Modelling and Structural Studies of a Gelling Polysaccharide: Agarose.

UNIVERSITY OF SURREY

A thesis presented to the University of Surrey in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Physical Sciences.

Nicholas M.W.Haggett MA, PGCE, (Cantab) GRSC

December, 1998

Abstract

This thesis details work carried out over a period of three years on the two gelling carbohydrates *agarose* and *carrageenan*. The major part of the work deals with agarose.

Two approaches have been used which yield information from different angles; these are the experimental (laboratory) and the simulation (computational) approaches.

There is a large field of interest in gelling carbohydrates from the point of view of the food industry. Their extraordinary ability to form stable gels and emulsions incorporating other food ingredients makes them important in many deserts and dairy products.

In the present work, models for agarose and carrageenan carbohydrates were developed using structural x-ray data and related carbohydrate literature. The models were treated with two different solvent simulation methods. It was found that the inclusion of individual solvent molecules (the closest approximation to a real solution) was extremely uneconomical when the demands on computing time were taken into account, and in fact the long term outcome of the simulation was the same for both methods. Inclusion of solvent simply reduces diffusion rates and the time constant for chain flexing.

Gel permeation chromatography and differential scanning calorimetry were used to prepare samples of agarose molecules of known size, and to probe temperature dependent phase transitions. This work was done at the UNILEVER laboratories at Colworth House, Sharnbrook in Bedfordshire. It was found that only molecules longer than fifteen residues displayed the molecular ordering transition typical of agarose polymer, and a value for the enthalpy of the transition of -1.5kcal per mole of residues was measured. It was predicted that in agarose itself, helical regions of a size of approximately 40 residues should exist.

Simulations were then done on several agarose molecules of different sizes in order to parallel the experimental work. The differences in energy between molecules in various conformations were compared. These results were also related to helix–coil transition theory. The modelling predicts an enthalpy per mole of residues for the agarose coil to helix transition of approximately -2kcal, and indicates that single agarose coils may be of some importance in agarose gel structure. The work illustrates the difficulty in modelling such complex systems, and in fact it remains impossible to observe agarose molecules undergoing the transition between a random coil and a helical conformation.

Acknowledgements

There are many people who deserve thanks for their help during my research and the production of my thesis.

Firstly I must thank my supervisors who have assisted me in my work and also in spirit over the past three years: Dr B.Howlin and Professor G.Webb at the University of Surrey, and Dr. R.Hoffmann and Professor A.Clark at UNILEVER Research, Sharnbrook, Bedfordshire. Thanks are also deserved by many other people at UNILEVER who assisted me during my work there, in particular Ms A.Russell who helped with the chromatographic analysis.

I wish to thank my parents and family for their support and encouragement, especially during the writing up of my work.

I am also indebted to numerous friends and colleagues in the Chemistry Department. Particularly, thanks go to Dr M.Parker for his help in managing the Silicon Graphics workstations, to Mr M.Read for many useful discussions concerning the finer points of technical writing using LaTeX and to Mr P.Branton.

I note here the efforts of the people who created and developed the LaTeX typesetting system (these people are not paid for their trouble!). The system was used to produce this thesis; it provides elegant solutions to many difficult formatting problems.

Lastly, I am most grateful to UNILEVER for financial support which has allowed this work to be done, and to be presented at the 1998 American Chemical Society Conference in Dallas, Texas.

Contents

Chapter 1

Introducing carbohydrate polymers

Contents

Overview

The introduction will be divided into two parts. Firstly, the term *polysaccharide* will be introduced. The word embodies all those molecular species which can be chemically classed as *saccharides* (a variety of carbohydrate) and *polymers*, and so a review of carbohydrates and polymers will be given.
The second part of the introduction will introduce the research presented in this thesis, and will deal specifically with the polymeric gels formed from agar and the carrageenans. At the end of this chapter the need to study these substances will be explained, and the structure of the remainder of the thesis set out.

1.1 Polymers

Polymers are a class of molecule[1] with a huge relative molecular mass achieved by joining many hundreds of repeating units known as *monomers* (or *mers* in the materials science world), together [1].

Polymers have probably existed ever since life began on the Earth, since at this time temperatures on the planet surface became low enough to permit large covalently bonded molecules to form. Inorganic polymers such as silicates may have existed since even earlier times; they can be found today in minerals [2, 3]. Nevertheless by far and away the largest variety of polymers is to be found in living systems, including those polymers derived from fossil fuels such as oil and coal.

Today the synthesis of polymers provides the basis for a vast industry, and a great deal of effort goes into the search for new polymers with properties tailored to meet the demands of modern technology. One part of this industry is that of food science where, amongst other chemicals, polysaccharides are used to improve the gastronomic properties of certain foods. The research presented in this thesis falls within the scope of this area.

The existence of polymers depends mainly on the stability and versatility of the carbon–carbon bond. In comparing the stability of the following bonds [4] we can readily see that for carbon, under normal conditions, a single bond between atoms of itself is just as stable as the bond between itself and oxygen. The same can hardly be said for silicon, or in fact for most other elements.

[1]The word "molecule" is taken to indicate a nanoscopic structure composed of atoms joined together by covalent bonds in a topologically unique arrangement.

Bond	Energy kJ per mol	Bond	Energy kJ per mol
C–C	347	Si–Si	226
C–O	336	Si–O	466

It could be argued that without this thermochemical fact, life as we know it would not be able to exist on Earth.

1.1.1 Polymer nomenclature

Mers

In general, any molecule formed from an infinitely repeating series of identical units (mers) is called a polymer. The mers must have at least two functional groups. The polymer chain grows as mers meet and react at their functional groups, and various catalysts may be employed to assist this process.

Some mers have no preference for which functional group reacts when polymerisation occurs, such as but-2-ene (see figure 1.1). Other mers have a specific *head* and *tail* location at which preferential reaction occurs so that all mers are joined up facing the same way.

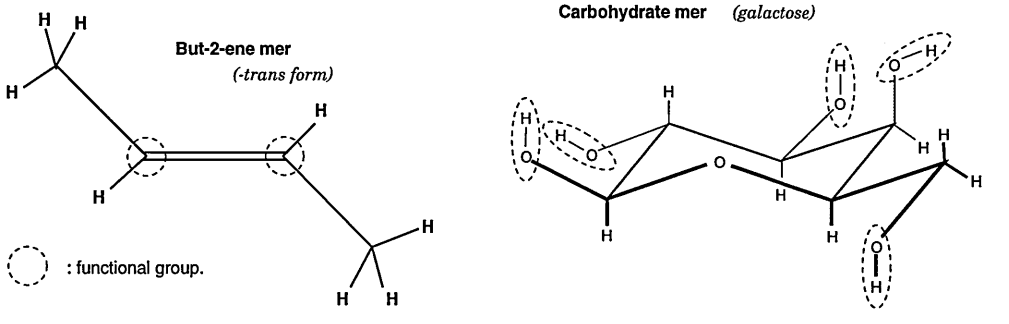


Figure 1.1: Two examples of a mer. The simple mer *trans*-but-2-ene can only give rise to one polymer. The more complex carbohydrate mer could produce many different polymers.

For example, cellulose is made from β-D-glucose joined specifically at carbon 1 (the head) and carbon 4 (the tail). More information about the naming of carbohydrates will be given in section 1.2.

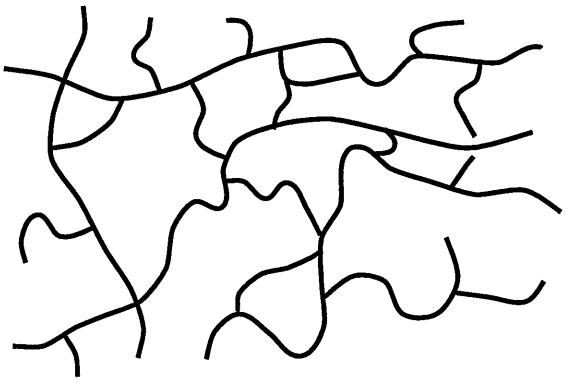


Figure 1.2: A crosslinked network polymer. Networks such as these tend to be hard and resistant to heat; they decompose before melting.

Polymer versatility

The simplest polymers are made from di-functional mers which polymerise with themselves, the classic example being poly(ethene). However there are many polymers which are much more complicated than this [1, 5]. A few examples are given below:

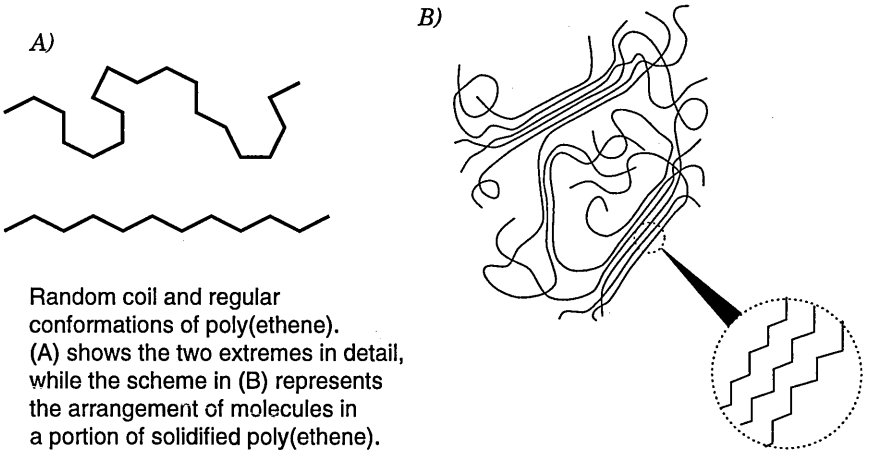
- **A-B** co-polymers. Here, two different types of mer are joined alternately to form a chain, as in nylon.
- Block co-polymers, such as the *Acrylonitrile-butadiene-styrene* (ABS) type. In this example there are three mers A, B and C. A typical portion of chain would be: **-B-B-B-A-A-A-A-C-C-B-B-B-B-**.
- Co-polymers can be altered by grafting side chains of another type of mer onto mers (having more than two functional groups) in the original chain. Many types of protein and polysaccharide have this form.
- Networks. Polymers with a high proportion of multi-functional mers can become networks. In effect, a material having a network structure is one huge molecule, because one can traverse the entire sample without having to jump between non-bonded atoms (see figure 1.2). The common industrial resin Araldite™ is an example, made from *bis*-phenol A-epichlorohydrin and tri-ethylenetetramine mers.

1.1.2 Polymer conformation

Polymers that do not undergo irreversible network formation are able to adopt a limitless variety of conformations, or chain shapes. This interesting feature of polymers will now be briefly reviewed, for it has a great bearing on the resulting properties of polymeric solutions, melts and solids.

Simple polymers

The simplest polymer of all is poly(ethene), a hydrocarbon. At each carbon atom the direction of the chain axis can be changed, so that in the molten state the overall chain shape can be a loop, multi-turn coil, wavy line or a plethora of other shapes. One particular conformation favoured due to the ease of stacking molecules in the solid, is the straight line, made up from a series of little zig-zags (see figure 1.3). Indeed, some poly(ethene) specimens can be as much as 90% crystalline, with very ordered regions joined by random loops of polymer chain.



Random coil and regular conformations of poly(ethene). (A) shows the two extremes in detail, while the scheme in (B) represents the arrangement of molecules in a portion of solidified poly(ethene).

Figure 1.3: An example of polymer chain shape variation in the solid.

These regions can be observed under crossed polars when the polymer sample is illuminated by plane polarised light (see figure 1.4). The ordered regions exhibit *birefringence*, and therefore appear as multi-coloured patches in the sample[2].

[2]The long axes of the chains retards one component of the plane polarised light more than the other, causing interference.

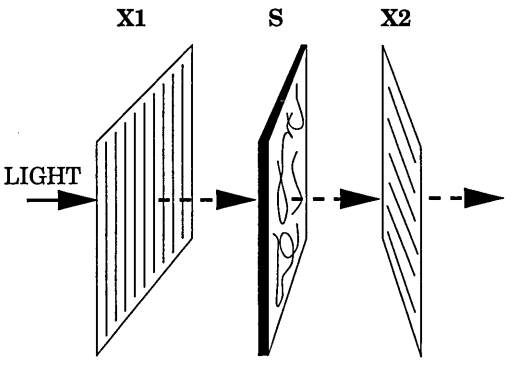


Figure 1.4: Observing birefringence. Ordinary light is passed through a polariser 'X1' and then the sample of interest 'S'. After passing finally though the analyser 'X2' (another polariser 90° to the first) the light can be observed by eye or by using an optical device.

