = \frac{1}{2}\sqrt{(m_{23}-m_{32})(m_{31}-m_{13})/(m_{12}+m_{21})}$, where m_{ij} is the matrix element from row i, column j.

Using the matrix elements of (xi), this simplifies to $a = \cos\frac{1}{2}(\Psi-\phi).\sqrt{\frac{1}{2}(1 + \cos\alpha)}$. Therefore, the two solutions, in terms of n, are

$$\left.\begin{aligned} \Psi-\phi &= 2\cos^{-1}[\cos(180/n).\sqrt{2}/(1 + \cos\alpha)] \\ \Psi-\phi &= 360-2\cos^{-1}[\cos(180/n).\sqrt{2}/(1 + \cos\alpha)] \end{aligned}\right\} \ldots\ldots \text{(xii)}$$

The projected residue length is calculated, rather than the repeat distance, since the latter is the product of the projected residue length and the screw-axis multiplicity, and hence a combination of two variables, which are more comprehensible if kept separate.

The projected residue length, D, is given by the projection on to l,m,n of the distance between equivalent atoms in adjacent residues. The direction ratios of the line joining the two atoms are $\underline{x}_r - M\underline{x}_{nr}$, where M is matrix (xi), and $\underline{x}_r, \underline{x}_{nr}$ are

the positions of corresponding atoms in the reducing and non reducing residues respectively. Hence, $D = L(\underline{x}_r - M\underline{x}_{nr})$ (xiii) where $L = [l\ m\ n]$, since, if d is the true distance between the atoms, the direction cosines of their join are $1/d\ (\underline{x}_r - M\underline{x}_{nr})$, and the cosine of the angle, γ, between this and the axis is $1/d\ (\underline{x}_r - M\underline{x}_{nr})$. Therefore, the projected distance, $D = d\cos\gamma = L(\underline{x}_r - M\underline{x}_{nr})$.

Since the relationship between Ψ and ϕ is linear for a given n, a screw-axis map was produced by calculation of $2\cos^{-1}[180/n.)\sqrt{2/(1 + \cos\phi)}]$ for various n, and drawing the corresponding straight lines (see Fig. 3.2).

The formula for D is more complex, and it was evaluated by a computer at various conformations, to give a residue-length map (Fig. 3.2). Results were similar to those of Ramachandran (55).

The experimental conformation of cellulose was calculated by assuming a two-fold screw-axis. The residue length is then 5.15 - 5.17 A. Using set A co-ordinates and bridge angle, this gives two results compatible with the experimental data, namely:

$$\phi = -32^\circ,\ \Psi = 148^\circ$$
$$\phi = 19^\circ,\ \Psi = 199^\circ$$

Similarly, the xylan results (n = 3, projected residue length = 4.93) are compatible with

$\phi = -80,\ \Psi = 172$ } left-handed
$\phi = -6,\ \Psi = 246$ }

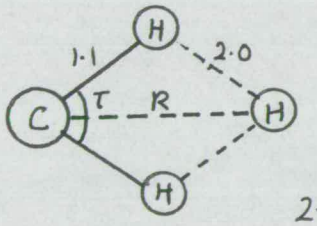
$\phi = -5,\ \Psi = 104$ } right-handed
$\phi = 65,\ \Psi = 174$ }

Calculation 3.8. Complete hard-sphere map for cellulose.

A hard-sphere conformational map was calculated for cellulose as in Calc. 3.5, except that all relevant interatomic distances were considered (see Fig. 3.3). The van der Waals radii which were used were:

atom pair	fully allowed	marginally allowed
C...C	3.2	3.0
C...O	2.8	2.7
C...H	2.4	2.2
O...O	2.8	2.7
O...H	2.4	2.2
H...H	2.0	1.9
H...CH_2OH	2.42	2.42
O...CH_2OH	2.90	2.90
C...CH_2OH	3.20	3.20

The X...CH_2OH distances were calculated as the minimum distance of approach to a methylene group, as shown for H...CH_2OH:

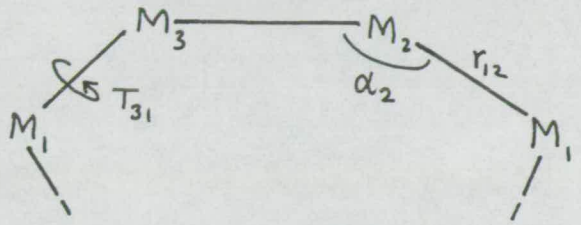


$$2{\cdot}0^2 = 1{\cdot}1^2 + R^2 - 2{\cdot}2\,R\cos\tau/2$$

The same distance was used for 'fully-allowed' and 'marginally allowed' contacts, and the C...C distance was used for the C...CH_2OH distance, since this was greater than the distance calculated on the basis of the diagram.

Calculation 3.9. Accurate calculation of projected residue length and screw-axis.

Miyazawa (68) gives expressions for n and D, based on a series of atoms linked in a cis conformation, thus:



T_{ij} is the angle of rotation of the helix about the bond between M_i and M_j, r_{ij} is the length of the bond M_i-M_j, and α_i is the inter-bond angle at atom M_i.

The initial problem, therefore, is to reduce cellulose to a linear cis conformation, i.e. atoms O_1,C_1,C_4, O_4,C_1',C_4', etc. must all lie in the xz plane. To bring C_4 into the plane, it must be rotated anticlockwise through an angle of $\tan^{-1}(y/z)$, where y,z are co-ordinates of C_4 with respect to the 'ϕ' axis. Thus, $T_{12} = 180 - \phi + \tan(y/z)$. The minus sign before ϕ is due to the fact that Miyazawa's rotation, and ours, are in opposite senses. The 180° is introduced because the chain is in the trans conformation at C_1.

T_{23} is given by the dihedral angle between C_1-O_1 and C_4-O_4 with respect to $C_1 \ldots C_4$. T_{23} is positive if O_4 is clockwise with respect to O_1.

The correction to ψ is a clockwise rotation of $\tan^{-1}(y'/z')$ where y',z' are co-ordinates of C_1 with respect to the 'ψ' axes. Therefore, $T_{31} = \psi - \tan^{-1}(y'/z')$ ψ is still positive because, unlike ϕ, it is a rotation of $-\psi$ of the non reducing end of the polymer, which is equivalent to a rotation of $-\psi$ of the non reducing end.

Maps were produced over the whole range of possible conformations, varying ϕ and ψ by 10° intervals, and over the allowed region as

given by Calc. 3.8 ($\phi = -80^\circ$ to 40°, $\psi = 120^\circ$ to 200°), varying ϕ and ψ by 2° intervals using set D co-ordinates.

Results (see Fig. 3.3.)

$T_{12} = 180 - \phi + 7.38$; $T_{23} = -9.08$; $T_{31} = \psi + 8.05$

$r_{12} = 1.397$; $r_{23} = 2.873$; $r_{31} = 1.446$

$\alpha_1 = 116.1$; $\alpha_2 = 152.83$; $\alpha_3 = 139.15$

Conformations compatible with the experimental data were:

Cellulose	$\phi = -25^\circ$	$\psi = 146^\circ$	
	$\phi = 25^\circ$	$\psi = 207^\circ$	
Xylan	$\phi = -79^\circ$	$\psi = 161^\circ$	left-handed
	$\phi = -25^\circ$	$\psi = 225^\circ$	left-handed
	$\phi = 13^\circ$	$\psi = 117^\circ$	right-handed
	$\phi = 82^\circ$	$\psi = 195^\circ$	right-handed

Discussion

The work described in this chapter has, first of all, provided sets of atomic co-ordinates which should represent the shape of residues in cellulose fairly accurately, and has calculated convenient parameters for sets of experimental conformations, by which the accuracy of theoretical results can be checked.

Four sets of atomic co-ordinates have been derived. Sets A, B and C (see Calcs. 3.1, 3.2) are derived from monosaccharides and disaccharides, and this means that C_1' and C_4', and their attached groups, are likely to be somewhat displaced from their positions in cellulose since, in the latter case, they are chemically different because they are attached to further sugar residues. These would alter the van der Waals and dipolar interactions in the regions of C_1' and C_4'. Set D (see Calc. 3.3) was constructed to allow for this, by using the co-ordinates of C_1 and C_4' and their attached atoms, from cellobiose, since the interactions about the bridge oxygen should be identical to those in cellulose. Set D co-ordinates should, therefore, represent the situation in cellulose most accurately.

Calculations of ϕ and ψ have also been made for the conformations of cellobiose; that from the work of Lipscomb *et al*. (65) gave results which were rather divergent from those from the later work (66,67). Lipscomb's structure was less refined than the later ones and is, therefore, probably rather less accurate. The structures by Brown (66) and Chu and Jeffrey (67) are more concordant, and indicate that the true conformation is in the region $\phi = -42^\circ$ to -44°, $\psi = 162^\circ$ to 168°. The

conformation of lactose ($\phi = -31^{\circ}$, $\psi = 155^{\circ}$) differs from this, but until the structure has been refined, it is impossible to say whether this is a significant difference.

The possible conformations of cellulose and xylan have been calculated in terms of ϕ and ψ, by both an approximate, and an accurate method using n and D values normally assumed for these polysaccharides. This shows that, in the allowed region, the two methods give quite similar results, although they are more divergent outside of this region.

Hard-sphere maps for cellulose (Calcs. 3.5, 3.8) enable a choice to be made between the various possible conformations of cellulose and xylan, especially when used in conjunction with the hydrogen bonding criterion (Calc. 3.6). This shows that the likeliest conformation for cellulose is $\phi = -25^{\circ}$, $\psi = 146^{\circ}$, which corresponds to the Hermans conformation (14). This is only on the edge of the allowed area, but is stabilised by an intramolecular hydrogen bond. The alternative conformation, however ($\phi = 25^{\circ}$, $\psi = 207^{\circ}$) and also the Meyer and Misch (73) conformation ($\phi = 0^{\circ}$, $\psi = 180^{\circ}$) are disallowed.

The allowed area is likely to be larger for xylan than the cellulose, but will otherwise be similar. It can therefore be seen that, in agreement with earlier work (57), one of the solutions ($\phi = -79^{\circ}$, $\psi = 161^{\circ}$) which lies on the left-handed 3-fold screw-axis is preferred on a hard-sphere basis; and this is also the solution which can form the best intramolecular hydrogen bond. Thus, the work described here is concordant with earlier experimental and theoretical studies.

The hard-sphere maps also suggest an explanation of the observed stiffness of cellulose and cellodextrins (74) in solution. When all relevant interatomic distances are

considered, only 1.9% of the conformations are fully allowed, and a further 1.6% marginally allowed. Even when only some of the distances are considered, as in Ramachandran's study (see Calc. 3.5) the corresponding figures are only 3.5% and 5.5%, Ø and Ø can thus vary over only small ranges, and the molecule is therefore stiff.

From these calculations, it is evident that useful results can be obtained from the hard-sphere approach. It is, however, obviously desirable to be able to express van der Waals interactions more quantitatively, and also to consider other forms of interaction. Having calculated accurate co-ordinates and experimental conformations, and devised computer techniques, in the present chapter, subsequent chapters will deal with quantitative approaches to this problem.

CHAPTER II

Introduction

The preliminary approach to conformational analysis, the hard-sphere approach, which was described in the previous chapter, depends for its validity on the repulsion which occurs between atoms at very short distances. This is due to van der Waals interactions, which are caused by the interaction of the electrons of the two atoms. At distances greater than the sum of the van der Waals radii, the electrons distribute themselves in such a way that the two atoms form a pair of fluctuating dipoles which attract one another. These attractive forces are referred to as London, or dispersion, forces. At smaller interatomic distances, the electron clouds overlap, and repulsion increases so rapidly that the hard-sphere is a good approximation. The two types of interaction can be described by a function of the form $V = -A/r^6 + f(r)$, where r is the interatomic distance, and A and f(r) are dependent on the identities of the atoms involved. The attractive potential, $-A/r^6$, is a theoretically derived relationship, but the repulsive potential, f(r) is empirical. There are two chief groups of van der Waals equation, differing in the analytical form of f(r); these are the Lennard-Jones, or "6 - 12" type, and the Buckingham, or "6-exp" type.

The Lennard-Jones equation has the form $V = A/r^6 + B/r^{12}$. This has the advantage of containing only two parameters, one of which can be calculated theoretically, so that only one remains to be calculated from experimental data. Lennard-Jones equations have been used by Scheraga and co-workers (75) in their studies of polypeptides.

The Buckingham equation has the form $V = -A/r^6 + B.\exp(-\mu r)$. This contains three parameters, only one of which can be calculated theoretically. In addition, it makes the unrealistic

prediction that the potential will be large and negative when r is close to zero. Nonetheless, it has been used more frequently in conformational studies. Two major versions of the Buckingham equation have been used, which differ in the value assigned to μ for different atoms. Brant and Flory (58) gave μ a constant value of 4.6, derived from molecular beam data. A was derived theoretically, by Pitzer's modification of the Slater-Kirkwood formula (76) and B was derived by assuming that $dV/dr = 0$ when r = the sum of the van der Waals radii. Other workers have used a formula containing three universal constants and two atomic parameters, of the form $V = e(-A/z^6 + B.\exp(-cz))$, where e is a characteristic energy, and $z = r/r_o$; r is the interatomic distance, and r_o is a characteristic radius. This type of equation was introduced by Hill (77) and a similar type has been derived by Kitaygorodsky (78).

Allinger et al. (62, 79) have used latent heat and conformational data to derive new values of e and r_o for carbon and hydrogen. They claim that, using these, various parameters can be predicted for small hydrocarbons with an accuracy comparable to experiment. Unfortunately, they did not consider oxygen, so that their equations are not immediately applicable to carbohydrate work. The method of deriving e and r_o can be extended to oxygen, but it is likely to be complicated by the fact that most oxygen compounds are dipolar, and hence dipolar interactions must be taken into account.

Liquori et al. (80, 81) have used more general functions, of the form $V = a.\exp(-br)/r^d - c/r^6$, where a,b,c,d are parameters which depend on the nature of the interacting atoms. Similar

functions have been used by Thompson et al.(82). This type of function reduces to a Lennard-Jones function, when b = 0 and d = -12, or to a Buckingham function, when d = 0. They have been successful in the investigation of polypeptides, but have recently been shown (J. Mark, quoted in ref. 83) to give poor agreement with experiment when applied to the isomerisation of butene-2. Liquori (80) also introduced functions of the type $V = ar^{-b}$, where $6<b<12$, for close-range interactions between some atoms, reserving the other type of function for more distant interactions ($r >$ sum of van der Waals radii).

The use of van der Waals functions in conformational analysis has been mostly confined to polypeptides and hydrocarbons, and there have only been two previous polysaccharide studies which employed them. Ramachandran et al.(60) studied the conformations of amylose, and used both Brant and Flory (58) and Kitaygorodsky (78) functions, thus providing a valuable comparison of the two types of function. It was found that the potential contours were fairly similar in shape, whichever type of function was used, but the Flory minimum was about 4 kcal./mole lower than the Kitaygorodsky minimum. The hard-sphere contour for amylose is closer to the Kitaygorodsky zero contour than to the Flory zero contour, which suggests that the Kitaygorodsky functions may be rather more accurate for polysaccharides. Both types of function, however, support the view (84) that V amylose occurs as a left-handed helix with, typically, six glucose residues per helix turn.

The work of Sarko and Marchessault (61) calculated the conformations of the glucose residues with respect to one another, and also the conformations of the acetate groups, in amylose

triacetate. Comparison of the energy minima for left- and right-handed helices showed that the former was far more stable; and calculations of the van der Waals interactions between chains, assuming parallel and antiparallel arrangements respectively, showed that the antiparallel arrangement was energetically favoured. Furthermore, the minimum was close to the conformation suggested by X-ray data only in the latter case. Liquori functions were used in energy calculations.

From both of these studies, it is evident that the calculation of van der Waals interactions can be of great value in the study of polysaccharide conformations.

This chapter describes the calculations of the positions and energies of conformational minima of cellobiose, and the energies of experimental conformations, in the hard-sphere allowed area, using the functions of Flory, Kitaygorodsky, and Liquori, and the various sets of co-ordinates derived in the previous chapter. It then describes the calculation of a potential map for xylan, based on cellobioses and Kitaygorodsky functions. Finally, it describes the derivation of a new van der Waals function for oxygen, based on latent heat and conformational equilibrium data, and its application, together with the functions derived by Allinger et al. (79).

EXPERIMENTAL

Calculation 3.10. Conformational maps for cellobiose in the hard-sphere allowed region.

Conformational maps were produced, over the allowed region of the hard-sphere map, using various bridge-angles, sets of co-ordinates, and sets of functions.

The sets of parameters for the cellobiose molecule (atomic co-ordinates plus bridge angle) were as follows:

Set 1. Set A co-ordinates, bridge angle, 117.5°

Set 2. Set B co-ordinates, bridge angle, 116.7°

Set 3. Set B co-ordinates, bridge angle, 117°

Set 4. Set C co-ordinates, bridge angle, 116.1°

Set 5. Set C co-ordinates, bridge angle, 117°.

[A sixth set, using C co-ordinates, and a bridge angle of 116.7°, gave results intermediate between sets 4 and 5, which have been reported elsewhere (85)]

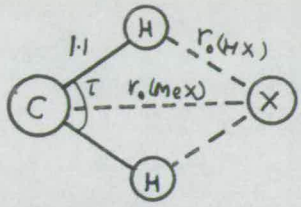
The functions which were used were the "Flory" functions (58,60), the "Kitaygorodsky" functions (60,78) and the "Liquori" functions (81,86, misprinted in ref. 80). The parameters used in the Flory functions were as given by Ramachandran (60), with additional parameters for interactions involving the hydroxymethyl group. This was regarded as a methylene "atom", and values of A were calculated from the date given by Flory, using the Slater-Kirkwood formula (76). Values of B were calculated by assuming that $dV/dr = 0$ when methylene was separated from the other atom by a distance equal to the sum of the van der Waals radii, again using the data of Flory (58). Values obtained were:

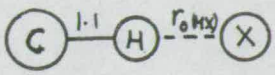
H...CH_2 interactions: $A = 718.5$, $B = 5.31 \times 10^5$

O...CH_2 " " $A = 2016$, $B = 3.10 \times 10^6$

C...CH_2 " " $A = 2605$, $B = 6.75 \times 10^6$

The r_o values used in the Kitaygorodsky functions were as given by Ramachandran (60). Values of r_o for interactions involving the hydroxymethyl group were calculated, using these values, and averaging the two distances obtained as shown:





$$r^2_{o(HX)} = 1.1^2 + r^2_{o(MeX)} - 2.2\, r_{o(MeX)} \cos \tau/2$$

$$r_{o(MeX)} = 1.1 + r_{o(HX)}$$

Values obtained were:

H...CH_2 interactions: r_o = 3.33

O...CH_2 " " r_o = 3.76

C...CH_2 " " r_o = 3.92

The Liquori functions already included functions involving methyl groups.

The calculations proceeded as in calc. 3.8, except that ϕ and ψ were varied in 2^o increments. At each conformation, the relevant interatomic distances were measured. If any of these showed the conformation to be disallowed, a zero was printed, and the next conformation was considered. If the conformation was allowed, its energy was calculated by summing the potentials between all relevant pairs of atoms, i.e. between C'_3, C'_4, C'_5, (CH_2OH), O'_3, H'_{3_1}, H'_4 and H'_5 on the reducing residue and C_1, C_2, O_2, O_5, H_1 and H_2 on the non-reducing residue, except that the $C_1...C'_4$ interaction, being the same for all conformations, was ignored.

Conformational maps were produced by drawing contour lines through points of equal potential (see Fig. 3.5). Table 3.4 lists the positions and energies of the minima, using different sets of co-ordinates and functions, and the energies of

Table 3.4

Energy functions	Conformation	Co-ordinates used															
		Set 1			Set 2			Set 3			Set 4			Set 5			
		ϕ°	ψ°	V (Kcal)	ϕ°	ψ°	V (Kcal)	ϕ°	ψ°	V (Kcal)	ϕ°	ψ°	V (Kcal)	ϕ°	ψ°	V (Kcal)	
Flory	Minima	-31	+181	-4.9	-27	+184	-5.2	-29	+184	-5.2	-24	+175	-5.0	-24	+181	-5.1	
		+19	+149	-5.0	+21	+147	-5.1	+22	+148	-5.2	+23	+141	-5.2	+24	+141	-5.3	
Kitay-gorodsky	Minima	-46	+180	0	-41	+181	-0.5	-43	+181	-0.6	-41	+175	-0.3	-40	+175	-0.5	
		+ 5	+147	+0.4	+ 1	+147	-0.3	+ 1	+147	-0.4	0	+143	-0.4	0	+143	-0.6	
Liquiri	Minima	-76	+175	-1.1	-70	+182	-2.2				-72	+176	-1.7				
		+31	+147	-1.0	+32	+146	-2.5				+32	+139	-2.8				
Flory	Cellobiose(B)	-44	+168	-4.4	-44	+168	-4.6	-44	+168	-4.6	-44	+168	-4.6	-44	+168	-4.6	
	Cellobiose(J)	-42	+162	-4.1	-42	+162	-4.3	-42	+162	-4.3	-42	+162	-4.4	-42	+162	-4.5	
Kitay-gorodsky	Cellobiose(B)	-44	+168	+0.2	-44	+168	-0.3	-44	+168	-0.4	-44	+168	-0.3	-44	+168	-0.4	
	Cellobiose(J)	-42	+162	+0.5	-42	+162	-0.1	-42	+162	-0.2	-42	+162	-0.2	-42	+162	-0.4	
Liquori	Cellobiose(B)	-44	+168	+2.0	-44	+168	+0.1				-44	+168	+0.2				
	Cellobiose(J)	-42	+162	+2.6	-42	+162	+0.5				-42	+162	+0.5				
Flory	Cellulose(H)	-32	+148	-3.6	-33	+147	-4.1	-33	+147	-4.0	-39	+141	-3.0	-40	+140	-3.5	
	Cellulose(M)	0	+180	+14.5	0	+180	+6.9	0	+180	+6.6	0	+180	+8.9	0	+180	+7.9	
Kitay-gorodsky	Cellulose(H)	-32	+148	+0.9	-33	+147	+0.7	-33	+147	+0.1	-39	+141	+0.8	-40	+140	+0.4	
	Cellulose(M)	0	+180	+5.2	0	+180	+2.3	0	+180	+2.2	0	+180	+2.6	0	+180	+2.3	
Liquori	Cellulose(H)	-32	+148	+3.3	-33	+147	+1.0				-39	+141	+0.3				
	Cellulose(M)	0	+180	+11.5	0	+180	+6.5				0	+180	+6.5				

B = Brown cellobiose conformation; J = Jeffrey cellobiose conformation;
H = Hermans cellulose conformation; M = Meyer and Misch cellulose conformation.

experimentally-determined conformations.

Calculation 3.11. Total conformational map for cellobiose.

A conformational map was produced as in Calc. 3.10, except that energies were calculated whether or not the conformation was allowed. ϕ and ψ were varied, in 10° increments, over the whole 360° range. Set 4 co-ordinates, and Kitaygorodsky functions, were used.

This map (Fig. 3.6) shows that, as well as the negative-energy area, examined in Calc. 3.10, there is a large area of low positive energy ($V<10$ kcal./mole), around $\psi = 170^{\circ}$, from $\phi = -180^{\circ}$, to $\phi = 50^{\circ}$; a smaller area around $\phi = 10^{\circ}$, $\psi = 0^{\circ}$; and, especially, a small minimum ($V<5$ kcal./mole) around $\phi = -160^{\circ}$, $\psi = 180^{\circ}$.

Calculation 3.12. Conformational map for xylan.

Set 4 co-ordinates were modified by shortening the $C_5' - C_6'$ distance to 1.1Å. A conformational map (see Fig. 3.6) was then produced, varying ϕ and ψ by 4° increments, and using Kitaygorodsky functions, those representing CH_2OH interactions being replaced by functions representing the corresponding H interactions. The new co-ordinates for H_5' (eq.) were:

$x = 1.693$ $y = 2.166$ $z = 0.02546$

The energies of the four conformations compatible with the experimental data ($n = 3$, $D = 4.93$ Å), as given by Calc. 3.7, were also calculated, using the Kitaygorodsky functions.

Results:

	ϕ	ψ	V
	-80	172	1.6
	-6	246	4.6
	-5	104	6.7
	65	174	15.2
Minima:	-32	204	-0.4
	16	151	-0.4

Calculation 3.13. A new van der Waals function for oxygen.

Van der Waals equations, as used by Allinger et al. (62,79), are of the form $V = e(-A/z^6 + B.\exp(-cz))$, where $z = r/r_o$. e and r_o are the only parameters which differ between atoms. It is obvious, from the form of the equation, that an equation giving e explicitly can readily be obtained. On the other hand, an explicit equation for r_o cannot be obtained. The procedure, therefore, in seeking values of e and r_o compatible with experimental data is to calculate $-A/z^6 + B.\exp(-cz)$ [f(z)] for various values of r_o, and then to calculate e as V/f(z), where V is the experimental energy. Thus a graph of e against r_o can be drawn. If this is done for more than one set of experimental data, the graphs should intersect, thus giving values of e and r_o agreeable to all sets of data.

It is difficult to find suitable experimental data for this work, and there are several possible sources of error, which will be described in more detail below. The three sets of data selected for study were: the latent heat of sublimation of solid carbon dioxide, the latent heat of sublimation of γ-oxygen, and the enthalpy change for the interconversion of the conformational isomers of cyclohexanol.

In all these calculations, e and r_o for interactions between two types of atoms, A and B, are assumed to be related by $e_{AB}^2 = e_{AA}.e_{BB}$ $r_{oAB} = \frac{1}{2}(r_{oA}+r_{oB})$. This assumption was also made by Allinger et al. (62,79). Values of e and r_o for carbon and hydrogen are those given by Allinger (79).

A. The latent heat of sublimation of carbon dioxide.

In order to relate the latent heat of carbon dioxide to the van der Waals equations, we need to know

(i) the crystal structure of CO_2;

(ii) the latent heat of sublimation of CO_2; and

(iii) since the molecule is polar, the dipole moment of C=O, and the dielectric constant of CO_2.

(i) The crystal structure of CO_2 is face-centred cubic, and has been determined by Mark and Pohland (87).

(ii) The latent head of sublimation is -6.068 kcal./mole (88).

(iii) The dipole amoment of C=O in CO_2 is estimated (89) as 1.9D. The dielectric constant of solid CO_2 appears not to have been determined, but that of liquid CO_2 is given as 1.585 (90). The dielectric constant of CS_2 increases by about 10% in passing from liquid to solid (90), so that, arguing by analogy, the dielectric constant of solid CO_2 should be about 1.75. This estimate is obviously a likely source of error.

A crystal was constructed mathematically, taking the origin as the carbon atom situated at the corner of the unit cell. From the co-ordinates of the molecule at the origin, and those in the three adjacent faces, the co-ordinates of any other molecule can be derived by the equations

$$x = x_o + hd,$$

$$y = y_o + kd,$$

$$z = z_o + ld,$$

where d is the length of the side of the unit cell, and h,k,l are the number of unit cell lengths by which the second atom is displaced, along each co-ordinate axis, from the first.

In order to calculate the dipole effect, the dipole moment was replaced by monopoles of -0.3416 electron charges on each oxygen atom, and 0.6832 on each carbon. The potential of the molecule at the origin, due to monopoles of other atoms, was then calculated using the formula $V = 332 q_1 q_2 / r$ kcal./mole where q1,q2 are the fractional charges (75). All molecules

were derived systematically, by varying h,k, and l from -9 to +9, and the resultant potentials added, by computer. (Actually, to economise on computer time, k was varied from 0 to 9, all potentials calculated when k = 0 being halved; the resultant was then doubled).

Allowance has to be made for the dielectric constant. This will be 1.75 for distant interactions, but 1.0 for neighbouring molecules. The potential due to neighbouring molecules was calculated by hand, and subtracted from the computed total. The difference was divided by 1.75 and the resultant energy was added to the hand-calculated figure, to give the effective dipolar potential.

The van der Waals potential of the molecule at the origin was computed in a similar way, except that van der Waals functions were used. If we put $e_c = c^2$, $e_o = o^2$, interaction between atoms A and B = f(AB), then the van der Waals potential, $V = \Sigma c^2 f(CC) + \Sigma co.f(CO) + \Sigma o^2.f(OO)$. c is known, and the functions f(AB) are calculated using various values of r_o for O. Hence, o is the only unknown, and is given by a quadratic equation:

$$o^2 \Sigma f(OO) + o\Sigma c.f(CO) + \Sigma c^2.f(CC) - (V_{experimental} - V_{dipolar}) = 0$$

$V_{experimental}$ is the energy required to take a molecule from a CO_2 crystal to infinity, i.e. it is the latent heat of sublimation.; Thus a value of o is obtained for every value of r_o used.

Results.

Total computed dipolar potential = -6.029

Neighbouring dipolar potential = -0.354

Effective dipolar potential = -3.597

van der Waals contribution = 3.597-6.068 = -2.471

r_o	$\Sigma c^2.f(CC)$	$\Sigma c.f(CO)$	$\Sigma f(OO)$	o
1.45	-0.39533	-6.2097	-27.027	0.1802
1.55	" "	-6.6883	-35.107	0.1660
1.65	" "	-6.9018	-40.275	0.1570

In order to check that a variation of h,k, l from -9 to 9 represents a large enough "crystal", the calculations were repeated, using a range of -6 to 6. This gave a variation of about 2% in the van der Waals calculation, and of less than 0.1% in the dipolar calculation. So the "crystal" appears to be large enough, considering the uncertainties involved in other parts of the calculation.

B. The latent heat of sublimation of γ-O_2.

In this case, the molecule is non-polar, so that we need to know only

(i) the latent heat of sublimation. This is given (91) as -0.8686 kcal./mole.

(ii) the crystal structure. This has been determined by Lipscomb et al. (92).

There is considerable disorder in this crystal structure, due either to random orientation of the molecules about their mid-point, or to free rotation about this mid-point.

The calculation was similar to that for carbon dioxide, but simpler, since there are neither dipolar effects nor atoms other than oxygen to be considered. To test the effect of the disorder, the calculation was performed with o = 1 and r_o = 1.5. and (a) with all molecules parallel to the y axis; (b) with the molecules in the yz face parallel to the x axis; and (c)

with the molecules arranged as in (b), but with the potential calculated with respect to one of the molecules in the yz face, instead of one of the corner molecules.

Results:

(a) -27.70 kcal./mole

(b) -30.326

(c) -28.16

Hence, different arrangements cause appreciable, but not major, changes in the potential. Arrangement (b) was used for further work.

The value of o was calculated from $o^2 = V/\Sigma f(00)$ for several values of r_o.

Results:

r_o	$\Sigma f(00)$	o
1.45	-27.317	0.1783
1.50	-30.326	0.1692
1.55	-32.445	0.1636
1.60	-33.059	0.1621

C. The enthalpy difference between the conformational isomers of cyclohexanol.

Eliel et al. (93) have reviewed the reported values of this enthalpy difference, and have shown that these vary widely, from 0.29 to 1.25 kcal./mole. The values for methoxycyclohexane are less variable, from 0.6 to 0.74 kcal./mole. Since, in the latter case, the methyl group is free to rotate away from the ring, it is not likely to have a major effect on the enthal

Therefore, a figure of 0.7 kcal./mole is a reasonable one for the enthalpy difference in both cases.

In order to calculate this enthalpy experimentally, the co-ordinates given by Eliel (ref. 93, p. 454) were used, with a C - O bond length of 1.425. The potential difference between the H_1 and the remainder of the molecule was first calculated, with Allinger functions and with the hydrogen axial, and then equatorial, by summing the potentials between the hydrogen and every other atom in turn. It was then calculated for the oxygen, with the latter equatorial, then axial. Putting $e = h^2$, and using other notation as for the CO_2 calculation, the enthalpy difference is given by

$$V = \Delta [\Sigma h^2 . f(HH) + \Sigma ch . f(CH) + \Sigma co . f(CO) + \Sigma oh . f(OH)]$$

Thus, a linear equation for o is obtained, which can be solved for various values of r_o.

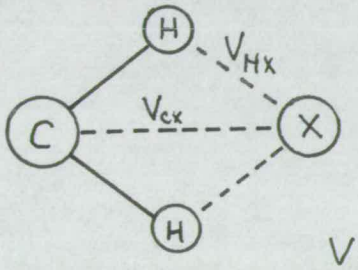
r_o	$\Delta[\Sigma h^2 f(HH) + \Sigma ch . f(CH)]$	$\Delta \Sigma c . f(CO)$	$\Delta \Sigma h . f(OH)$	o
1.45	0.532	0.295	0.591	0.1896
1.55	0.532	0.153	0.736	0.1890
1.65	0.532	-0.092	0.903	0.2071

Using the three sets of r_o and o values, three curves could be drawn. These, if extrapolated, almost met at a single point, namely $r_o = 1.4$, $e_o = 0.0361$. Since, however, the extrapolation of curved lines is liable to error, a pair of values, namely $r_o = 1.45$, $e_o = 0.0324$, within the range examined, was chosen. Using these values, theoretical values of the latent heats and enthalpy difference, compared with the experimental values are:

	Experimental	Theoretical
latent heat, CO_2	-6.068	-5.986
" γ -O_2	-0.8686	-0.8851
enthalpy	0.7	0.6925

Calculation 3.14 A new van der Waals function for methylene

Values of r_o and e which will adequately represent methylene interactions can be calculated, using the Allinger functions. An atom,X, with an r_o of 1.5, is assumed to interact with a methylene group as shown:



$$V = V_{CX} + 2V_{HX}$$

The C-X and H-X interactions, at various values of r, were added to give a total interaction. The curve given by plotting the values of r and the corresponding potential values can itself be closely approximated by a single van der Waals function. The values of e and r_o for this new function were derived from the equations $dV/dr = 0$, and $V = V_{composite}$, using the values of $V_{composite}$ and r at the minimum of the composite curve. The values found were:

$$r_o = 1.91, \quad e = 0.8858$$

Calculation 3.15. Van der Waals conformational maps, using Allinger-type functions

Conformational maps for cellulose were constructed, using set D co-ordinates, a bridge angle of 116.1°, the functions

derived by Allinger et al. for hydrogen and carbon, and the functions derived above for oxygen (calc. 3.13) and methylene (Calc. 3.14).