Glasses and elastomers

Most polymers do not contain significant amounts of crystalline regions: they are known as amorphous polymers. The extended nature of amorphous polymer chains causes such a high degree of entanglement that crystallisation is not able to take place when the melt is cooled. Therefore, as the thermal energy becomes less and molecular motions are reduced, a point is reached where the random liquid conformations become "frozen" into a glass. This polymer "freezing point" is analogous to the freezing point of metals and other crystalline materials, and is given the special term *glass transition temperature*, or T_g. A comparison of amorphous and crystalline materials is given in figure 1.5.

Amorphous polymers may exhibit elastomeric behaviour when between the melt and glass transition temperature. In this phase, a force applied to the polymer specimen results in a reversible deformation proportional to the force.

Helices

A large number of polymers, both synthetic and biological, exist in the form of helices, where the polymer chain forms a regular spiral shape. More than one polymer chain may be present in a helix, all coiled around a central axis. The helix is a way of minimising free energy in long molecules such as proteins. Helices are therefore an extremely significant structural feature, and consequently are very common in biological molecules [6, 7]. They can even be coiled again to form multi-stranded *super-helices*, or biological cables [6],

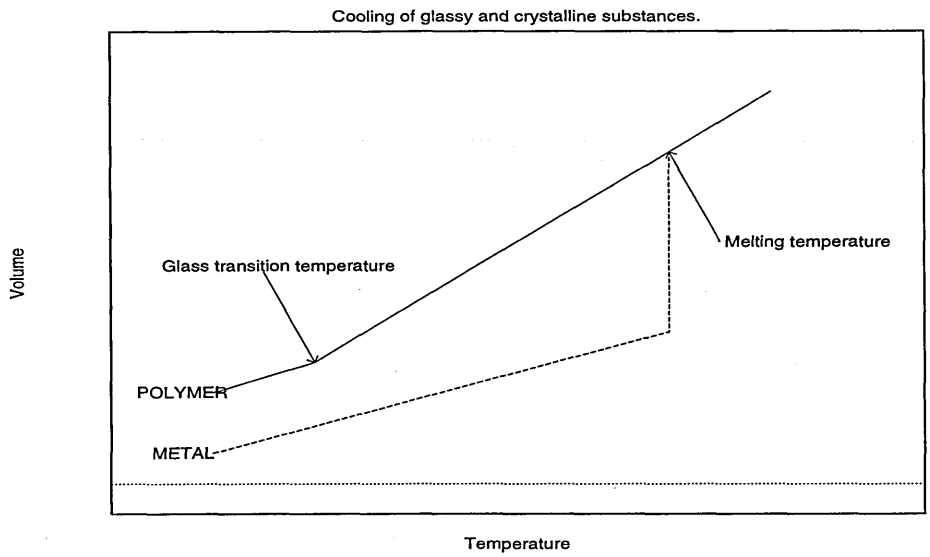


Figure 1.5: As the temperature of a liquid is reduced, its volume becomes smaller. For crystalline substances an abrupt change occurs at the melting point and the crystalline solid appears. For large polymer molecules however there can be significant super-cooling of the liquid phase, and in this region the substance can be elastic in nature. Then at the glass transition temperature the material finally becomes a hard solid: a glass.

rather like the tungsten filament of a light bulb.

In solutions of polymers at high temperatures helices are normally completely disrupted, and the conformations of long, flexible polymer chains can best be described as random coils and loops. The formation of helices from random coils as the temperature of such solutions is reduced is a most interesting conformational transformation, being associated with subtle but crucial enthalpic and entropic energy changes. It has been described as a form of crystallisation since the result is sometimes a precipitate, although helices can also exist as colloidal particles in solution.

As we shall see in the following chapters, the helix plays a central role in the physical properties of the marine polysaccharides *agar* and *carrageenan*. Before these substances can be fully described, some introduction to the nomenclature of carbohydrates is needed.

1.2 Carbohydrates

Carbohydrates are made from carbon dioxide and water by green plants, which absorb the two gases directly from the air. The reaction is catalysed by various enzymes, and the required energy is provided by sunlight. Higher plants, such as trees, are able to synthesise carbohydrates not only in the form of *mono-saccharides*, but also as *oligosaccharides* (a few monosaccharides joined together) and *polysaccharides*. Polysaccharides can be broken down into monosaccharides by hydrolysis.

The main advantages for the plant in devoting this effort into creating carbohydrates are to:

- build structure (in the form of e.g. cellulose)
- store food for future seedling plants to use.

The synthesis of carbohydrates in the laboratory is one of the biggest challenges facing organic chemists. Fortunately many different carbohydrates can be isolated from natural sources, but a few have been synthesised that are not known in nature.

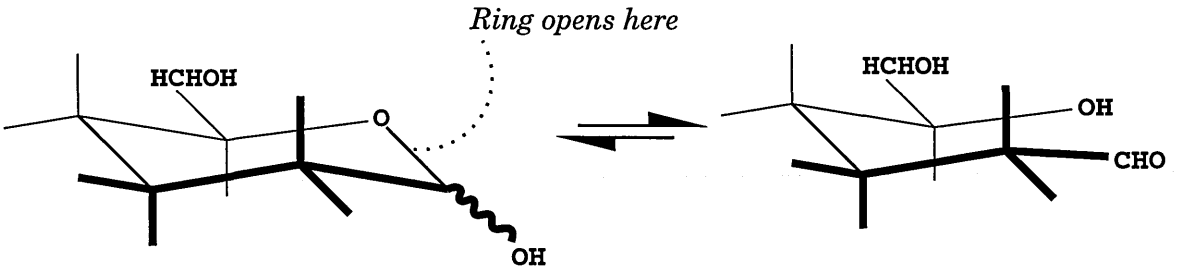
1.2.1 Nomenclature

Definition and how to draw them properly.

What characterises a carbohydrate? In fact, there is no simple definition for a carbohydrate as they can exist in various forms, especially when dissolved

in water. Generally though, carbohydrates are poly-hydroxy aldehydes or ketones [6]. Carbohydrates can exist as straight–chain or ring molecules having four, five or six carbon atoms, or they can be polymers consisting of many individual rings linked together. Rings and straight chains are interconvertible, the ring forms being properly described as *hemi-acetals*:

Ring form (hemi-acetal) *Chain form*



Carbohydrates contain *chiral centres* [6][3]. The problem therefore arises of how best to represent them on paper, and two methods are commonly employed. Firstly, one could try to devise a set of rules to be applied to a relatively simple schematic type drawing. Whenever this drawing is used the rules allow one to elucidate the absolute configuration at each chiral centre. An example is the *Fischer formula* [8] (see figure 1.6). This method is best suited to straight-chain carbohydrates. Secondly an attempt could be made to draw the structure as a direct two-dimensional representation of the three-dimensional reality. An example here is the *boat–chair* representation, best suited to ring structures.

Carbohydrate rings: atom numbers

The most stable length for the carbon chain in a carbohydrate is five or six carbon atoms. In this way a six-membered ring is formed using the oxygen from the –OH group on atom five as a hetero-atom in the ring, as shown above. The hemi-acetal carbon atom is known as the *anomeric carbon*. The anomeric carbon has special properties, described in section 1.2.2. The accepted numbering of carbon atoms and hydroxy-groups is given in figure 1.7.

[3]A chiral centre is an atom with four (in the case of carbon) different groups attached to it. Molecules containing chiral centres are optically active, meaning that they rotate the plane of plane polarised light.

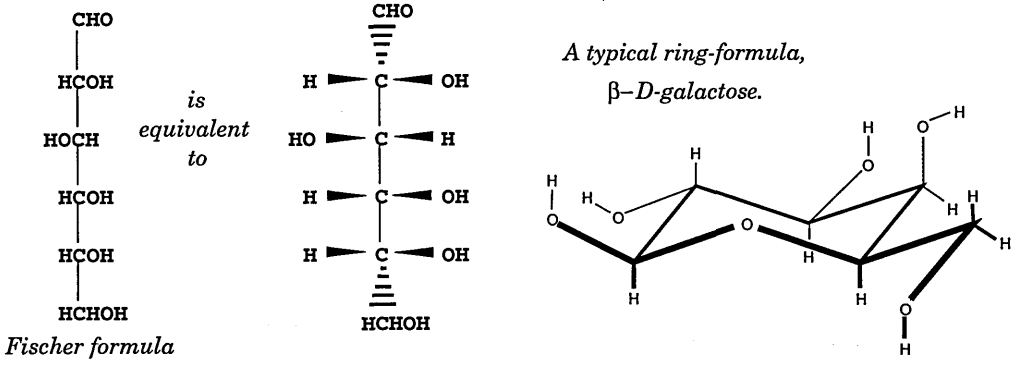


Figure 1.6: Representations of carbohydrate structure.

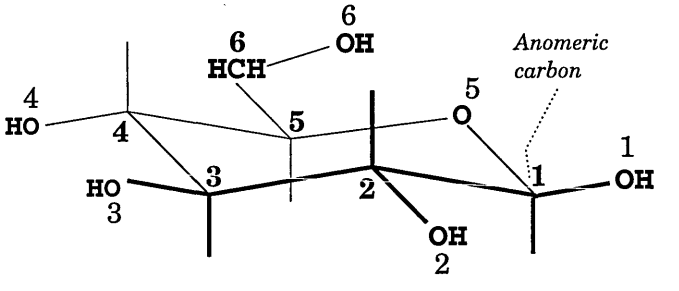
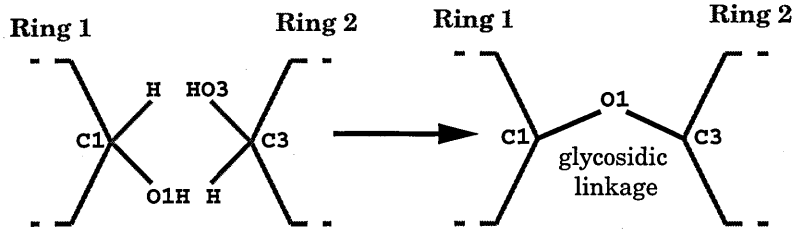


Figure 1.7: Carbohydrate rings have an accepted numbering system, as shown here. The **bold** numbers refer to carbon atoms, and the oxygens attached to them generally carry the same number.

1.2.2 Glycosidic Linkages

So far we have dealt with very small carbohydrates. In nature most carbohydrates contain thousands of atoms and are referred to as polysaccharides.

In order to make polysaccharides, monosaccharides are linked together, each monosaccharide being one sugar ring. The joining of two rings creates a *glycosidic linkage* between the anomeric carbon of one ring and any carbon *except* the anomeric carbon of the next ring. The glycosidic linkage therefore consists of three atoms describing two dihedral angles, as shown below:



Note that carbon 1 becomes a full *acetal*, not just a hemi acetal.

The anomeric carbon

The acetal carbon in ring forms of carbohydrates (labelled carbon) exhibits the so called *anomeric effect*. In order to appreciate this it is necessary to recall that in six membered carbon rings, any groups more bulky than a simple hydrogen atom prefer to be equatorially positioned as this reduces steric hindrance with neighbouring groups. For the specific case of an oxygen atom attached to the anomeric carbon in a carbohydrate ring however this preference is reduced or even reversed. This has implications for the properties of glycosidic linkages.

The reasons for the anomeric effect are electronic, being based on an overlap between the lone pair on the ring oxygen (O5) and the empty σ^* orbital on the anomeric carbon-oxygen bond [9, 10]. Overlap of filled and empty orbitals can sometimes result in a covalent bond, but here a bond already exists and the net effect is a favourable reduction in energy of the system. The overlap can be optimised by two factors: using orbitals close in energy, and using orbitals which are aligned and in the right positions. Here, the exo-cyclic C–O σ^* orbital is the closest in energy to the oxygen lone pair. The most favourable overlap is achieved when this orbital is on the same side of the ring as the oxygen lone pair, causing the C–O bond to be axially positioned. Therefore an unusual preference for this axial position is seen (figure 1.8).