The maps were calculated as in Calc. 3.10, except that the total energy was calculated even for conformations outside of the allowed area. Two distinct maps were derived (see Fig. 3.7), the first excluding, the second including the $O'_3 \ldots O_5$ interaction. The reason for this is that these two oxygens can be joined by a hydrogen bond, in which case the normal van der Waals repulsion, which occurs when the oxygens are close, is replaced by an attraction. It is therefore, instructive to compare the van der Waals maps in the two cases.

Previous conformational maps were printed directly by the line printer of the computer. In the present calculations, however, the potentials were printed on punched tape, which could be read by the computer, and printed by the line printer, at a later stage. The first program calculated the potential, excluding the $O'_3 \ldots O_5$ interaction, and punched the results on tape. This tape was then read by a subsequent program, and the corresponding $O'_3 \ldots O_5$ interaction added to each conformation. This method of printing data economises on computer time, and thus avoiding the need to repeat calculations by giving data in a form which can readily be used in further work.

Results:

A. Excluding $O'_3 \ldots O_5$

Minima	Ø	Ψ	V
	-79	135	-1.0
	28	141	-0.5
H (see table 3.4)	-25	146	1.2
B	-44	168	0.5
J	-42	162	0.7
M	0	180	5.4

B Including $O'_3 \ldots O_5$

Minima	-74	157	-0.7
	-33	121	-0.3
	28	142	-0.5
H	-25	146	1.2
M	0	180	5.4
B	-44	168	0.5
J	-42	162	0.6

DISCUSSION

Calc. 3.10 provided a survey of the variations caused by the use of different co-ordinates sets, bridge angles, and types of functions. It was found that the use of different bridge angles gave almost negligible changes in the positions of the potential minima, whilst the use of different sets of co-ordinates gave changes of up to about 10°, which is about twice the discrepancy between experimentally-determined conformations of cellobiose (the Brown conformation of cellobiose is $\phi = -44^{\circ}$, $\psi = 168^{\circ}$, whilst that of Jeffrey is $\phi = -42^{\circ}$, $\psi = 162^{\circ}$; a difference of about 6.5°). The largest variations, however, are caused by the use of different sets of functions. Corresponding minima given by the Flory and Kitaygorodsky functions are separated by about 15°, whilst the Liquori and Kitaygorodsky minima are separated by about $25\text{-}30^{\circ}$. The tendency, previously noted in the study of amylose (60; see Introduction to Chapter), for the Flory functions to give lower energies than the Kitaygorodsky functions, is also seen in the present case.

As was noted in the Introduction to the Chapter, Allinger et al. (62,79) have obtained values, which fall within the limits of error, of a wide range of experimentally-determined quantities for small hydrocarbons. Their success argues strongly for the soundness of their approach, but, unfortunately, they did not derive an oxygen potential function. This has now been derived (Calc. 3.13) by the Allinger method. The experimental data which were used all involve some uncertainty; the dielectric constant of CO_2 is not accurately known, the molecules of $\gamma\text{-}O_2$ are somewhat disordered, even in the crystal, and measurements of the enthalpy difference between the conformational isomers

of cyclohexanol differ widely. Nonetheless, the selected values of e and r_o for oxygen give close agreement with the best experimental data (see Calc. 3.13), and they seem to be at least of the right order. The small size of e (0.0324) is at first surprising, since it predicts that O-O interactions are relatively weak. However, the small latent heat of oxygen and the crystal disorder, both suggest that this is indeed the case.

The conformational maps produced by using these Allinger-type functions resemble, in general shape, those given by Liquori functions. The similarity appears to be coincidental, however. The corresponding parameters used in the Allinger and Liquori functions are quite different, and one of the conformational minima is, infact, different by about 20° in the two cases.

One notable feature of the Allinger map ($O'_3 \ldots O_5$ interaction included) is the presence of a third minimum, centred at $\phi = -33^{\circ}$, $\psi = 121^{\circ}$. This is outside of the hard-sphere allowed area, and near the edge of the map, so that it may have passed un-noticed in the Flory and Liquori cases, but appears to be absent in the Kitaygorodsky case (see Fig. 3.6).

Despite their differences, all the conformational maps also show basic similarities. All have two main minima in or near the allowed area, one in the region of the 2.5-fold or 3-fold left-handed screw axes, and the other in the region of the corresponding right-handed axes. They are separated by a relatively high-energy "col" in the region of the 2-fold axis, and all except the Allinger map give the right-handed minimum as the lower of the two.

The cellobiose conformation is in the region of the left-handed minimum, and is particularly close to it in the

Kitaygorodsky map. All the maps suggest that the Brown conformation is marginally preferable to the Jeffrey conformation. The Hermans conformation is situated on the high-energy col, at an energy about 0.8 kcal./mole higher than the left-handed minimum. This, however, could be easily offset by a more favourable $O'_3 \ldots O_5$ hydrogen bond, or by more favourable intermolecular interactions. On the other hand, it has been suggested recently (64) that cellulose may have a conformation similar to cellobiose, rather than a 2-fold screw-axis, in which case its conformation would not be situated on the col. The Meyer and Misch conformation is shown to be energetically unfavourable.

The xylan conformational map, using Kitaygorodsky functions, is fairly similar to that of cellulose. Its main difference is that the ϕ and ψ values of the minima are increased, in particular the ψ value of the left-handed minimum (175° to 204°). The suggested conformation for xylan ($\phi = -80^\circ$, $\psi = 172^\circ$) is 1.95 kcal./mole above the minimum, although it is stabilised by a hydrogen bond. Since the equivalent of the Hermans conformation ($\phi = -40^\circ$, $\psi = 140^\circ$) would be only 1.29 kcal./mole above the minimum, it appears that intermolecular interactions, particularly hydrogen bonding, may be the cause of the 3-fold screw axis in xylan. The fact that xylan diacetate crystallises with a 2-fold screw axis supports this suggestion.

Since the xylan chain thus appears to be somewhat more strained than that of cellulose in the crystalline state, it might be expected that xylan would be more readily dissolved than cellulose, and this is, in fact, so (85). Until, however, all other types of interaction have been studied, conclusions

which are drawn from van der Waals conformational maps can only be tentative.

The calculations described in this chapter have shown the general validity of the hard-sphere approach, and have attempted to improve on it by enabling quantitative comparisons to be made. The comparative solubilities of xylan and cellulose seem to be explicable on the basis of these calculations. The following chapters deal with hydrogen-bonding, and dipolar and torsional potentials, and firmer conclusions can be drawn when the effects of these are known.

CHAPTER III

Introduction

Van der Waals forces are the only interactions which need to be considered in the conformational analysis of alkanes and cycloalkanes, since these contain only carbon and hydrogen atoms. In molecules containing heteroatoms, however, another type of interaction needs to be considered. This is due to the fact that atoms in the hydrocarbons are presumed to have virtually no net charge. In molecules such as polypeptides and polysaccharides, on the other hand, the heteroatoms have permanent net negative charges, which induce corresponding net positive charges on the carbon and hydrogen atoms. There is therefore, an array of dipoles within the molecule which, by their mutual interactions, alter the potential of the molecule.

Arridge and Cannon (95) were the first to realise the importance of dipolar interactions, in the conformational analysis of macromolecules. They showed that these interactions lowered the potential energy of nylons by 4.6 to 4.8 kcal./mole. They also showed that, whereas the β-structure of poly-L-alanine was stabilised by 5.7 kcal./mole due to dipolar interactions, the α-helix was stabilised by only 1.8 kcal./mole. Brant and Flory (58) found it necessary to include dipolar interactions in theoretical calculations of the mean end-to-end distance of polypeptides in solution in order to give agreement with experimental results. This distance, calculated on the basis of van der Waals and torsional potentials only, was too short by a factor of about 2.5. Hence, it is clear that dipolar interactions need to be taken into account in the conformational analysis of molecules containing heteroatoms.

This study by Brant and Flory showed that, whilst dipolar interactions modify the shape of the potential contours and the positions of minima, their most obvious effect was an overall lowering of potential.

In both of these studies, the potential was calculated from the dipolar vectors, and the vector joining them, using a standard formula (96). Scheraga et al. (75) used a monopole approximation in order to calculate the dipolar potential. In this method, the bond dipole is replaced by point charges on the bonded atoms. This method has the advantage of simplicity, and was used in the previous chapter (Calc. 3.13), in the derivation of the oxygen van der Waals function from carbon dioxide experimental data.

The dipole moments which were used in these studies were derived from experimental data for amides etc. Poland and Scheraga (97) have, more recently, used the method of Del Re (98), which is based on quantum theory, for the calculation of atomic charges.

Dielectric constants are also needed in the calculation of dipolar potentials. For two atoms which are fairly close, with no other atoms coming between them, the dielectric constant is taken as one. For pairs of atoms which have other atoms between them, the dielectric constant is taken as that of the whole molecule. This is obviously an approximation, since different intervening atoms are likely to have "dielectric constants" which differ from that of the whole molecule. However, Scheraga et al. (75) have varied the dielectric constant from 1 to 4 for several polypeptides, and have shown that this variation does not produce major changes in the final result. When conformations in solution are being calculated, the dielectric

constant of the solvent is also needed. When the solvent is water, all interactions between atoms separated by more than one layer of water molecules can be disregarded (83), since the dielectric constant of water is about 80.

Dipolar effects have been calculated for various polypeptides and nylons (58, 75, 95), but they have apparently not been calculated for any polysaccharide. Marchessault and Liang (49) have shown, however, that methylal occurs in a conformation similar to that of the atoms about the bridge oxygen in β-1,4 xylan, and that this preferred conformation is likely to be due to dipolar interactions within the molecule. The anomeric effect in monosaccharides and glycosides can also be explained (99) as a dipolar effect. Axial substituents are normally less stable than their equatorial counterparts, but the reverse is true in the case of glycosides, and even more markedly, in the case of glycosyl halides. Van der Waals interactions would favour the equatorial anomers, but the interactions between the dipoles due to the lone-pairs of the ring oxygen and to the O_5-C_5 bond, and that along the C_1-aglycone bond, more strongly favour the axial anomers. It is therefore evident that dipolar interactions are likely to affect the conformation of cellulose and other polysaccharides.

The work described in this chapter investigates the effects of dipolar interactions. Partial charges are calculated by the method of Del Re (98), and the α:β ratios of some monosaccharides are considered in the light of these. Conformational maps are derived for cellulose over the allowed area, and comparisons are made with allied molecules.

Experimental

Calculation 3.16. The partial charges of atoms in some carbohydrates.

Partial σ charges were calculated by Del Re's method (98) as outlined by Poland and Scheraga (97). The partial charge on each atom is related to the partial charges on the atoms to which it is attached, so that, in theory, it is dependent on the partial charges of every other atoms in the molecule. In practice, it is found that any variation in the charge of an atom is unlikely to have any effect more than two or three atoms away.

An atom A, covalently bonded to another atom B, has a partial charge, q_{AB}, due to the bond AB, which is given by $q_{AB} = P_{AB}/(1+P_{AB}^2)^{\frac{1}{2}}$, where $P_{AB} = (\delta_B - \delta_A)/2 e_{AB}$, and where $\delta_A = \delta_A^{\circ} + \Sigma \gamma_{AB} \delta_B$ .... xiii) δ_A° is characteristic of atom A, and e_{AB}, γ_{AB} are characteristic of bond AB. For a molecule containing n atoms, there will generally be n simultaneous equations of the type (xiii) which need to be solved in order to obtain the n values of δ_A. For a glucose molecule, n = 24, and, although many of the coefficients would be zero, the solution of 24 equations is obviously a laborious procedure. It was therefore accomplished by computer for various pyranoses and their derivatives, using the values of δ_A°, e_{AB}, and γ_{AB} given by Poland and Scheraga. For the calculation of partial charges in the hexosamine cation, the following values for N^+ suggested by Del Re (98) were used:

$$\delta_{N^+}^{\circ} = 0.31; \quad e_{N+C} = 1.33; \quad e_{N^+H} = 0.60.$$

Values of γ were the same as for neutral N. In a neutral molecule, the total charge on an atom, $Q_A = \Sigma q_{AB}$. For the nitrogen

Table 3.5

atom	a	b	c	d	e	f	g	h
C (methyl)	-	-0.024		-0.033	-0.024	-0.024	-0.024	-0.024
C_1	0.190	0.178	0.177	0.178	0.178	0.175	0.173	0.179
C_2	0.114	0.113	0.113	0.113	0.113	0.110	0.065	0.076
C_3	0.106		0.104			0.103	0.101	0.107
C_4	0.104		0.092	0.105	0.102	0.103	0.103	
C_5	0.095		0.093	0.035	0.082	0.104		
C_6	0.045			-	-0.107	0.214		
O_1	-0.446	-0.260	-0.264	-0.260	-0.260	-0.260	-0.261	-0.260
O_2 or N_2	-0.457						-0.526	0.131
O_3	-0.458						-0.459	
O_4	-0.458		-0.264		-0.459			
O_5	-0.262	-0.264			-0.266		-0.265	-0.264
O_6	-0.457			-	-	-0.253		
O_6(2)	-					-0.286		
H_1	0.061	0.060	0.060	0.060	0.060	0.060	0.059	0.060
H_2	0.053						0.048	0.054
H_3	0.052						0.051	
H_4	0.052		0.050		0.051			
H_5	0.051		0.050	0.052	0.049	0.052		
H_6	0.053			-	0.040	-		
H (O-methyl)	-	0.053		0.055	0.053	0.053	0.053	0.053
H (O_1)	0.303	-	-	-	-	-	-	-
H(O_2 or N_2)	0.302						0.219	0.248
H (O_3)	0.301							
H (O_4)	0.301		-					
H (O_6)	0.302			-	-	0.304		

Figures in subsequent columns are only included when they disagree with those in column a.
Sugars represented (all in pyranose form): a: hexose; b: methyl hexoside; c: 1,4 hexosan; d: methyl pentoside; e: methyl 6-deoxy-hexoside; f: methyl hexuronoside; g: methyl 2-amino-2-deoxyhexoside; h: cation og g.

of the cation, it was assumed that $Q_A = 1 + \Sigma q_{AB}$.

Partial π charges, used in the calculation of partial charges in methyl hexuronoside, were the same as those used by Poland and Scheraga (97) for aspartic and glutamic acids, namely

C_6 + 0.036 ; O_6 (single bond) + 0.189 ; O_6(2) (double bond) - 0.286.

These were added to the σ charges, calculated in the normal way. For the results, see Table 3.5.

The partial charges on the atoms of pentose, in all likely cyclic and acyclic forms, have recently been published (100). Except for C_1, O_1, etc. the figures for pentopyranose are very similar to those reported here for methyl pentopyranoside, although several appear to be about 2% higher.

Calculation 3.17. Dipolar effects in monosaccharides

Dipolar effects in monosaccharides were calculated, using the monopole approximation, with the partial charges obtained from Calc. 3.16, and interatomic distances measured from molecular models. The interactions which were calculated were C_1 and O_1 with C_3, H_3, C_5, and H_5; in addition, in for instance, the comparison of mannose and glucose, O_1 and H_1 interactions with O_2 and H_2 were calculated. The differences in potentials of the α and β anomers were thus obtained.

Results

C_2 epimers. The calculated differences in dipolar potential (β-α) between the anomers of glucose and of mannose were:

glucose 7.47 kcal./mole; mannose 11.00 kcal./mole.

This predicts that both should occur almost entirely as the α anomer, which is not true, although the α anomer is preferred.

More realistic results can be obtained by using the net charge on the hydroxyl groups, rather than merely the charge on the oxygens. If this was done, the differences were:

glucose 2.29 kcal./mole; mannose 2.76 kcal./mole

This result suggests that the α anomer should be preferred in both cases, but more so in the case of mannose. This is found to be the case (ref. 93, p. 408). The β methyl glycosides are also more easily hydrolysed than the α anomers, and, again, this is more marked in the case of mannose (99).

The greater preponderance of α anomer in the case of mannose has been explained by an effect, known as the Δ 2 effect (ref. 93, p. 377), which is assumed to be due to dipolar interactions, and which favours the α anomer by 0.45 kcal./mole more in mannose than in glucose. The α anomer is favoured by 0.47 kcal./mole more in mannose than in glucose in the above calculation.

C_3 epimers. The difference in potential between anomers of allose was found to be 0.76 kcal./mole. If the O_3-βO_1 repulsion was ignored, this difference (β-α) became -1.03 kcal./mole. It is probably justifiable to ignore the O_3-βO_1 repulsion, or to give it less weight, since C_1 intervenes as a dielectric. These calculations therefore suggest that sugars with O_3 in the α configuration should have the anomers in roughly equal proportions, or with the β anomer somewhat predominant. In fact, such sugars appear to exist very largely in the β form. Ribose is probably about 90% β anomer in solution (ref. 93, p. 407). The α anomers of allose and idose also appear to be very unstable (101). This instability must be at least partially due to van der Waals repulsions between O_3 and αO_1. It may

also be that, since these oxygens are so close, the approximation of using the net charges of the hydroxyl groups is no longer valid.

C_4 epimers. The difference in potential between anomers of galactose was found to be 2.20 kcal./mole. This suggests that the α anomer is the more abundant in galactose, though less so than in glucose. This appears to be so (ref. 93, p. 408).

Calculation also showed that α anomers were stabilised, not only by the dipolar interaction with the O_5 lone pairs, but also by attraction to C_3-H_3 and C_5-H_5, to the extent of 0.81 and 1.04 kcal./mole, respectively.

These dipole calculations could not be used by themselves to give quantitative estimates of the α/β ratio, since van der Waals forces and the dielectric effect of intervening atoms would also have to be taken into account. Nonetheless, they seem to give qualitative agreement with experimental data.

Calculation 3.18. Conformational maps of dipolar potential in cellulose.

Conformational maps (see Fig. 3.8) of dipolar potential in cellulose were produced in a similar way to the Allinger conformational maps (Calc. 3.15). The potential was calculated by summing the monopole interactions between all the rigidly-held atoms of one residue and all the rigidly-held atoms of the other, excluding the $C_1 \ldots C_4'$ interaction. The $O_3' \ldots O_5$ interaction was also excluded in the first map, but included in the second. For the interactions between C_1, C_2, H_1, H_2, O_2 and O_5, and C_3', C_4', C_5', C_6', O_3', H_3', H_4' and H_5' the dielectric constant was taken as 1. For other interactions, the dielectric constant was taken as 4. Cellulose is reported (102) to have a dielectric constant varying from 3.9 to 7.5, and sucrose (103) a dielectric constant of 3.32. A value of 4 for carbohydrate materials therefore seems to be a reasonable estimate.

Calculation 3.19. Conformational maps for cellobiose, lactose and mannan.

Conformational maps ($O'_3 \ldots O_5$ interactions excluded) for these cases were produced (cf. Fig. 3.8), varying ϕ and ψ by 2° increments, over the range $\psi = 152^\circ$ to 172°, $\phi = -48^\circ$ to -28°. Co-ordinates of O_4, H_4, O'_1, and H'_1 in lactose, and of O_2 and H_2 in mannan, were calculated by lengthening and shortening the C-H and C-O bonds appropriately. The gradients of the three maps were measured at the Jeffreys conformation ($\phi = -42^\circ$, $\psi = 162^\circ$).

Results

	Slope (cal. mole^{-1} deg.$^{-1}$)	Direction (clockwise of ψ axis)
Cellobiose	50.8	5°
Lactose	58.5	18°
Mannan	91.1	17°

The van der Waals torsional and hydrogen bonding interactions in lactosecan be assumed to be virtually identical to those in cellobiose; therefore, any difference in conformation due to intramolecular forces will be caused by dipolar effects. The larger gradient of the lactose map, and its different direction, suggest that lactose should have a conformation with a rather lower value of ψ, and an appreciable higher value of ϕ, than cellobiose. This appears to be the case. Lactose has an approximate conformation of $\phi = -31^\circ$, $\psi = 155^\circ$, although these figures may be modified somewhat when the structure has been refined.

The van der Waals interactions in mannan will not be identical to those in cellobiose, so that dipolar effects will not be the only influence modifying the conformation. On the basis of the dipolar effects, however, one would expect the conformation

to be displaced towards the Hermans conformation ($\phi = -25^{\circ}$, $\psi = 146^{\circ}$).

Discussion

The partial charge calculations of Calc. 3.16 show that the oxygen and nitrogen atoms make the largest contributions to dipoles in carbohydrates. The contributions of carbon and hydrogen are not, however, negligible. The α/β ratios of some monosaccharides can be qualitatively explained on the basis of monopole interactions (Calc. 3.17); quantitative agreement is not to be expected, however, unless van der Waals interactions are also included. The extra stability of α-mannose, relative to α-glucose, has been called the Δ 2 effect (ref. 93, p. 377). This effect appears to be due to dipolar interactions between C_1-O_1 and C_2-O_2. By the reasonable approximation of using the net hydroxyl charges, rather than the oxygen charges, the α anomer is calculated to be favoured by 0.47 kcal./mole more in mannose than in glucose. The estimated value of the Δ 2 effect is 0.45 kcal./mole. Although such close agreement is likely to be fortuitous, it nonetheless, suggests strongly that the Δ 2 effect can be adequately explained by C_1-O_1, C_2-O_2 dipolar inte

The conformational maps o dipole potential in cellulose (Fig. 3.8) show that, whilst this potential will modify the positions of the van der Waals minima, its most obvious effect is likely to be a general lowering of potential, similar to that found by Brant and Flory (58) in the case of polypeptides. The differences between the two maps emphasize the importance of the $O'_3 \ldots O_5$ interactions, since the repulsion between them causes the increasing potential towards $\phi = -80^\circ$, $\psi = 120^\circ$, and is the dominant influence in the shape of the second dipole-potential map.

Comparison of an area of the cellobiose map with those for lactose and mannan shows the sort of variations that can be expected from minor changes in the positions of dipoles (Calc. 3.19). The case of lactose is noteworthy since this disaccharide differs from cellobiose only in its configurations at C_1' and C_4. Van der Waals forces between residues should be virtually the same in the two cases, but dipolar forces may still have some effect. This is because the van der Waals; interactions decrease much more rapidly with distance than do dipole interactions; van der Waals interactions are inversely proportional to the sixth power of the distance, whereas dipole interactions are inversely proportional only to the first power of the distance. The dipole maps suggest that the conformation of lactose should have a lower Ψ value, and a higher ϕ value, than cellobiose, and this appears, from unrefined X-ray data, to be the case.

Comparison between mannan and cellobiose must be more approximate, since the differing configurations at C_2 will alter the van der Waals, as well as the dipole, map. On the basis of the Allinger functions, it is to be expected that the van der Waals forces would favour a relatively close approach of H_2 and $(CH_2OH)'$. This would be best achieved by a low value of ϕ. The dipole interactions, on the other hand, would favour a close approach of O_2 and $(CH_2OH)'$. The steep slope of the dipole map for mannan at the Jeffrey conformation ($\phi = -42^\circ$, $\Psi = 162^\circ$) suggests that the dipole effects are the more important, in which case the equilibrium conformation would be closer to the Hermans conformation ($\phi = -25^\circ$, $\Psi = 146^\circ$).

On the basis of the partial charges calculated in Calc. 3.16, the conformations of some other polysaccharides can be qualitatively considered. In xylan, the attraction between O_2

and $(CH_2OH)'$ is replaced by a weaker $O_2...H_5'$ attraction. This would allow the xylan to assume a left-handed 3-fold screw-axis more readily than cellulose. Thus, dipole effects may be at least part of the reason for xylan's 3-fold screw-axis, since intramolecular van der Waals interactions, discussed in the previous chapter (Calc. 3.12 and Chapter II, Discussion), do not explain it.

The dipole interaction between N_2 and $(CH_2OH)'$ in chitosan will be similar to the $O_2...(CH_2OH)'$ interaction in cellulose. Hence it is to be expected that their structures are similar. This is probably so, since the X-ray fibre pattern of chitosan (104) is very similar to that of cellulose II. In the case of the chitosan polycation, however, strong repulsion would occur between N_2^+ and $(CH_2OH)'$, which would tend to force the molecule into a left-handed 3-fold screw-axis, or to an even higher screw symmetry. Although polycationic chitosan, or possibly some of the corresponding dextrins, is known (105), it has not, apparently, been examined by X-ray diffraction.

It has already been pointed out that the $O_3'...O_5$ dipolar interaction, if included in the total dipole potential, is of over-riding importance. It is, however, more reasonable to exclude this interaction, since these two atoms are involved in a hydrogen-bond. This hydrogen bond will have an effect on the conformation which is likely to be quite distinct from the $O_3'...O_5$ van der Waals and dipolar interactions, and which will be considered in the next chapter.

CHAPTER IV

Introduction

Electronegative atoms frequently form a bond between themselves, through a hydrogen atom covalently attached to one of them. This type of bonding is termed hydrogen bonding, and can lower the potential of a system by several kcal./mole. Such bonding, involving nitrogen and oxygen atoms, is common in biological macromolecules. Of particular relevance to the conformational analysis of cellulose and allied polysaccharides is the intramolecular bond which is known to occur in these between O_3' and O_5.

Coulson, in describing the hydrogen bonds in ice (27), divided the interactions between the two oxygens and one hydrogen into four inter-related interactions. These were (A) electrostatic, (B) delocalization, (C) repulsive and (D) dispersion. Since interactions C and D (van der Waals interactions) and A are generally recognised interactions (which have been considered in the previous two chapters), hydrogen bonding appears to be essentially a delocalization effect, although this will itself modify effects A, C and D. Rough calculations of the energies involved suggested contributions of (A) -6 kcal./mole, (B) -8 kcal./mole, (C) 8.4 kcal./mole and (D) -3 kcal./mole, a total of -8.6 kcal./mole (experimental value -6.1 kcal./mole). This shows that all the interactions need to be considered in a quantitative description of hydrogen bonding.

Estimates of the energy of hydrogen bonding in alcohols (106) range from about -4 to -6 kcal./mole. A value of -5 kcal./mole therefore seems the best compromise. Bonds of maximum strength are found to occur at an O...O distance of about 2.7 Å. An examination of crystal structures shows that in hydrogen bonding, the hydrogen occurs, as nearly as possible, within a

lone-pair orbital of the other oxygen or other heteroatom. Deviations of up to 24° from the theoretical H...O-C angle are known, however, so that this is not an unduly restrictive requirement.

Lippincott and Schroeder (107) have described hydrogen bonding potential as the sum of four functions, $V = V_1 + V_2 + V_3 + V_4$. $V_3 + V_4 = A \exp(-bR) - B/R^m$, where R is the O...O distance, and A, b, B and m are constants. m can be 1, 3, or 6; if it is 1, then B/R is equivalent to electrostatic interaction, whilst, if it is 6, B/R^6 is equivalent to dispersion interactions. m = 3 is an intermediate case. Thus, $V_3 + V_4$ in effect combines van der Waals and electrostatic interactions. $V_2 = -D \exp[-n(r-r_o)^2/2r]$, where r is the O...H distance, r_o is the length of the O-H bond (taken as 0.97 Å), and D and n are constants. $V_1 = D' [1-\exp(-n'(R-r-r_o)^2/2(R-r))]$, which represents the increase in potential due to the lengthening of the O-H bond.

The Lippincott and Schroeder equation gives good agreement with experiment for the correlation of the shift in frequency of the O-H stretch peak with r, R, and V. It has some theoretical justification (108), although it is essentially semi-empirical. A more rigorous treatment of the hydrogen-bond (109) shows that, for a constant value of R, the V:r curve has a main minimum, and a marked shoulder at low values of r, and the Lippincott and Schroeder treatment (110) gives a very similar curve. It appears, therefore, that the Lippincott and Schroeder equation gives a reasonable description of hydrogen bonding potential.

The earliest attempt to estimate hydrogen bonding in conformational

analysis was made by Liquori et al (80) in their study of polypeptides. They used an equation originally proposed by Stockmayer (111) to represent the interactions of polar gas molecules. This, however, seems to be unsatisfactory, for two reasons. Firstly, Stockmayer himself pointed out that the equation would be unsatisfactory for condensed phase interactions, and for interactions between molecules such as carboxylic acids. It is therefore likely to be invalid for polypeptides. Secondly, the Stockmayer equation is essentially a description of dipolar, and not hydrogen-bonding, interaction. Subsequent analyses (75) have used the Lippincott and Schroeder equation, but Poland and Scheraga (97) now use empirical hydrogen bond equations, their criticisms of the Lippincott and Schroeder equation being that it does not merge into the van der Waals + dipole curve at high values of r, and that it is insufficiently angle-dependent.

Various attempts have been made to introduce corrections for non-linear hydrogen bonding into the Lippincott and Schroeder equation. Moulton and Kromhout (112) gave a correction for C-O-H bending. Scheraga et al. (75) gave a correction for cases where the hydrogen does not lie on the axis of the lone-pair orbital of the oxygen to which it is hydrogen-bonded. They also ignored the hydrogen bonding contribution for O-H...O angles of less than 150^{o}, although Lippincott and Schroeder (107) suggested that angular dependence in this case was sufficiently represented by the altered value of $V_3 + V_4$.

Estimates of N-H...O=C bonding in polypeptides have the advantage that the two groups are held rigidly with respect to the backbone of the molecule. In polysaccharides, however, bonding involves hydroxyl groups, and in this case, the hydrogen

can rotate about the C-O bond independently of the conformation of the polysaccharide chain. This introduces a complication into hydrogen bonding calculations for polysaccharides, since the hydrogen has to be placed in the best position for hydrogen bonding at each conformation. The intramolecular hydrogen bonding which occurs in cellulose, between O'_3-H' and O_5 is, fortunately, a simple case, since there is only one new degree of freedom, namely the rotation of H about C'_3-O'_3, to be allowed for. Other cases could be very complex, however. For instance, a hydrogen bond between O_2 and O'_6, which has also been suggested for cellulose (56), would involve three new degrees of freedom, namely the rotations about C_2-O_2, C'_5-C'_6, and C'_6-O'_6.

This chapter describes the calculation of suitable $V_3 + V_4$ terms for the Lippincott and Schroeder equations. A method is derived for fixing the hydrogen in the most favourable hydrogen bonding position, and a conformational map is produced, using this method. Finally, a map is produced of the angle between the O'_3-H bond and the O_5 lone pair orbitals, as a further guide to the position of most favourable hydrogen bonding.

Experimental

Calculation 3.20. Derivation of parameters for the Lippincott and Schroeder equation.

To derive parameters for the Lippincott and Schroeder equation, it was assumed that the O-H bond was inflexible, i.e. $R-r=r_o$, so that $V_1=0$. The values of D, n and r_o were those used by Lippincott and Schroeder (107), so that $V_2 = -81.34 \exp(-6.66(r-0.97)^2/r)$ In the equation $V_3 + V_4 = A \exp(-bR) - B/R^m$, the value of b used by Lippincott and Schroeder, i.e. b = 4.8, was used. A value of m = 3 was also used, making $V_3 + V_4$ a potential intermediate between van der Waals and dipolar potentials.

A and B were fixed empirically to give a minimum potential of -5 kcal./mole at 2.7 Å. This was done by drawing the curve of V_2, and finding the values of V_2 and dV_2/dr at r = 2.7 Å. Then,

$$V_2 + V_3 + V_4 = -5$$

$$dV_2/dr + d(V_3+V_4)/dR = 0$$

These two equations can be solved for A and B.
The total equation then becomes

$$V = -81.34 \exp[-6.66(r-0.97)^2/r] - 100.3/R^3 + 3.908 \times 10^6 \exp(-4.8R) \quad \text{(xiv)}$$

This gives a minimum of -4.85 kcal./mole at 2.68 Å, the slight deviation from the assumed value of -5 kcal./mole at 2.7 Å is presumably due to inaccuracies in reading V_2 and dV_2/dr from a graph.

Calculation 3.21. Conformational map for the $O'_3 \ldots O_5$ hydrogen bond in cellulose.

As was pointed out in the Introduction to the Chapter, hydrogen bonding in cellulose involves hydrogens which are free

to rotate about the C-O bond. The initial problem in calculating hydrogen bond potential is, therefore, to place the hydrogen at the position of strongest bonding. Since in the present case, the bonding occurs through a lone-pair orbital of O_5, the strongest bond will occur when the H'...orbital distance is minimum. An attempt was made to calculate the potential on this basis, using the standard equation for minimum distance between a point and a line (70). This gave a quartic equation which will generally have four solutions, two maxima and two minima. An attempt was made to solve this by using Ferrari's method for the solution of quartics, and comparing the four solutions with one another in order to find which was the required minimum. It was found, however, that this method was prohibitively slow in giving a conformational map. Accordingly, an approximate method was used, namely the placing of H' to give the minimum H'...O_5 distance (unless this was less than 2.5 Å), and the calculation of the potential on this basis.

Co-ordinates for H(O_3)' were derived from the Jeffrey cellobiose structure (67). These were then modified to correspond to set D co-ordinates by adding or subtracting the differences between the O'_3 co-ordinates in set C and set D. H(O_3)' co-ordinates: x = 0.983 y = -2.513 z = 0.470.

The conformational map was produced in the same manner as in Calc. 3.15. In this case, however, an additional calculation was performed in order to find the minimum distance between H' and O_5.

We need to calculate three unknowns, namely the new co-ordinates of H'. The O'_3-H' bond length, and the C'_3-O'_3-H' bond angle, can be calculated from the original co-ordinates. We, therefore, have three conditions from which the new H' co-ordinates can be

calculated, namely:

O'_3-H' = known length d

C'_3-O'_3-H' = known angle α

H'...O_5 = is minimum.

The simplest method of performing the rotation about C'_3-O'_3 is, at each conformation, to rotate the axes so that Ox is parallel to C'_3-O'_3, and move the origin to a point on the extrapolation of C'_3-O'_3, so that the new co-ordinates of O'_3 are x = -d. cos α y = O, z = o. The locus of H' is now a circle in the yz plane. This immediately fixes the x co-ordinate of H' as zero, so that only y and z need to be calculated.