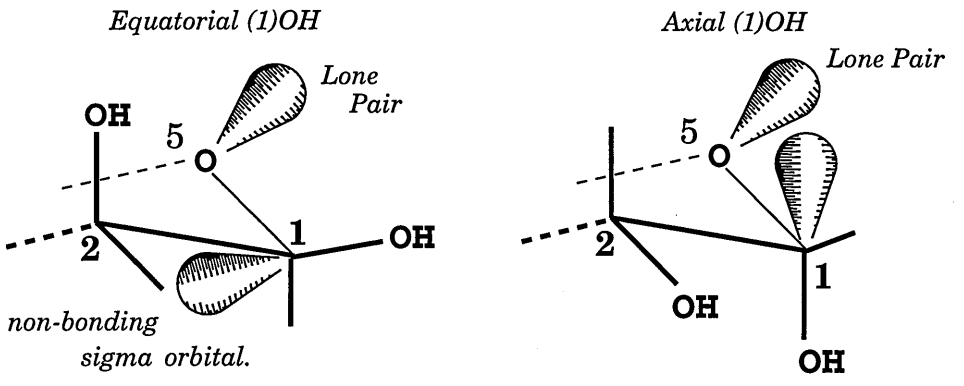


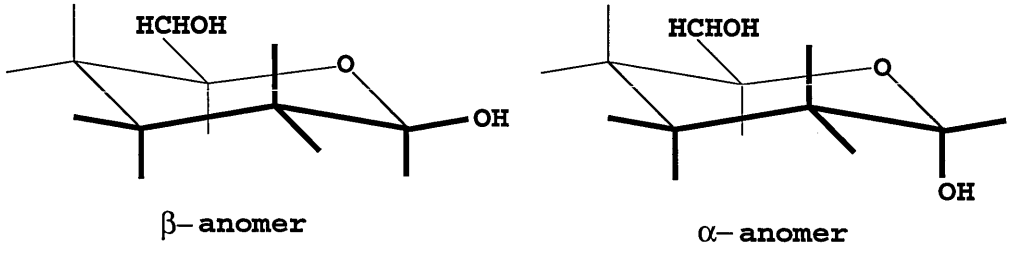
Figure 1.8: Scheme to demonstrate overlap of orbitals at the anomeric carbon (labelled '**1**'). When the (1)OH is in the axial position, there is a favourable stabilisation energy due to orbital overlap.

Most of the other groups on the carbohydrate ring are unaffected by this, and retain the usual equatorial preference. Therefore in spite of the anomeric effect the –O(1)H group may still be equatorial since achieving a ring flip to the all-axial conformation may be too energetically unfavourable.

These special characteristics of anomeric carbons are taken into account in programs which model carbohydrates.

Configuration at glycosidic linkages

The glycosidic linkage between two sugar rings can be designated either α- or β-, depending on the stereochemistry at the anomeric carbon (C1). The two stereo-conformers are known as *anomers* [6].



Note that anomers are inter-convertible by the process of hydrolysis at the anomeric carbon, followed by rotation of the C1 group and re-formation of the sugar ring. They are not inter-convertible *via* ring–flipping. The designation 'α' or 'β' is also dependent on whether the sugar is the D or L

form. For aldohexoses (which cyclise to six membered rings) the designations in the scheme above are for the D forms[4].

1.3 Polymeric gels

Biological polymers, with their large numbers of associated polar groups (-OH or -NH_2 for example), tend to be soluble in water. Examples such as egg white (albumin) and cornflour (starch grains) are well known. However at the molecular level these polymers are generally not fully solvated but tend to be folded or looped onto themselves and surrounded by solvent molecules. Sometimes the molecules associate into clumps in solution [7, 11].

In special cases, solutions of polymers can form *gels* when left to stand or cooled below the *gelation point*. This is not due to precipitation of polymer molecules or crystallisation of the solvent, but is a distinct phase by itself.

Definition of a gel

Gels can not be readily defined. They can have various compositions (polymer/solvent ratios) and the mechanical properties of gels vary enormously. The point at which a gel melts into a viscous fluid is sometimes not at all obvious. However, most sources in the literature are agreed on the basic properties required for a substance to be classed as a gel [12].

- Gels are commonly solids, meaning that in practise a gel holds it shape without the need for confinement by some container.
- Gels may be deformed by application of a force, but they can return to their original shape when the forces are removed; gels are therefore elastic.
- Furthermore, gels are soft: even the most rigid ones have fairly low Young's moduli.
- A large proportion of a gel consists of small molecules. The solid nature of the gel can be ascribed entirely to the small proportion (typically only a few percent) of a macromolecular ingredient.
- The macromolecular ingredient in a gel is assembled into a network, with fixed or non–mobile junctions and regions of "netting" which retain the overall turgidity by their tendency to hold solvent.

[4] D and L labels are attached to sugars depending on the conformation about carbon **5** *in the straight chain form.* To be strictly accurate, the R and S notation for chiral centres should be used, but the other method has stuck for historical reasons.

- Over long periods of time solvent can leak out from pores in the gel network leading to shrinkage and cracking (known as *syneresis*). Gels thus tend to have a limited 'shelf life'.

The internal structure of a gel is therefore akin to that shown in figure 1.2 in many ways. However due to the inclusion of large amounts of solvent, there are fewer cross-links and a larger number of open voids in the gel network.

1.3.1 Aqueous gels

The biggest variety of gelling systems is found in the aqueous environment.

The roles of gels in living systems

Gels are frequently found in living organisms [13], where they fulfill purposes such as cushioning from physical blows, protection from microbial attack, transport of ions and support in aqueous environments.

Formation

Typically gels are produced when the solution of a gelling polymer is subjected to physical change. The easiest approach is to reduce the temperature; this causes conformational changes in the polymer and can produce a gel [12]. More subtle are changes in the ability of the solvent to dissolve the polymer molecules. This ability is often termed *water quality* when referring to aqueous systems. Water quality is affected when hydrogen bonding between water molecules is disrupted, so the addition of small solutes such as sugars, ionic substances such as salts, can effect a conformational change in the dissolved polymer and subsequent formation of a gel [14, 15].

Mixed gels

Some gels are formed when two polymers, whose solutions would not normally gel, are mixed together. The two components may be of a different chemical nature, such as a protein and a polysaccharide, or they can be chemically similar but physically different, such as two varieties of carbohydrate. If both components are of the same type, the name *binary synergistic* gel is used.

Mixed gels are common in foodstuffs where many different biopolymers come into contact; for example we have the xanthan (a bacterial polysaccharide) and carob gum (a plant polysaccharide) gel, and gels formed between carrageenans and carob gums [16, 17]. Another interesting example is the

so-called "milk reaction" [18–20], in which the mixing of bovine casein with the polysaccharide carrageenan results in a gel.

Internal structure of aqueous gels

Many proposals about the internal structure of gels have been put forward [12, 21] but due to the difficulty in examining gels[5] few have been supported with anything more than rather tentative evidence. Some proposals for internal structure are given pictorially in figure 1.9.

1.4 Agarose and the carrageenans

In this thesis the focus will be on the gelling materials *agar* and *carrageenan*; these are extracted from marine algae of the family *Rhodophyta* (red algae). The usual preparation method is boiling of the algae and precipitation of polysaccharide from solution using added ions [22].

It is thought that the role polysaccharides play in marine algae is one of buoying up and extending the body of the plant in the water to permit maximum photosynthesis to occur, while still allowing the flexibility to withstand movements in the water without breaking up [23]. Other useful properties imparted to surface cells of the plant body may be anti-dehydration at low tide and an ionic permeability, allowing ion-exchange mechanisms to occur with dissolved minerals in the sea water.

1.4.1 Composition

Both agar and the carrageenans are polymers made from disaccharide mers.

One sugar ring in the mer is always of the ordinary six-carbon ring form, while the other has an additional ether oxygen bridge between carbons three and six, creating a double ring structure. The correct names given to the two sugars are respectively *galactopyranose* and *anhydro-galactopyranose*. In agar the galactopyranose (or galactose) sugar ring is 1,3 linked and is of the β-D conformation, and the anhydro-galactose ring is 1,4 linked and of the α-L conformation. In the carrageenans the anhydro- sugar has the α-D form instead.

The most significant difference between agar and the carrageenans is that the carrageenans are polyelectrolytes, and have ionic sulphate groups are covalently bound at regular points along the chain. Differences in the frequency

[5] See the final part of this chapter for more details on the methods of looking at gelation.

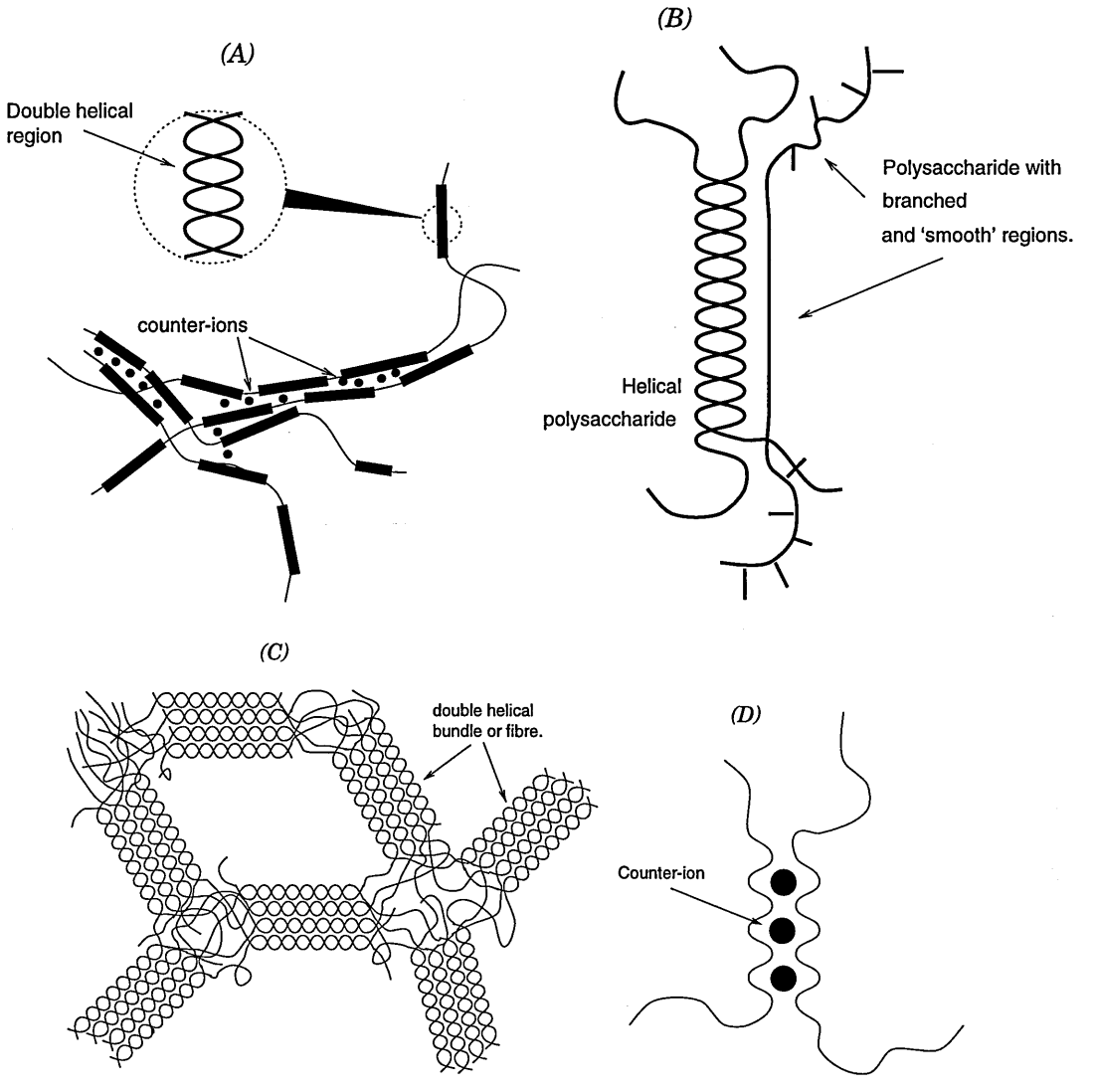


Figure 1.9: Some proposals for gel structure. These structures are thought to be appropriate for different kinds of gel, but all of them contain junction zones where polymers are joined non-covalently. The spaces in between the immobile polymer network are filled with solvent molecules. (A) Filaments of polymer with regions of double helix bound together by counter-ions. This structure is thought to exist in gellan. (B) A binary synergistic gel. Here, the junction zone is formed by a 'smooth' polymer backbone binding to a different polymer which has formed a double helix. (C) A gel constructed from bundles of several double helices linked by randomly coiled polymer chains. This is one proposal for the structure of agarose. (D) Single helices, bound by counter-ions nestling in between the loops. This cation specific model is commonly called the 'egg–box' model, and has been proposed for carrageenans.

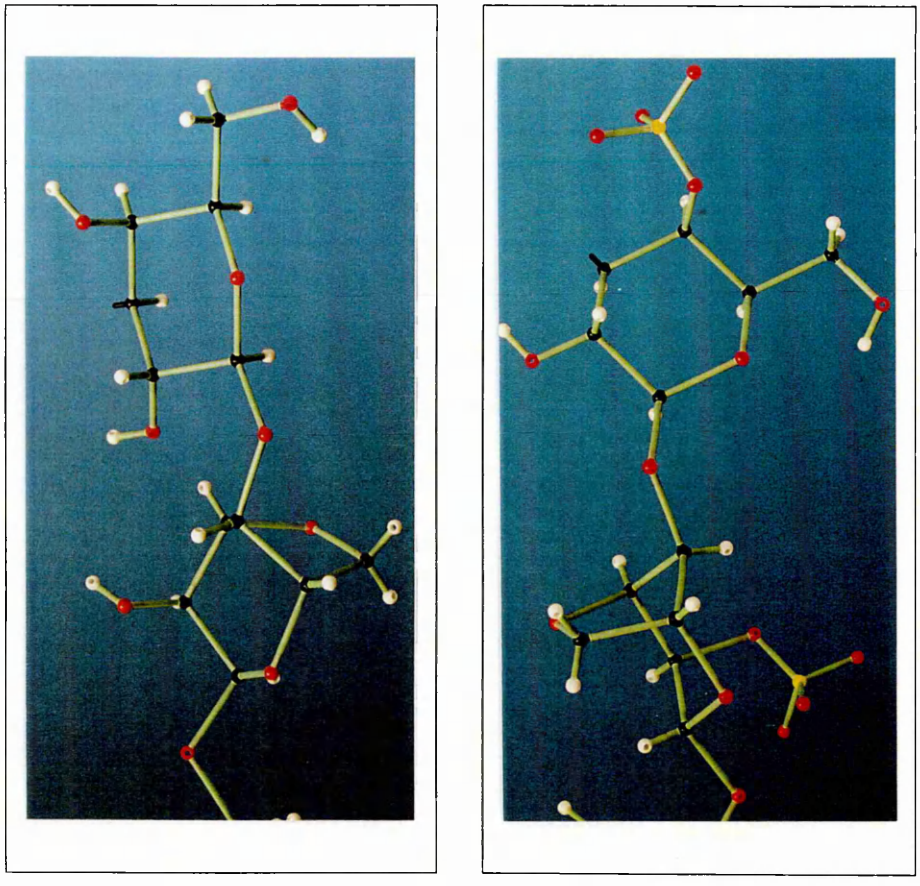


Figure 1.10: Ball and stick models of agarose and iota-carrageenan (see text). Two sugar units from each chain are shown, and these two constitute the basic repeating units in each case. Iota-carrageenan has one sulphate group on each sugar ring, as shown here.

of occurrence of these substituents result in three different carrageenans. Kappa (κ-) carrageenan has one sulphate group per repeating mer; iota (ι-) carrageenan has two sulphate groups and lambda (λ-) carrageenan has between two and three sulphate groups per repeating mer[6]. The repeat units for agar and ι-carrageenan are illustrated in figure 1.10, and in chapter 3.

The quality and composition of agars and carrageenans obtained from different sources varies significantly. Natural agars contain varying proportions of sulphate and methyl groups which occur on the basic polysaccharide chain at random points. These substituents affect the gelling ability of the polymer, altering the physical characteristics (or rheology) of the gel. Hence, there are agar gels which are hard and brittle, gels which are rubbery, and gels which remain stable even at 100°C! Similar variations are found in carrageenans; sulphate, methyl and pyruvate substitution onto the polymer chain is known [24, 25]. The variation in sulphate substitution has a big effect on the gelling ability of the carrageenans, to the extent that λ-carrageenan does not form gels at all [13].

The "ideal" agar has no sulphate or methyl substituents at all and the disaccharide units are all identical. This form of agar rarely occurs naturally and must be made in the laboratory by purification of crude extractions, after which process it is known as *agarose* [22].

1.4.2 Levels of structure in agarose

As with proteins, several levels of structure can be identified in the gelling polysaccharides, which together result in a complete and unambiguous description of their composition and shape. Some of these levels are better understood than others.

Primary structure

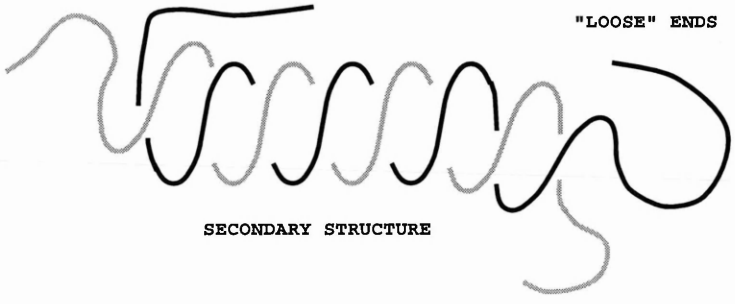
To begin the description of agarose, the basic order of sugar units (the primary structure) within each molecule needs to be specified. The arrangement of atoms within the sugar units in this context is taken to be already known, as it can be ascertained using the nomenclature of the sugar unit. In agarose the primary structure is:

$$-\Big[(\beta\text{-D-galactopyranose})\text{–}(\alpha\text{-L-anhydro-galactopyranose})\Big]_n-.$$

[6] Another non-sulphated carrageenan, β-carrageenan, has been made artificially.

Secondary structure

Next we come to the *secondary* structure. Here a description of the shapes of individual agarose chains is added to our picture of the bulk agarose material. Agarose molecules favour the helical conformation at temperatures below 40°C, and are thought to exist in the form of *double helices* [26]. Evidence for this takes the form of secondary reflections observed in x-ray photographs of agarose helices; a representation of the double helix is given below.



Parameters are usually given when helices are being described. The important ones are the height per helical turn, which is the distance between successive loops *of the same chain*, and number of polymeric repeat units per turn. For agarose the height per turn is 19Å and there are three repeat units (disaccharides) per turn. In addition the helix is left-handed [26].

Tertiary and quarternary structure

In order to arrive at the final macroscopic gel structure, two further levels of structure must be mentioned. After the coiling of agarose chains into double helices we have *double helix aggregation*, which forms *tertiary* structure. Here, double helices line up side by side to form bundles with a diameter of several nanometres. It has been estimated that these bundles can contain six or seven double helices [27]. Finally the bundles are linked into a three dimensional arrangement by flexible randomly coiled chains of agarose to form a gel network: this is *quarternary* structure.

1.4.3 Polysaccharides and the microstructure of foodstuffs

Microstructure and texture in food

Foodstuffs are probably the most complex products which are commercially available. The food industry has a turnover measurable in billions of pounds per annum, and eating foods is one of the greatest of social activities.

The diversity of foodstuffs is enormous, ranging from fruit and vegetables which need virtually no processing through to frozen dairy products and ready-meals, into which goes much research and development. Foods contain innumerable chemicals from simple substances such as water and inorganic salts to highly complex and delicate flavourings and proteins.

Apart from the flavour of foods, the texture experienced during chewing and swallowing is probably one of the most important factors to be taken into account in the design of special food products, particularly for the convenience market. Texture is imparted to the food through the physical interaction of various microscopic elements in food. It obviously has to be acceptable to the consumer, relevant to the type of food (*e.g.* cereal) and stable during storage. The principal components of food (although not all may be present at the same time) can be itemised as [28]:

- Fibres. These are composed of indigestible or semi-indigestible solids such as cellulose fibres, and animal protein.
- Fat crystals.
- Fat or oil droplets.
- Droplets of aqueous solutions
- Protein globules and fibres
- Gelatinous matrices. The substances responsible for these are typically polysaccharides and proteins.
- Gas bubbles. These are absolutely crucial to the texture of whipped creams and mousses.

Uses of agar and carrageenan

The major commercial use of agar and the carrageenans is in the dairy foods area as gelling and bulking agents [24, 29]. Both chilled and room temperature deserts contain stabilisers and gelling components which are blended with sugars, cream and ice or water. Some of the properties imparted to these foods are only realised when two components are in contact, such as the carrageenan–protein "milk reaction" [28].

Agar is also used in large quantities in the medical world, as it provides an ideal medium for growing microbial cultures. The gel can be poured as a liquid into trays and allowed to set. It is easy and clean to dispose of and nutrients can be added to it.

Other polysaccharides

Gellan, a gum secreted as an extra cellular slime by the yeast *Pseudomonas elodea*, is used in food also as a stabilising agent [21, 30]. Another polysaccharide, *gelatin*, originates from bovine cartilage [31] and is the main ingredient of jellies. Yet another widely used polysaccharide is *pectin*, used in jams and preserves. It is in fact a group of related polysaccharides and occurs in many land - growing plants [32–34].

Starch is one of the most well known of gelling materials.

It occurs naturally as granules in plant tubers and seeds. It is in fact two polysaccharides; *amylose* (1→4 linked α-D-glucose) and *amylopectin* (a 1→4 linked α-D-glucan with extensive branches achieved by 1→6 linkages). These two polysaccharides constitute crystalline and amorphous regions within starch granules [35, 36]. When immersed into boiling water, the granules take up water and expand to become small particles with a jelly-like nature. Upon further boiling these particles burst and a homogeneous solution is formed. The process is greatly accelerated if the granules are broken (by grinding) beforehand [6]. These properties are essential to processes such as bread making, and also in the paper industry.

Other polysaccharides are also employed in the paper and textiles industries; these will not be described here.

1.5 Aims of the current work

This research project set out with one main objective, which can be summarised in a single sentence:

> The aim of the work is to add to the current understanding of the nanoscopic structure of marine algal polysaccharide gels through the application of computer modelling and laboratory analytical techniques.

An increased knowledge of these gels is crucial in the drive towards improved design of foods which contain them.

1.5.1 Why use computer modelling to probe gel structure?

Gels are notoriously difficult to examine. Currently the biggest area of gel research focuses on *rheology*. Here, a host of physical tests measure such properties as birefringence, shear resistance and Young's Modulus in order to describe and characterise a gel. However, no information concerning the internal structure of the gel is gathered. Standard techniques are all but useless for the study of internal structure: NMR necessitates sample spinning, which destroys most gels; x-ray crystallography only works for materials with a high degree of order. X-ray diffraction patterns have been obtained for certain preparations of agar and ι-carrageenan but these were not gels, but dried and extruded fibres[7]. Electron microscopy (EM) is useless because the sample preparation results in severe degradation of the original gel structure.

Scattering measurements of x-rays and visible light are more useful. During gelation, these techniques are able to provide information on the size of aggregated particles in solution, but they still do not deliver a picture of the final three-dimensional network [12, 37].

Optical rotation devices measure the degree of rotation imparted by a solution of molecules to plane polarised light, in a similar way to the scheme given in figure 1.4. Again, during gelation, this allows conformational changes of molecules to be followed but no information on the final structure is produced. There is therefore considerable evidence which supports structural elements in gels such as double helices and aggregated particles, but there is little information on the complete picture.

More recently, some more complex techniques have been developed which attempt to probe the internal structure of gels. If the gel sample is 'slammed' onto a polished copper block kept at liquid nitrogen temperatures, water vitrification occurs. The water can then be sublimed off, preserving the gel structure, and metal shadowing techniques applied [38]. After this the usual EM imaging process is used. The resulting images show open pores and linked strands reminiscent of the structure of bread. Strand thicknesses can be measured at less than ten nanometres.

Another method uses *birefringence* measurements to study the agarose gel matrix [27]. Intrinsic birefringence measurements, taken in a similar fashion to the method shown earlier in figure 1.4, and other measurements taken under the effects of an electric field applied to the sample again indicate

[7]Note these do give a picture for double helices present in the samples; see chapter three.

the presence of fibres bound into a matrix. There is also evidence for the existence of domains using this method.

The role for molecular modelling

If the internal structure of the agarose gel system is to be discovered, more methods which can detect it directly must be developed and used.

One technique stands out by virtue of its very intimate portrayal of the behaviour of biomolecules: molecular modelling is already well developed and faster computers are enabling the study of larger and larger systems. The use of this technique in gaining new insight into the gelation of agarose will be described in this thesis.

Molecular modelling however must be used in conjunction with the understanding that it is a theoretical method, and the results might be inaccurate or even incorrect. With this in mind some laboratory work has also been included for this thesis, and work published in the literature has been constantly referred to in order to try and ensure that the results from computer modelling are consistent with currently known facts. The whole project can indeed be termed 'structural studies on agarose', as reflected in the title.

This thesis is constructed in a similar fashion to the way in which the actual progression of the work itself unfolded. Before the results are presented an overview of this will be given.

1.6 The thesis tree

Investigate the background

In the introduction and in the second 'Principles of modelling' chapters, the background to the project is described. The existence of agarose, a gelling polysaccharide, is placed into context beside foodstuffs and carbohydrates. The theory behind molecular modelling is reviewed.

Build models of the polysaccharides.

A chapter is devoted to the transfer to computer of data currently known for agarose and carrageenan structures. This utilises several key literature references and serves as a good introduction to the application of computer modelling. Some simple investigations into the properties of these new models are presented.

The treatment of solvent

Chapter four deals with the task of discovering the significance of solvent in the modelling of the carbohydrates. To this end, several experiments are presented and analysed. Results for cases where solvent was included and omitted (*in vacuo*) are compared. It is to be expected that solvent is of some importance in determining the properties of agarose, and may be even more important in the modelling of charged polysaccharides such as carrageenans.