If H' is at O, y, z, and O_5 is at x_5, y_5 z_5, then

$$r^2 = x_5^2 + (y-y_5)^2 + (z-z_5)^2$$

$$\therefore \quad 2r \frac{dr}{d\theta} = 2(y-y_5) \frac{\partial y}{\partial \theta} + 2(z-z_5) \frac{\partial z}{\partial \theta}$$

where θ is the angle of rotation, from the original position, of H'.

If O, y', z' is the original position of H' then, by the standard formula for rotation about Ox,

$$y = y'\cos\theta - z'\sin\theta$$

$$z = y'\sin\theta + z'\cos\theta$$

$$\therefore \quad \frac{\partial y}{\partial \theta} = -y'\sin\theta - z'\cos\theta = -z.$$

$$\frac{\partial z}{\partial \theta} = y'\cos\theta - z'\sin\theta = y$$

$$\therefore \quad 2r \frac{dr}{d\theta} = -2z(y-y_5) + 2y(z-z_5) = 0 \text{ when r is minimum.}$$

Also, $y^2 + z^2 = d^2 \sin^2 \alpha$

Hence, from these two equations, y or z can be obtained, from a quadratic equation. In practice, the two roots of the quadratic in y were obtained, and the negative one was used, since, with the present axes, this will always be the minimum value, whereas the positive value will always be the maximum. Using the minimum y value, the corresponding z value was calculated.

Having obtained the values of x, y and z, r could easily be calculated. If R was greater than, or equal to, 2.5 Å, the potential was calculated from equation (xiv). H...O distances of less than about 1.5 Å seem to be disallowed in hydrogen bonding therefore, if R was less than 2.5 Å, it was assumed that H' would rotate, not to the minimum $H'...O_5$ distance, but to a distance of 1.5 Å. For results, see Fig. 3.9, and the next calculation.

Calculation 3.22. The angles between the $O'_3...O_5$ hydrogen bond and the lone-pair orbitals of O_5.

In order to be able to calculate the angles between the hydrogen bond and the lone-pair orbitals of O_5, the direction cosines of the latter are needed. These can be derived by assuming that the angles between the lone-pair orbitals and the C_1-O_5 and C_5-O_5 bonds are all tetrahedral. The direction-cosines of the bonds can be calculated from the co-ordinates of C_1, C_5 and O_5. If we call these direction-cosines $\underline{l}_1$ and $\underline{l}_2$, and those of the lone-pair orbitals $\underline{l}$, where $\underline{l}_1$, $\underline{l}_2$ are row vectors, and $\underline{l}$ is a column vector, then $\underline{l}_1 \cdot \underline{l} = T$, $\underline{l}_2 \cdot \underline{l} = T$, where T is the cosines of the tetrahedral angle.

Also, $l^2 + m^2 + n^2 = 1$, where l, m, n are the direction-cosines of a lone-pair orbital.

Thus we have three equations which can be solved for l, m, and n. The solution involves a quadratic, so that there will be two sets of values of l, m and n corresponding to the two lone-pair orbitals.

The angle between the hydrogen bond and the equatorial orbital was derived, choosing the set of direction-cosines for which the signs of m and n are opposite, since these will represent the equatorial orbital. These direction-cosines were treated as dummy atoms; that is, the same rotation operations were performed on them as on the atoms in the non-reducing residue. If R was less than 2.5 Å, the hydrogen-bond was regarded as "buckled" or non-existent, and the bond-orbital angle was not calculated. Otherwise the position of H' was calculated as in Calc. 3.21. The angle between H'...O_5 and the equatorial orbital was then calculated. If this angle is β, the relationship is $r \cos \beta = -lx_5 + m(y-y_5) + n(z-z_5)$.

The calculation was then repeated for the axial orbital, for which the signs of m and n are the same.

From these two maps, a composite map was produced, which was divided into three areas (see Fig. 3.9). The top left-hand corner (low ϕ and ψ) is an area of buckled or non-existent hydrogen bonding. The lower left-hand corner (low ϕ, high ψ) is an area in which the H'...O_5 axial orbital angle is smaller. The main area is the area in which the H'...O_5 - equatorial orbital angle is smaller. Assuming that hydrogen-bonds are strongest when linear, and remembering that the approximation of minimum H'...O_5 distance, rather than H'...orbital distance, was used, it appears that cellulose, cellobiose, and lactose, are hydrogen bonded equatorially, whereas xylan is hydrogen bonded axially.

Results for Calc. 3.21 and 3.22

Conformation		hydrogen bond potential (kcal./mole)	$H'...O_5$- orbital angle
ϕ	ψ		
-79°	161° (xylan)	+ 1.2	37° (axial)
-25°	146° (cellulose)	- 1.6	13° (equatorial)
-42°	162° (Jeffrey cellobiose)	- 2.0	22° (" ")
-44°	164° (Brown cellobiose)	- 2.5	32° (" ")
-31°	155° (lactose)	- 2.1	6.5° (" ")

Discussion

The work described in this chapter has provided an estimate of the $O'_3 \ldots O_5$ hydrogen bonding potential in cellulose, calculated on the basis of the minimum $H' \ldots O_5$ distance. This, as was pointed out in the calculation (Calc. 3.21), is an approximation, since the hydrogen would be expected to lie at a minimum distance from a lone-pair orbital. The approximation, however, is likely to be a fairly good one, since shortening of the $H' \ldots O_5$ distance (unless this is less than about 1.5 Å) should strengthen the bond. It also appears (106) that lone-pair orbitals will bend fairly readily from their normal positions in order to form hydrogen bonds. Therefore, one might expect that, in cellulose, the lone-pair orbital will bend when necessary, in order to give a closer $H' \ldots O_5$ distance, and hence a stronger hydrogen bond. It is to be noted that, in instances where a hydrogen-bond is formed between two secondary hydroxyl groups, e.g. the $O_2 \ldots O'_3$ bond in V amylose and cyclohexamylose, the use of the minimum H'...O distance will not involve any approximation, since rotation about both C-O bonds will ensure that the hydrogen is at the minimum distance from the oxygen, and also at the minimum distance from the orbital.

Accepting that some approximation is involved in the calculation of $O'_3 \ldots O_5$ hydrogen bond potential, the resultant map seems to be of a reasonable shape, with the strongest bonds at a given $O'_3 \ldots O_5$ distance formed when the $H' \ldots O_5$ - orbital angle is minimum. Nonetheless, the angular dependence does seem insufficient. The potential at $\phi = -27^\circ$, $\psi = 166^\circ$ is -2.75 kcal./mole, with a $H' \ldots O_5$-orbital angle of 13°, whilst that of $\phi = -58^\circ$, $\psi = 180^\circ$ (with an approximately equal $O'_3 \ldots O_5$ distance)

is -2.6 kcal./mole, with a $H'...O_5$ orbital angle of 54°. It seems improbable that such a change in $H'...O_5$ - orbital angle would cause a rise in potential of only 0.15 kcal./mole or less than 6%. Another criticism of this conformational map is that it predicts that xylan will not have a $O_3'...O_5$ bond. It therefore seems that the calculations of hydrogen bonding potential can be considerably improved, and that the present results can only be regarded as a first approximation.

It would be particularly desirable to be able to calculate hydrogen bonding potential accurately, since it is recognised (ref. 27, p. 385; ref. 106) that the strength of a hydrogen bond is proportional to the shift in frequency of the hydroxyl stretch peak, from the unbonded position, in the infrared spectrum. Therefore, it should be possible to calculate the strength of hydrogen bonds in polysaccharides from infrared data. If, in addition, an accurate hydrogen bonding conformational map can be produced, it is obvious that this will suggest conformations at which such hydrogen bonds could form. The use of infrared data together with such a conformational map could, therefore, be a valuable method of determining polysaccharide conformations, complementary to X-ray diffraction.

Some qualitative comparisons can already be made on the basis of infrared data. The hydroxyl peaks which, on the basis of polarization studies, are assigned to the $O_3'...O_5$ hydrogen bond, are at 3440 $cm.^{-1}$ in cellulose II (see Fig. 2.1) and at 3345 $cm.^{-1}$ in cellulose (I) (see Fig. 1.3). Peaks which appear to correspond to these occur at 3500 $cm.^{-1}$ in xylan (49), and at 3415 $cm.^{-1}$ in cellobiose (45). The unbonded hydroxyl frequency is 3700-3750 $cm.^{-1}$. Thus, xylan has the weakest hydrogen bond, and cellulose I the strongest.

Cellulose II and cellobiose hydrogen bonds are roughly equal, and intermediate in strength. It has recently been suggested (64) that cellulose II has a conformation very similar to cellobiose and the infrared data are certainly compatible with this suggestion. This would mean that cellulose II does not have a 2-fold screw axis, but it has been pointed out (69) that the apparent screw symmetry may be due to the relative positions of the cellulose molecules, rather than to the conformation of the residues within the molecule. Some such explanation seems to be necessary to reconcile the apparently identical screw symmetries and repeat distances of cellulose I and cellulose II with their possession of hydrogen bonds of different strengths. If cellulose II has a conformation like cellobiose then it may be that the Hermans conformation ($\phi = -25^{\circ}$, $\psi = 146^{\circ}$) represents cellulose I. This should have the strongest hydrogen bond, though, by the results obtained in this chapter, the Hermans hydrogen bond is less strong (-1.6 kcal./mole) than the cellobiose hydrogen bond (-2.0 to -2.5 kcal./mole). On the other hand, the $H'...O_5$ -orbital angle is more favourable (13° as against 22° to 32°), and this factor might reverse the order of hydrogen bond strengths in a more accurate calculation.

Xylan is distinct from cellulose and cellobiose in that the hydrogen on O_3' seems more likely to bond through the axial lone-pair orbital of O_5. Sundaralingam (63) has made a survey of X-ray data for mono- and oligosaccharides and their derivatives, and finds only one instance of an axial hydrogen bond at O_5, in 2-amino-2-deoxyglucose hydrochloride. He, therefore, suggests that only equatorial hydrogen bonds are normally formed. However, an axial hydrogen bond seems to be more likely in xylan,

since the H'...O_5 orbital angles are 37° (axial) and 68.5° (equatorial). Admittedly, the hydrogen is placed by an approximate procedure, but the size of the difference in angles strongly suggests that an axial bond will be formed, and molecular models support this conclusion. A determination of the crystal structure of a xylodextrin might help to decide whether it is a valid conclusion.

The work described in this chapter has demonstrated that fairly good calculations of hydrogen bonding potential in polysaccharides are feasible, although, due to free rotation about the C-O bonds, more difficult than in proteins. Improvements, perhaps in the mathematical methods, and certainly in the potential function, would, however, be desirable.

This chapter, and the previous two, have studied separately the effects of van der Waals interactions, dipolar interactions, and hydrogen bonding. The next chapter examines other factors influencing the conformations of molecules and the net effect of all these interactions when added together.

CHAPTER V

Introduction

In the previous potential calculations, it was assumed that the pyranose rings, the glycosidic links, and bonds C_2-O_2, C_3-O_3, C_5-C_6, were rigid. Whilst this is a reasonable approximation, it is not strictly true, since all are capable of distortion by stretching and bending. It is possible to calculate force constants for these distortions from vibrational spectra, and hence they can be included in potential calculations. Calculations which allow for these distortions have been applied most widely to small hydrocarbons (62, 78, 79, 82) though also to a dipeptide, glycylalanine (113).

The deformation potential is given by $V = \frac{1}{2}k(p-p_o)^2$, where k is the force constant, and p is the variable (bond length or angle). Hence the total potential is given by

$$V = \sum_i f(r_i) + \tfrac{1}{2} \sum_j k_j(p_j - p_{jo})^2,$$

where $\sum_i f(r_i)$ is the total, over i atom-pairs of non-bonded interactions, and j is the number of possible deformations. For the most stable conformation; this potential is minimum. Hence, at every conformation, the atomic co-ordinates are adjusted so that

$$\frac{dV}{d\theta_n} = \sum_i f'(r_i) \frac{\partial r_i}{\partial \theta_n} + \sum_j k_j(p_j - p_{jo}) \frac{\partial p_j}{\partial \theta_n} = 0,$$

where θ_n represents the modification of the co-ordinates of atom n.

The general effect of introducing deformation functions is to lower the resultant potential. Gibson and Scheraga's study (113) of glycylalanine showed that, whilst the positions of the

minima were not significantly different, their relative depths were altered, so that the inclusion of deformation functions may alter the relative importance of minima.

Calc. 3.10 (Chapter II) tested the effect of varying the bridge-angle in cellulose. This tended to confirm the findings of Gibson and Scheraga (113), since potentials were lowered by up to 0.5 kcal./mole for an increase of 1° in the bridge angle, but the positions of the minima were, in general hardly changed.

Cellulose contains 15 relatively rigidly-held atoms per glucose residue. Thus, if deformation functions are introduced into the calculation, there will be 15 equations of the form $dV/d\theta_n = 0$ to be solved at each value of ϕ and ψ in order to find the new atomic positions. This is obviously a slow process, and Gibson and Scheraga's work (113) suggests that it would give only minor corrections to the conformational map. It has, therefore not been attempted, although it might be feasible, at a later stage, to include a few of the possible deformation functions. For instance, it seems likely that the most important deformations will be in the system $H_1-C_1-O-C_4'-H_4'$, and it would be relatively simple to treat these. It needs to be remembered that, although deformations in C_1-O-C_4' may not alter the positions of minima, they will alter the projected residue length and the screw symmetry. The ϕ, ψ conformations of cellulose and xylan, which are calculated from these, will therefore be only approximate if a rigid model is used in their derivation, as was the case in Calcs. 3.7, 3.9.

Torsional potential is another factor which is of importance in conformational analysis. It is found, for instance, in ethane, that there is a potential barrier to free rotation about

the C-C bond. This barrier is a maximum (2.9 kcal./mole) when two hydrogens are at their closest distance of approach, i.e. when the dihedral angle between the two C-H bonds is 0°. This is known as the eclipsed conformation. When the angle between the two bonds is 60°, the molecule is in the staggered conformation, and the barrier disappears. This type of variation can be described by a function $V = \frac{1}{2} V_o (1 + \cos 3\theta)$, where Vo is the height of the potential barrier, and θ is the dihedral angle between two hydrogens.

The cause of torsional potential is not definitely known. Van der Waals forces between three pairs of opposed hydrogens would give rise to a potential of only 0.36 kcal./mole (ref. 93, p. 8), so that these do not provide a sufficient explanation. Other causes which have been suggested are dipolar interactions, hyperconjugation, and some hybridization of the C-H bonds with d and f orbitals. The last explanation seems to be the most promising in explaining experimental data.

Rotational barriers have been measured for a variety of small molecules (ref. 93, p. 140). One noticeable feature of torsional potential is that the height of the rotational barrier does not vary greatly with the types of substituent. Thus, for instance, the rotational barrier in ethane, for three hydrogen-hydrogen interactions, is 2.9 kcal./mole. In propane, with two hydrogen-hydrogen, and one methyl-hydrogen, interactions, the barrier is only 3.3 kcal./mole. Thus, to a first approximation at least, the situation about C_1-O-C_4' in cellulose is similar to that in dimethyl ether, for which the rotational barrier (2.7 kcal./mole) is known.

This chapter describes the derivation of a suitable torsional function for cellulose, and the conformational map given by this function. All the different types of interactions are then added together, and composite maps are produced.

Experimental

Calculation 3.23. The torsional potential in cellulose

Dimethyl ether was chosen as the best available model for the $H_1-C_1-O-C_4'-H_4'$ area of cellulose. The rotational barrier for this is 2.7kcal./mole, of which about 0.3 kcal./mole (ref. 93, p. 8) is likely to be due to van der Waals interactions. The true torsional barriers was therefore, taken as 2.4 kcal./mole. It seems reasonable to assume that this corresponds to a rotational barrier of 1.2 kcal./mole for each C-O bond. Confirmation of this value is given by its similarity to the rotational barrier in methanol (1.07 to 1.6 kcal./mole). The torsional potential at conformation ϕ, ψ is therefore

$$V = 0.6\ (1 + \cos 3\phi) + 0.6\ (1 + \cos 3(180+\psi)) \quad \ldots(xv)$$

$180 + \psi$ was used, rather than ψ, because the $C_4'-H_4'$ and O-C bonds are staggered at $\psi = 0^\circ$. 180°, or an equivalent figure, needs to be added to ψ in order to give this potential minimum at $\psi = 0^\circ$.

A conformational map was produced for the allowed area, as in Calc. 3.15 except that the result was printed directly by the line-printer of the computer, instead of being punched on to paper tape. See Fig. 3.10.

Calculation 3.24. Total potential maps for cellulose

The paper tape of the van der Waals interactions excluding $O_3'...O_5$, obtained in Calc. 3.15, was read by the computer, and the potential at each conformation was stored as an element of an array. Similarly, the tape of the dipolar potential excluding $O_3'...O_5$ (Calc. 3.18) was read, and the potential at each conformation was added to the corresponding van der Waals potential,

stored in the array. The tape of the hydrogen bonding potential (Calc. 3.21) was added in a similar manner. Finally, the torsional potential at each conformation was calculated, using equation (xv), and added to the sum of the other potentials. The result was then printed, for each conformation, to give a total potential map for cellulose.

The torsional potential was calculated at each conformation, rather than being read from tape as in the other cases, because it was found to be quicker to calculate equation (xv) each time than to read a fourth tape.

This procedure was repeated, using the van der Waals and dipolar data including $O'_3 \ldots O_5$ interactions, and excluding hydrogen-bonding interactions, and a second total potential map was produced. See Fig. 3.11.

Results

Conformation	Including hydrogen-bonding			Excluding hydrogen bonding		
	ϕ°	ψ°	V kcal/mole	ϕ°	ψ°	V kcal./mole
Minima	-58	172	-14.3	-62	181	1.7
	- 7	125	-16.0	35	145	-3.4
Hermans	-25	146	-13.5	-25	146	2.8
Meyer & Misch	0	180	- 9.8	0	180	5.1
Jeffrey cellobiose	-42	162	-13.6	-42	162	3.1
Brown cellobiose	-44	168	-13.9	-44	168	2.7
Lactose	-31	155	-13.2	-31	155	3.4
Xylan	-79	161	-11.5	-79	161	2.8

Discussion

It is now possible to assess the relative importance of the various types of interaction which influence the conformation of cellulose. The van der Waals conformational maps (Chapter II) had two main minima, in the regions of the 2.5- to 3- fold screw axes of either sense, separated by a higher-energy col roughly coinciding with the two-fold screw-axis. This shape is still present in the total potential maps (Fig. 3.11), so that van der Waals forces appear to be the predominant factor in determining the conformation of cellulose. Dipolar interactions are also important, since they cause a general decrease in potential, and modify the positions of the minima.

The chief effect of the hydrogen-bond seems to be a lowering of the col relative to the two minima, and a favouring of the left-handed minimum relative to the right-handed one. This can be seen from the following results, taken from the total potential maps:

	left-handed minimum	col	right-handed minimum
with hydrogen bond	-14.3 kcal/mole	-13.2	-16.0
without hydrogen bond	1.7	3.3	-3.4

A comparison of the results for the total potentials with and without hydrogen bonds strengthens, incidentally, Poland and Scheraga's argument (97) that the hydrogen bonding function should merge into the normal van der Waals and dipole functions. Use of an empirical $V_3 + V_4$ term (see Chapter III) to represent these interactions gives a general lowering of potential of 10 to 15 kcal./mole in the conformational map which includes

hydrogen bonding potential. Since the maximum lowering of potential due to hydrogen bonding is only 5 or 6 kcal/mole, it is obvious that the $V_3 + V_4$ term gives values which are generally too low.

Torsional potential probably has relatively little effect on the positions of the minima. This, however, is largely due to the fact that the maximum ($\phi = 0^\circ$, $\psi = 180^\circ$) and minimum ($\phi = -60^\circ$, $\psi = 120^\circ$) coincide with areas of high van der Waals and dipolar potential, and it is conceivable that it might have more effect in other polysaccharides.

The conformations of xylan and the disaccharides are centred on the left-handed minimum, but it is noticeable that the right-handed minimum is the more stable, both in the total potential maps (Fig. 3.11), and in the van der Waals potential maps (Chapter II). This is particularly pronounced in the case of the potential map which excludes hydrogen bonding, and strongly suggests that cellulose derivatives which are incapable of hydrogen bonding, e.g. cellulose acetate or methyl cellulose, will occur with a right-handed screw symmetry.

There would appear to be three likely conformations for cellulose; in the area of the left-handed minimum, in the area of the right-handed minimum, and on the two fold screw axis. The latter is a less stable conformation than the other two, as an isolated molecule, but is likely to be amenable to more efficient crystal packing, and could, therefore, be comparably stable.

It is probable that at least some of the polymorphism of cellulose can be explained in terms of these three likely conformations. On the basis of the conformational maps, the following properties are likely for a form of cellulose

represented by the Hermans conformation ($\phi = -25^{\circ}$, $\psi = 146^{\circ}$) i.e. having a 2-fold screw-axis:

(i) a fairly strong $O'_3 \ldots O_5$ hydrogen bond, or possibly an $O_2 \ldots O'_6$ bond;

(ii) a strained conformation so that, if the crystal lattice is weakened by, for instance, a swelling agent, or if the lattice is destroyed, by dissolving the cellulose, this conformation would be lost, and a more stable conformation, corresponding to one of the two minima, would be assumed.

Cellulose I, as was pointed out in the previous chapter, has a stronger intramolecular hydrogen bond than cellulose II, so that it satisfies condition (i). It appears, however, to be less stable than cellulose II, since it is converted into the latter by treatment with alkali, or by precipitation from solution in Schweizer's reagent. Thus it also satisfies condition (ii).

Cellulose II appears to be a more stable form, which can be recovered unchanged, from solution. Its infrared spectrum is similar to that of cellobiose (45), and it may also resemble it in conformation, since cellotetraose, which would be expected to be very similar in conformation to cellulose, appears to do so (64). Thus cellulose II is likely to have a conformation close to the left-handed minimum. It may be significant that the cellobiose Brown conformation ($\phi = -44^{\circ}$, $\psi = 168^{\circ}$) has a 2.5 fold screw symmetry, and a projected residue length of about 5.15 Å. Thus, cellulose with the cellobiose conformation would have a stable conformation, and also be amenable to fairly efficient crystal packing.

β-1,4 Mannan is known (28) to have two crystalline modifications. The native form has an X-ray fibre pattern similar to that of

cellulose I. This form can be largely dissolved in alkali and recovered in the same crystalline form. The fraction which does not dissolve, however, changes its crystalline form to one very similar to cellulose II. It appears, therefore, that mannan has two crystalline modifications equivalent to celluloses I and II, but that "mannan I" is relatively more stable than cellulose I. It was noted in Chapter III (Calc. 3.19) that, due to the different dipolar interactions, the Hermans conformation in mannan should be relatively more stable than that in cellulose.

There are several indications, therefore, that cellulose I has a 2-fold screw-axis, with the Hermans conformation ($\phi = -25^{\circ}$, $\psi = 146^{\circ}$), whereas cellulose II has something like a 2.5-fold left-handed screw axis. The evidence is only circumstantial, and further study is required. In particular, a more thorough comparison of mannan and cellulose should provide useful information.

There is no obvious reason why the left-handed minimum is preferred to the right-handed one, although the cellobiose, lactose, and xylan conformations occur in this area. Possibly the right-handed minimum corresponds to one of the other modifications of cellulose.

The work described in the past five chapters has studied the conformations of cellulose and allied polysaccharides and disaccharides. The final chapter is an initial study of polysaccharides which, unlike cellulose, are likely to curve back on themselves, causing close contacts between distant residues.

CHAPTER VI

Introduction

In turning from the consideration of cellulose to the consideration of polysaccharides which can form helices, the most logical step is to study the conformations of β-1,2 and β-1,3 glucans. This is because, in the first place, co-ordinates are available for a β-glucose residue, which can be used with the minimum of modification. Secondly, β-linked glucose has all the bulky ring substituents in the equatorial position, so that the configurations, close to the glycosidic link are expected to be similar whether the linkage is 1,2 or 1,3, or 1,4. In the 1,3-linked case, for instance, the adjacent bulky substituents are O_2 and O_4, whereas in the 1,4-linked case, they are O_3 and C_6. The only appreciable difference, therefore, will be in the longer C_5-C_6 bond, and the bulkier C_6 substituent. This means, therefore, that β-1,3 and β-1,3 glucans will have interactions very similar to those in cellulose between adjacent residues, and will differ only in the interactions between non-adjacent residues.

Pachyman, from the fungus Poria cocos (114), callose, from plants (115), and paramylan, from flagellates, are examples of β-1,3 glucans. Laminaran, from various brown seaweeds, also contains a considerable proportion of β-1,3 links. In insoluble laminaran, from Laminaria hyperborea (116), the great majority of the linkages are β-1,3 but there is evidence of a small proportion (less than 1%) of β-1,6 linkages and branch points. In soluble laminaran, from L. digitata, the proportion of 1,6 linkages is rather higher, whilst in laminaran from Eisenia bicyclis (117), this proportion rises to about a third of the total. Branched glucans with β-1,3 and β-1,6 linkages are

also synthesised by yeast (118) and Pullularia pullulans (119).

The cereal ß-glucans are another class of polysaccharides containing ß-1,3 linked glucose. In this case, however, ß-1,4 linkages are also present (120).

ß-1,3 glucans have not, apparently, been examined by X-ray diffraction methods. On the basis of infrared and NMR studies, however, Casu et al. (121) suggest that the strong hydrogen bond found in laminaran (frequency-shift 410 $cm.^{-1}$) is either an $O_2'...O_5$ or $O_4'...O_5$ bond. The former could give a 2-fold, the latter a 3-fold screw-axis.

ß-1,3 Xylan, which is likely to have a conformation identical to ß-1,3 glucan, occurs in Caulerpa filiformis (122) and other green seaweeds, replacing cellulose as the structural polysaccharide. This xylan has been investigated by X-ray diffractio (123). It was thought to occur as a double right-handed helix with a 3-fold screw axis, and a repeat distance of 5.85 Å when dry, and 6.12 Å when moist, but it now appears (124) to be a triple helix, with a different repeat-distance. The diameter of the helix is 13.7 Å when dry, and 15.4 Å when moist. Water appears to enter into the crystal structure, since the material gives a stronger diffraction pattern when moist.

Rhodymenan is an allied xylan from Rhodymenia palmata. It contains ß-1,3 and ß-1,4 linkages in the approximate ratio of 1:2 (125).

Glucans and xylans which are wholly or partially ß-1,3 linked are therefore widespread and important in nature, so that their conformation is of considerable interest.

The ß-1,2 link, on the other hand, is rare. The glucose disaccharide, sophorose, occurs in a few glycosides (126).

The β-1,2 glucan is produced as an extracellular polysaccharide by various species of Agrobacterium (127). It appears to be unbranched, and to contain only β-1,2 linked glucose, except that A. pseudotsugae polysaccharide also contains some galactose. This glucan is difficult to methylate completely, especially on O_3, and it is suggested that this is due to (i) steric hindrance, and (ii) the presence of an intramolecular $O_3'...O_5$ hydrogen bond. The frequency shift of the hydroxyl stretching peak is given as 195 $cm.^{-1}$, a half of that of cellulose I, so that it appears to be only weakly hydrogen bonded.

There are several mathematical methods for treating the rotations of sugar residues in a molecule apart from that used in previous chapters. One method would be that of Settineri and Marchessault (57; see Fig. 3.1); the conformation is defined by a single rotation about $O_1...O_4$ (in β-1,4 xylan). The Miyazawa method is similar (68), but the conformation is defined in terms of the two rotations ϕ and ψ. Sugeta and Miyazawa (128) have recently described this method in a more general and compact form, which can be readily incorporated in computer programs.

The calculations described in this chapter use a different method developed by Thompson (129). Given a single set of atomic co-ordinates, $\underline{x}j$ for, say, a glucose residue with respect to the ϕ axis, the co-ordinates $\underline{x}_j'$ of a second residue, attached to atom k, can easily be calculated. Normally, these would be represented by $\underline{x}_j' = M\underline{x}_j + \underline{x}_k$ ..... (xvi). M is a 3 x 3 matrix which represents the rotation of the second residue with respect to the first, and $\underline{x}_k$ represents the translation of the co-ordinates from the origin to atom k. Thus, two matrix operations are performed in order to generate $\underline{x}_j'$; firstly, a rotation operation,

which is a multiplication, and secondly, a translation operation, which is an addition. The Thompson method simplified this process by combining the two operations in one. If $M = [m_{ab}]$ and $\underline{x}'_j$, $\underline{x}_j$ and $\underline{x}_k$ are the Cartesian co-ordinates of atoms, then (xvi), written in full, is

$$\begin{bmatrix} x'_j \\ y'_j \\ z'_j \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \cdot \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} + \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}$$

This can be written as a single multiplication, thus:

$$\begin{bmatrix} x'_j \\ y'_j \\ z'_j \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & x_k \\ m_{21} & m_{22} & m_{23} & y_k \\ m_{31} & m_{32} & m_{33} & z_k \\ . & . & . & 1 \end{bmatrix} \cdot \begin{bmatrix} x_j \\ y_j \\ z_j \\ 1 \end{bmatrix} \quad \ldots\ldots \text{(xvii)}$$

In a regular helix, such as would occur in a crystalline polysaccharide, the third residue is related to the second in the same way that the second is related to the first. Thus, if $x'_j{}'$ is the 4 x 1 column vector representing an atom in the third residue, if M now represents the 4 x 4 matrix, and if $\underline{x}_j^{(n)}$, x'_j, x'_j are 4 x 1 column vectors, corresponding to equation (xvii), then

$$x'_j{}' = M\underline{x}'_j = M^2\underline{x}_j$$

or, generally,

$$\underline{x}_j^{(n)} = M^{n-1}\,\underline{x}_j \qquad \ldots\ldots \text{(xviii)}$$

Hence, the co-ordinates of the atoms in the nth residue can easily be calculated from those of the first residue, with respect to the same axes.

The work described in this chapter is the application of this formula in the conformational analyses of laminaribiose (3-O-ß-D-glucopyranosyl-D-glucose) and sophorose (2-O-ß-D-glucopyranosyl-D-glucose), and the corresponding dextrins up to, and including, the octasaccharide. Laminaribiose will also be studied by the methods of Chapter I, so that the two procedures can be compared.

Experimental

Calculation 3.25. The conformational map of laminaribiose.

Set D co-ordinates were modified, by operations similar to those described in Calcs. 3.2, 3.3, in order to give co-ordinates for a new reducing residue, with origin at O_3, and the Ox axis lying along O_3–C_3. $\not\phi = 0$ was defined as the conformation at which H_3 lay in the xz plane, with its z co-ordinate negative. The non-reducing residue was represented by set D non-reducing co-ordinates and the bridge angle was taken as 116°. Set D co-ordinates were derived for the express purpose of representing β-1,4 linked glucose, so that they will not be as accurate when applied to other linkages. They should not be a major source of inaccuracy, however, and are adequate for an initial, introductory, study such as the present chapter.

A hard-sphere conformational map was produced for laminaribiose, in exactly the same way as for cellobiose (see Calc. 3.8). Result, see Fig. 3.12.

Calculation 3.26. Conformational maps for β-1,2 and β-1,3 glucans.

The procedure will be described for β-1,3 glucan. The initial residue was defined by Set D non-reducing co-ordinates. This residue is held stationary whilst the ϕ and $\not\phi$ rotations are performed on a second residue attached to O_3. If ϕ and $\not\phi$ are measured in the same senses as for cellobiose, the rotation matrix for ϕ will be the same, but the rotation of $\not\phi$ will be in the opposite sense. This is because, in the case of cellobiose, $\not\phi$ was a rotation of the reducing residue. But a clockwise rotation, $\not\phi$, of the reducing residue, about O–C_4 is exactly the same as an anticlockwise rotation $\not\phi$ of the non-reducing residue about the same bond. Thus, in the present case,

where the initial residue is held stationary, and the

second residue is rotated, the Ψ rotation matrix must be in the opposite sense in order to correspond to the cellobiose results.

The co-ordinates of the second residue were derived from those of the first by the operation represented in (xvii). As well as the ϕ, Ψ and α rotations, we need to rotate the residue so that the Ox axis is parallel to O_3-C_3. This rotation is represented by the R^{-1} matrix [eqn (ix)] derived in Chapter I. The matrix R rotates Ox so that it lies along the former direction **l**, m, n, i.e. the direction **l**, m, n becomes **l**, 0, 0. But in this case the reverse is required; we want **l**, 0, 0, to become parallel to **l**, m, n, where **l**, m, n, represents O_3-C_3. Therefore, R^{-1} is used.

If P is the ϕ rotation matrix, A the α rotation matrix ($\alpha = 64^{\circ}$), and S the Ψ rotation matrix, then the overall 3 x 3 rotation matrix is $R.^{-1}S.A.P.$ This is converted to the 4 x 4 matrix by inserting the co-ordinates of O_3, as in eqn. (xvii).

By systematic variations of ϕ and Ψ, a conformational (hard-sphere) map can be produced, as in Chapter I.

Two separate programs were used for the complete analysis. The first derived a conformational map for the disaccharide; the second derived conformational maps for the trisaccharide and higher dextrins. There are two reasons for this. Firstly, as with cellobiose, the interactions in the disaccharide disallow all but a few percent of the conformations. Subsequent maps are therefore only derived over the small area allowed by the disaccharide, thus giving a considerable saving in computer time. Secondly, some interatomic distances, e.g. $C_1 \ldots C_3'$, O (bridge) ... H_1 and H_3', are always within the sum of the van der Waals radii. These therefore must not be

measured in the disaccharide case, but they need to be measured for the higher dextrins. Thus, it is convenient to use two programs.

The second program gives a conformational map, over the disaccharide allowed area, for interactions between the initial and third residues. It then does the same for interactions between the initial and fourth residues and so on, using eqn. (xviii).

The conformational map for the disaccharide can then be further limited by excluding any conformations which are disallowed by too close contacts with more distant residues.