Laboratory work

The work detailed in chapter five covers the bench chemistry and analytical work done at the food research laboratories of UNILEVER, Sharnbrook, Bedfordshire. This work is of considerable use in supporting and guiding the modelling aspects of the thesis, and represents a significant portion of the research.

The following are described:

- Hydrolysis of agarose, and the optimised conditions.
- High performance anion exchange chromatography (HPAEC), used to characterise the hydrolysis mixtures (and the products of other tests).
- Gel permeation chromatography, which was employed in the isolation of samples of agarose of known chain size.
- Thermodynamic changes during cooling of hot solutions of agarose chains, observed and quantified using differential scanning calorimetry (DSC).

The reversible coil to helix transformation

Using the previous understanding, an investigation on agarose random coil to double helix ordering is described in chapter six. Two main approaches are used. Firstly, the coil to helix theories of Zimm and Bragg [39] are used to predict some characteristics of the transformation.

Secondly, molecular modelling of long agarose chains is carried out, using various initial conformations. This approach starts from double helices and observes their disruption.

Conclusion

The conclusion will bring all the results together, forming the final, seventh, chapter.

The principal achievements are:

- an understanding of molecular modelling and carbohydrate polymers;
- the conclusion that, in modelling the carbohydrate agarose, the inclusion of solvent reduces the amplitudes of molecular motion but does not affect the final result significantly;
- experimental determination of estimated values for the minimum size of junction zones in agarose, and the enthalpy of phase transitions.
- an understanding of of helix–coil theory with regard to agarose.

Some appendices follow the conclusion. These contain details of `FORTRAN` programs used, and information on the theory behind high performance anion exchange chromatography.

Chapter 2

Principles of modelling and application to agarose

Contents

2.1 Introduction

This chapter will cover the theory and application of computer - based molecular modelling methods to carbohydrates [9] and in particular, agarose. Key words such as *mechanics, dynamics* and *forcefields* will be introduced.

All computer aided chemistry packages are designed to assist the chemist by providing a mathematical model which can evaluate the energy of molecules as a function of the atomic nuclei positions. Some of these models are extremely accurate (such as those based upon *ab initio* molecular orbital methods) but can only be applied to systems containing a few tens of atoms. Others are less accurate at the electronic level but are able to cope with large proteins and bulk solutions [40–42].

Although molecular modelling can be used equally well in the study of ionic lattices, micro-composites and molecular materials, the method of its application differs slightly between these fields. In the current work we are interested only in the modelling of carbohydrates and the solvent: water.

At the end of this chapter a simple example of the building of an anhydro–sugar ring is given.

2.2 Molecular modelling: the basics

The previous few years have witnessed an explosion in the availability and concurrent usage of molecular modelling packages, greatly assisted by the increase in power of workstations and desktop machines. Molecular modelling at a basic level at least is now routinely applied by many chemists, to assist in the visualisation and study of almost any type of molecule. Packages are available for all the common computer platforms: UNIX and Linux, and even MS Windows and the Macintosh. Some of these are freely available on the world wide web, while others are extremely sophisticated (and expensive) having as they do very powerful graphical interfaces and analysis modules. The availability of this arsenal of tools has helped meet the demand for non-experimental methods of investigating molecular structures.

Some programs which have been used for this thesis include Rasmol, Molden and MOLMOL for visualisation, HYPERChem™ for the PC, Amber for larger simulation jobs, and Gaussian94.

The theory of molecular modelling is treated comprehensively in many texts [43–46], and over the years many papers have been published which cover the development of codes and parameters for the theory. Indeed, whole journals have sprung up which cover the subject exclusively. In the

present circumstances it will be sufficient to give an overview of the basic pre-requisites for molecular modelling.

2.2.1 Forcefields

The essential goal of molecular modelling is to describe molecules in terms of an *energy surface* whose variables are the *internal parameters* of the molecule. Internal parameters are coordinates, bond lengths and bond angles, which together describe the molecule.

The forcefield is the marriage of mathematical expressions and pre-defined molecular parameters which enable the energy function to be evaluated with respect to atomic coordinates[1].

Other modules and algorithms in modelling packages are all concerned with various ways of changing molecular structure and then analysing results.

Mathematical expressions for evaluating the energy function are based on classical (Newtonian) mechanics. The atoms are treated as simple spheres, and the bonds are treated as elastic rods. The size of the spheres and the force constants of the rods are dependent on the types of 'atoms' involved. Thus, parameters for carbon–oxygen bonds are not the same as those for carbon–carbon bonds. Forcefields are more accurate if they are restricted to certain types of molecule such as proteins or metallo-complexes. Hence a whole range of forcefields is available, but in the present work the AMBER forcefield is used exclusively because it has been parameterised specifically for proteins and carbohydrates, and has been well documented [47–50].

Parameters for the forcefield are found from both experimental and theoretical sources. Quantum mechanical methods are used to derive partial charges. X-ray experiments yield geometrical information. Spectroscopic methods allow the determination of force constants for bonds. In addition, modelling is frequently used in conjunction with NMR coupling constants and nuclear overhauser effects [51] (NOEs) in order to arrive at biopolymer configurations in solution [52, 53].

It is considered that for the purpose of obtaining well - optimised structures, a simple energy expression *with good parameters* does the best job [54].

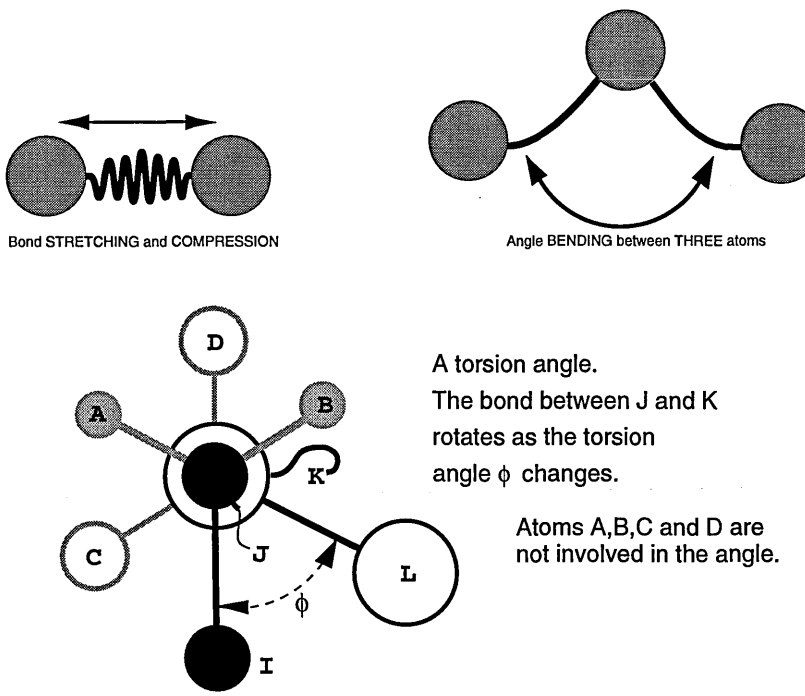


Figure 2.1: Schematic showing the origin of the energy terms for bonds (two - body), angles (three - body), and torsions (four - body) in the AMBER forcefield.

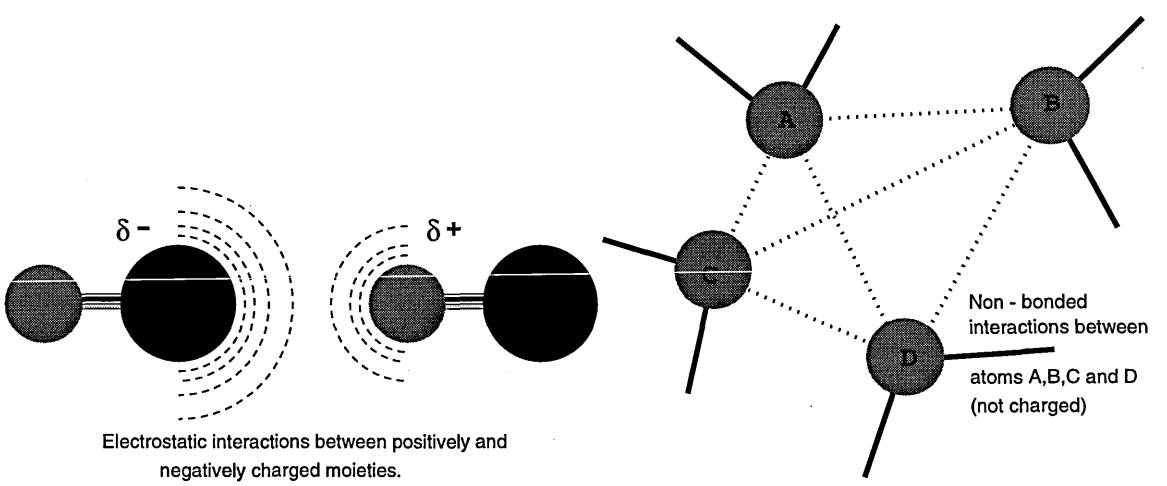


Figure 2.2: Two non-bonded interactions contributing to the total energy in the AMBER forcefield. All atoms experience the so-called Van der Waals force, but may not feel electrostatic forces.

The AMBER forcefield

The total energy of a molecule in AMBER is found from six basic terms: atom-atom bonds, angles and torsion angles as well as two types of non-bonded interaction and a hydrogen bonding term.

The expression for **V**, the total energy is given as [55]:

$$\mathbf{V} = \sum_b K_2(b - b_0)^2 + \sum_\theta H_\theta(\theta - \theta_0)^2 + \sum_\phi \frac{V_n}{2}[1 + \cos(n\phi - \phi_0)]$$

$$+ \sum_{i,j} \epsilon[(r^*_{ij}/r_{ij})^{12} - 2(r^*_{ij}/r_{ij})^6] + \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon r_{ij}} + \sum_{i,j}\left[\frac{C_{ij}}{r^{12}_{ij}} - \frac{D_{ij}}{r^{10}_{ij}}\right] \qquad (2.1)$$

These terms will be described separately in the following text, and are represented schematically in figures 2.1 and 2.2. Note that a change in molecular structure may affect several terms at once.

Bond stretching

In molecular modelling bonds exist according to how the system is initially described to the modelling program. They cannot be altered once defined. The total energy from all defined bonds is given by the relation:

$$\mathbf{E}_{\text{bonds}} = \sum_b K_2(b - b_0)^2 \qquad (2.2)$$

This is term 1 in Equation 2.1. It is of the form of a simple harmonic oscillator (SHO), with force constant K_2, bond length b and equilibrium bond length b_0. The sum is over all bonds in the system. Some forcefields use other mathematical forms for the term, such as Morse potentials and inverse power potentials [54], but they are computationally more expensive and the SHO is adequate for most purposes.

Angle bending

The valence angles are considered to deform in a similar fashion to bonds; *i.e.* the energy is represented by an SHO with force constant H_θ:

[1] Atomic positions can be identified by *internal coordinates*, where the positions are described within the molecule's own frame of reference, or in *Cartesian coordinates* where atoms have (x, y, z) values with respect to an external frame.

$$\mathbf{E}_{\text{angles}} = \sum_{\theta} H_\theta(\theta - \theta_0)^2 \tag{2.3}$$

Here, the force constant is H_θ, and the deviation from equilibrium is $(\theta - \theta_0)$.

The torsion angle

This four body term has a periodic angular dependence. See figure 2.1 for a representation of the four atoms involved. The dihedral angle ϕ between bond `I-J` and `K-L` can rotate though 360° until it arrives back at the starting point, and the energy varies throughout this cycle. Of course, some kind of command needs to be input to cause such a rotation (see section 2.2.2).

The energy expression is term three of Equation 2.1:

$$\mathbf{E}_{\text{torsions}} = \sum_{\phi} \frac{V_n}{2}[1 + \cos(n\phi - \phi_0)] \tag{2.4}$$

$\frac{1}{2}V_n$ is the rotational barrier height, with a multiplicity (over 360°) given by n in the round brackets. ϕ_0 is the phase angle where the function is a minimum, and ϕ is the torsion angle itself.

Non-bonded interaction (1)

Non bonding energy terms are applicable to all atoms, whether they are bonded or not. The van der Waals (vdW) interaction is one such term. It is thought to arise from the interactions of polarisable electron clouds around atoms [22, 46]. In molecular modelling the electrons are not represented, so the attractive vdW interaction is accounted for by a dispersive term proportional to $\frac{1}{r^6}$, where r is the atomic separation. The repulsion, felt by atoms if they approach one another too closely, is accounted for by a $\frac{1}{r^{12}}$ term. A *Lennard–Jones* potential, or 6–12 term (equation 2.5) can be used to represent the vdW energy.

$$\mathbf{E}_{\text{VdW}}(r) = \sum_{r_{ij}} \epsilon \left[\left(\frac{r^*}{r}\right)^{12} - 2\left(\frac{r^*}{r}\right)^{6} \right] \tag{2.5}$$

r^* is the collision diameter (sum of atomic radii: see figure 2.3), and ϵ is the minimum attainable energy. A schematic plot of the function is given in figure 2.3.

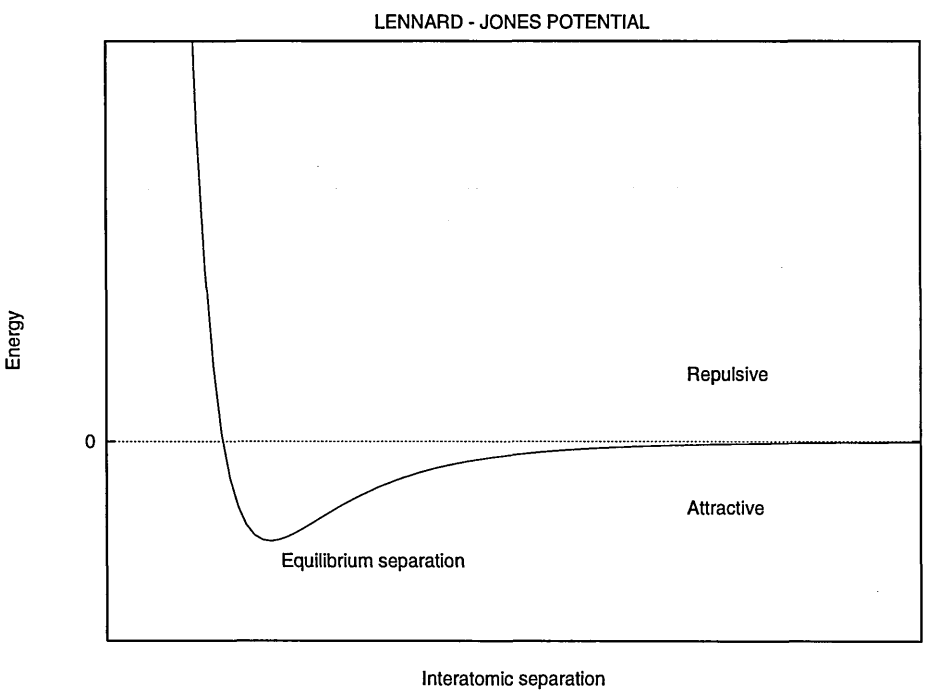


Figure 2.3: The Lennard–Jones potential. If the inter-atomic separation is the sum of the atomic radii, r^*, the function becomes zero, and then it increases rapidly if the atoms are pushed together. At the equilibrium inter-atomic separation the function is at a minimum value (ϵ), and beyond that it slowly returns to zero.

Non-bonded interaction (2)

The second non-bonding energy term arises from the forces existing between charged atoms, and is evaluated using Coulomb's Law. The only factors are the sizes of the partial charges (q_i, q_j), their separation r and the dielectric constant ε:

$$\mathbf{E}_{\text{elec.}}(r) = \sum \frac{q_i q_j}{4\pi\varepsilon\, r_{ij}} \tag{2.6}$$

Non-bonded interaction (3)

This term, which is not available in all forcefields, gives an extra contribution to the total energy for atoms which can hydrogen-bond.

These atoms are commonly oxygen, nitrogen and hydrogen, although recently work has emerged indicating that specific carbon–hydrogen to oxygen non-bonded interactions may exist [56, 57]. However these are not included in most forcefields. The form of the hydrogen-bonding term in AMBER is a modified Lennard–Jones potential:

$$\mathbf{E}_{\text{H.bond}}(r) = \sum \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right] \tag{2.7}$$

C_{ij}, D_{ij} are constants derived from the ideal hydrogen bonding distance (where the energy is a minimum), and r is the separation between the two hydrogen bonding atoms. The expression forms term six of Equation 2.1.

2.2.2 Mechanics

Molecular mechanics (MM), dynamics (MD, see section 2.2.3) and Monte Carlo dynamics (see section 2.3.4) together provide most of the techniques required for molecular modelling. MM is a static approach, with the two most common strategies being *minimisation* and *conformational searching*. Both of these are designed to find conformational arrangements of the molecular system which have low energy, and are therefore likely to exist in real systems.

The lowest energy of all, over the complete energy surface describing all possible conformations, is the *global minimum*. Any conformational change whatever from this will result in a higher energy structure. For complex proteins the global minimum may be impossible to determine, due to the vast number of individual contributions to the total energy. However, any number of *local minima* may exist which are lower in energy than the immediate surrounding energy surface, but not as low as the global minimum.

The process of finding energy minima is known as *minimisation*. A computer algorithm for minimising needs to be able to determine the slope of the energy surface at the specified starting point, and alter atomic coordinates so that it begins to explore down the slope. The method of determining the direction in which to explore depends on the minimiser chosen, while the search for the minimum along that direction is usually done using a *line search* algorithm.

Note that the magnitudes of the energies so calculated by MM (or MD) are not to be compared to those obtained either experimentally or by *ab initio* studies, even for identical molecules.

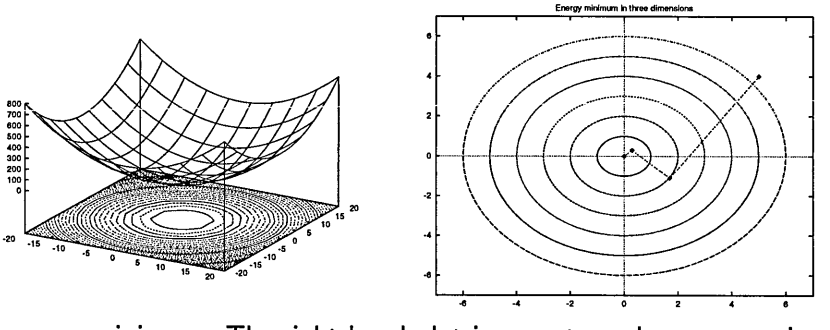
Steepest Descents

Steepest descents [45, 46, 58] employs the first derivative of the potential function:

$$\frac{\partial V}{\partial r}$$

The slope of the energy surface is determined at the given starting point and the minimiser begins to alter the molecular geometry in this direction (literally, the steepest descent), until the first derivative is zero (see the diagram below). The minimiser then determines whether there is still a slope and if so it sets off again: the second direction will be perpendicular to the first. The method is robust (it always finds a lower energy structure) but not very efficient since it is continually overshooting the shortest route to the minimum.

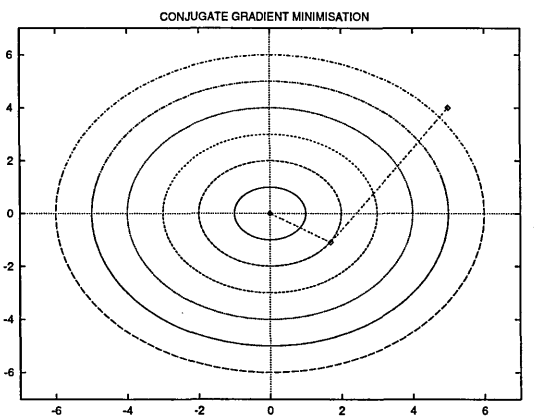
This is illustrated schematically below for a very simple energy minimum:



A simple energy minimum. The right hand plot is a contoured representation of the left hand surface plot. The location of the minimum by inspection is clearly at the origin, (0,0). The minimisation routine reaches it in three steps, starting at (5,4) (right hand plot). This is a typical "steepest descents" result: see text.

Conjugate Gradients

This method is more advanced than steepest descents, as it makes use of direction information from the previous step in order to make the next adjustment towards the minimum. It is able to find the exact (local) minimum of the energy function in M steps, where M is the dimensionality of the energy function [58]. For the simple function shown earlier therefore the minimum is found in two steps:



This is particularly advantageous in a long narrow shaped minimum, where the steepest descents method might require many iterations to arrive at the bottom.

Newton–Raphson

The previous methods employed the first derivative of the potential energy function in order to find the direction to search for a minimum. In general however for a harmonic function in M dimensions, one should be able to use the second derivative also, and so predict where the minimum is likely to be without searching for it. At the minimum point on the gradient, the function will change direction and so:

$$\mathbf{r}_{\min} = \mathbf{r}_0 - \mathbf{A}_0^{-1} \nabla V(\mathbf{r}_0)$$

The predicted minimum, $\mathbf{r}_{\min}$, is found using the matrix of second partial derivatives, $\mathbf{A}_0$, of the energy at the starting point $\mathbf{r}_0$. $\nabla V(\mathbf{r}_0)$ here denotes the gradient of the potential energy at $\mathbf{r}_0$.

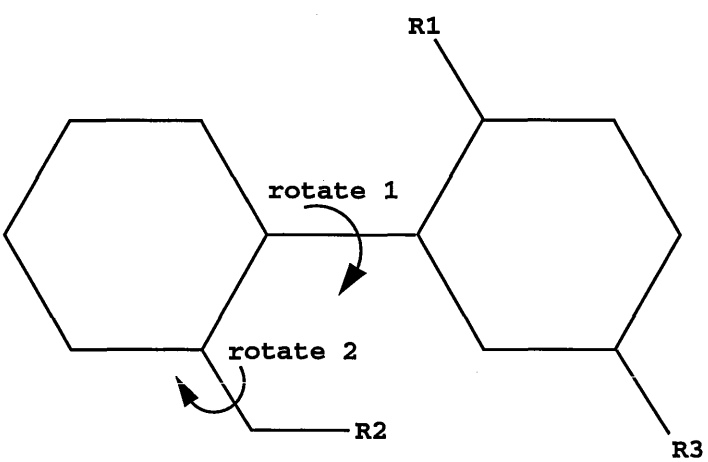
Although this method seems rather attractive at first sight, it has some intractable drawbacks [58]. Firstly the calculation and inversion of large matrices is very demanding on computer memory, and so it cannot be applied

to large systems (more than 1000 atoms). Secondly the minimiser will become unstable in some situations where there is no minimum close by. It therefore tends to be reserved for cases where very accurate minima are desired.

In the majority of cases (and in this thesis) the method of steepest descents is initially used, followed by conjugate gradients. Since even local minima can be elusive, control parameters to limit the use of computational resources are usually entered. These controls may be of the form of a maximum number of steps to be executed, or a particular convergence level to be reached, in kilo-calories per mole per Angstrom (kcal/A).

Conformational searching

In many computer modelling experiments the global minimum is a good point from which to start, and in addition often reflects conformations found in crystals [9, 52]. The global minimum can only be found by a minimisation algorithm for very simple molecules. For more complex molecules a systematic search using key conformational features can be done in order to identify it. This process is called *conformational searching*. As an example, consider a molecule which has two rings connected by a central bond, each ring having different substituents:



A simple minimisation of energy is unlikely to find the global minimum of this molecule. The key conformational features are the freedom about the central bond and the bond connecting R2 to the ring. A conformational search would rotate the central bond and plot out the variation in energy. Each rotation increment would be accompanied by an energy minimisation. Once a full 360° had been completed, the second rotation (around the bond connecting group R2) would be incremented once, and the first rotation repeated. Slowly a complete map of the potential energy *versus* conformation

would emerge. From this the absolute minimum can be identified. The method is fail-safe, but for even more complex molecules with more key conformational features the process may be too lengthy to contemplate!

2.2.3 Dynamics

Molecular dynamics (MD) methods are very attractive to physical chemists. Using them, the behaviour and interaction of molecules with time can be predicted.

In MD, atoms are assigned velocities and the equations of motion are solved by numerical methods. This introduces a fundamental *time step* for each iteration. In order to give sufficient accuracy the time step must be assigned a small value (of the order of 1×10^{-15} seconds), and this means that only nanoseconds of time can be simulated: certainly not enough in which to observe for example protein folding. Nevertheless any point on the potential energy surface can in principle be accessed during molecular dynamics, so it is a powerful tool for the study of molecular motions.

Theory

Since all of the atoms in the system are assigned a velocity, and all have a mass with a force acting on them, the acceleration $\vec{a}_i$ of each atom can be calculated according to Newton's First Law of motion:

$$\mathbf{F}_i = m_i \vec{a}_i$$

The force arises from the derivative of the potential energy surface:

$$-\frac{\partial V}{\partial \mathbf{r}_i} = m_i \frac{\partial^2 \mathbf{r}_i}{\partial t_i^2} \tag{2.8}$$

where $\mathbf{r}_i$ are the coordinates and m_i and $\vec{a}_i$ are the mass and acceleration of atom i.