If any interatomic distance was less than the sum of the outer-limit van der Waals radii, the conformation was disallowed, and the computer printed 0. If the conformation was allowed, but there was an interatomic distance less than the sum of the fully-allowed van der Waals radii, the computer printed 1, unless two oxygens were between 2.5 and 2.8 Å, apart, thus permitting hydrogen bonding, when it printed 3. If all interatomic distances were fully allowed, the computer printed 2, or 3 if hydrogen bonding was possible.

The same procedure was followed for β-1,2 glucan. For results, see Fig. 3.12.

Discussion

The calculations described in this chapter have shown that, as expected, laminaribiose and sophorose have conformational maps which correspond closely to that of cellobiose. Both have somewhat larger allowed areas than cellobiose; laminaribiose has 61 allowed conformations (4.71%), and sophorose has 51 (3.94%), whilst cellobiose has only 45 (3.47%). On the other hand, laminaribiose appears to have a smaller fully-allowed area, and no conformation is fully allowed for sophorose. It seems likely, then, that β-1,2 and β-1,3 linked glucans will have relatively rigid chains, like cellulose.;

There are some minor discrepancies between the conformational maps produced for laminaribiose by the methods of Chapter I (Calc. 3.25) and by the use of the Thompson matrix (Calc. 3.26). The outline of the allowed area is different, and there is a different number of fully-allowed conformations. The reason for this is that, at the zero conformation in the first method, C_1-H_1 is eclipsed by O-C'_3, and C'_3-H'_3 is eclipsed by O-C_1. In the second method, at $\phi = \psi = 0$, C_1-H_1 will still be eclipsed, but C'_3-H'_3 may not be. This is because, instead of using two sets of axes, and defining the zero points of each by H_1 and H'_3, respectively, the Thompson method uses only one set of axes, and defines zero for both ϕ and ψ by H_1. The deviation should not be large, however, and a comparison of the two maps suggests that it is less than 10°.

When ϕ and ψ are varied by 10° increments, laminaribiose has 61 allowed conformations (4.71%). Close contacts with subsequent residues in a laminaridextrin reduce this number to 57 (4.40%). The conformations which are disallowed by close

contacts with subsequent residues all, by analogy with cellobiose, lie on the high-energy col, so that their exclusion is less important than would be the exclusion of an equal number of conformations near a potential minimum. From the conformational analysis of lamimaridextrins, then, it appears that β-1,3 glucans will be stiff molecules, but that this stiffness is largely due to interactions between adjacent residues.

On the basis of the conformational maps, a choice can also be made between the alternative conformations for laminaran suggested by Casu et al. (121). An attempt was made to construct a model which resembled the conformation illustrated by Casu. This model had an approximate conformation of $\phi = -50^{\circ}$, $\psi = -80^{\circ}$, which is not an allowed conformation, and does not have the 2-fold screw-axis suggested by Casu's illustration. This conformation has several close contacts, in particular an $O'_2 \ldots H_1$ distance of 1.75 Å. Thus this conformation, which might have had an $O'_2 \ldots O_5$ hydrogen bond, must be discounted.

The alternative conformation, suggested by Casu, would have an $O'_4 \ldots O_5$ hydrogen bond. A helix model with a conformation of $\phi = -50^{\circ}$, $\psi = 20^{\circ}$ has a 3-fold right-handed screw-axis, and a repeat distance of 7 Å. This may be close to the conformations of β-1,3 glucans although it is not quite correct, since the $O'_4 \ldots O_5$ distance (3.15 Å) suggests a weak hydrogen bond, although the infrared data (121) show it to be strong.

The predicted rigidity of these β-1,3 linked polysaccharides is consistent with the occurrence of the xylan as a skeletal polysaccharide, replacing cellulose (123). The β-1,6 links,

which occur to some extent in laminaran, are likely to increase the flexibility and the irregulatiry of the molecule, which explains why the form with the lower percentage of these linkages is also the insoluble form (116).

Sophorose is rather more rigid than laminaribiose, having only 51 (3.94%) allowed conformations. Close contacts with subsequent residues in sophorodextrins decrease this to 34 (2.62%). The conformations which are disallowed by close contacts with subsequent residues comprise, by analogy with cellobiose, one of the potential minima. It appears, therefore, that β-1,2 glucan is an extremely rigid molecule, more so than cellulose, and that a good deal of this rigidity is caused by interactions between non-adjacent residues.

It has been suggested (127) that the difficulty experienced in methylating O_3 of β-1,2 glucans is due to steric hindrance, and to an $O'_3 \ldots O_5$ hydrogen bond. The very small allowed area of the conformational map, and models which were built to correspond to allowed conformations, suggest that steric hindrance may well cause difficulty in complete methylation. The suggested hydrogen bond is shown, by the examination of models, to be impossible, but two others, namely $O'_3 \ldots O_2$, and $O'_3{}' \ldots O_5$ are possible. At the allowed conformation $\phi = 0^\circ$, $\psi = 180^\circ$, both O...O distances, measured from models, are about 2.5 Å, but the alignment, in the case of $O'_3 \ldots O_5$, is poor. Therefore, at this conformation, the most likely hydrogen bond would be a fairly strong $O'_3 \ldots O_2$ one. At another allowed conformation, $\phi = -40^\circ$, $\psi = 150^\circ$, both possible bonds have rather poor alignment, but the $O'_3 \ldots O_2$ distance is 4.25 Å, whereas the $O''_3 \ldots O_5$ distance is 3.25 Å, and alignment is rather better in the latter case. Hence the $O''_3 \ldots O_5$ bond would be preferred at this conformation. The shift in frequency of the OH stretch peak

195 $cm.^{-1}$ indicates that the hydrogen bond is weak, so that this does not enable us to decide between the two possibilities. Sundaralingam (63) notes that there is no known instance of a hydrogen bond to a bridge oxygen in carbohydrates. There is no obvious reason why this should be impossible, however, when, as in the case of $O'_3 \ldots O_2$, it is sterically feasible.

It is evident, from the calculations described in this chapter, that conformational analysis of polysaccharides, using the Thompson method, is useful in explaining experimental data. These calculations are, however, only introductory. Their usefulness would be greatly enhanced by the calculation of screw axis and repeat distance data, and the Thompson method (129) might be better replaced by the Miyazawa method (128) for copolymers such as agarose, carrageenans and mucopolysaccharides.

Further Work

The work described in this section has attempted to describe and correlate polysaccharide conformations in terms of interatomic interactions. These initial studies have shown that conformational analysis has considerable potential in explaining and supplementing experimental data. The results of these studies have already been discussed at the end of each chapter. It is worthwhile to consider now some of the weaknesses of the calculations reported here, and some of the areas where further work could be done.

The mathematical method of rotation of sets of co-ordinates, used in Chapters I-V, has the advantage of simplicity. It is also useful in that molecular models can easily be built having conformations described in terms of ϕ and ψ as defined in Chapter I. On the other hand, if the xz planes were defined by the planes $O\text{-}C_1 \ldots C_4$ and $O\text{-}C'_4 \ldots C'_1$, instead of $O_1\text{-}C_1\text{-}H_1$ and $O\text{-}C'_4\text{-}H'_4$ (in cellulose), this would avoid the use of the hydrogens, whose positions are less accurately known than are the positions of other atoms. It would also simplify screw-axis and repeat-distance calculations by the Miyazawa (68) method. Furthermore, although the method of Chapters I-V is suitable for the study of cellulose, it is seen (Chapter VI) to be less so for other polysaccharides. It may well be that the general Miyazawa formulae (126) are the most suitable for more complicated cases. They have the advantage that the same basic mathematics will be involved, whether they are used for rotations of atomic co-ordinates or for calculations of screw symmetry and repeat distance. It might be possible to combine the Miyazawa (126) and Thompson (127) methods, which would be even better.

The work described in this section has generally regarded polysaccharides as frameworks of rigidly-held atoms. This is a useful approximation, but a more accurate analysis would need to take account of free rotations which are allowed, about C_2-O_2, C_3-O_3, C_6-O_6, and, particularly, C_5-C_6, in cellulose. $H(O_2)$, $H(O_3)$, $2H_6$, O_6, and $H(O_6)$ will adjust to positions of minimum potential at each conformation. They will, however, affect the total van der Waals and dipolar potentials of the molecule, and therefore they may also affect the positions of minima.

It would, theoretically, be possible to rotate each of these atoms in turn, at every conformation, until the potential of the whole molecule is minimum. This has been done by Ramachandran et al. (60) in the case of glucose, but it would be prohibitively slow and costly to apply this method to the calculation of complete conformational maps. Approximate methods are therefore needed. One such method is the replacement of a group by an equivalent "atom". Thus, Flory (58) derived a van der Waals equation for a methylene "atom". by assigning appropriate values to the variables in the Slater Kirkwood equation. An Allinger-type function was derived for the methylene group, in this present study, by summation of the separate interatomic interactions at a reasonable spatial distribution of the individual atoms (see Calc. 3.14). Dipolar interactions in the third chapter (see Calc. 3.17) were found to give more reasonable results if the net monopole charge of the hydroxyl group was assumed to be centred on the oxygen atom. Thus, there are two methods of replacing a group, which is capable of free rotation, by an artificial "atom" viz:

(i) by assuming that interactions of the separate atoms of the group all act from the position of the rigidly-held atom, e.g. Flory' van der Waals equation, or the net monopole charge of a hydroxyl group;

(ii) by assuming a reasonable mean geomtery for the system, summing all the interactions, and deriving parameters for a new "atom", e.g. the Allinger-type methylene function.

Another method of approaching this problem would be to derive two sets of conformational maps. The first set would be calculated on the basis of the rigidly held atoms, as in the present study. The second set would include the atoms which are free to rotate, but would assume them to be rigidly-held in, for instance, the conformations found in cellobiose. If the potential at any conformation in the second set is significantly lower than at the corresponding conformation in the first set, this indicates that the conformations of the rotatable atoms would be fairly favourable, and therefore the second set of maps more nearly represents the true situation than does the first. If, on the other hand, the potential in the second set is higher than in the first, this indicates that the rotatable atoms are in unfavourable conformations, The true potential would be lowered by rotation of these atoms. Therefore, in this case, the first set of maps would probably be more accurate than the second. It would therefore, be possible to obtain an approximate total map from these two, by selecting the conformations of each map which are lower than the corresponding conformations in the other.

The calculations in the second chapter included the derivation of an Allinger-type function for oxygen van der Waals interactions. There were several possible sources of error in the derivation of this function, so that it would be advantageous if more accurate data could be found in order to derive the function more accurately. Fortunately, oxygen van der Waals forces appear to be weak, so that the effect of inaccuracy in the function is less than it might be.

Several improvements could be made in the calculation of dipolar interactions. Point dipoles, as used by Brant and Flory (58) and Arridge and Cannon (95), would be more accurate than the monopole approximation which was used in the present study. The dipoles of oxygen lone-pairs might also be calculated, rather than assuming that the charges centred on the oxygen nuclei. It has been noted (99) that the anomeric effect inglucosides may be at least partially due to interactions between the aglycone and the charge of the ring oxygen. Such interactions could therefore, be important.

Since the hydroxyl groups in cellulose, particularly O'_3-H, are hydrogen bonded, the monopole charges of their atoms will be somewhat different from those calculated by Del Re's method. The possible delocalization structures (27) give some indication of the direction of the change which can be expected, and estimates of the magnitudes of the contributions from these delocalization structures, together with calculations of the monopole charges for such structures, by Del Re's method, should enable improved values to be obtained for these monopoles.

The effects of the hydroxyl hydrogens are particularly important in dipolar calculations. The charge on a hydroxyl oxygen is about -0.45 , whereas the charge on a hydroxyl hydrogen is about 0.3 Therefore, the net charge on the group is only -0.15, so that calculations of net potential due to a hydroxyl group could be two-thirds smaller than the same calculations, using only the charge on the oxygen. It may well be, therefore, that dipolar effects have been overestimated in the present study. On the other hand, Arridge and Cannon (95) showed that, in nylons, the polarizability of the covalent bonds could often increase the importance of the dipolar effects by 50%. Hence,

the neglect of hydroxyl hydrogens and of the polarizability of covalent bonds, in the present study, would tend to offset each other, although it is unlikely that they would cancel each other completely.

It has already been noted, in Chapter IV, that the Lippincott and Schroeder equations which were used to calculate hydrogen-bonding potential were insufficiently angle-dependent, and predicted no hydrogen bond in xylan, where one is known to occur. A good case can be made, therefore, for adopting Poland and Scheraga's idea (97) of calculating an empirical function for hydrogen bonding potential, which would run smoothly into the van der Waals + dipolar potentials at large interatomic distances. It would seem reasonable, however, to retain the analytical form of the Lippincott and Schroeder equation (eqn. xiv), and also to attempt to allow for the lengthening of the O'_3-H bond.

Another source of inaccuracy in the calculation of hydrogen bonding potential was the use of H...O_5 minimum distance, rather than H... lone-pair orbital minimum distance, as a criterion for fixing the position of H. It might be worthwhile to continue to look for a practicable method of employing the latter criterion.

Also, in the matter of hydrogen bonding, it is apparent that an equation linking hydrogen bonding potential to the shift in frequency of the O-H stretch peak in the infrared spectrum would be very useful in helping to determine polysaccharide conformations.

Deformation (stretching and bending) in the polysaccharide chain is the only internal effect (i.e. not caused by solvent) which has not been quantitatively estimated in the present study. It was pointed out in Chapter V, though, that deformations, in particular the bending of C_1-O-C'_4, will not only alter the potential, but

will also affect screw symmetry and repeat distance calculations. It seems, therefore, that some study of this is advisable.

The present study has not considered 1,6-linked polysaccharides. There will not be any fundamental new difficulty in treating these. Their study will, however, be more time-consuming, since they have one more degree of freedom than polysaccharides linked through secondary oxygens. Cellulose, for instance, can change its conformation by rotation about C_4'-O and O-C_1, whereas pustulan (β-1,6-glucan) can change its conformation by rotation about C_5'-C_6', C_6'-O, and O-C_1. In addition, the fact that the residues in pustulan will be separated by three bonds instead of two, is likely to mean that there will be a larger number of allowed conformations to be considered.

Intermolecular interactions are another problem which needs to be studied. This has already been done, in the case of amylose triacetate, by Sarko and Marchessault (61). It will be even more essential to study them in a case such as β-1,3-xylan, which occurs as a triple helix (124).

Many polysaccharides are branched, and conformational analysis could also be applied to these. This is likely to be time-consuming, since there will be four degrees of freedom (a ϕ and ψ value for each residue) at a branch point. Sarko and Marchessault (61) can be considered to have studied this problem also, although, in their case, the "branches" were acetate groups. Also, many polysaccharides contain furanose residues, and the conformational analyses of these would be an entire new field.

Finally, it may be possible to calculate the interactions of solvents with polysaccharide chains. The possibility of hydrophobic bonding in polysaccharides was mentioned in Section I. It may be possible, on the basis of data such as those of Némethy (30)

to calculate the energy of hydrophobic interactions in polysaccharides. Recently, Scheraga (83) has also discussed analytical functions for calculating the energy of hydration of a molecule.

It is evident that there is much ground, both for improvement of methods in the calculations reported here, and for expansion into new areas, of the conformational analysis of polysaccharides. It is probably not possible, or even desirable, in terms of improved results, to achieve all the aims suggested here. Nonetheless, the success already achieved by the methods described in this Section suggests that further work would be well repaid, by providing a powerful means of insight into the physical structure and properties of polysaccharides.

APPENDIX

Flow Diagrams of Computer Programs

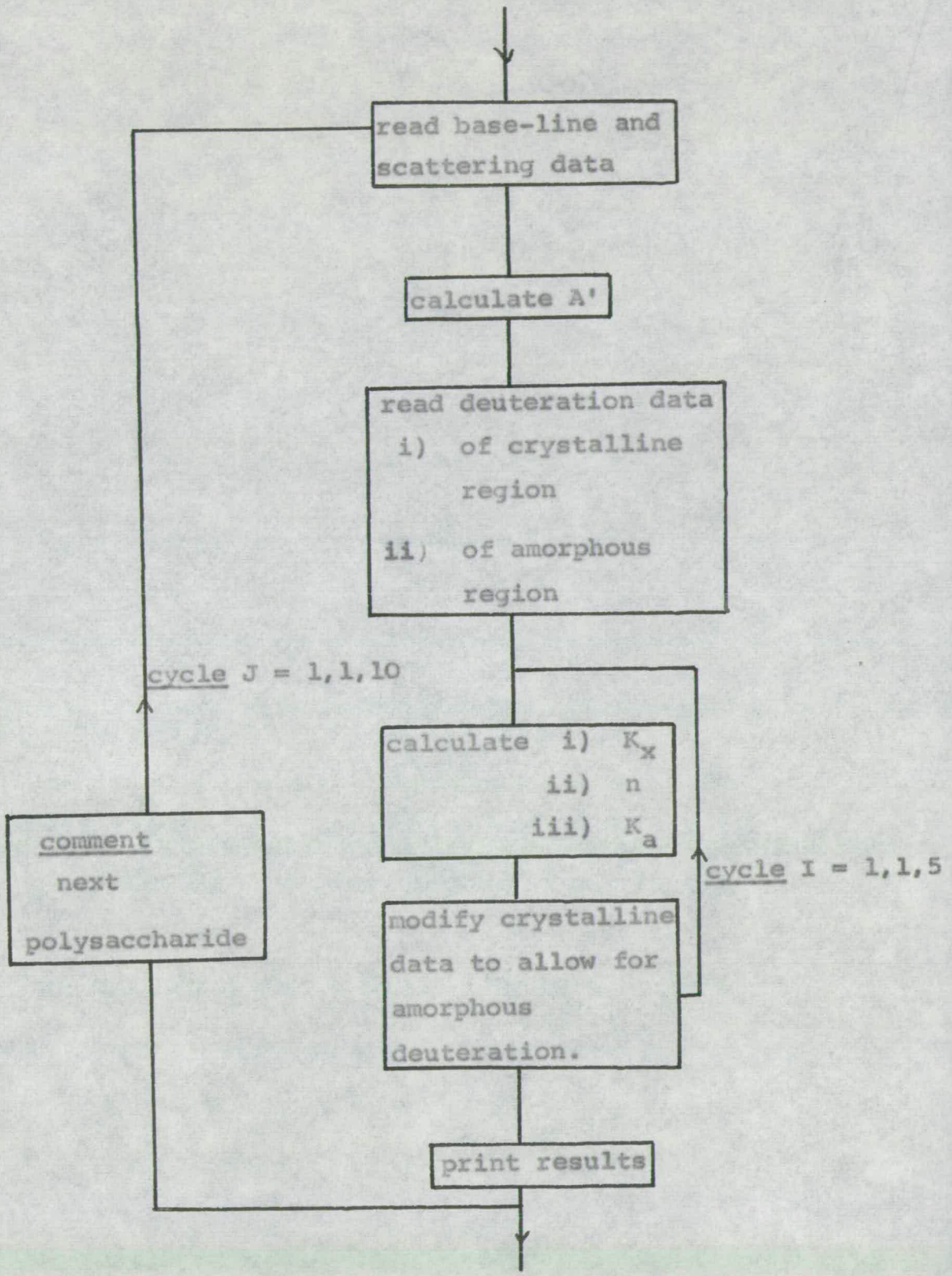
This appendix contains flow diagrams of the main types of program used in these studies.

Declarations and terminators are omitted. The following symbols are used:

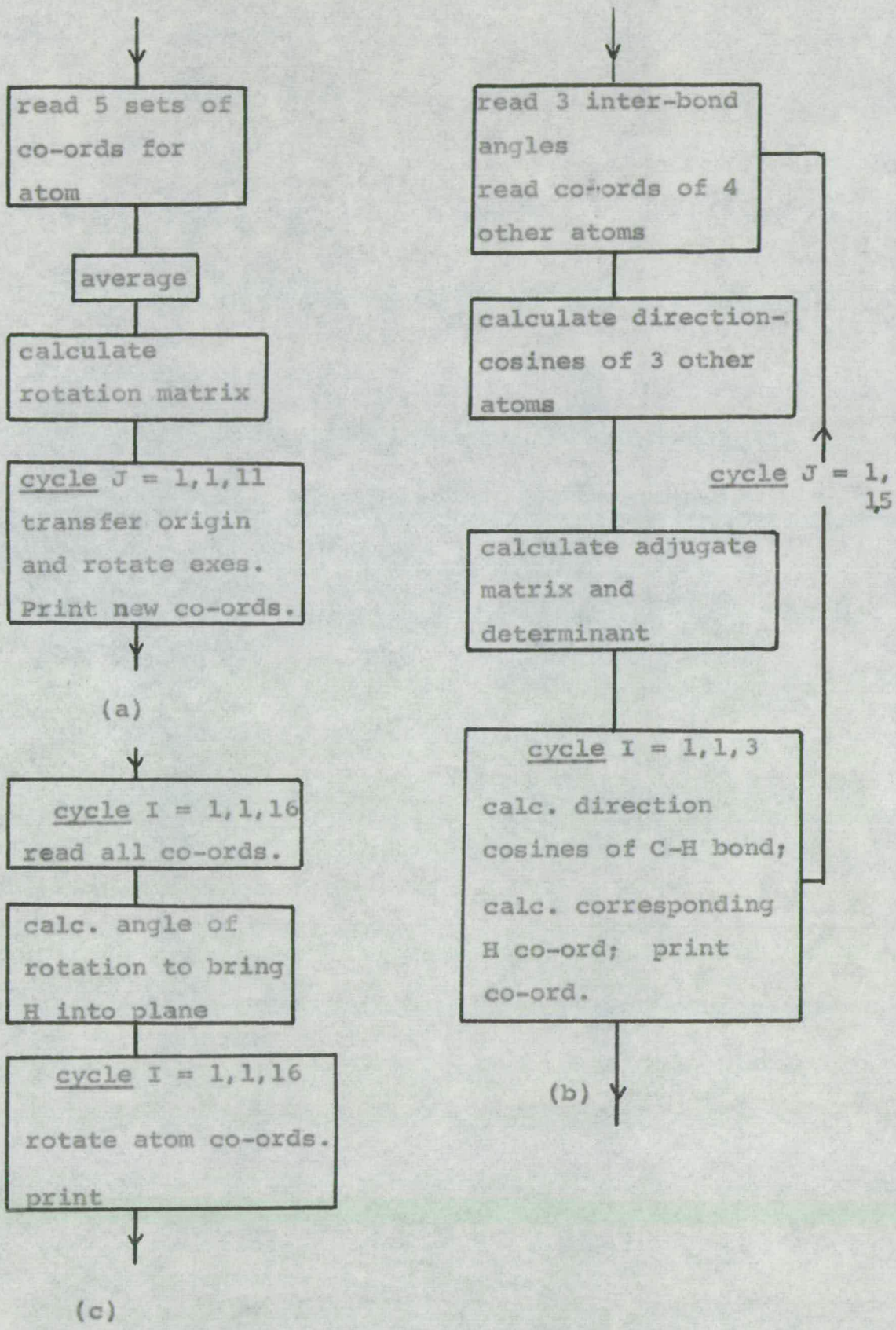
separator, function, routine or switch.

instruction or comment

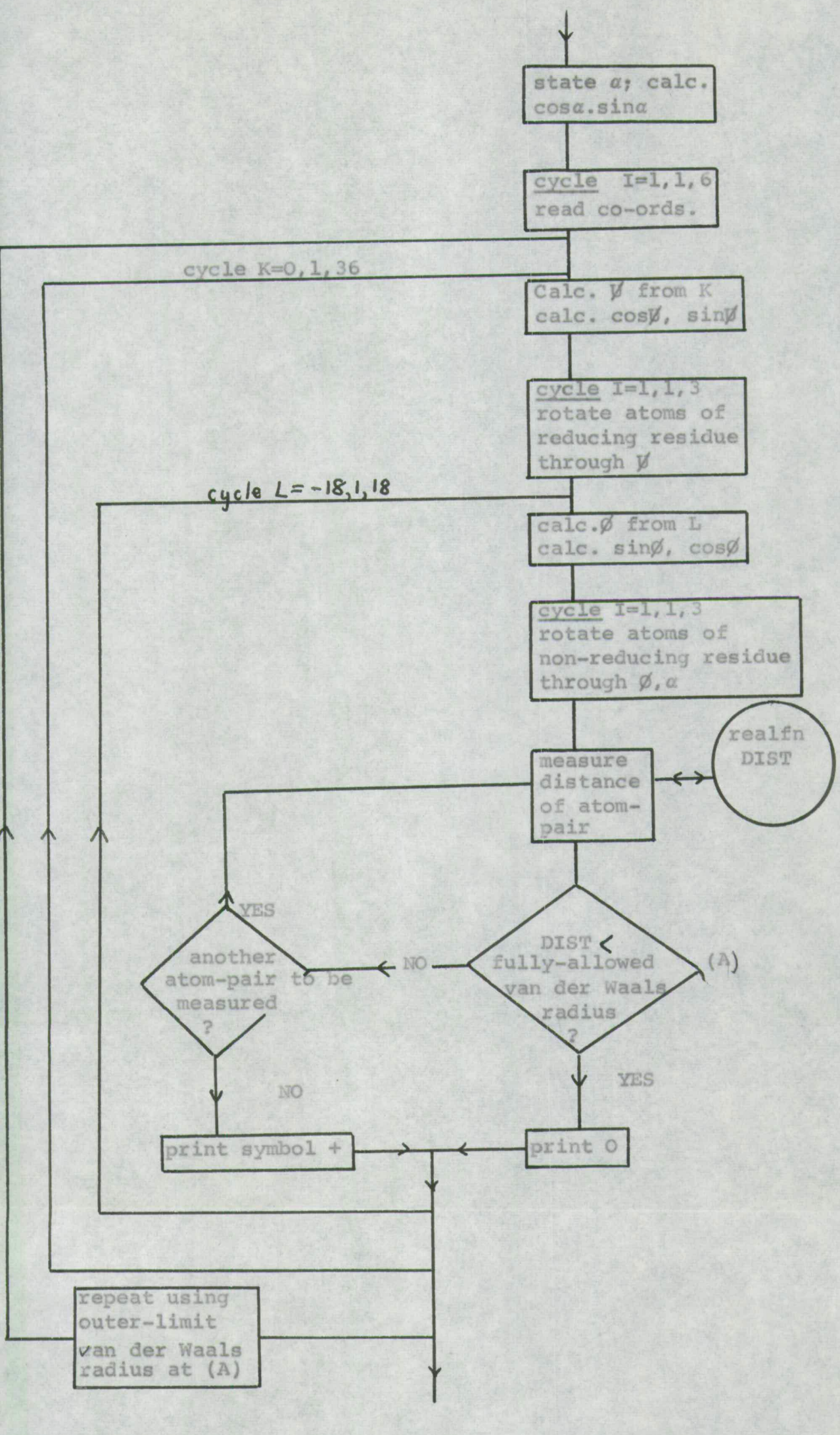
question.



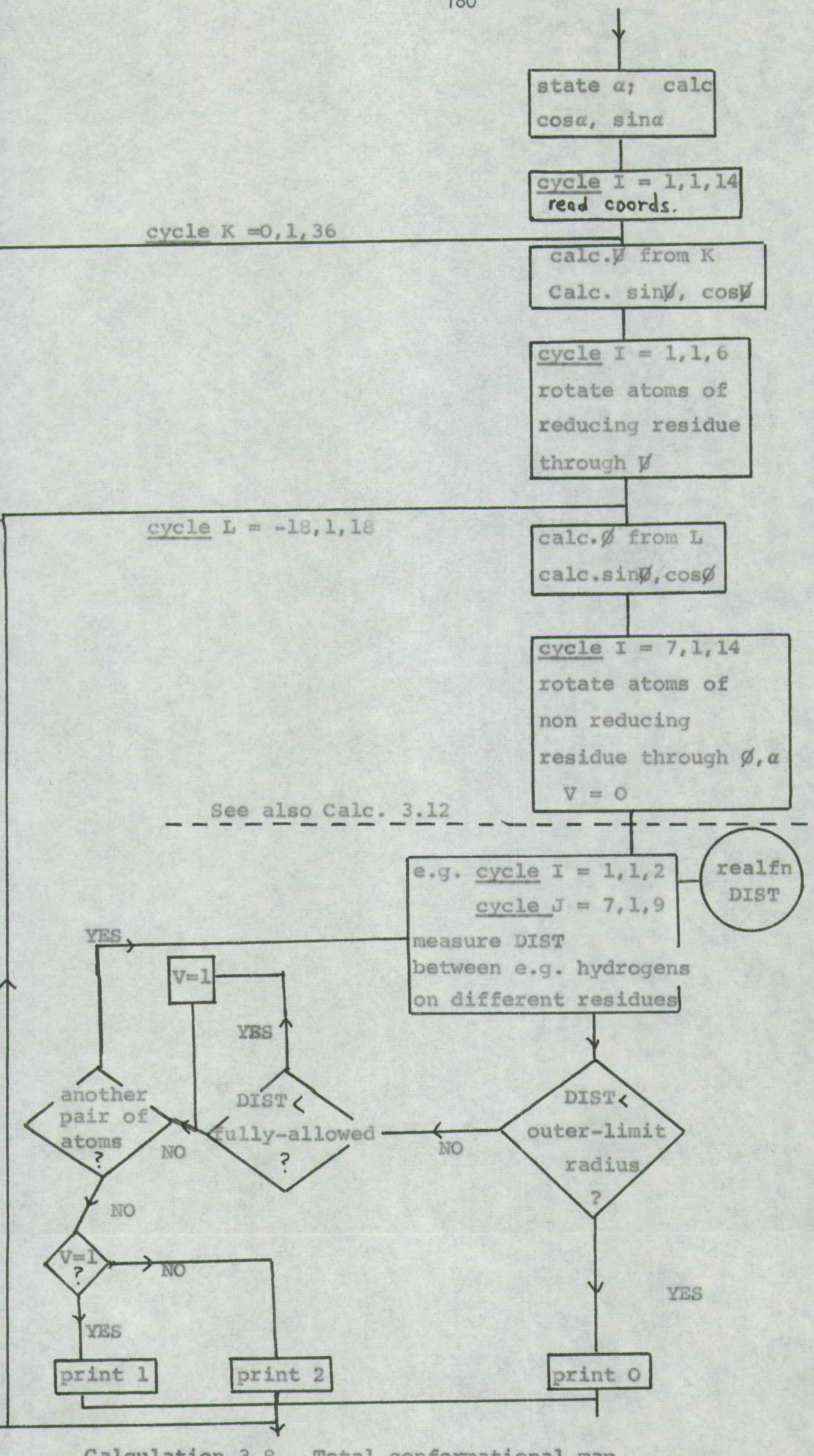
Experiment 2.5 The deuteration of polysaccharides.



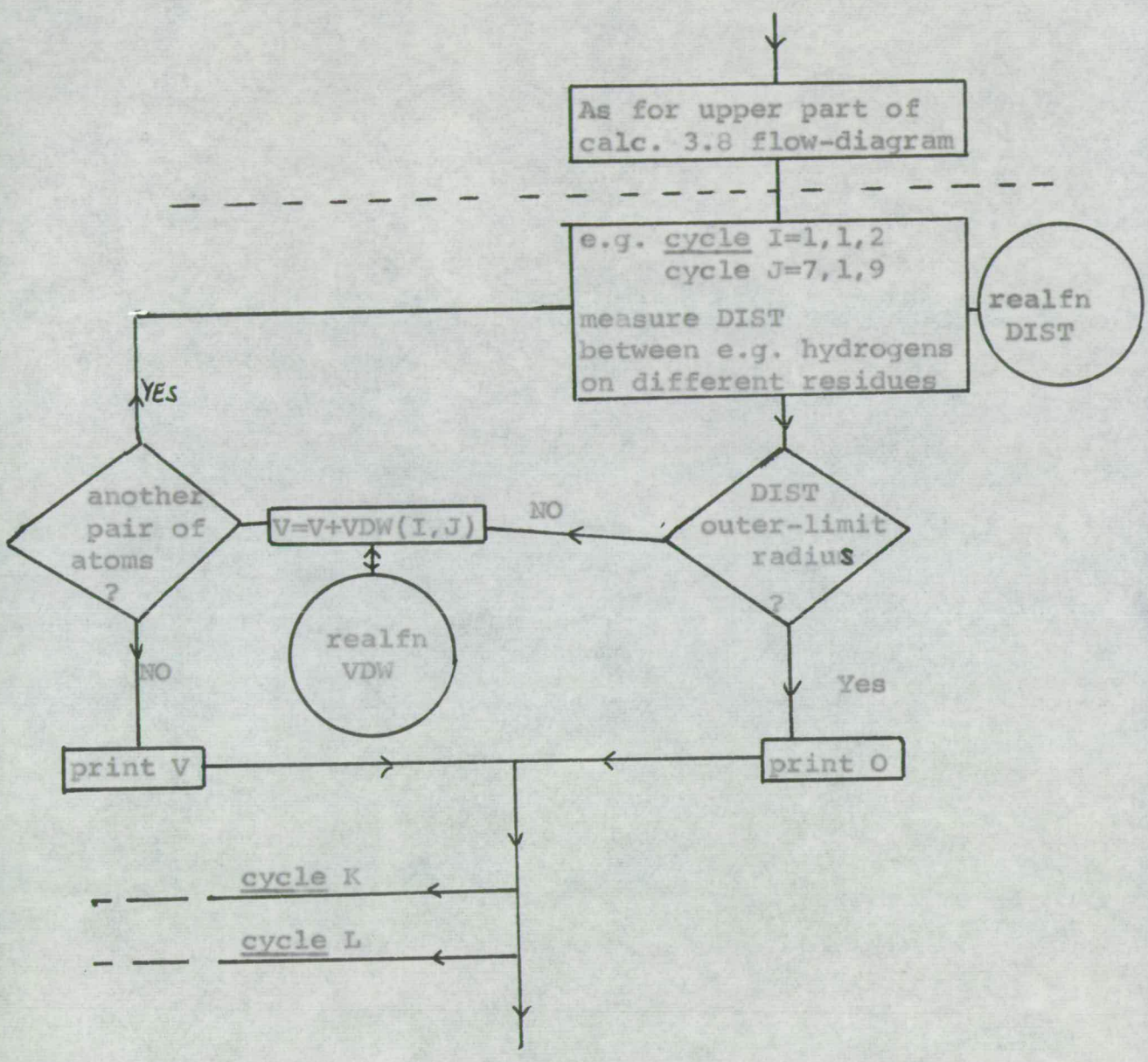
Calculation 3.1. Derivation of Ramachandran co-ordinates.