The determination of future positions and velocities of any atom in the system does not however follow immediately. Exact solutions to the equation cannot be elucidated for more than one or two particles, and so numerical methods must be applied. For any particle, the motion can be expressed using a Taylor series description of its position $\mathbf{r}(t+\delta t)$, a very short time δt after the starting time t:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \frac{\partial \mathbf{r}}{\partial t}\delta t + \frac{\partial^2 \mathbf{r}}{\partial t^2}\frac{\delta t^2}{2} + \cdots \tag{2.9}$$

In order to solve the equations of motion, the starting coordinates are entered, random initial velocities are assigned using a Maxwell–Boltzmann distribution, the acceleration $\partial^2\mathbf{r}/\partial t^2$ is found from equation 2.8, and finally a suitable approximation to the higher terms in the Taylor expansion is used. This is the basis for the *Verlet algorithm* [59] for calculating time dependent trajectories.

Two further points should be amplified:

1. The time step, δt, needs to be carefully chosen. If too small then unnecessary computer time is wasted in achieving the desired simulation. If too large, fast vibrations such as bond stretches involving light atoms will not be properly reproduced, and atoms on a collision path may actually intersect each other because they have moved too far during the time step. This can result in abnormally large forces and "system blow-up". In most cases a time step of one femtosecond (1 fs, 1×10^{-15}s) is adequately small.

2. The Maxwell–Boltzmann distribution of velocities v, for N atoms at a temperature T Kelvin is given by:

$$3kT = \sum_{i=1}^{N} \frac{mv^2}{N} \quad \ldots \text{where } k \text{ is the Boltzmann Constant.}$$

   In other words, the temperature is directly related to the atomic velocities (and the kinetic energy). This is a result from the kinetic theory of gases.

Temperature control

In MD, the temperature of the system must be initialised (when the velocities of the atoms are assigned) and then maintained at the desired value during the calculation. A *heat bath* is therefore needed to keep the temperature correctly adjusted. During the first stages of MD, the initial random velocities (kinetic energies) of atoms become partitioned into other forms of energy, mainly potential, as the molecules warm up and bonds become stretched. Hence the initial assignment of velocities is insufficient to maintain the target temperature, and a velocity scaling algorithm is used to warm the system up quickly.

At the target temperature the distribution of velocities is maintained using a coupling algorithm, allowing heat exchange between the system and an imaginary bath. The Berendsen method [60] is one such algorithm. Each

velocity is scaled by a factor λ, determined according to how far from the desired temperature, T_0, the actual temperature T of the system is:

$$\lambda = 1 + \frac{\delta t}{2\tau}\left(\frac{T_0}{T} - 1\right)$$

Here, τ is a characteristic relaxation time controlling the tightness of coupling and δt is the time step. The temperature T is evaluated using the sum of all the current atomic velocities.

2.3 Molecular modelling: more features

2.3.1 Cutoff values

If no limits are set, the calculation involving the non-bonded energy terms in equation 2.1 can become very demanding on computer time. The number of included terms is therefore limited by ignoring interactions between atoms which are separated by more than the *cutoff limit*. For systems containing a few hundred atoms it is not worth applying cutoffs, but as the number is increased beyond a thousand the cost in terms of computing resources forces cutoffs in the region of 6 to 10Å to be applied. The reduction in the accuracy of the non-bonded energy term is acceptable for most work[2].

2.3.2 Periodic boundary conditions

Although MD methods using forcefields and classical mechanics are able to handle many more atoms than those methods which employ quantum mechanics, they still suffer limitations. Workstations found in most Universities can cope with perhaps a thousand or so atoms. On supercomputers this may be increased to several thousand. However, large biomolecules such as proteins might be even larger than this, and if solvent is included then the number of atoms will increase still further.

The usual approach to reducing this limitation is to apply periodic boundary conditions (PBC). A typical method is as follows:

- Build the solute of interest;
- define a box around the solute (this can be any crystallographic shape, but is commonly cubic or tetragonal);

[2]The non-bonded interactions are switched off smoothly over a distance of a few Å, to prevent non conservation of energy.

- fill the box with solvent using a predefined solvent–generating algorithm to reproduce the correct density;
- reproduce the defined box in three dimensions;
- start MD calculations.

See figure 2.4 for an example of this procedure.

The technique can also be applied to regular infinite polymers, where the defined box contains a building unit which has a specific head and tail.

Using PBC "bulk" properties can be calculated. The MD calculations are only performed on the original box, the results being copied into all other boxes. The molecules in the original box feel the effects of their "ghost replicants" in the neighbouring boxes, so they behave as if the environment was an infinite solution. If a molecule drifts out of the original box, a replicant simultaneously enters from the opposite side (see figure 2.4).

The caveat to using PBC is that the cutoff should be set to no more than half of the cell side length to prevent atoms from interacting with replicants of themselves. Additionally it is no good trying to simulate a very large molecule using a very small box, because the contents of the box must adequately represent the smallest building unit of the system in question. PBC by their very nature introduce an inherent periodicity into the system; care should be taken that this does not lead to artifacts.

2.3.3 Trajectories

MD calculations generate new conformations for the system with every time step which is completed. Most MD programs allow the modeller to specify how often the data is to be written to a file: for example if the only factor of interest is long term inter-molecular interactions it is hardly necessary to save data after every single time step. The final output is of the form of a *trajectory* or *history* file.

History files are generally very large, containing as they do a long list of structures and energies. Analysis programs extract and create visual graphs for properties such as geometry, energy and temperature *versus* time or other suitable regularly increasing dimensions. These data are the final result of the investigation.

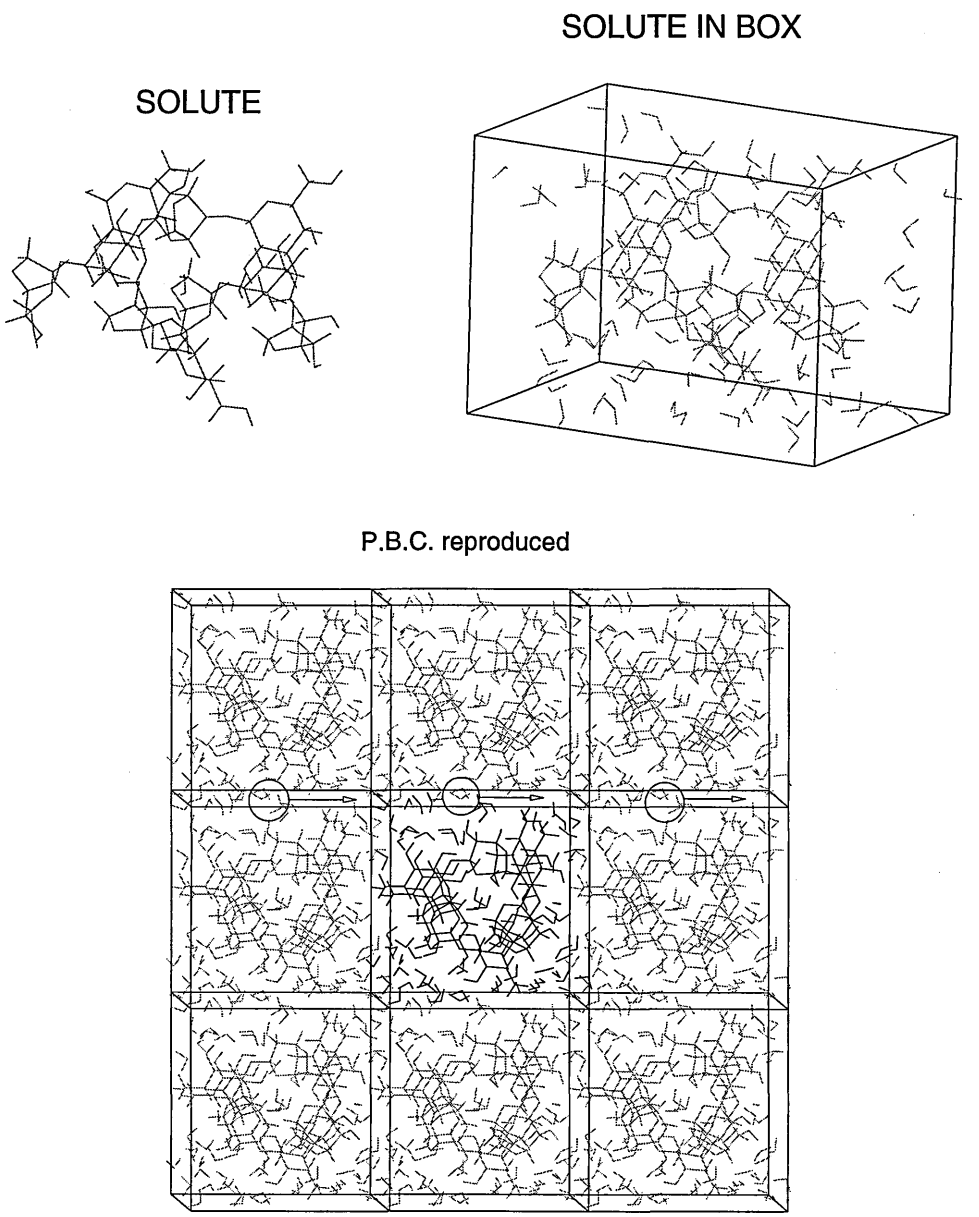


Figure 2.4: Solvation of a solute and application of periodic boundary conditions. The two pictures at the top show a solute (here consisting of two short carbohydrate chains) being enclosed in a box and solvated. The lower image demonstrates the reproduction of the original box, which is at the centre. The circles and arrows indicate how the movement of one molecule is copied into other boxes.

2.3.4 Monte Carlo dynamics

The Monte Carlo method is another way of exploring conformational space. Instead of using atomic velocities to provide the next conformation, new structures are generated by random rotations of key conformational features rather similar to the conformational searching described on page 36. High energy conformations involving superposition of atoms are rejected. Monte Carlo dynamics will access more of the molecule's conformational space than can be achieved using time-dependent molecular dynamics, but the drawback is that the information bears no relation to how a real molecule might diffuse and change shape since any energy barrier can be overcome. Monte Carlo dynamics are therefore best suited to the estimation of thermodynamical quantities such as free energy, since they can be used to build up a statistical sample of the relative populations of different conformations.

2.4 Hardware and software

The modelling work described in this thesis was carried out on a Silicon Graphics Iris Indigo XZ 4000 workstation. These computers combine fast processors with the highest quality graphics hardware.

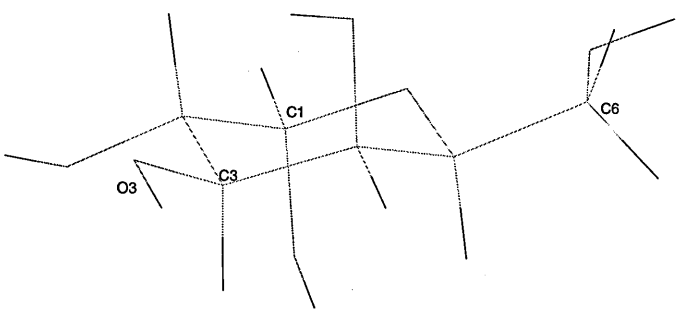
The modelling program used was DISCOVER version 2.95 by Biosym Technologies, which has a command based interface and also a graphical interface in the Biosym INSIGHTII software (version 2.3.0 was used). INSIGHTII was also used to build the polysaccharides, using the biopolymer add–on module. The forcefield used was AMBER, with Homans' modification for use with polysaccharides [47, 61].

Further calculations were done on the DEC supercomputer at the Rutherford Appleton Laboratories again using the AMBER forcefield. The results were analysed in this case using `amber5/carnal`, a structural analysis program.

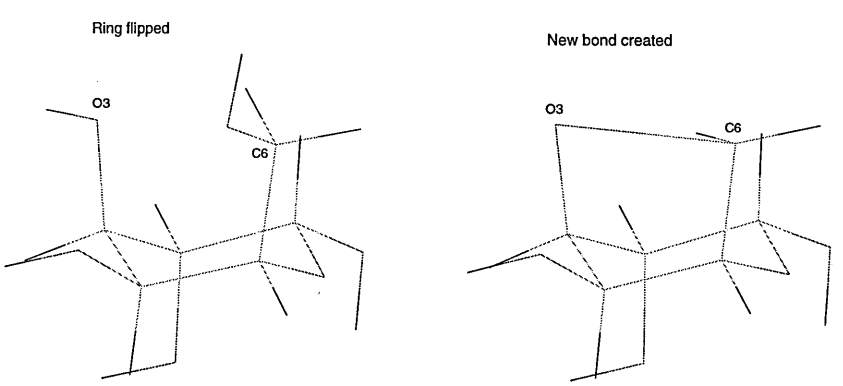
Other analysis was done using INSIGHTII and the graph plotting utilities GNUPLOT (GNU software foundation) and ORIGIN. Visualisation of structures and creation of images was done using INSIGHTII, Rasmol and MOLMOL. Colour images were produced using 35mm KODAK Ektachrome film and a still camera.