Calculation 3.5 Ramachandran conformational map.



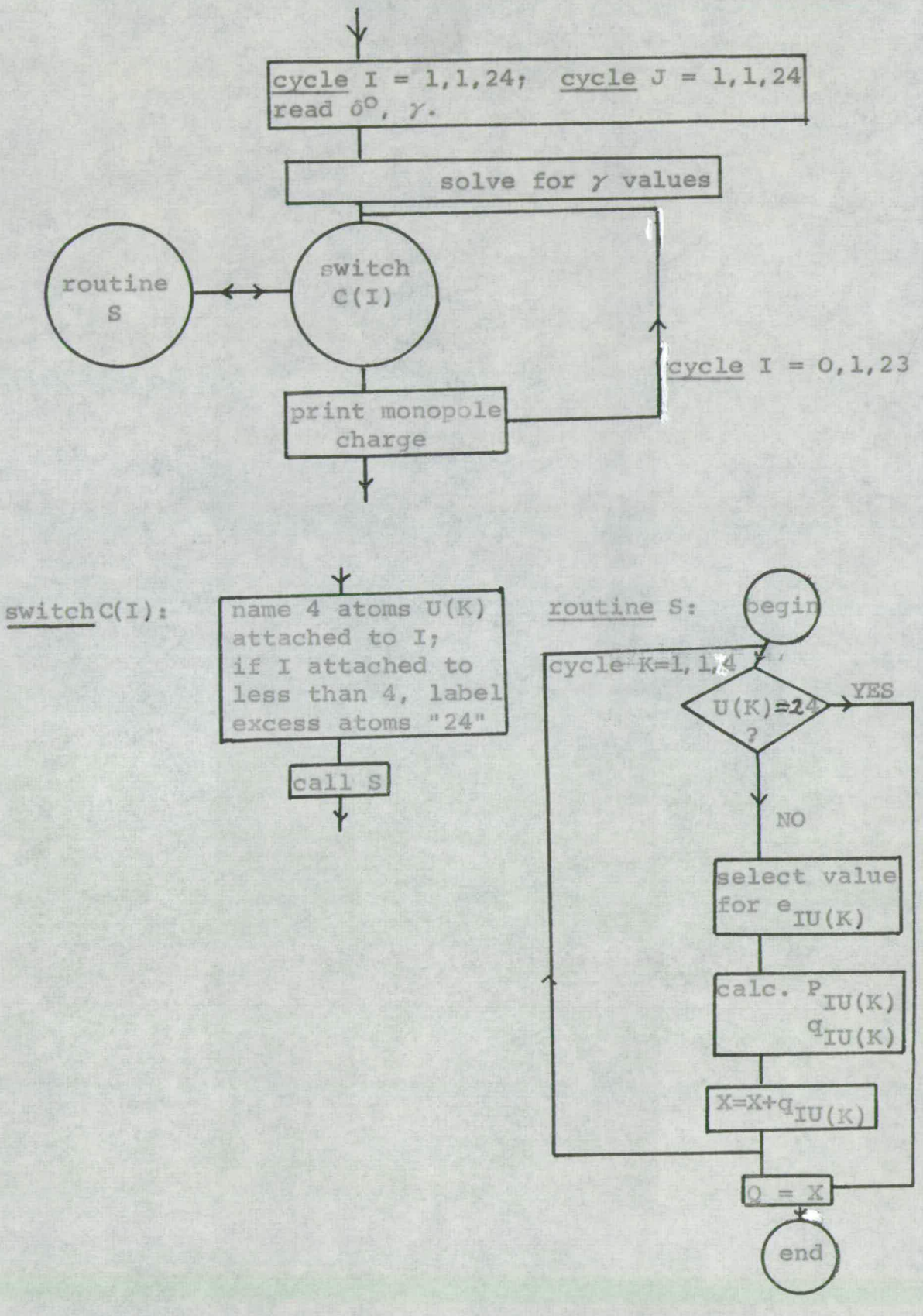
Calculation 3.8 Total conformational map



Calculation 3.12. Van der Waals conformational map.

realfn DIST measures distance between two atoms.

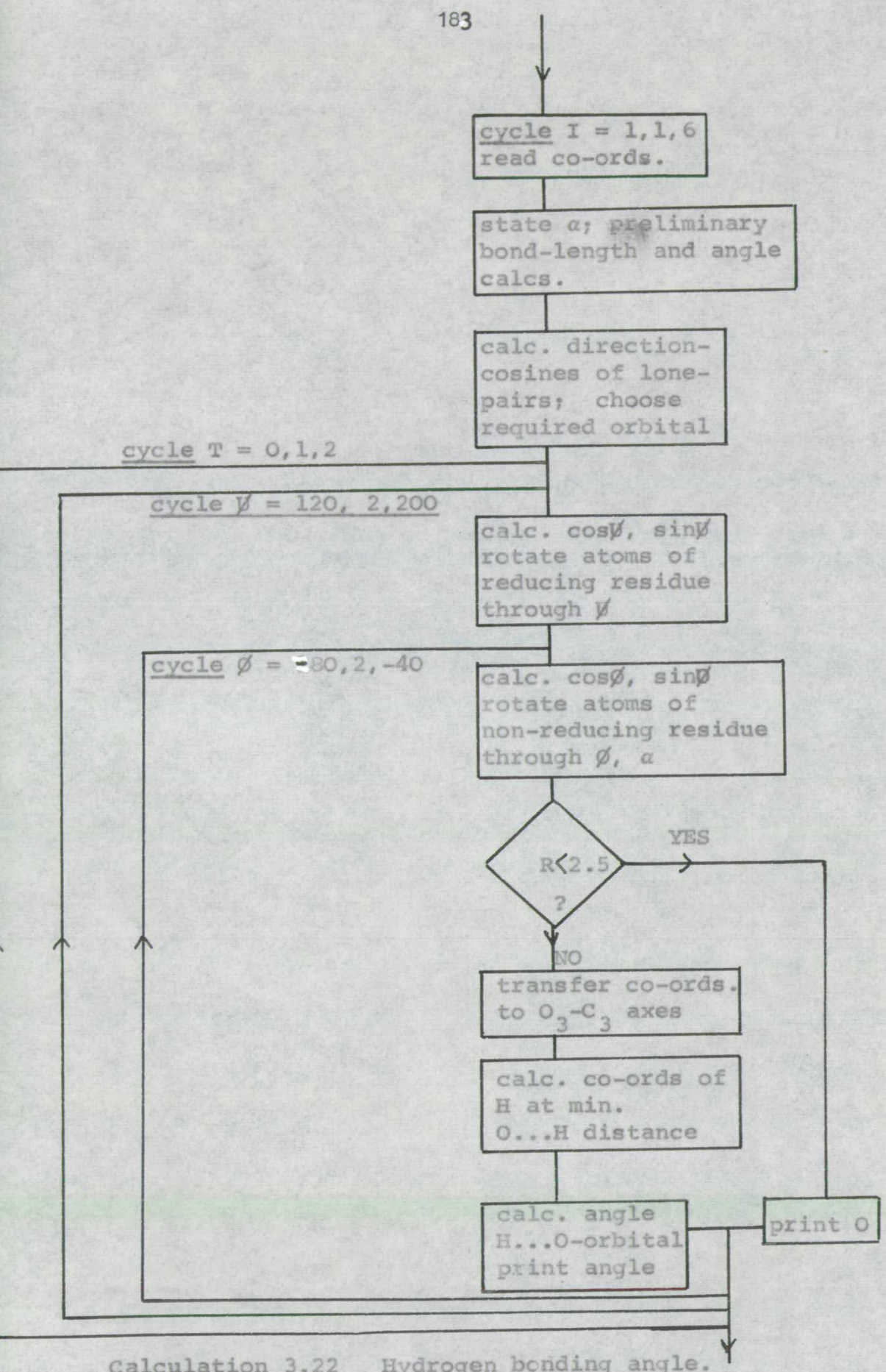
realfn VDW measures van der Waals potential between two atoms.



Calculation 3.16. Calculation of monopole charges.

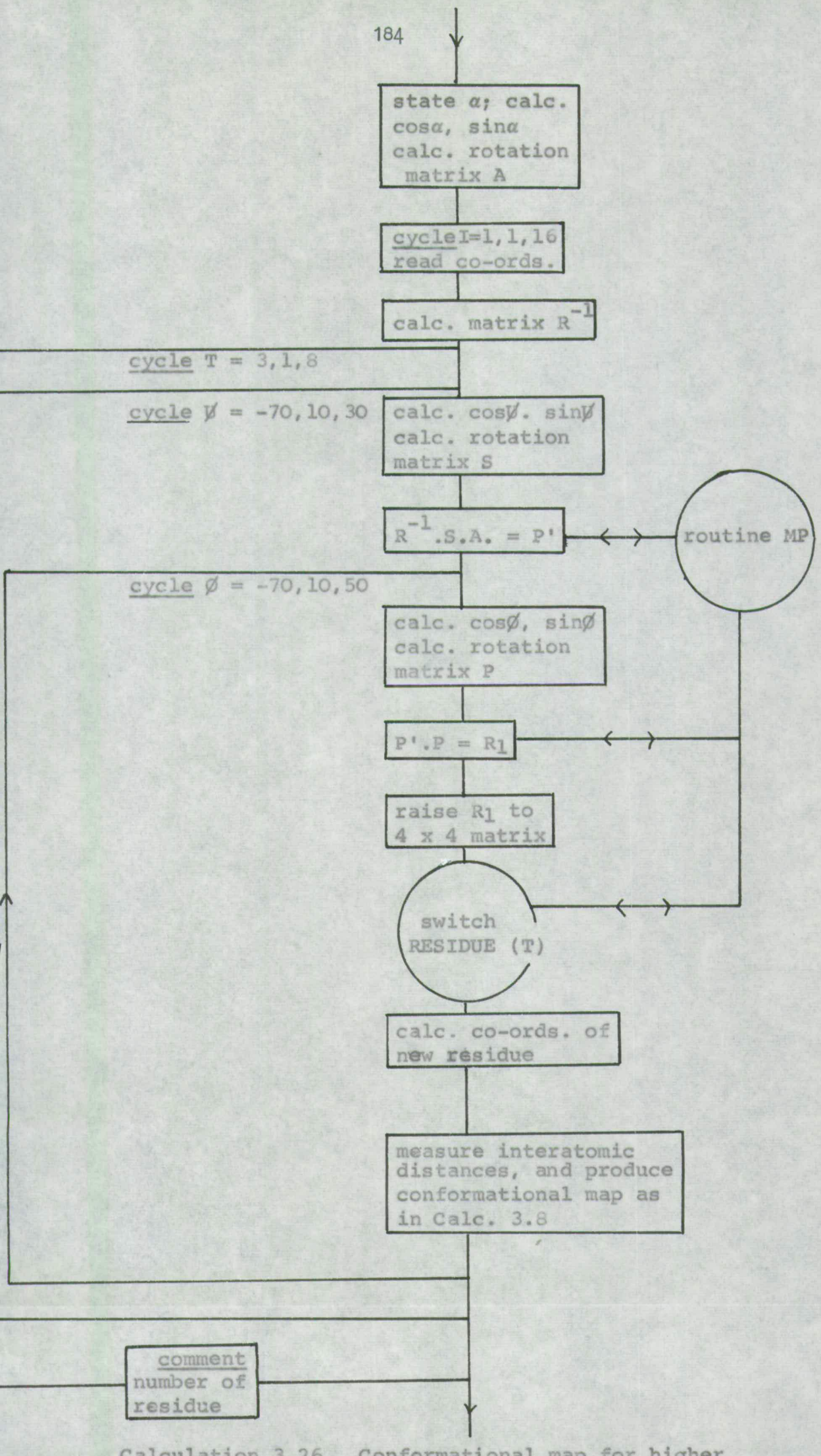
switch C(I) names atoms attached to atom I

routine S calculates charge from δ value.



Calculation 3.22 Hydrogen bonding angle.

The map has to be produced in three sections, by means of the T-cycle, because of the size of the output paper.



Calculation 3.26. Conformational map for higher residues in laminaran.

routine MP multiplies square matrices

switch RESIDUE (T) multiplies R_1 the appropriate number of times.

References

1. P. A. Roelofsen, The Plant Cell Wall, Encyclopedia of Plant Anatomy, Gebruder Borntraeger, Berlin 1959, Vol. 3, Part 4; a) p.54; b) p.63; c) p.69; d) p.40; e) p.70; f) p.59; g) p.63; h) p.65; j) p.62; k) p.10; l) p.198.
2. H. Meier, Biochem. Biophys. Acta, 28, 229 (1958).
3. H. H. Martin, Ann. Rev. Biochem. 35, 457, (1966).
4. C. T. Bishop, Canad. J. Chem., 33, 1521 (1955).
5. E. L. Hirst, J. K. N. Jones and W. O. Walder, J. Chem. Soc., 1225, (1947).
6. G. O. Aspinall and A. Nicolson, J. Chem. Soc., 2503 (1960).
7. F. Smith and R. Montgomery, The Chemistry of Plant Gums and Mucilages, Reinhold Publishing Corpn., New York, 1959; a) p.281; b) p.306; c) p.350; d) p.355.
8. D. A. Rees and N. G. Richardson, Biochemistry, 5, 3099, (1966).
9. E. G. V. Percival, Structural Carbohydrate Chemistry, J. Garnet Miller, London, 1962.
10. P. A. Roelofsen and D. R. Kreger, J. Exp. Bot., 2, 332, (1951).
11. A. Frey-Wyssling and K. Mühlethaler, Ultrastructural Plant Cytology, Elsevier, Amsterdam, 1965, p. 275.
12. K. Mühlethaler, Exp. Cell. Res., 1, 341, (1950).
13. M. H. F. Wilkins, Nobel Lectures, Physiology and Medicine, Elsevier, New York, 1964, p. 754.
14. P. H. Hermans, Kolloid-Z., 102, 169 (1943).
15. G. W. Monier-Williams, J. Chem. Soc., 119, 803 (1921).
16. M. Dubois, K. A. Gillies, J. K. Hamilton, P. A. Rebers and F. Smith, Analyt. Chem., 28, 350 (1956).
17. C. M. Wilson, Analyt. Chem., 31, 1199 (1959).
18. E. A. McComb and R. M. McCready, Analyt. Chem., 24, 1630 (1952).
19. J. Mann and H. J. Marrinan, J. Appl. Chem., 4, 204 (1954).

20. J. Mann and H. J. Marrinan, Trans. Faraday Soc., 52, 481, 487, 492 (1956).
21. R. Nodzu and T. Tomita, Chem. Abs. 48, 8568 (1954).
22. R. D. Preston, Polymer, 3, 511, (1962).
23. H. J. Marrinan and J. Mann, J. Polymer Sci., 21, 301 (1956).
24. P. A. Roelofsen, V. C. Dalitz and C. F. Wijnman, Biochem. Biophys. Acta 11, 344, (1953).
25. H. Spedding and J. O. Warwicker, J. Polymer Sci. A2, 3933 (1964).
26. B. Rånby and R. W. Noe, J. Polymer Sci. 51, 337 (1961).
27. C. A. Coulson in Hydrogen Bonding, Ed. D. Hadzi, Pergamon, London, 1959, p. 350.
28. R. H. Marchessault and A. Sarko, Adv. Carbohydrate Chem., 22, 473, (1967).
29. N. S. Anderson, Ph.D. thesis, Edinburgh, 1966.
30. G. Némethy, Angew.Chem. International Ed., 6, 195 (1967).
31. K. Bailey and F. Norris, Biochem. J., 26, 1609 (1932).
32. K. Bailey, Biochem. J., 29, 2477 (1935).
33. J. Tyler, J. Chem Soc., 5288 (1965).
34. ibid., p. 5300.
35. A. Erskine and J. K. N. Jones, Canad. J. Chem., 35, 1174 (1957).
36. K. Hunt and J. K. N. Jones, Canad. J. Chem., 40, 1266 (1962).
37. S. James and F. Smith, J. Chem. Soc., 739 (1945).
38. G. Aspinall and J. Baillie, J. Chem. Soc., 1702 (1963).
39. ibid., p. 1714.
40. K. Hamaguchi and E. Geiduschek, J. Amer. Chem. Soc., 84, 1329 (1962).
41. D. Kay (ed.) Techniques in Electron Microscopy, Blackwell, Oxford (1961).

42. G. Ackers and R. Steere, Nature, 194, 114 (1962).
43. S. Gould, D. Rees, N. Richardson and I. Steele, Nature, 208, 876 (1965).
44. H. Spedding, Adv. Carbohydrate Chem. 19, 23, (1964).
45. J. Mann, Pure Appl. Chem., 5, 91 (1962).
46. M. Tsuboi, J. Polymer Sci., 25, 159 (1957).
47. J. Mann and H. Marrinan, J. Polymer Sci., 27, 595, (1958).
48. C. Liang and R. Marchessault, J. Polymer Sci., 39, 269, (1959).
49. R. Marchessault and C. Liang, J. Polymer Sci., 59, 357, (1962).
50. R. Fraser, J. Chem. Phys., 21, 1511 (1953).
51. I. W. Steele, Ph. D. Thesis, Edinburgh, 1967.
52. J. Cumberbirch, and H. Spedding, J. Appl. Chem., 12, 83 (1962).
53. G. Jeffrey and R. Rosenstein, Adv. Carbohydrate Chem., 19, 7, (1964).
54. L. Pauling, R. Corey and H. Branson, Proc. Natl. Acad. Sci. U.S.A., 37, 205, (1951).
55. G. Ramachandran, C. Ramakrishnan and V. Sasisekharan, in G. Ramachandran (Ed.), Aspects of Protein Structure, Academic Press, New York, 1963, p. 121.
56. D. Jones, J. Polymer Sci., 32, 371 (1958).
57. W. Settineri and R. Marchessault, J. Polymer Sci. part C,11, 253 (1965); [W. Settineri, Ph.D. thesis, Syracuse, 1966].
58. D. Brant and P. Flory, J. Amer. Chem. Soc., 87, 2791 (1965).
59. A. Liquori, J. Polymer Sci., part C,12, 209 (1966).
60. V. Rao, P. Sundararajan, C. Ramakrishnan and G. Ramachandran in G. Ramachandran (Ed.), Conformation of Biopolymers, Academic Press, New York, 1967, p. 721.

61. A. Sarko and R. Marchessault, J. Amer. Chem. Soc., 89, 6454 (1967).

62. N. Allinger, M. Miller, F. Vancatledge and J. Hirsch, J. Amer. Chem. Soc., 89, 4345 (1967).

63. M. Sundaralingam, Biopolymers, 6, 189 (1968).

64. B. Poppleton and A. Mathieson, Nature, 219, 1048 (1968).

65. R. Jacobson, J. Wunderlich and W. Lipscomb, Acta Cryst., 14, 598 (1961).

66. C. Brown, J. Chem. Soc., (A), 927 (1966).

67. S. Chu and G. Jeffrey, Acta Cryst., B24, 830 (1968).

68. T. Miyazawa, J. Polymer Sci., 55, 215 (1961).

69. K. Ellis and J. Warwicker, Nature, 181, 1614 (1958).

70. W. McCrea, Analytical Geometry of Three Dimensions, Oliver and Boyd, Edinburgh 1960.

71. A. Aitken, Determinants and Matrices, Oliver and Boyd, Edinburgh 1964.

72. C. Ramakrishnan, Proc. Indian Acad. Sci., sect. A,12. 327 (1964).

73. K. Meyer and L. Misch, Helv. Chim. Acta, 20, 232 (1937).

74. M. Ihnat and D. Goring, Canad. J. Chem., 45, 2353, 2363 (1967).

75. T. Ooi, R. Scott, G. Vanderkooi and H. Scheraga, J. Chem. Phys., 46, 4410 (1967).

76. K. Pitzer, Adv. Chem. Phys., 2, 59 (1959).

77. T. Hill, J. Chem. Phys., 16, 399 (1948).

78. A. Kitaygorodsky, Tetrahedron, 14, 230 (1961).

79. N. Allinger, J. Hirsch, N. Miller, I. Tyminski and F. Vancatledge, J. Amer. Chem. Soc., 90, 1199 (1968).

80. P. de Santis, E. Giglio, A. Liquori and A. Ripamonti, Nature, 206, 456 (1965).

81. J. de Coen, G. Elefanti, A. Liquori and A. Damiani, Nature, 216, 910 (1967).

82. E. Jacob, H. B. Thompson and L. Bartell, J. Chem. Phys., 47, 3736 (1967).
83. H. Scheraga, Adv. Phys. Org. Chem., 6, 103 (1968).
84. A. Hybl, R. Rundle and D. Williams, J. Amer. Chem. Soc., 87, 2779 (1965).
85. D. Rees and R. Skerrett, Carbohyd. Res., 7, 334 (1968).
86. P. de Santis, E. Giglio, A. Liquori and A. Ripamonti, J. Polymer Sci., A1, 1383 (1963).
87. H. Mark and E. Pohland, Zeit. Kryst., 61, 239 (1925).
88. J. Andrews, J. Amer. Chem. Soc., 47, 1597 (1925).
89. R. Le Fèvre and D. Narayana Rao, Austral. J. Chem., 7, 135 (1954).
90. International Critical Tables, 6, 76 (1929).
91. ibid, 1, 102 (1929).
92. T. Jordan, W. Streib, H. Smith and W. Lipscomb, Acta. Cryst. 17, 777 (1964).
93. E. Eliel, N. Allinger, S. Angyal and G. Morrison, Conformational Analysis, Interscience, New York, 1965, p. 437.
94. R. Marchessault, Chimie et Biochimie de la Lignine, de la Cellulose, et des Hemicelluloses, Les Imprimeries Réunies de Chambéry, 1965, p. 287.
95. R. Arridge and C. Cannon, Proc. Royl. Soc. (London), Ser.A 278, 91 (1964).
96. C. Coulson, Electricity, 3rd ed., Oliver and Boyd, Edinburgh, 1953, p. 157.
97. D. Poland and H. Scheraga, Biochemistry, 6, 3791 (1967).
98. G. del Re, J. Chem. Soc., 4031 (1958).
99. J. Edward, Chem. and Ind., 1102 (1955).
100. Yu. Zhdanov, V. Minkin, Yu. Ostroumov and G. Dorofeenko, Carbohyd. Res., 7, 156 (1968).
101. L. and M. Fieser, Organic Chemistry (3rd ed.) Reinhold,

New York, 1956, p. 373.

102. International Critical Tables, 2, 310 (1929).

103. Handbook of Chemistry and Physics (46th ed.) Chemical Rubber Co. (1965) E52.

104. G. Clark and A. Smith, J. Phys. Chem., 40, 863 (1937).

105. K. Meyer and W. Wehrli, Helv. Chim. Acta, 20, 353 (1937).

106. G. Pimentel and A. McClellan, The Hydrogen Bond, Freeman, San Francisco, 1960. p. 212.

107. E. Lippincott and R. Schroeder, J. Phys. Chem., 61, 921, (1957).

108. E. Lippincott, J. Chem. Phys., 26, 1678 (1957).

109. M. Weissmann and N. Cohan, J. Chem. Phys., 43, 119 (1965).

110. E. Lippincott, J. Finch and R. Schroeder, in ref. 27, p. 361.

111. W. Stockmayer, J. Chem. Phys., 9, 398 (1941).

112. W. Moulton and R. Kromhout, J. Chem. Phys., 25, 34 (1956).

113. K. Gibson and H. Scheraga, Biopolymers, 4, 709 (1966).

114. S. Warsi and W. Whelan, Chem. and Ind., 35, 1573 (1957).

115. G. Aspinall and G. Kessler, Chem. and Ind., 35, 1296 (1957).

116. S. Peat, W. Whelan and H. Lawley, J. Chem. Soc., 729 (1958).

117. N. Handa and K. Nisizawa, Nature, 192, 1078 (1961).

118. S. Peat, W. Whelan and T. Edwards, J. Chem. Soc., 3862 (1958).

119. H. Bouveng, H. Kiessling, B. Lindberg and J. McKay, Acta Chem. Scand., 17, 1351 (1963).

12o. S. Peat, W. Whelan and J. Roberts, J. Chem. Soc., 3916 (1957).

121. B. Casu, M. Reggiani, G. Gallo and A. Vigevani, Tetrahedron, 22, 3061 (1966).

122. I. Mackie, and E. Percival, J. Chem. Soc., 1151 (1959).

123. E. Frei and R. Preston, Proc. Roy. Soc. (London) Ser. B. 160, 293 (1964).

124. R. Preston, Scientific American 218(6), 102, (1968).

125. E. G. V. Percival and S. K. Chanda, Nature, 166, 787 (1950).

126. R. Bailey, Oligosaccharides, Pergamon Press, Oxford, 1965, p. 57.

127. P. Gorin, J. Spencer and D. Westlake, Canad. J. Chem., 39, 1067 (1961).

128. H. Sugeta and T. Miyazawa, Biopolymers, 5, 673 (1967).

129. H. Thompson, J. Chem, Phys., 47, 3407 (1967).

REPRINTED FROM

CARBOHYDRATE RESEARCH

ELSEVIER PUBLISHING COMPANY
AMSTERDAM

Carbohydrate Research
Elsevier Publishing Company, Amsterdam
Printed in Belgium

CONFORMATIONAL ANALYSIS OF CELLOBIOSE, CELLULOSE, AND XYLAN

D. A. REES AND R. J. SKERRETT
Chemistry Department, University of Edinburgh, West Mains Road, Edinburgh 9 (Great Britain)
(Received January 1st, 1968)

ABSTRACT

All of the likely conformations of cellobiose, cellulose, and xylan have been explored systematically by using an electronic computer, assuming the ring conformations and (C-1)–O–(C-4′) angle for each pair of residues to be fixed and derivable from known crystal structures. The absolute van der Waals energies, but not the relative energies of different conformations, are sensitive to the choice of energy functions and atomic coordinates. The conformation of cellobiose in the crystal occurs near the minimum for intramolecular van der Waals interactions, perhaps slightly displaced to allow formation of an inter-residue hydrogen bond. The Hermans chain conformation for cellulose corresponds to a somewhat higher van der Waals energy, but this is probably offset by efficient crystal packing resulting from the two fold screw axis, and the hydrogen bond remains. The Meyer and Misch conformation is unlikely, because of its high van der Waals energy and for other reasons. The results lead to possible explanations of the known conformational stiffness of cellulose and its solubility properties in alkali. The characteristics of xylan conformations have been compared with cellulose.

INTRODUCTION

The molecular structure of cellulose was determined over thirty years ago, to a large extent through the brilliant researches of Sir Edmund Hirst and his colleagues[1]. It was one of the first polymer structures to be established and was important for subsequent development of polymer theory. In contrast, and despite much research effort, the conformation and crystal structure of cellulose are still unsolved. Most workers favour the "Hermans" or "bent chain" conformation[2] rather than the conformation proposed earlier by Meyer and Misch[3], but other forms have also been suggested[4], and it is possible that none is correct because each is based on the assumption that the cellulose chain has a two-fold screw axis which, strictly, is unnecessary[5]. Outside carbohydrate chemistry, the calculation of conformational energies is becoming accepted as a valid approach to the determination of conformations. Except for very simple molecules, such as hydrocarbons[6], the methods are not yet refined to the extent that they can give final answers without the need to invoke

obtained by averaging the coordinates listed by Ramachandran *et al.*[13] for various β-D-glucose residues, and taking a value of 117° for the (C-1)–O–(C-4′) angle. *Set 2* was derived from the crystal structure coordinates of Brown[19], for which the (C-1)–O–(C-4′) angle is 116.7°. *Set 3* was obtained by modification of Brown's coordinates[19] to correspond to a (C-1)–O–(C-4′) angle of 117°. *Set 4* was derived from the refined crystal structure coordinates of Chu and Jeffrey[20], for which the (C-1)–O–(C-4′) angle is 116.1°. *Set 5* was obtained by modification of Chu and Jeffrey's coordinates to correspond to a (C-1)–O–(C-4′) angle of 116.7°. *Set 6* was obtained by modification of Chu and Jeffrey's coordinates to correspond to a (C-1)–O–(C-4′) angle of 117°.

Each set was transferred to new axes which had the bridge oxygen as origin; the O–(C-4′) bond was the x axis for the reducing residue and the (C-1)–O bond produced was the x axis for the nonreducing residue. The xz planes were defined by (O-4′)–(C-4′)–(H-4′) and (O-4′)–(C-1)–(H-1), respectively, with the y axes perpendicular to these planes to form right-handed systems of axes. These transformations were done in stages by using standard expressions for the rotation and translation of axes. The systematic rotations about bonds to the bridge oxygen, followed by calculation of all interatomic distances and the projected residue length, h, were performed on the English Electric KDF9 computer with programs written in Atlas Autocode. These calculations, and those for n, were based on the principles outlined by Ramakrishnan[12]. For calculations of n and h, (C-1)–(O-1) in each residue was taken to be parallel to (C-4)–(O-4), and (C-1)–H to be coplanar with (C-4)–H. This is inexact, but leads to simple expressions for n and h and makes no difference to the energy maps. The exact approach is, in any case, impossible when coordinates are taken, as in this work, from disaccharide structures in which the residues differ in detailed geometry. Comparison with results derived rigorously showed that any error was likely to be within the accuracy of experimental data for polysaccharide fibres, at least for conformations examined here. From their published results[13], it would seem that Ramachandran *et al.* also used this approximation. The exact expressions would have to be used for higher screw symmetries, because serious errors would otherwise result.

The increments of rotation were 2° for energy calculations, but otherwise 10°. The angles of rotation ϕ and ψ (see **1**), were the dihedral angles between (C-1)–(H-1) and (O-4′)–(C-4′), and between (O-4′)–(C-1) and (C-4′)–(H-4′), respectively. The conformation $\phi = \psi = 0$ is defined as having H-4′ and H-1 in the same plane as (C-1)–(O-4′)–(C-4′), and the apex of the (C-1)–(O-4′)–(C-4′) angle pointing in the same sense as the (C-4′)–(H-4′) bond and in the opposite sense from (C-1)–(H-1). To convert a model of cellobiose, in this form, into any conformation (ϕ, ψ), the reducing residue is held stationary and, viewed from the nonreducing end, the nonreducing residue is rotated anticlockwise through ψ about (C-4′)–(O-4′) and clockwise through ϕ about (O-4′)–(C-1).

The interatomic distances, which were used to classify the conformations, are listed in Table I. The computer program was written in such a way that *one* inter-

experimental evidence, but accurate predictions have been possible for some synthetic polymers[7] and polypeptides[8-12]. Pioneer work with polysaccharides by Ramachandran *et al.*[13,14], and related approaches[15,16], are described in more detail below. In this paper, we report on calculations which confirm that the method has promise for polysaccharides and which add further evidence in favour of the Hermans conformation for cellulose. Some unusual physical properties of cellulose are discussed in the light of the results.

PRINCIPLE OF THE METHOD

We assume that solid-state and solution conformations can be understood from consideration of intramolecular interactions. This is reasonable, because the same assumption has led to useful results for other polymers[7-12] and because the properties of cellulose suggest that intramolecular interactions are especially important[17]. Present evidence suggests that there are standard values which can be assumed for bond angles and bond lengths in carbohydrates[13,18], and that the common sugar residues in polysaccharides will favour the *C1* conformation. The overall conformation of the chain would then depend only on the angles of rotation about the two bonds to each glycosidic oxygen atom. The rotational states about these bonds were therefore explored systematically, as described below, as an aid to selecting the most likely conformations for testing against experimental evidence.

A set of coordinates was chosen for cellobiose, based on crystal structure data (see below). New coordinates were then calculated from these to correspond to various angles of rotation about the (C-1)–O and O–(C-4′) bonds. This was done by computer, in such a way that all possible conformations were sampled. The new sets of coordinates were used to characterise each conformation in several ways: (*i*) The distances between all relevant pairs of atoms were calculated, and, on this basis, each conformation was classified as either *fully allowed* (implying no infringement of van der Waals radii), *marginally allowed* (implying slight compression of some atoms or groups), or *disallowed* (implying that atoms or groups are brought so close that severe van der Waals repulsion is expected). (*ii*) To find the van der Waals energy minimum, each interatomic distance was used to calculate the corresponding energies of van der Waals attraction and repulsion, and these values were summed over all interactions for each allowed conformation. (*iii*) The possibility of intramolecular hydrogen bonding was assessed from the (O-5)–(O-3′) distance. (*iv*) For all conformations, the two D-glucose residues can be considered to be related by an *n*-fold screw axis, where *n* need not be an integer. Values of *n* were calculated for selected conformations. (*v*) The length of the disaccharide projected on the screw axis, *h*, was also calculated for selected conformations.

PARAMETERS AND MATHEMATICAL METHODS

The sensitivity of the results to inaccuracy in the atomic coordinates was assessed by repeating the calculations for six sets which were as follows: *Set 1* was

We used the constants given by Rao *et al.*[14], together with the following values which were calculated using the methods of earlier workers:

for H ... CH_3 interactions, $A = 718.5$, $B = 5.31 \times 10^5$
for O ... CH_3 interactions, $A = 2016.0$, $B = 3.10 \times 10^6$
for C ... CH_3 interactions, $A = 2605.0$, $B = 6.75 \times 10^6$.

These will be referred to as the "Flory functions". (*iii*) The functions proposed by Kitaygorodsky[22], which are of the Buckingham type, but expressed thus:

$V = 3.5(-0.04/z^6 + 8.6 \times 10^3 e^{-13z})$, where $z = r/r_0$.

Only one constant, r_0, is dependent on the particular atom pair. The values of r_0 listed by Rao *et al.*[14] were used, together with the following values which were derived geometrically:

for H ... CH_3 interactions, $r_0 = 3.33$
for O ... CH_3 interactions, $r_0 = 3.76$
for C ... CH_3 interactions, $r_0 = 3.92$.

These will be referred to as the "Kitaygorodsky functions".

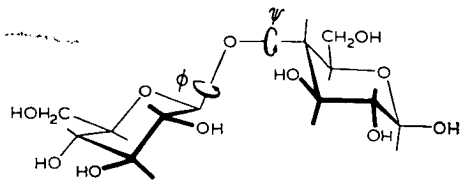
The dihedral angles ϕ and ψ calculated for cellobiose in the crystalline state from the data of Brown[19], and of Chu and Jeffrey[20], will be referred to as the cellobiose (*B*) and cellobiose (*J*) conformations, respectively. The values for a cellulose chain with a two-fold screw axis and projected residue repeat of 5.15 Å were taken as those at the intersections of the curves which represent $n = 2$ and $h = 5.15$, respectively (see below). Two solutions exist for each set of coordinates. One resembles the bent-chain conformation[2] in general features and will be referred to as cellulose (*H*). The Meyer and Misch conformation[3] is referred to as cellulose (*M*); although this is quite close to the second solution, the dihedral angles are taken as (0°, 180°) for all sets of coordinates, for reasons given below.

All potential energies are given in Kcal/mole per disaccharide residue.

RESULTS AND DISCUSSION

Several attempts at exploration of polysaccharide conformations have been reported previously. Jones[15] attempted to refine the crystal structure of cellulose by an ingenious approach to systematic model building. Ramachandran *et al.*[13] have used mathematical methods to examine some of the possible cellulose conformations, with interatomic distances as criteria. Calculation was by hand, however, and it was therefore feasible to include only a few atoms and to cover only part of the conformational map. More recently, the same group[14] has made an important study of amylose, in which conformational energies were calculated by computer. Setterineri and Marchessault[16] have used computer methods to search for the xylan conformation which best satisfies stereochemical criteria and experimental evidence from X-ray diffraction and infrared dichroism.

atomic distance within the "fully allowed" limit was sufficient to classify a conformation as marginally allowed, unless any distance was within the "marginally allowed" limit, when it was classified as disallowed.



Scheme for variation of bridge-angle geometry in cellobiose
$\phi = \psi = 0°$
Rotations are positive when each residue is rotated in the direction shown about its bond to the bridge oxygen.

TABLE I

INTERATOMIC DISTANCES USED FOR THE HARD SPHERE CALCULATIONS[a]

Atom pair	*Minimum distance (Å) allowed for:*	
	Fully allowed contact	*Marginally allowed contact*
C ... C	3.20	3.00
C ... O	2.80	2.70
C ...H	2.40	2.20
O ... O	2.80	2.70
O ... H	2.40	2.20
H ... H	2.00	1.90
H ... CH_2OH	2.42	2.42
O ... CH_2OH	2.90	2.90
C ... CH_2OH	3.20	3.20

[a]Hydrogen bonding distances are given separately in the Figure captions.

Van der Waals interactions are normally expressed by functions of either the Lennard-Jones type (potential energy, $V = -A/r^6 + B/r^{12}$, where r is the interatomic distance, and A and B are constants), or the Buckingham type ($V = -A/r^6 + Be^{-\mu r}$, where A, B, and μ are constants). The constants may be different for each atom pair and have to be derived empirically. We have made the additional approximation that the hydroxymethyl group is treated as equivalent to a methyl group in its van der Waals interactions, on the grounds that the hydroxyl could probably rotate away from any region of steric overcrowding. The following three sets of functions were used to calculate the conformational energies reported in this paper: (*i*) Those given by Liquori and his coworkers[8], which include functions of both types. These are referred to as "Liquori functions". (*ii*) The functions proposed by Brant and Flory[21], which are of the Buckingham type with $\mu = 4.6$ and A and B derived for each atom pair.

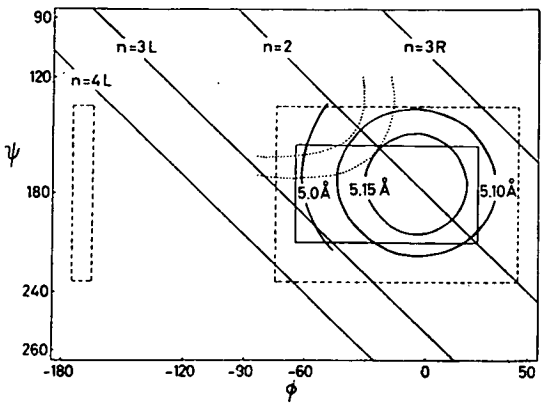


Fig. 1. Restriction of cellulose conformations by interactions between C-1 and C-3′, C-1 and C-5′, O-5 and C-4′, and C-2 and C-4′, based on *Set 1* coordinates modified to correspond to a (C-1)–O–(C-4′) angle of 117.5°. Fully and marginally allowed conformations lie within the solid and broken rectangles respectively. The limits of (O-5) ... (O-3′) hydrogen bonding, assuming that this requires the O ... O distance to be 2.50–2.80 Å, are shown by dotted curves. Conformations with the same screw axis lie on the diagonals and those with the same projected repeat distance on the solid curves. The coordinates differ slightly from those used earlier[13] (see text). The angle of rotation, ϕ, is equivalent to the ϕ of Ramachandran *et al.*[13], and $\psi = 360° - \phi'$.

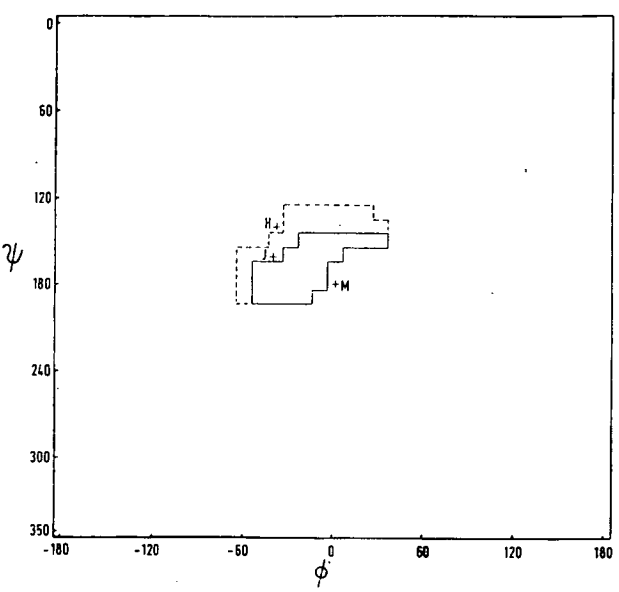


Fig. 2. Allowed (enclosed by solid line) and marginally allowed (broken lines) cellulose conformations, using *Set 4* coordinates with all relevant contacts considered. Cellobiose (*J*), cellulose (*H*), and cellulose (*M*) are marked as *J*, *H*, and *M*, respectively.

Waals interactions (especially repulsion) might be particularly important. At this stage of the investigation, we have therefore attempted a quantitative approach to van der Waals potential energy, while considering other interactions qualitatively.

A preliminary assessment of the types of conformation which are possible for a given polymer is most easily and quickly made by treating the atoms as "hard spheres" which are incompressible beyond certain limits. The results are expressed in the form of a "conformational map" which shows the combinations of ϕ and ψ permitted by these criteria. The map may also include the calculated properties of each range of allowed conformations, such as screw symmetry, projected repeat distance, and likelihood of hydrogen bonding within the chain. Fig. 1 shows results which represent the restriction of possible cellulose conformations by interactions between C-1 and C-3′, C-1 and C-5′, O-5 and C-4′, and C-2 and C-4′. The same contacts were examined over a limited region of the map by Ramachandran *et al.*[13], and our results agree with theirs. When *all* relevant contacts were included, the allowed conformations occupied smaller areas (Fig. 2). Conformations were sampled at 10° intervals about each of the C-O bonds, so that a total of $36^2 = 1296$ representative conformations were examined. Only 24 fully allowed conformations were found, and only a further 21 were marginally allowed, *i.e.*, 96% of the conformations are excluded! The results seem to explain the stiffness of the cellulose chain, which has often been noted[17], and are consistent with other available evidence. The accepted crystal conformation of chitin[23] and the most-favoured cellulose conformation[2] each correspond to (−40°, 140°), which is almost within the allowed area. In Fig. 2, this point is outside by only one 10° increment in ϕ, and recalculation at 2° intervals showed it to be less than 2° outside. With other sets of coordinates, it was actually inside. The crystal conformation of cellobiose[19,20] is in the allowed area (Fig. 2). Although atomic coordinates have not been published, inspection of models and photographs[24] of models of the lysozyme complex with tri-*N*-acetylchitotriose [*O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1→4)-*O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1→4)-2-acetamido-2-deoxy-D-glucose] shows that each glycosidic linkage in the trisaccharide has a conformation which is close to cellobiose in the crystal and therefore within the allowed region. Finally, p.m.r. spectroscopy[25] suggests an intramolecular hydrogen bond between O-5 and H-3′ for cellobiose in methyl sulphoxide solution, and this is also consistent with calculation because this hydrogen bond is possible for allowed conformations (compare Figs. 1 and 2). Reasons why all of these conformations should be "marginally allowed", rather than "fully allowed", are discussed below.

The scope of the hard-sphere approach is limited because it does not show which parts of an allowed area are likely to be preferred, and therefore the polymer conformation is only defined approximately. To go further, it would be desirable to search for free energy minima, by estimating and summing the free energies of all intramolecular interactions for each conformation, including dipole interactions, hydrogen bonding, torsional strain, and van der Waals attraction and repulsion. Such intermolecular interactions as solvation and crystal packing ought also to be considered. This is a formidable undertaking which, it seems to us, would best be approached by successive approximation. The severe steric restrictions indicated by hard-sphere calculations and by experiment[17] suggest that, for cellulose, intramolecular van der

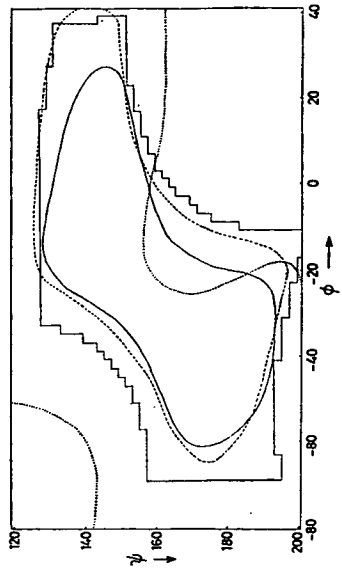


Fig. 3

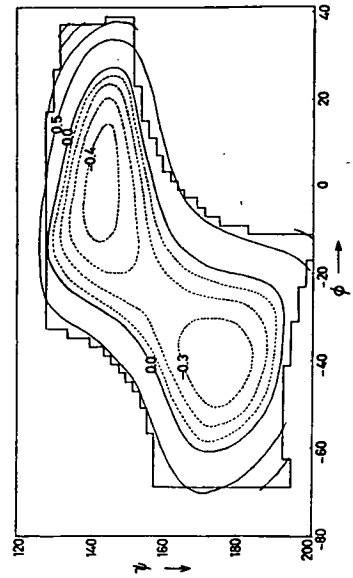


Fig. 4

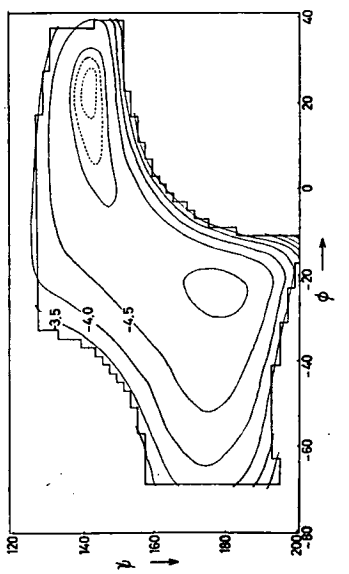


Fig. 5

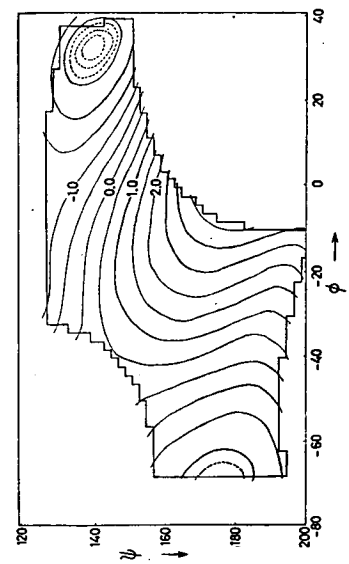


Fig. 6

Fig. 3. Entire, allowed area for cellulose conformations (boxed by solid straight lines), compared with 0 Kcal Kitaygorodsky contour (solid curve), −4 Kcal Flory contour (broken curve), and +2 Kcal Liquori contour (dotted curve), for *Set 4* coordinates.

Fig. 4. Van der Waals energy map for cellulose, using Kitaygorodsky functions and *Set 4* coordinates. The entire allowed area, predicted by hard-sphere calculations, is boxed by solid straight lines. Broken curves represent contours at intervals of 0.1 Kcal. Solid curves represent contours at intervals of 0.5 Kcal.

Fig. 5. Van der Waals energy map for cellulose, using Flory functions and *Set 4* coordinates. Other details as for Fig. 4.

Fig. 6. Van der Waals energy map for cellulose, using Liquori functions and *Set 4* coordinates. Other details as for Fig. 4.

All of the available van der Waals functions are semi-empirical, having been derived by fitting the constants in a theoretical equation to experimental data. The three sets of equations used here were derived by somewhat different fitting procedures and are therefore to some extent independent. Each has been used successfully in conformational analysis before[7,8,14,21,26]. Calculations for cellulose, with each function in turn, gave the results shown as energy maps in Figs. 4, 5, and 6. The positions of Flory and Kitaygorodsky minima agree tolerably (though not exactly), and also seem reasonable because they lie well within the allowed region. The outer energy contours calculated with these functions are similar in shape to each other and to the outline of the allowed region (Fig. 3), which also makes the results seem plausible. In contrast, the Liquori minima lie on the edge of the allowed region (Fig. 6), and the contours have an unexpected shape (Fig. 3); these results therefore seem less reliable. The most serious discrepancies between the Flory and Kitaygorodsky functions are in absolute energies at the minima and other conformations (Table II). Earlier workers have also found that Flory functions give substantially lower energies, both for carbohydrate and peptide conformations[14,26]. By comparison with experiment, energies calculated with Kitaygorodsky functions would seem to be more accurate for carbohydrates[14]. However, our chief concern at present is with *relative* energies of different conformations, for which all three sets of functions fortunately lead to similar predictions (see below). The results are also influenced by the choice of atomic coordinates. With different sets of crystal coordinates and a given set of functions, the energy minimum varied in position by up to 8° in ϕ and ψ, and energies corresponding to particular conformations varied by several tenths of 1 Kcal and sometimes more (Table II). The trends in the results do not suggest that the variations could be attributed simply to changes in the (C-1)–O–(C-4′) bond angle (Table II). Several conclusions emerge irrespective of the coordinates and energy functions however, and will now be discussed.

There seem to be two minima within the allowed area, and these have similar energies (Figs. 4–6; Table II). The cellobiose conformation of Chu and Jeffrey[20] and that of Brown, which is only slightly different[19], both correspond to energies that are above the minimum by small amounts, which are calculated with the Flory and Kitaygorodsky functions as a fraction of 1 Kcal. Estimates with the Liquori functions are larger, but otherwise in agreement with this trend. The equivalent result with the hard-sphere method was that crystal conformations occurred in the "marginally" rather than "fully" allowed region. However, calculation also showed (*e.g.*, Fig. 7) that O-3′ and O-5 are sufficiently close to hydrogen bond; the equilibrium distance between these atoms is therefore smaller than for the general case, and the conformational energy should be corrected accordingly. Although the correction turns out to be negligible for the Flory functions, it is -0.2 to -0.5 Kcal for the Kitaygorodsky functions, and this brings the van der Waals energy very close to the calculated minimum. With the added advantage of a strong, inter-residue hydrogen bond, this conformation would seem to have more favourable intramolecular interactions than any alternative. It can also form a fully hydrogen-bonded lattice[19,20]. It seems likely

that, for this and perhaps other di- and oligo-saccharides, possibilities for favourable intra- and inter-molecular interaction are so restricted that the *structure* must allow each to be simultaneously near its minimum if crystallisation is to occur with a favourable free-energy change. From this point of view, it is not surprising that some isomers should be syrups which obstinately resist crystallisation.

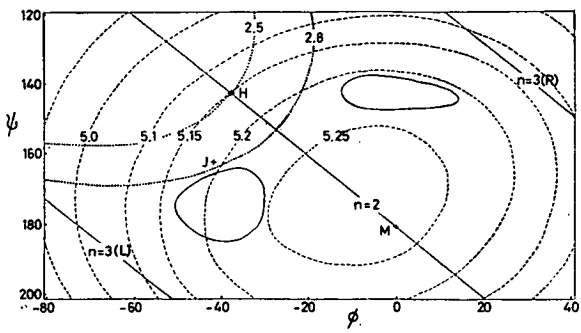


Fig. 7. Van der Waals energies and other properties of cellulose and cellobiose conformations. The lowest Kitaygorodsky contours (compare Fig. 4) are represented by solid curves, and the limits of (O-3′) ... (O-5) hydrogen bonding are represented by dotted curves, assuming that the O ... O distance must be 2.50–2.80 Å. Broken curves define conformations with the same projected repeat distance, and the diagonals define those with the same screw symmetry. $H \equiv$ cellulose (H), $M \equiv$ cellulose (M), $J \equiv$ cellobiose (J).

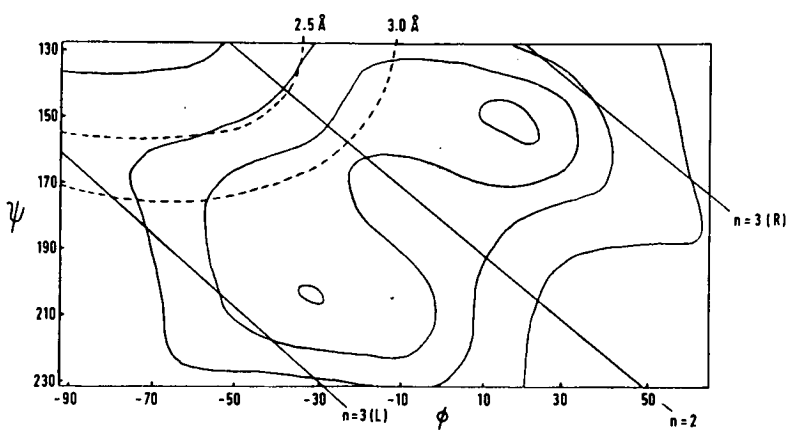


Fig. 8. Conformational energy map for xylan based on *Set 4* coordinates and Kitaygorodsky functions. The limits of (O-3′) ... (O-5) hydrogen bonding are shown by the dotted curves, assuming that the O ... O distance must be 2.50–3.00 Å. Van der Waals energy contours are shown by solid curves for −0.4, 0, 1.0, and 10.0 Kcal. Conformations having integral screw axes lie on the diagonals.

With a given set of coordinates, there are two cellulose conformations having a two-fold screw axis together with the observed fibre-repeat distance. Mathematically, these correspond to the intersections of the curves which represent $n = 2$ and $h = 5.15$Å (see Fig. 7). As Jones has pointed out[15], other solutions would be possible with a variable (C-1)–O–(C-4′) bond angle, but indications from disaccharide[19,20,27,28]

TABLE II

DIHEDRAL ANGLES AND VAN DER WAALS ENERGIES CALCULATED FOR CELLOBIOSE AND CELLULOSE CONFORMATIONS

Energy functions	*Conformation*	*Coordinates used*																		
		Set 1			Set 2			Set 3			Set 4			Set 5			Set 6			
		ϕ,°	ψ,°	V (*Kcal*)	ϕ,°	ψ,°	V (*Kcal*)	ϕ,°	ψ,°	V (*Kcal*)	ϕ,°	ψ,°	V (*Kcal*)	ϕ,°	ψ,°	V (*Kcal*)	ϕ,°	ψ,°	V (*Kcal*)	
Flory	Minima	−31	+181	− 4.9	−27	+184	−5.2	−29	+184	−5.2	−24	+175	−5.0	−24	+181	−5.1	−24	+181	−5.1	
		+19	+149	− 5.0	+21	+147	−5.1	+22	+148	−5.2	+23	+141	−5.2	+23	+142	−5.3	+24	+141	−5.3	
Kitay-gorodsky	Minima	−46	+180	0	−41	+181	−0.5	−43	+181	−0.6	−41	+175	−0.3	−40	+174	−0.4	−40	+175	−0.5	
		+ 5	+147	+ 0.4	+ 1	+147	−0.3	+ 1	+147	−0.4	0	+143	−0.4	0	+143	−0.5	0	+143	−0.6	
Liquori	Minima	−76	+175	− 1.1	−70	+182	−2.2				−72	+176	−1.7							
		+31	+147	− 1.0	+32	+146	−2.5				+32	+139	−2.8							
Flory	Cellobiose (*B*)	−44	+168	− 4.4	−44	+168	−4.6	−44	+168	−4.6	−44	+168	−4.6	−44	+168	−4.6	−44	+168	−4.6	
	Cellobiose (*J*)	−42	+162	− 4.1	−42	+162	−4.3	−42	+162	−4.3	−42	+162	−4.4	−42	+162	−4.5	−42	+162	−4.5	
Kitay-gorodsky	Cellobiose (*B*)	−44	+168	+ 0.2	−44	+168	−0.3	−44	+168	−0.4	−44	+168	−0.3	−44	+168	−0.4	−44	+168	−0.4	
	Cellobiose (*J*)	−42	+162	+ 0.5	−42	+162	−0.1	−42	+162	−0.2	−42	+162	−0.2	−42	+162	−0.3	−42	+162	−0.4	
Liquori	Cellobiose (*B*)	−44	+168	+ 2.0	−44	+168	+0.1				−44	+168	+0.2							
	Cellobiose (*J*)	−42	+162	+ 2.6	−42	+162	+0.5				−42	+162	+0.5							
Flory	Cellulose (*H*)	−32	+148	− 3.6	−33	+147	−4.1	−33	+147	−4.0	−39	+141	−3.0	−39	+141	−3.3	−40	+140	−3.5	
	Cellulose (*M*)[a]	0	+180	+14.5	0	+180	+6.9	0	+180	+6.6	0	+180	+8.9	0	+180	+8.2	0	+180	+7.9	
Kitay-gorodsky	Cellulose (*H*)	−32	+148	+ 0.9	−33	+147	+0.7	−33	+147	+0.1	−39	+141	+0.8	−39	+141	+0.5	−40	+140	+0.4	
	Cellulose (*M*)[a]	0	+180	+ 5.2	0	+180	+2.3	0	+180	+2.2	0	+180	+2.6	0	+180	+2.4	0	+180	+2.3	
Liquori	Cellulose (*H*)	−32	+148	+ 3.3	−33	+147	+1.0				−39	+141	+0.3							
	Cellulose (*M*)[b]	0	+180	+11.5	0	+180	+6.5				0	+180	+6.5							

[a]The values of h calculated for these conformations were, from left to right across the Table, 5.25, 5.25, 5.26, 5.26, 5.27, and 5.27 Å, respectively.
[b]For the values of h calculated for these conformations, see the corresponding entry for the Flory or Kitaygorodsky functions.

TABLE III

XYLAN CONFORMATIONS DERIVED FROM *Set 4* COORDINATES TO FIT $n = 3$, $h = 4.93$ Å

Helix sense	ϕ,°	ψ,°	V (*Kcal, from Kitaygorodsky functions*)
Left handed	−80	+172	+ 1.6
Left handed	− 6	+246	+ 4.6
Right handed	− 5	+104	+ 6.7
Right handed	+65	+174	+15.2

Kitaygorodsky functions and *Set 4* coordinates suggest that the van der Waals energy differences between (−40°, 140°) and (−80°, 172°) are also similar (0.70 Kcal for cellulose and 0.66 Kcal for xylan). Crystal packing must be the decisive influence, even though it seems probable that the conformation having a two fold screw axis is inherently more stable. In xylan diacetate, intermolecular hydrogen bonding would be replaced by weaker van der Waals and dipole interactions, and this polymer seems to crystallise with a two fold screw axis[32], in agreement with our conclusions.

In solution, configurational (= conformational) entropy effects will cause both xylan and cellulose conformations to be more disordered than in the crystal[33]. Even apart from the effect of the lower molecular weight of xylan and the weakening of its crystal lattice by occasional branching, it might be expected to be more soluble than cellulose for two reasons. Firstly, the xylan hydrate crystal is probably less favoured by potential energy, relative to an unconstrained state such as solution, because there are fewer intermolecular hydrogen bonds per sugar residue, and the chain conformation seems inherently less stable; and, secondly, hard-sphere calculations show a larger allowed area for xylan, which should allow a bigger entropy gain when it dissolves. It is well known that xylans dissolve in alkali, whereas cellulose does not normally do so, in agreement with these predictions. Nevertheless, the extent of the insolubility of cellulose is, at first, surprising. The *swelling* reactions of cellulose in alkali can be understood in terms of the important concepts introduced by Warwicker and Wright[34]. The cellulose chains, each in a ribbon-like conformation with equatorial bonds close to the plane of the ribbon, are considered to be laid on top of each other in stacks (called "sheets" by Warwicker and Wright). In the intact crystal structures of cellulose I and cellulose II, these stacks are joined by hydrogen bonding between equatorial hydroxyl groups. Evidence from X-ray diffraction indicates that swelling in sodium hydroxide causes rupture of hydrogen bonds and separation or relative movement of the stacks. The reason why cellulose of high molecular weight does not dissolve at room temperature is therefore that the stacks themselves remain intact. In the dry state, cohesion within each stack would be chiefly owing to van der Waals attraction[34]. However, forces of this type are so weak that, on this basis alone, it is difficult to imagine why chains should not separate completely and enter solution, because the conditions cause ionisation of the hydroxyl groups and therefore

and other[16,29] structures are that this does not deviate far from 117°. One solution (19°, 199°) is quite close to the Meyer and Misch conformation[3], which is known[23] to be sterically unfavourable. The energy calculated with Kitaygorodsky functions and *set 4* coordinates is 49.5 Kcal per disaccharide residue, compared with 0.7 Kcal for the other solution which resembles the Hermans[2] conformation! A more lenient assessment of the Meyer and Misch conformation was made by using the original definition, according to which the glycosidic oxygens lie on the two fold screw axis [*i.e.*, it is (0°, 180°)], and ignoring the discrepancy of about 0.1 Å in h (see Table II). Even so, a series of calculations with fifteen different combinations of coordinates and energy functions showed this conformation to be consistently less stable (by several Kcal) than the alternative (Table II). Another energetic disadvantage is the absence of an (O-5) ... (O-3′) hydrogen bond which occurs in the Hermans conformation, and for which there is evidence from infrared dichroism[30]. Our calculations therefore support the view[23] that the Hermans conformation is more likely than the Meyer and Misch version for cellulose. The Hermans conformation also seems very likely in itself: it represents a displacement of the experimentally proved cellobiose conformation to slightly higher van der Waals energy, without sacrifice of the hydrogen bond. The increase in intramolecular energy is estimated, without correction for the changed (O-5) ... (O-3′) contribution, as about 0.8 Kcal (see Table II). This could easily be offset by more efficient packing such as might be possible with an integral screw axis along the chain. Chain folding of the type proposed by St. John Manley[4] would be so expensive energetically (Figs. 4–6) that it is difficult to imagine any source of sufficient intermolecular compensation: this structure would therefore seem unlikely.

Some comments can also be made on Setterineri and Marchessault's[16] xylan hydrate structure. Fig. 8 shows the map with Kitaygorodsky functions, and xylan coordinates derived from cellobiose by replacement of each hydroxymethyl group by hydrogen, with appropriate shortening of the bond length. The Flory map was similar in its main features. The conformation proposed[16] corresponds to (−80°, 172°), which is certainly more likely than any alternative with $n = 3$ and $h = 4.93$ Å, both in van der Waals energy (see Table III) and because no other solution has the (O-5) ... (O-3′) hydrogen bond. It has several unfavourable features, however. The van der Waals energy is 1.95 Kcal above the Kitaygorodsky minimum and 0.66 Kcal above a xylan conformation of the Hermans type (−40°, 140°), and the hydrogen bond is rather long (3.0 Å). These disadvantages might be partly offset by favourable interactions between the atomic dipoles of (O-5) and (O-4′)[31]; furthermore, the 3.0 Å hydrogen bond might be stronger than would first appear, because models show the (O-3′)–H bond to be collinear with the atomic dipole of O-5, and that shortening of the O ... O distance causes distortion of this favourable geometry. We presume that crystal packing favours a structure having an integral screw axis for both cellulose and xylan. There is no obvious reason, in terms of intramolecular interactions, why these polymers should crystallise with different screw symmetries. Hydrogen bonding and dipole effects should be similar for each and calculations with

REFERENCES

1 J. C. IRVINE AND E. L. HIRST, *J. Chem. Soc.*, (1923) 518; W. N. HAWORTH, E. L. HIRST, AND H. A. THOMAS, *ibid.*, (1931) 824.
2 P. H. HERMANS, *Physics and Chemistry of Cellulose Fibres*, Elsevier, New York, 1949; C. Y. LIANG AND R. H. MARCHESSAULT, *J. Polymer Sci.*, 37 (1959) 385.
3 K. H. MEYER AND L. MISCH, *Helv. Chim. Acta*, 20 (1937) 232.
4 R. ST. J. MANLEY, *Nature*, 204 (1964) 1155.
5 K. C. ELLIS AND J. O. WARWICKER, *Nature*, 181 (1958) 1614.
6 N. L. ALLINGER, M. A. MILLER, F. A. VANCATLEDGE, AND J. A. HIRSCH, *J. Am. Chem. Soc.*, 89 (1967) 4345.
7 A. M. LIQUORI, *J. Polymer Sci., Part C*, 12 (1966) 209.
8 P. DE SANTIS, E. GIGLIO, A. M. LIQUORI, AND A. RIPAMONTI, *Nature*, 206 (1965) 456.
9 T. OOI, R. A. SCOTT, G. VANDERKOOI, AND H. A. SCHERAGA, *J. Chem. Phys.*, 46 (1967) 4410.
10 P. R. SCHIMMEL AND P. J. FLORY, *Proc. Natl. Acad. Sci. U.S.*, 58 (1967) 52.
11 M. GOODMAN AND M. FRIED, *J. Am. Chem. Soc.*, 89 (1967) 1264; J. E. MARK AND M. GOODMAN, *ibid.*, p. 1267.
12 C. RAMAKRISHNAN, *Proc. Indian Acad. Sci., Sect. A*, 12 (1964) 327; C. RAMAKRISHNAN AND G. N. RAMACHANDRAN, *Biophys. J.*, 5 (1965) 909.
13 G. N. RAMACHANDRAN, C. RAMAKRISHNAN, AND V. SASISEKHARAN, in G. N. RAMACHANDRAN (Ed.), *Aspects of Protein Structure*, Academic Press, New York, 1963, p. 121.
14 V. S. R. RAO, P. R. SUNDARARAJAN, C. RAMAKRISHNAN, AND G. N. RAMACHANDRAN, in G. N. RAMACHANDRAN (Ed.), *Conformation of Biopolymers*, Academic Press, New York, 1967, p. 721.
15 D. W. JONES, *J. Polymer Sci.*, 32 (1958) 371.
16 W. J. SETTINERI AND R. H. MARCHESSAULT, *J. Polymer Sci., Part C*, 11 (1965) 253.
17 P. J. FLORY, *Proc. Roy. Soc.* (London), *Ser. A*, 234 (1956) 60; H. MORAWETZ, *Macromolecules in Solution*, Wiley, New York, 1965, p. 111.
18 G. A. JEFFREY AND R. D. ROSENSTEIN, *Advan. Carbohyd. Chem.*, 19 (1964) 7.
19 C. J. BROWN, *J. Chem. Soc.* (*A*), (1966) 927.
20 S. S. C. CHU AND G. A. JEFFREY, *Acta Cryst.*, in press.
21 D. A. BRANT AND P. J. FLORY, *J. Am. Chem. Soc.*, 87 (1965) 2791.
22 A. I. KITAYGORODSKY, *Tetrahedron*, 14 (1961) 230.
23 R. H. MARCHESSAULT AND A. SARKO, *Advan. Carbohyd. Chem.*, 22 (1967) 421.
24 C. C. F. BLAKE, L. N. JOHNSON, G. A. MAIR, A. C. T. NORTH, D. C. PHILLIPS, AND V. R. SARMA, *Proc. Roy. Soc.* (London), *Ser. B*, 167 (1967) 378.
25 B. CASU, M. REGGIANI, G. G. GALLO, AND A. VIGEVANI, *Tetrahedron*, 22 (1966) 3061.
26 C. M. VENKATACHALAM AND G. N. RAMACHANDRAN, in G. N. RAMACHANDRAN (Ed.), *Conformation of Biopolymers*, Academic Press, New York, 1967, p. 83.
27 C. A. BEEVERS AND W. COCHRAN, *Proc. Roy. Soc.* (London), *Ser. A*, 190 (1947) 257.
28 S. S. C. CHU AND G. A. JEFFREY, *Acta Cryst.*, 23 (1967) 1038.
29 A. HYBL, R. E. RUNDLE, AND D. E. WILLIAMS, *J. Am. Chem. Soc.*, 87 (1965) 2779.
30 J. MANN AND H. J. MARRINAN, *J. Polymer Sci.*, 27 (1958) 595.
31 R. H. MARCHESSAULT AND C. Y. LIANG, *J. Polymer Sci.*, 59 (1962) 357.
32 R. H. MARCHESSAULT, *Chimie et Biochimie de la Lignine, de la Cellulose et des Hémicelluloses*, Les Imprimeries Réunies de Chambéry, 1965, p. 287.
33 P. J. FLORY, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, New York, 1953.
34 J. O. WARWICKER AND A. WRIGHT, *J. Appl. Polymer Sci.*, 11 (1967) 659.
35 H. STAUDINGER AND I. JURISCH, *Kunstseide und Zellwolle*, 21 (1939) 6.
36 W. KAUZMANN, *Advan. Protein Chem.*, 14 (1959) 1.
37 F. FRANKS AND D. J. G. IVES, *Quart. Rev.* (London), 20 (1966) 1.
38 G. NÉMETHY, *Angew. Chem., Intern. Ed. Engl.*, 6 (1967) 195.
39 A. SARKO AND R. H. MARCHESSAULT, *J. Am. Chem. Soc.*, 89 (1967) 6454.
40 M. IHNAT AND D. A. I. GORING, *Canad. J. Chem.*, 45 (1967) 2353, 2363.

electrostatic repulsion, and our calculations suggest that van der Waals repulsion would be significantly relieved in solution. An explanation of the stability of the stacks can be deduced from the observation that cellulose fractions of low molecular weight are more soluble in alkali if the temperature is lowered[35]. This implies that the stacks are *less* stable at lower temperature, and therefore that $T\Delta S°$ favours stacks rather than separated chains. Such positive entropy changes for association reactions are rare, except[36,37] where "hydrophobic interactions" stabilise the product. This type of interaction arises from the unique properties of water as solvent and is quite distinct from van der Waals forces. During the past few years, it has become accepted as an important source of molecular cohesion in biological and other systems[38], but there are, as yet, no proved examples that sugar rings can associate in this way. We suggest that the (H-1)–(H-3)–(H-5) surfaces of D-glucose residues in cellulose, and perhaps the (H-2)–(H-4)–(O-5) surfaces, can interact with water in a similar fashion to, for example, the nonpolar side groups of proteins or the hydrocarbon chains of detergents. The result is that their association is entropically favoured and that cellulose crystallites gain stability in the presence of water*.

ACKNOWLEDGMENTS

It is a pleasure to be able to thank Sir Edmund Hirst, in a special way, by contributing this paper to mark his seventieth birthday. His constant encouragement, interest, and material help have sustained our research for the past eight years. We thank the Pioneering Research Committee of the Institute of Paper Chemistry, Appleton, Wisconsin, U.S.A., for a research grant, and Dr. N. S. Anderson and Mr. J. W. Campbell for their collaboration in preliminary investigations. Drs. R. H. Marchessault and G. A. Jeffrey kindly gave us copies of important papers in manuscript.

**Note added in proof (14 May 1968)*—In this Department, C. A. Beevers and H. N. Hansen have recently solved the crystal structure of α-lactose monohydrate and very kindly made the preliminary coordinates available as a further test of our calculations. This disaccharide is extremely interesting in relation to cellobiose: intramolecular influences on conformation are expected to be identical because of the structural similarity around the bridge oxygen but packing influences should be quite different because configuration is inverted at C-1′ and C-4 and the crystal is a monohydrate. Hydrogen atoms have not yet been placed by Beevers and Hansen but we estimate from the ring atom positions that the conformation is (−31°, 155°). In Fig. 7, this corresponds to a rather small displacement of the cellobiose conformation ("*J*") with little change of intramolecular energy, in the direction of the two fold axis and the second minimum and approximately following the 2.8 Å hydrogen bond contour (actual 0-5 ... 0-3′ distance, 2.84 Å). The (C-1)-O-(C-4′) angle is 118.5°. The lactose structure therefore provides further support for our assumptions and conclusions.

A very sophisticated stereochemical study of amylose triacetate has been published[39], and evidence now exists that, as expected from our calculations, cellodextrins have rather rigid extended conformations in aqueous solutions[40].

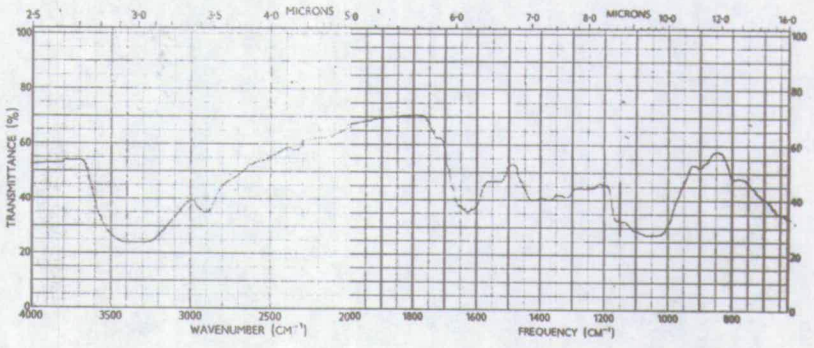


Figure 1.1. IR spectrum of mucilage. See Expt. 1.2

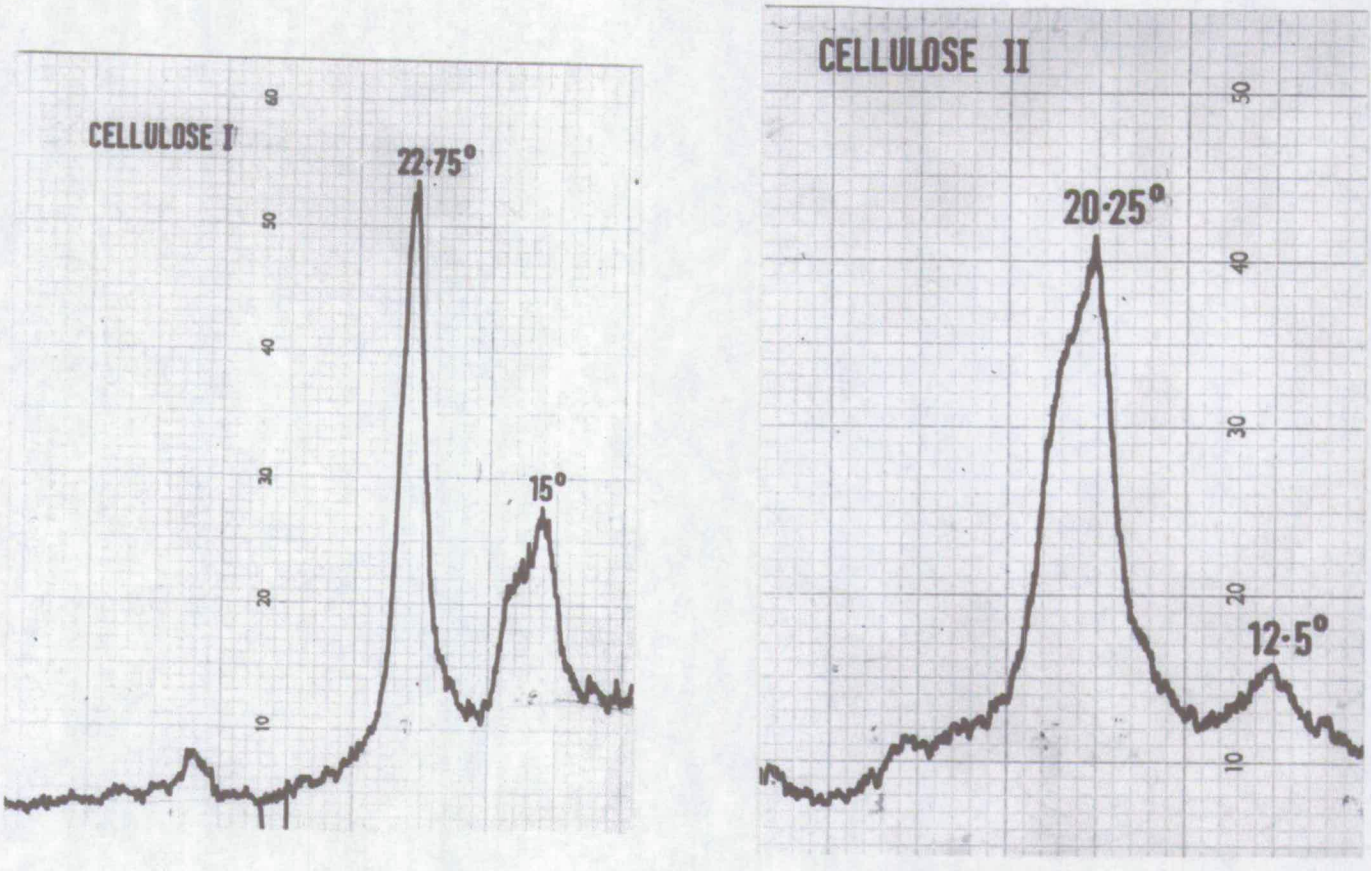
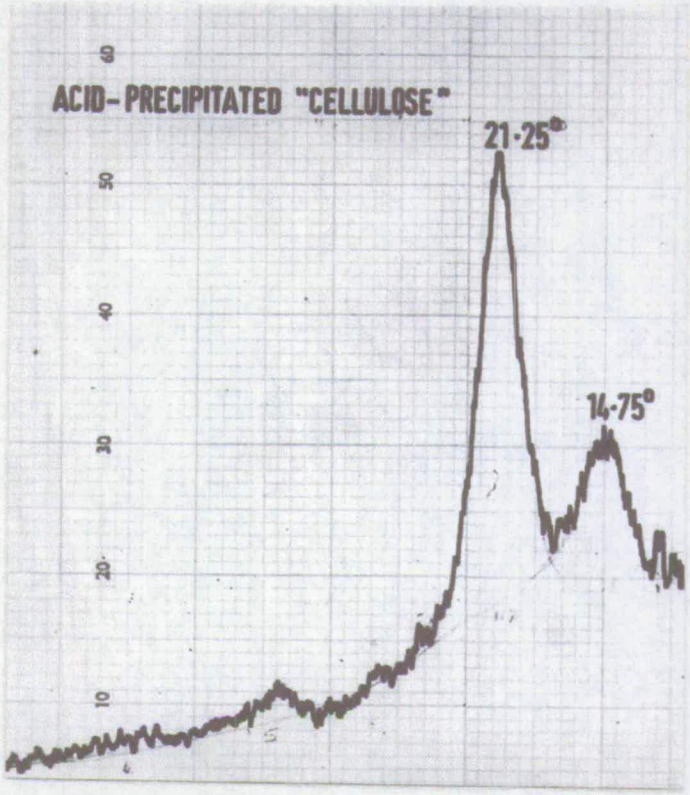


Figure 1.2. X-ray powder diffraction patterns. See Expt. 1.11.



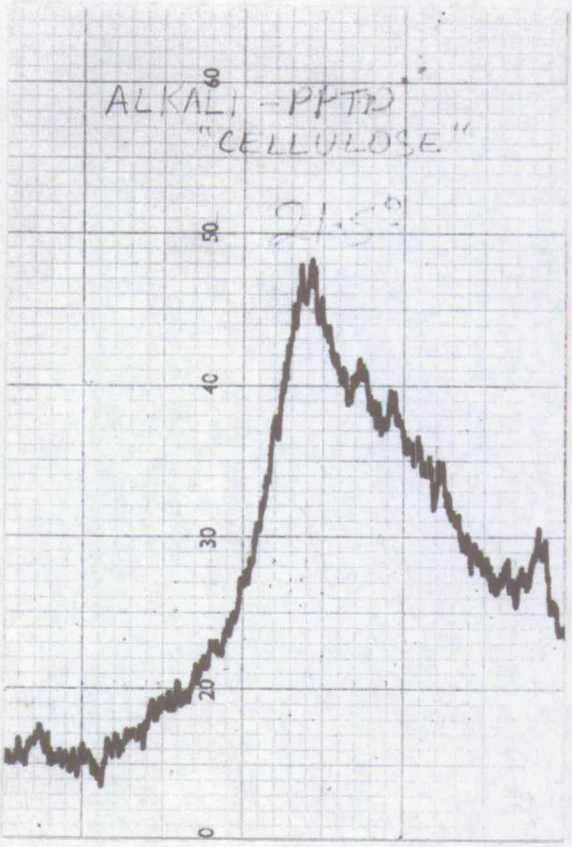


Figure 1.2. X-ray powder diffraction patterns. See Expt. 1.11.

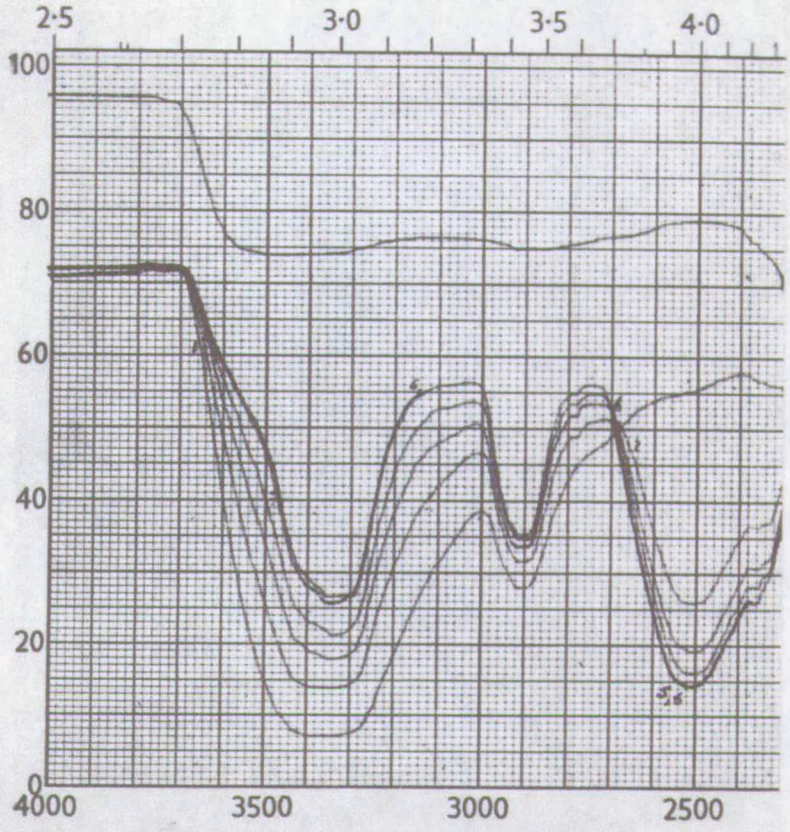
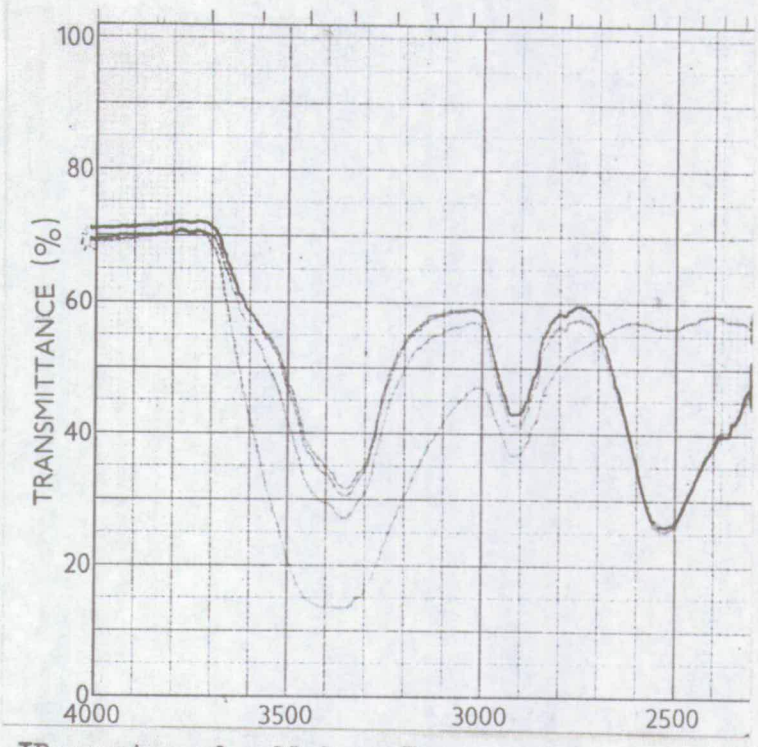
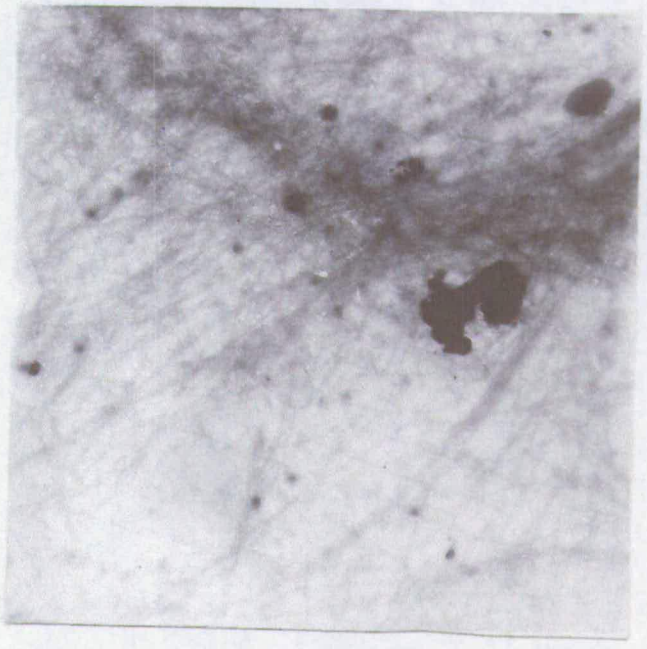


Figure 1.3. IR spectra of mucilage during deuteration. See Expt. 1.12.

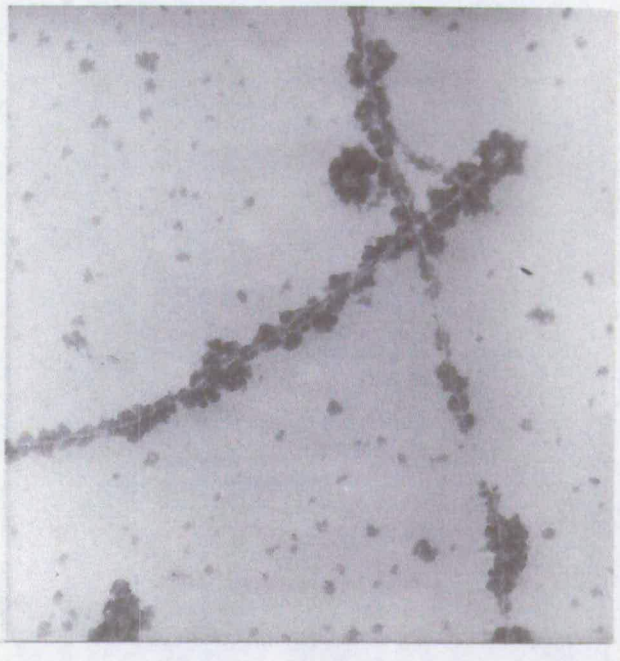


IR spectra of cellulose I

Figure 1.4. Electron micrographs. See Expt. 1.13.



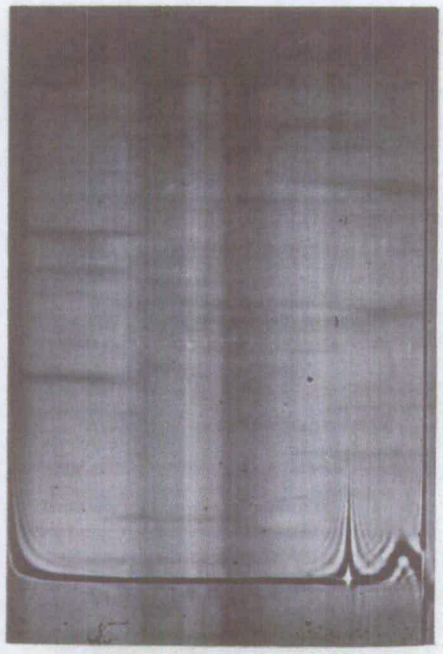
mustard mucilage (x20,000)



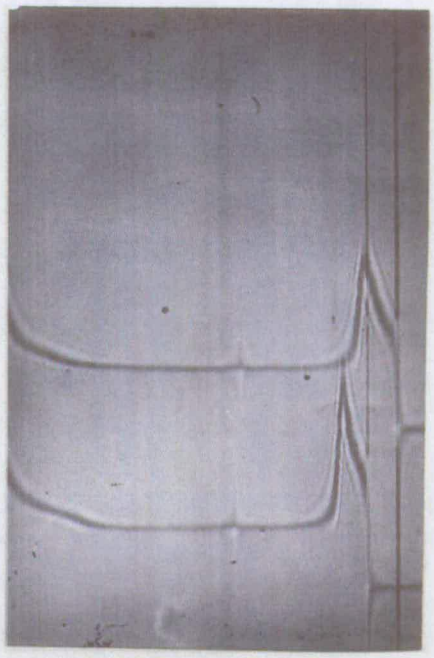
cress mucilage (x 60,000)

gum tragacanth (x 60,000)

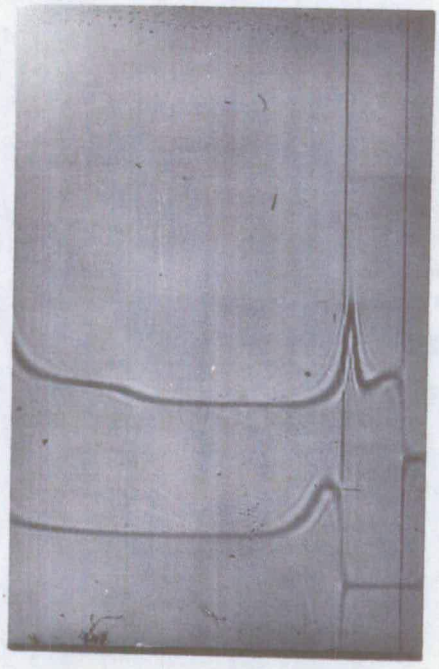
Figure 1.5. Ultracentrifuge sedimentation patterns. See Expt. 1.15,1.19



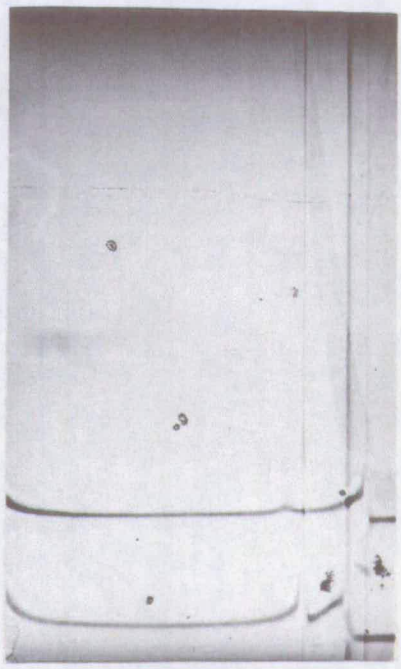
mucilage frn. A,
after 50 minutes at 20,410 rpm.



frns. A (upper) and C (lower);
15 min. at 50,740 rpm.

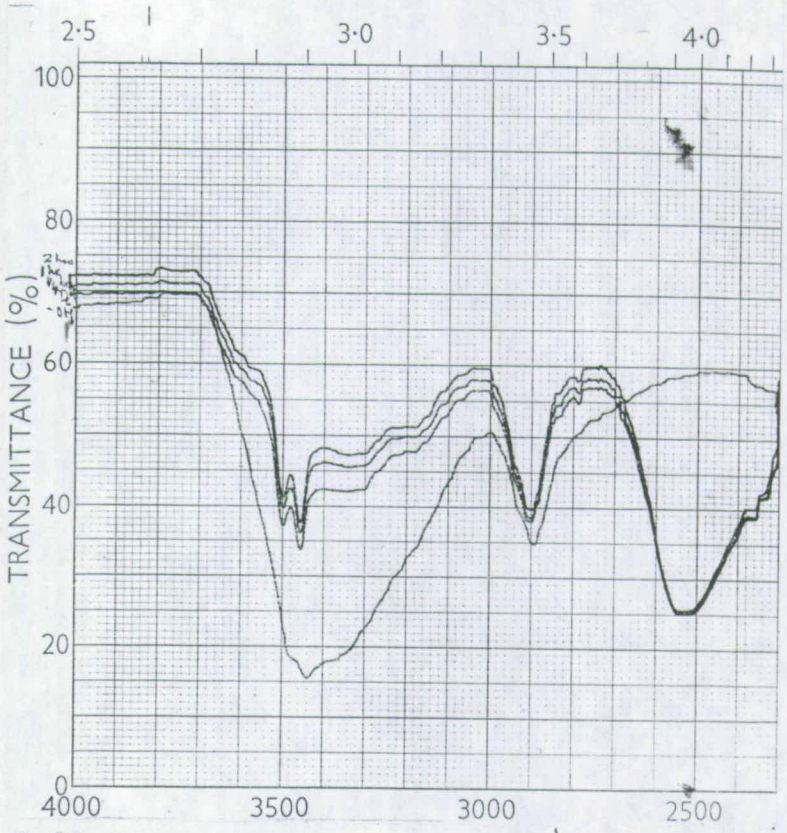


frn. E (upper), and
cellulose-free (lower);
25 min. at 50,740 rpm.

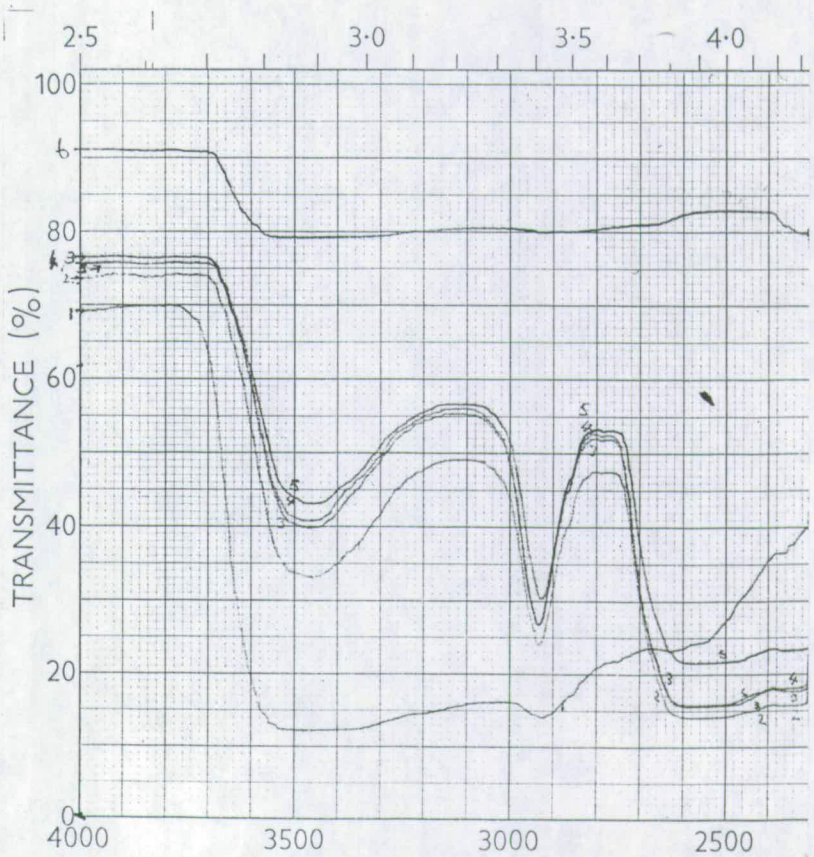


cress (upper) and mustard (lower);
60 min. at 20,410 rpm.

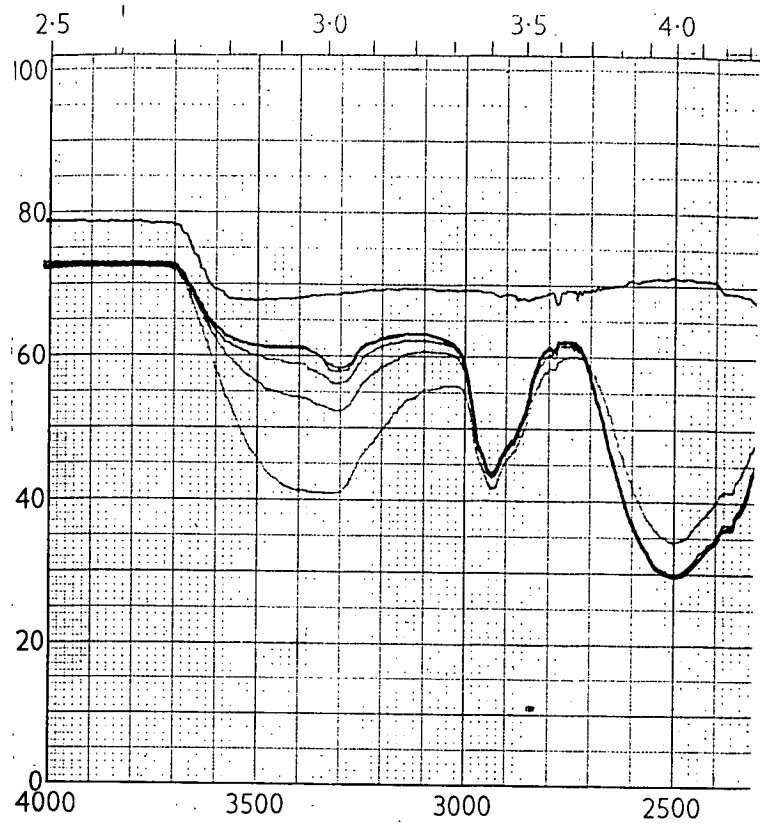
Figure 2.1. IR spectra. See Expt. 2.1.



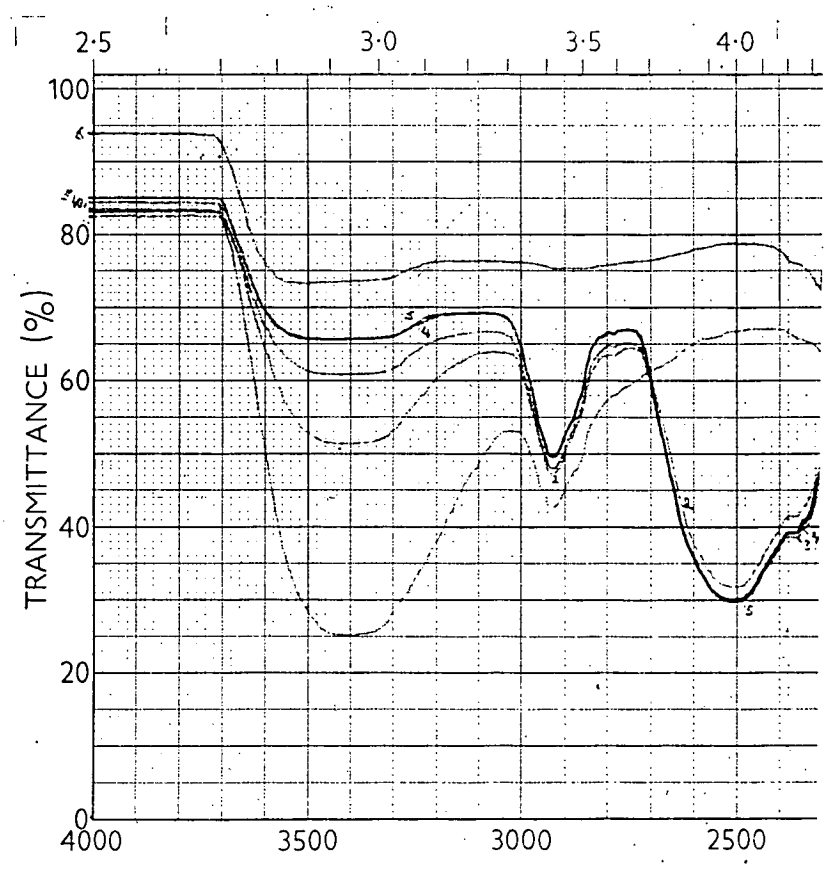
Cellulose II



Alginic acid from L.hyperborea



Germinated-mustard pectin



Gum tragacanth

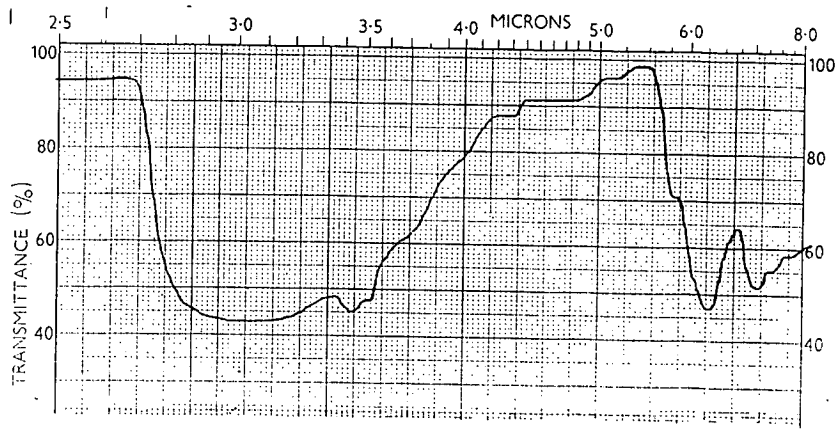
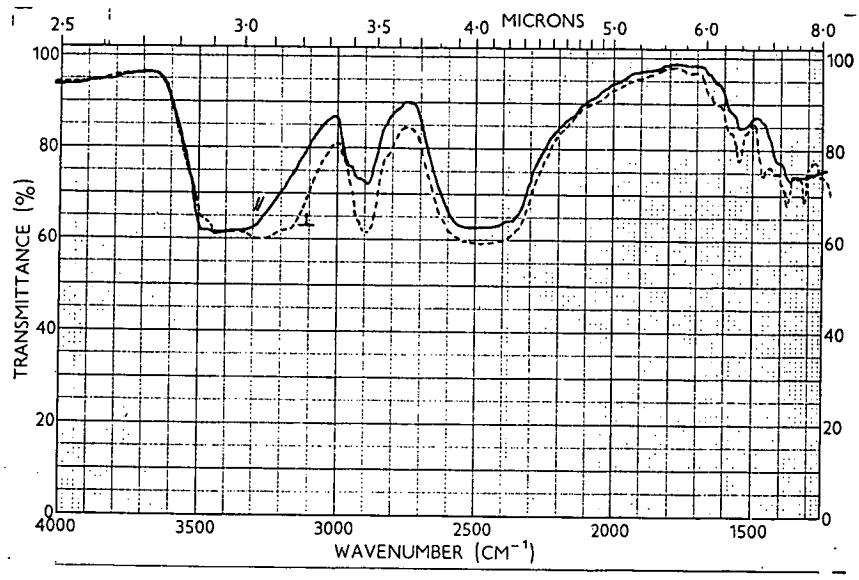
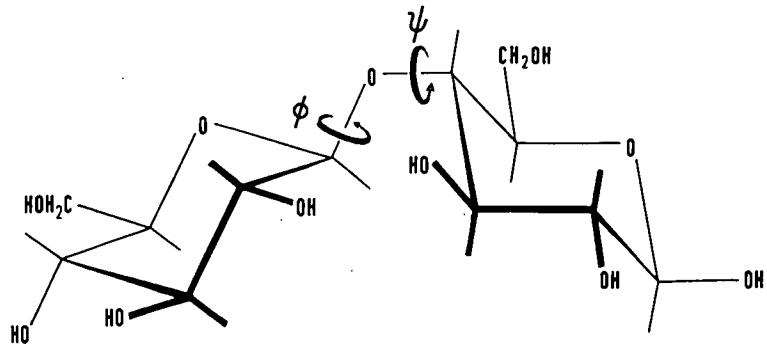


Figure 2.2. IR spectrum of calcium β-D-glucopyranosiduronate. See Expt. 2.3.



Polarised IR spectra of oriented Fortisan. See Expt. 2.6.

Figure 3.1.

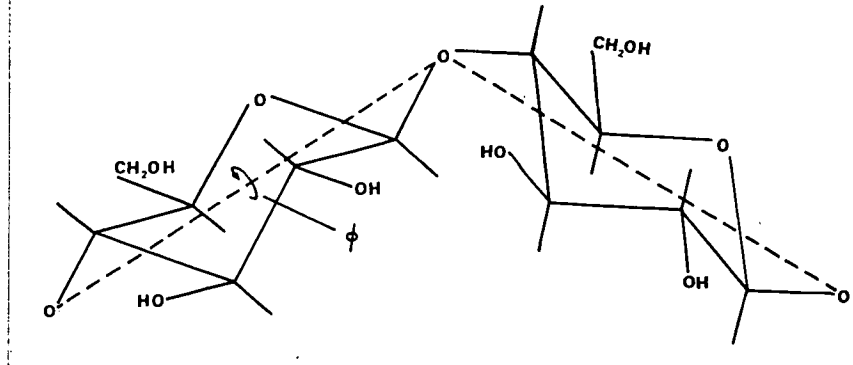


SCHEME FOR VARIATION OF BRIDGE-ANGLE GEOMETRY IN CELLOBIOSE

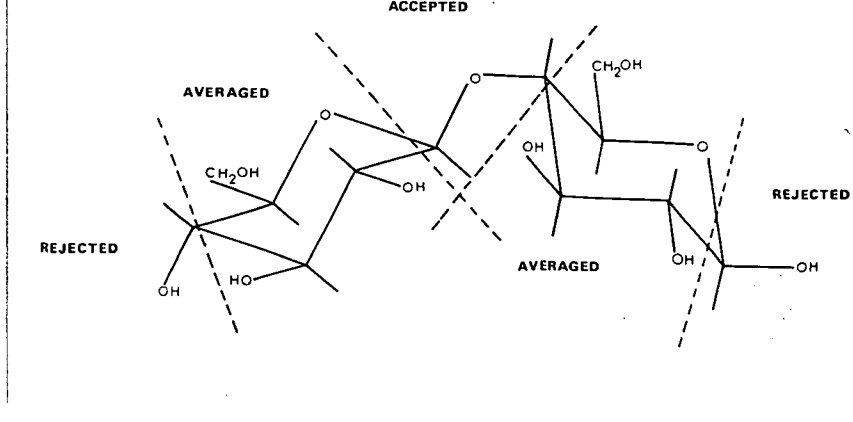
$\phi = \psi = 0°$

Rotations are positive when each residue is rotated in the direction shown about its bond to the bridge oxygen.

See Introduction.



Jones and Settineri method of rotation. See Introduction.



Scheme for derivation of cellulose co-ordinates. See Calc. 3.3.

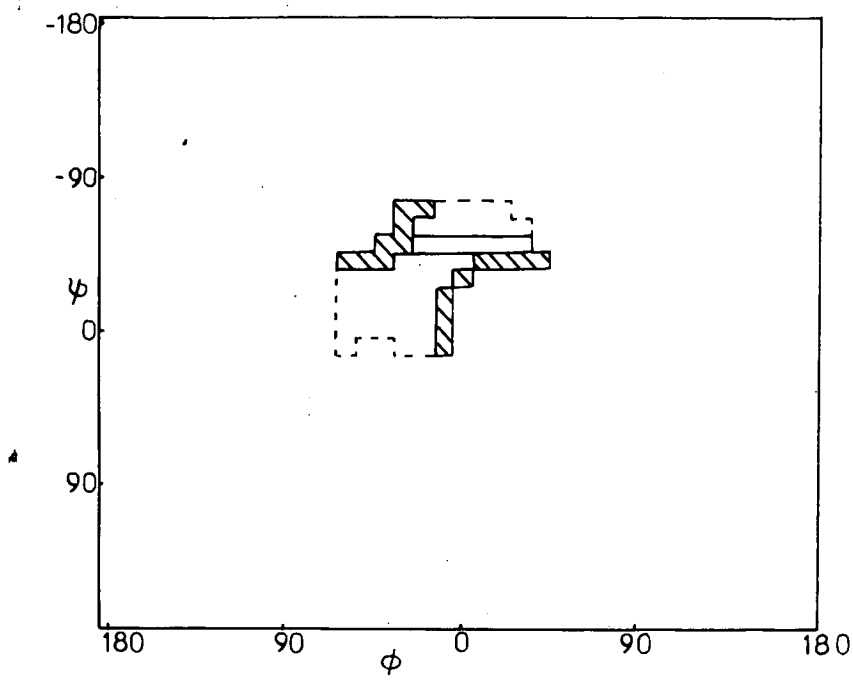
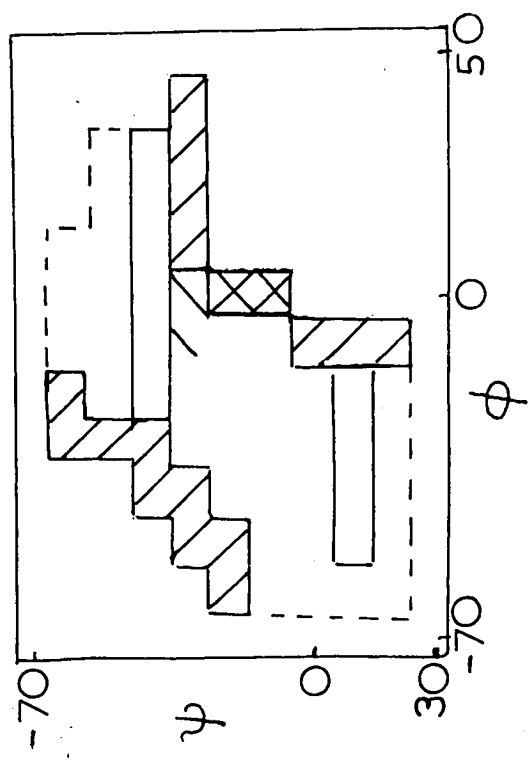
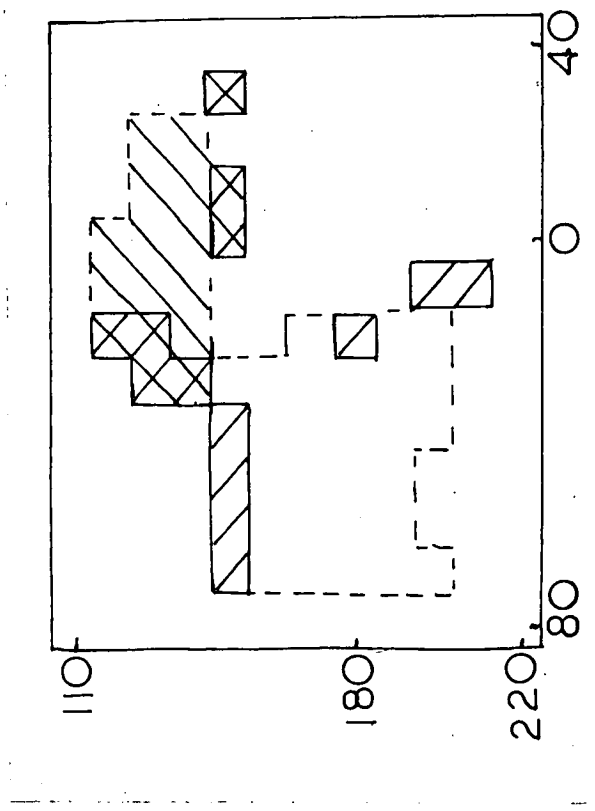


Figure 3.12. Hard-sphere map for laminaribiose. See Calc. 3.25



Hard-sphere map for laminaridextrins



Hard-sphere map for sophorodextrins

See Calc. 3.26

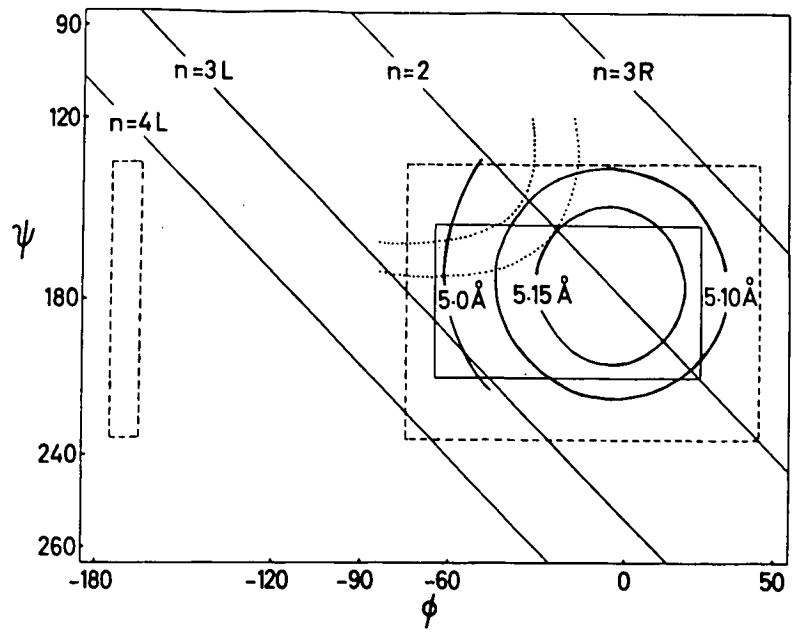


Figure 3.2. Conformational map, corresponding to Ramachandran's study. See Calc. 3.5.

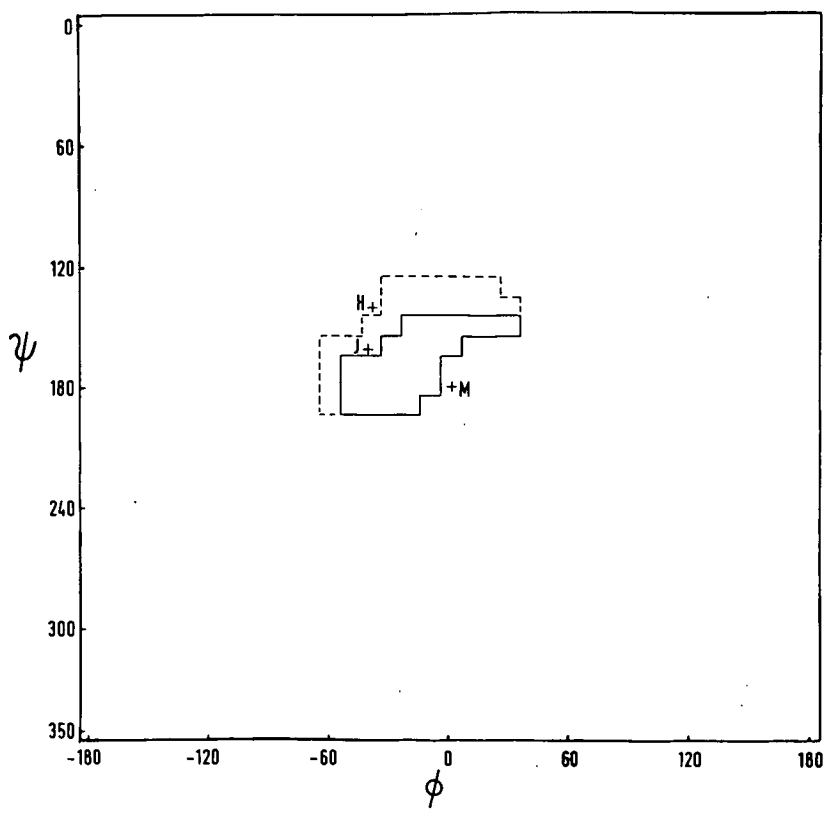
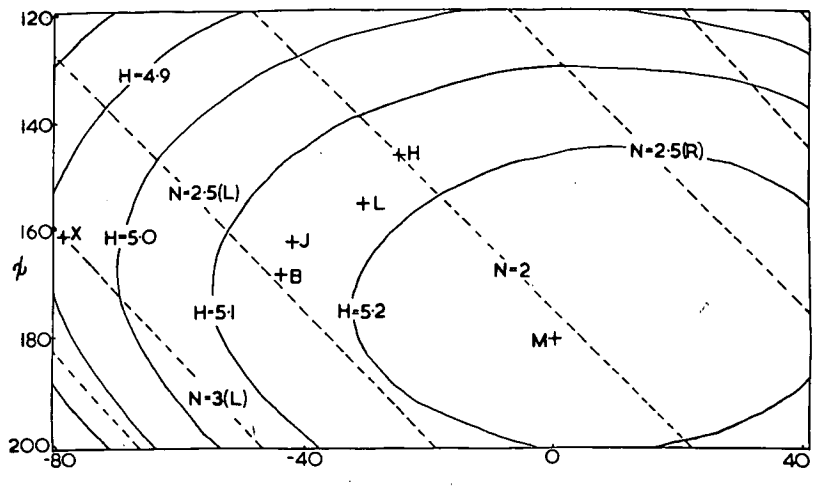


Figure 3.3. Hard-sphere conformational map for cellulose. See Calc. 3.8.



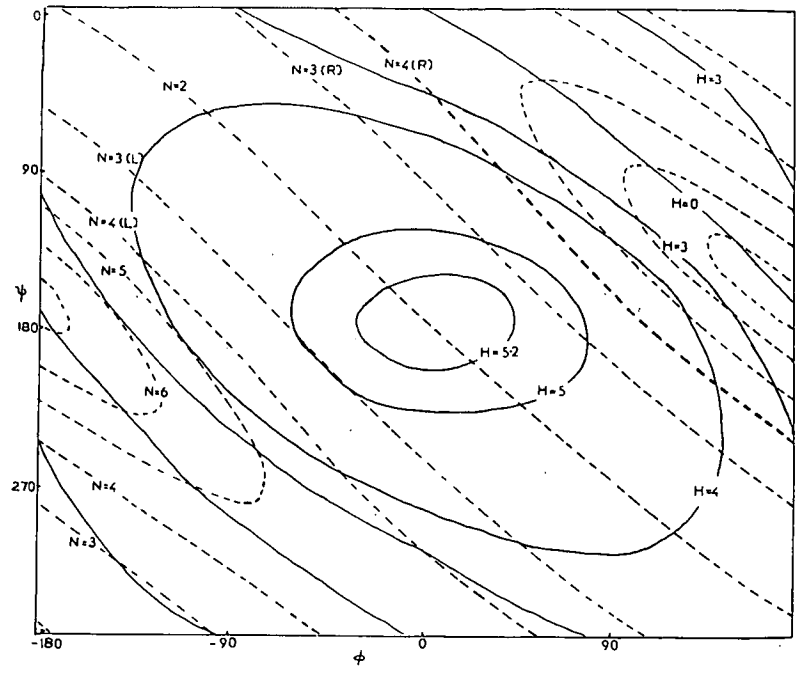
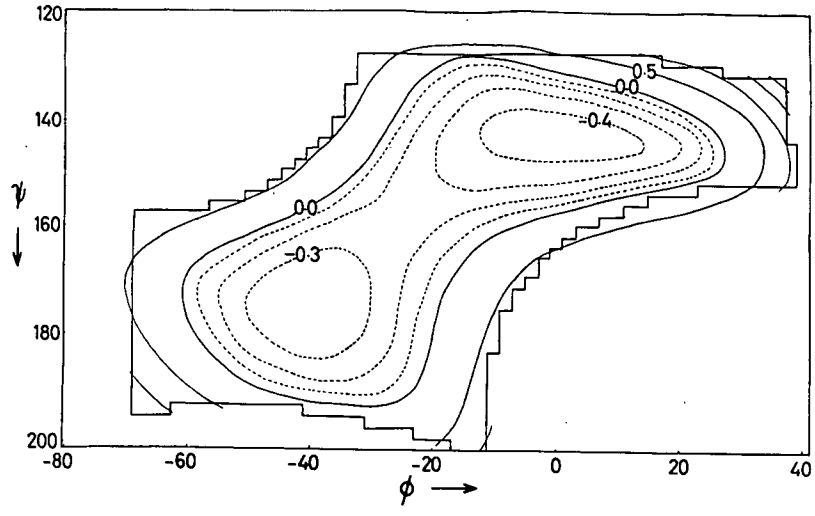
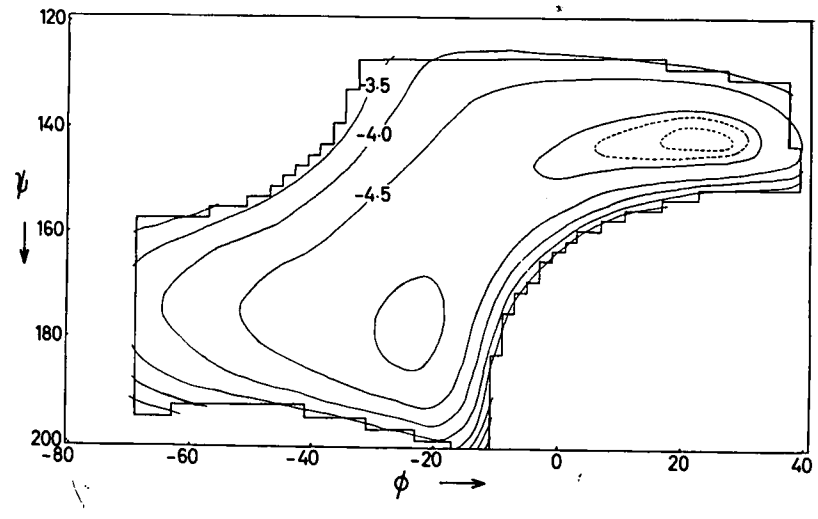


Figure 3.3. Residue-length and screw symmetry maps. See Calc. 3.9.

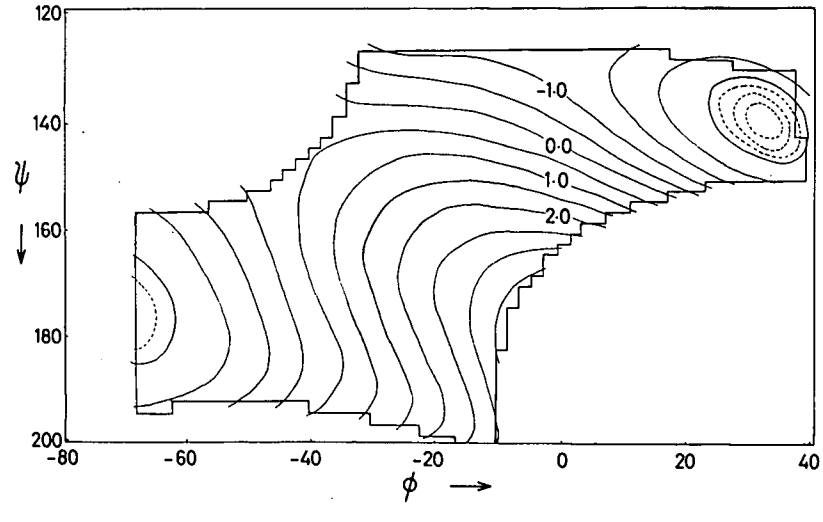
Figure 3.5 Van der Waals conformational maps. See Calc. 3.10.



Kitaygorodsky



Flory



Liquori

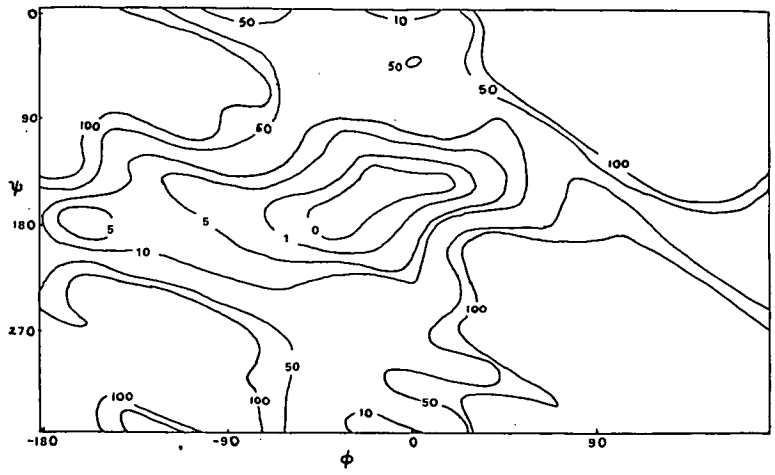
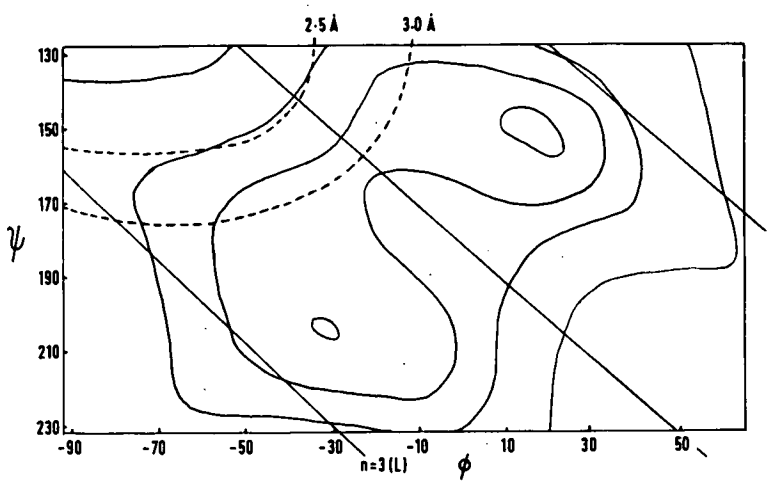
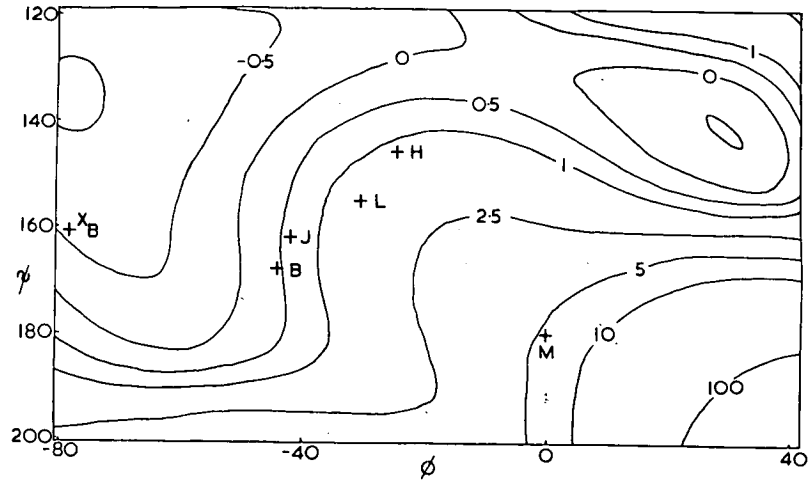


Figure 3.6 Van der Waals map over the whole area. See Calc. 3.11

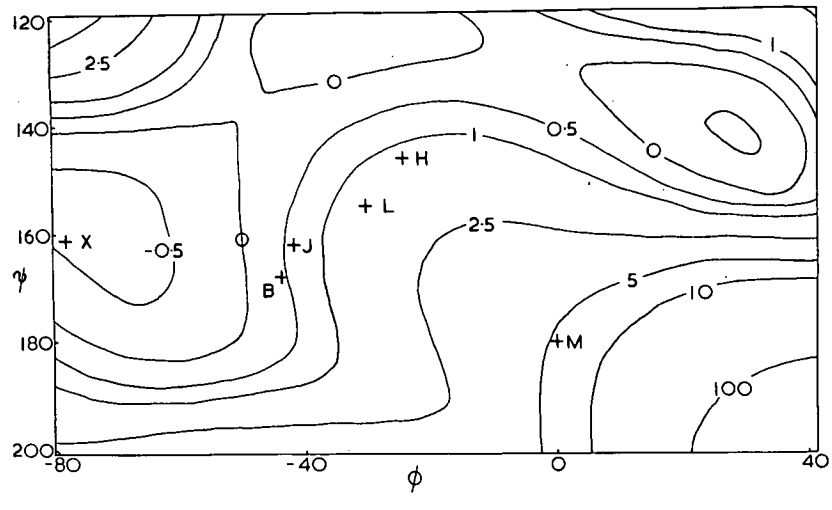


Xylan conformational map. See Calc. 3.12.

Figure 3.7. Van der Waals maps with Allinger functions. See Calc. 3.15.

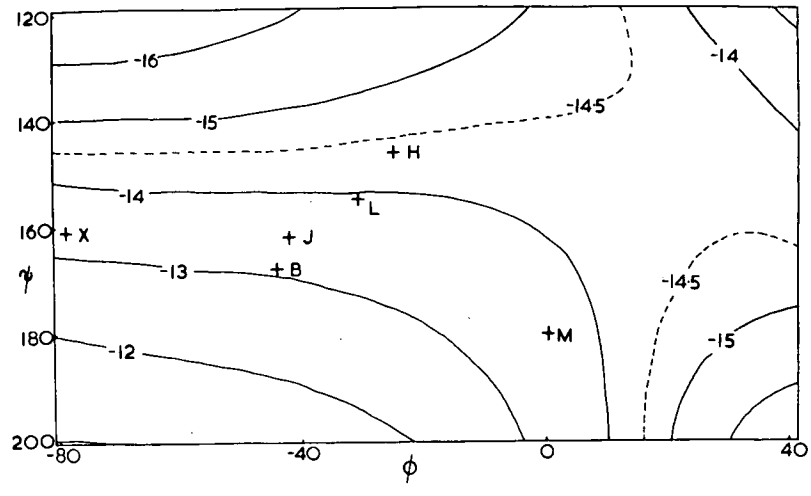


Excluding $O'_3 \ldots O_5$ interaction

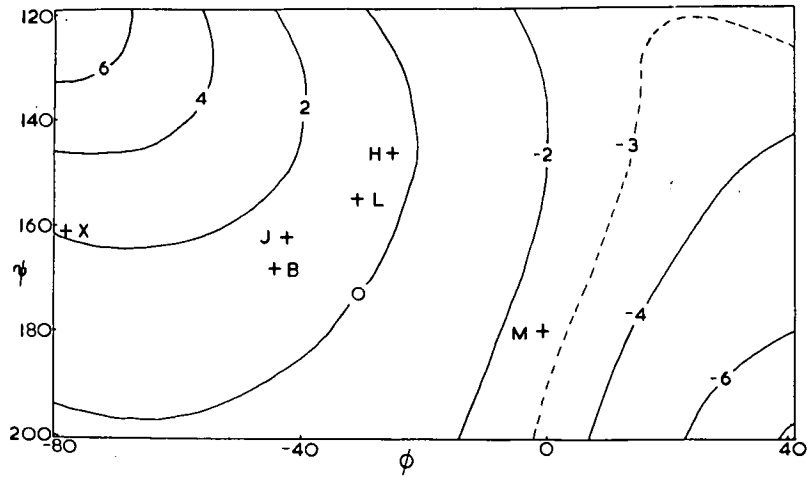


Including $O'_3 \ldots O_5$ interaction

Figure 3.8. Dipole maps. See Calc. 3.18

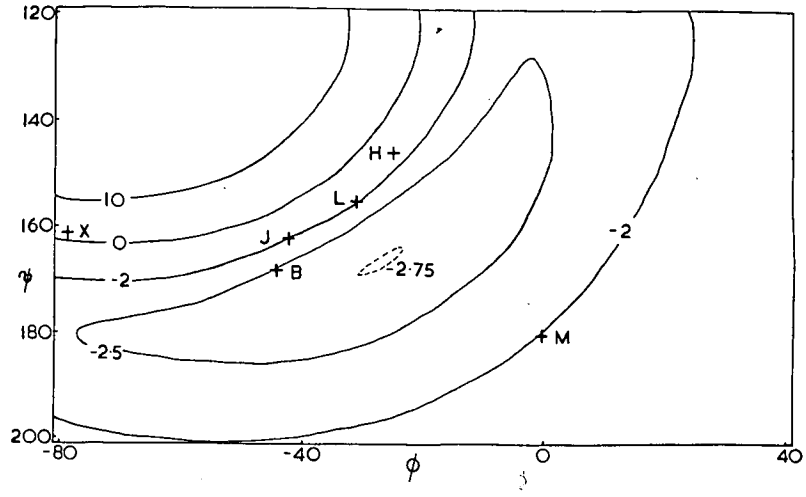


Excluding $O'_3 \ldots O_5$ interaction

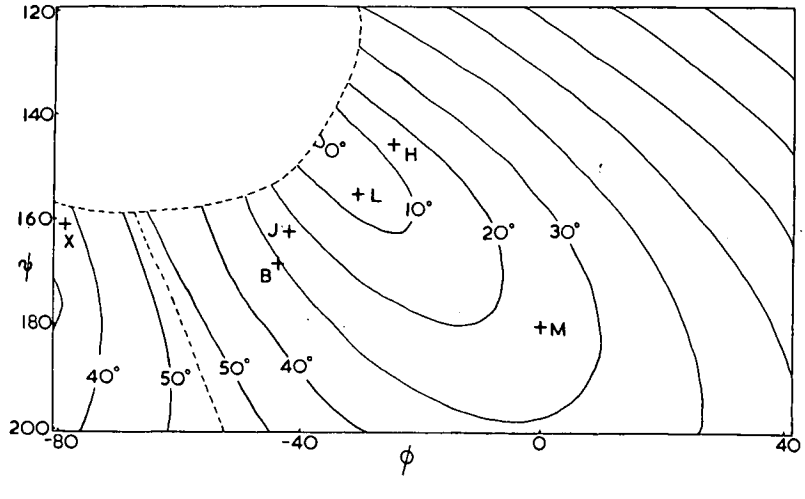


Including $O'_3 \ldots O_5$ interaction

Figure 3.9. Hydrogen bonding



Potential map. See Calc. 3.21.



Hydrogen-bond angle. See Calc. 3.22

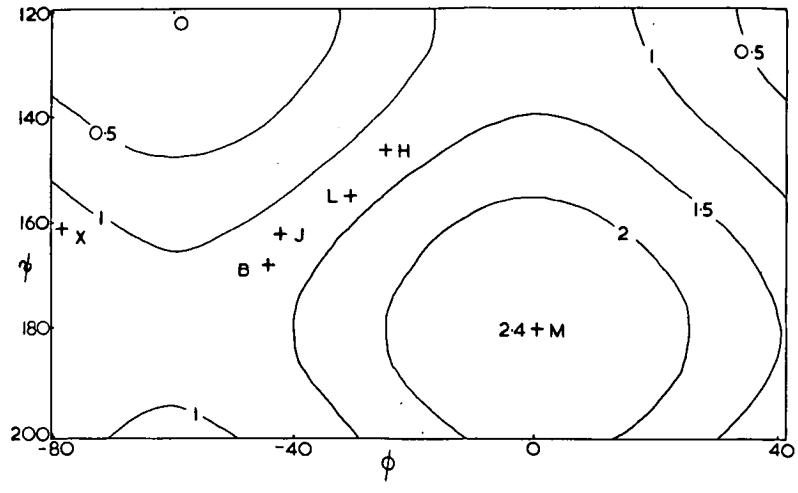


Figure 3.10. Torsional potential map. See Calc. 3.23.

Excluding hydrogen-bond

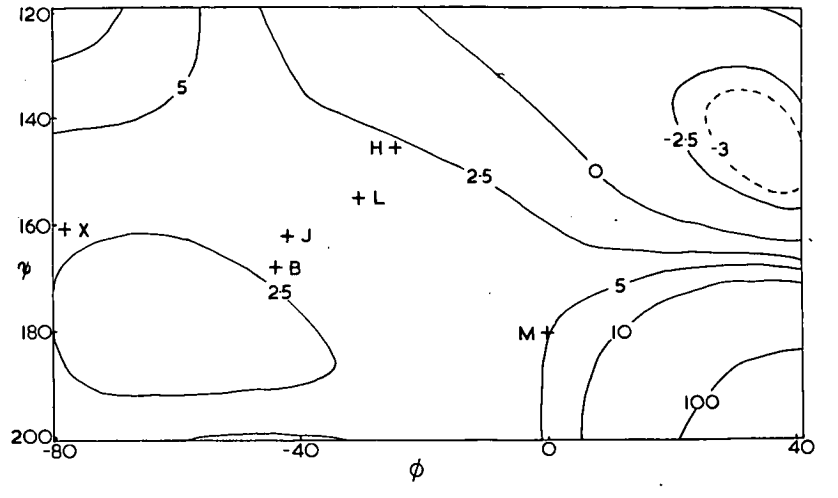


Figure 3.11. Total potential maps. See Calc. 3.24.

Including hydrogen-bond