2.5 A simple calculation

A simple illustration is provided by the building of the sugar α–L–anhydro-galactopyranose, which is a constituent monomer of agarose. Within INSIGHTII the basic galactopyranose ring is obtained using the `carbo_get` command:



In order to create the anhydro ring it is necessary to condense two hydroxy-groups, eliminating a water molecule. The relevant atoms are removed using the `delete_atom` command, and significant distortion of the molecule is required in order to bring the remaining atoms into bonding positions. The new bond between C3 and O6 is made using the `modify_bond_create` command.

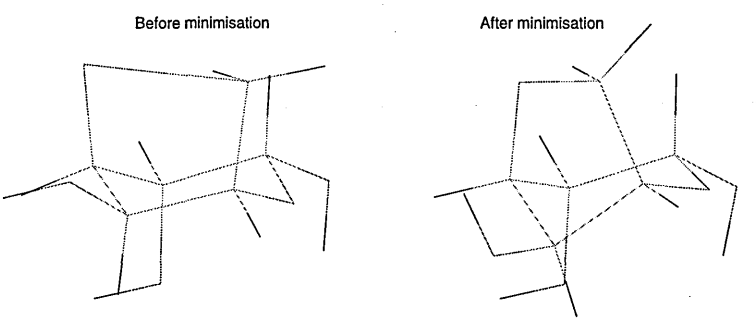


The molecule now contains two rings, both having a hetero–atom: oxygen. It must be minimised before it can be used for building agarose. This is done with the DISCOVER module, setting the minimiser algorithms and number of iterations to be executed. In the present small molecule case the following is likely to produce the global minimum:

```
!
  Minimize
 *    for 500 iterations
 *    using steepest descents
 *    until the maximum derivative is less than 1.000000 kcal/A
!
  Minimize
 *    for 500 iterations
 *    using conjugate gradients
 *    until the maximum derivative is less than 0.100000 kcal/A
!
```

The units 'kcal/A' represent energy per mol per Å.

The 'before and after' states of α-L-anhydro-galactopyranose are shown below:



The energy minimisation has produced a relaxed structure in which bond lengths and angles are as close to their equilibrium values as possible, producing a mutually satisfactory arrangement. In particular the bond between C6 and O3 has shortened dramatically compared to its initial state, and the six membered ring has been distorted slightly to compensate for the new five membered ring (note that two bonds are part of both rings).

2.6 Summary

In this chapter we have seen that molecular modelling is able to contribute to the study of a large variety of molecular species. The basic theory is based on classical mechanics and numerical methods and it is possible to predict the behaviour of complex systems containing several thousand atoms. The accuracy of the technique relys strongly on the quality of parameterisation of the forcefield.

As a simple example an unusual sugar unit was built using libraries of standard molecules and application of molecular mechanics. The sugar, α-L-anhydro-galactopyranose, is a building block needed for the construction of an agarose polysaccharide chain, described in the next chapter.

Chapter 3

Building and validation of models

Contents

3.1 Introduction

Building models and validating them is the next step in the investigation of the agarose polysaccharide. We have seen how experimental results can give insight into what agarose is and the role it plays in foodstuffs and also in pure agarose gels. We have set out the principles behind molecular modelling, the technique which is to be used here in an attempt to throw more light on the nature of the agarose polysaccharide.

Now this polysaccharide chain must be "built", *i.e.* a model must be created in readiness for computer aided experiments. In addition the sulphated carbohydrate *carrageenan* is to be built for comparison, since this is closely related to agarose. In order to build the three main types of carrageenan some new forcefield parameters for AMBER are required. These have already been calculated by other workers and are available in the literature.

The models will need to be *validated*; that is to say, we must ensure that they will behave in the same way that their real – life counterparts do when subjected to conformational changes. The most satisfactory way to validate models is to compare them with experimental data. Two approaches commonly used for this are:

1. Find the global minimum of the subject molecule and compare it to the conformation found in crystals as determined by x-ray diffraction [43, 52, 62]

2. Compare the heights of energy barriers separating different energy minima on the potential energy surface to rotational energy barriers between similar conformations determined by NMR studies [63–68] (these are typically solution studies). Other internal motions can also be probed by NMR.

In principle both of these could be used in the case of polysaccharides, since they are soluble (when boiled in water) and the constituent disaccharides may be expected to crystallise. In fact it turns out that, somewhat surprisingly, very few di- and tri- saccharides can be induced to crystallise and so it is somewhat meaningless to model crystals of these compounds.

In addition, the strongly gelling nature of the polysaccharides hinders solution NMR experiments. For determination of rotational barrier energies the temperature of the solution must be lowered until rotation stops, but while this is happening other re-organisations (i.e.: gelation) dominate. Even the short chain oligosaccharides tend to form syrups in solution (instead of gels), and so are typically too viscous for solution NMR studies.

Therefore in order to validate the models in the present case, a systematic conformational search over the two glycosidic linkages found in agarose (these being responsible for most of the shape changes of an entire chain) was done. The minimum energy conformations thus found were used to generate helices. This procedure is often reported by other workers [49, 69].

The low energy (favourable) conformations of each type of linkage were also compared with the conformations proposed for extended agarose fibres determined by x-ray diffraction [26].

To extend the validation further, the same conformational searching was done for the carrageenans, and the results of all the tests were compared to work done before using other forcefields [70], so that the suitability of the AMBER forcefield could be established.

Finally, it is worth noting here that the potential energy maps obtained by conformational searching can be compared with conformations accessed during dynamics calculations. The maps represent base–line conformations depending only on steric and electrostatic effects, whereas dynamics calculations show conformations which can be accessed when the system is given additional energy in the form of atomic velocities. These results are given in chapter four.

3.2 New forcefield parameters for sulphate.

The AMBER forcefield, as supplied, does not have any atom types or partial charges necessary for the correct description of sulphate groups. The partial charges have been found by *ab initio* studies and the bond force constants from infra red data by other workers [48]. The $-OSO_3^-$ group contains a tetrahedral sulphur atom with three partially double–bonded oxygen atoms and one ether oxygen which links the group to the polysaccharide; see figures 3.1 and 3.2.

In order to model sulphur in a sulphate group, the atom type `S`, which formerly represented sulphur in a sulphide linkage for proteins, was used as it is not required for carbohydrates. The relevant parameters [48] were used instead of the sulphide ones. The `amber.frc` forcefield data file was edited accordingly, and all the new parameters such as bond lengths and angles (see Table below) were added. The file was converted to binary format for use with the modelling software using the `binff` utility.

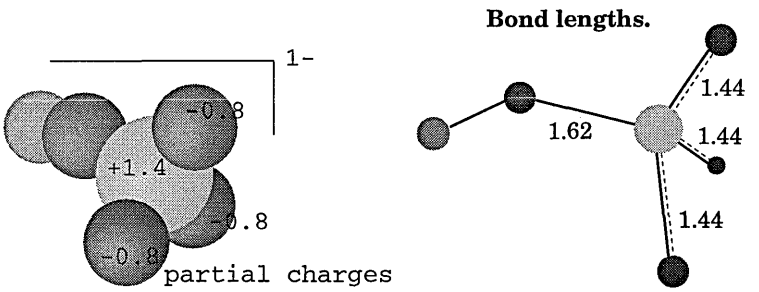
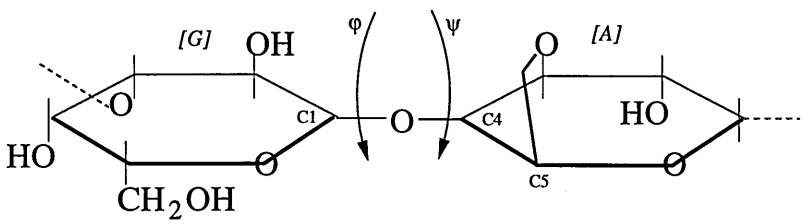


Figure 3.1: The carrageenans contain covalently bound sulphate groups (shown above), for which new parameters were added into the AMBER force-field. On the left are shown the partial charges assigned to the atoms in the sulphate group, and on the right are the equilibrium bond lengths in Å. The green atom represents the carbon to which the sulphate group is attached. Note that the overall charge on the group is only -1.

Parameter	Connected atoms	Values	
		Equilibrium	Force constant
Bonds	S–OS	1.62Å	230 kcal/A
	S–O2	1.44Å	525 kcal/A
Angles	S–OS–CT	120.5°	100 kcal/A
	O2–S–O2	119.9°	140 kcal/A
	OS–S–O2	108.2°	100 kcal/A
	OS–S–OS	102.6°	45 kcal/A

3.3 Glycosidic linkages: systematic searching

All of the glycosidic dihedral angles referred to in this work are defined in the following way: Ring 1[ring oxygen → carbon 1 → linking oxygen] → Ring 2[carbon n → carbon $(n+1)$]. In this way the first four atoms define angle 1 and the second four define angle 2. An example of this is shown below for a glycosidic linkage in agarose. The linkage has directionality since it is made by joining the head of one sugar residue to the tail of the next one. The direction of propagation of the chain follows the vector joining carbon 1 of the first residue to the tail carbon of the next residue.



Path for dihedral angles:
ring-O/C1/bridge-O/C4/C5.

A dihedral angle of zero degrees means that all four atoms concerned in that dihedral angle are equi-planar.

The two possible glycosidic linkages in agarose are those between the anhydro-galactose ring and the galactose ring (denoted link **A–G**) and *vice versa* (denoted link **G–A**); these are quite distinct from one another.

For κ–carrageenan, which has a single sulphate group on the galactose ring, there are similarly two possible glycosidic linkages. The same is true for ι–carrageenan, which has a sulphate group on both rings. λ–carrageenan is somewhat more complicated, because the galactose ring can have one or two sulphate groups. Hence there are no less than four possible glycosidic linkages between pairs of different sugar rings.

A macro for the DISCOVER simulation language (see Appendix A) was written which assigned names to the two torsion angles in each glycosidic linkage so that they could be fixed in space while the rest of the molecule was minimised, but could be incremented by any specified amount in between minimisations.

The searching sequence was constructed using the `constraint_torsionforce` command from within the DISCOVER simulation module. This forces the dihedral angle to the angle required by the current loop of the program. The force constant can be defined in terms of energy; 1000 kcal/mol was used here (the contribution of this is always subtracted from the final energy) which is sufficient to hold the conformation of the linkage to within 0.5 degrees, even in strained molecules.

To produce the three dimensional energy map, a complete revolution of one torsion angle was performed for each value of the other torsion angle, systematic increments through 360 degrees in steps of 12 degrees being implemented. A dielectric constant of $\epsilon = 80.0$ was used to represent an aqueous medium. This results in a smoother surface plot than simulations *in vacuo*, which are rough due to large variations in energy [1].

[1]The actual values of *phi* and *psi* at energy minima were found to be unchanged by the use of different dielectric constants.

The energy of the molecule was minimised using the steepest descents and conjugate gradients methods for every combination of dihedral angles. Possible effects of next - neighbour sugar residues (as in an infinite chain) were not taken into account, as the effect due to such remote residues was assumed to be small. After systematically searching the linkage in question the total energy *versus* dihedral angle combination was plotted, producing a so-called Ramachandran plot[2]. These are often calculated for glycosidic linkages in the modelling of carbohydrates as they give a characteristic three dimensional map for each linkage [49, 71, 72].

The result is a 30 × 30 grid of energy values, which can be viewed in three-dimensional mode using the INSIGHTII analysis module and exported to other graph plotting programs. This grid is fine enough to permit all of the minimum energy conformations to be detected to within a few degrees. If the precise position of the minimum is required then even finer grids would have to be used so that the area around the minimum energy configuration would be explored in greater detail.

3.4 Results of conformational searching

The glycosidic linkages were examined in two sections: firstly kappa and lambda carrageenan and secondly iota carrageenan and agarose. Fibre X-ray data is available for the second two polysaccharides, permitting comparison with experimental values of *phi* and *psi*. Iota carrageenan and agarose glycosidic linkages were therefore examined using more detailed search grids. The four polysaccharides are depicted in figure 3.2.

3.4.1 Kappa and lambda carrageenan

κ–carrageenan contains two glycosidic linkages in the ideal polymer (see figure 3.2):

anhydro-galactose–galactose(sulphated) (or **A–G**)

and:

galactose(sulphated)–anhydro-galactose (or **G–A**)

[2]Ramachandran plots were originally developed for characterising peptide linkages in proteins.

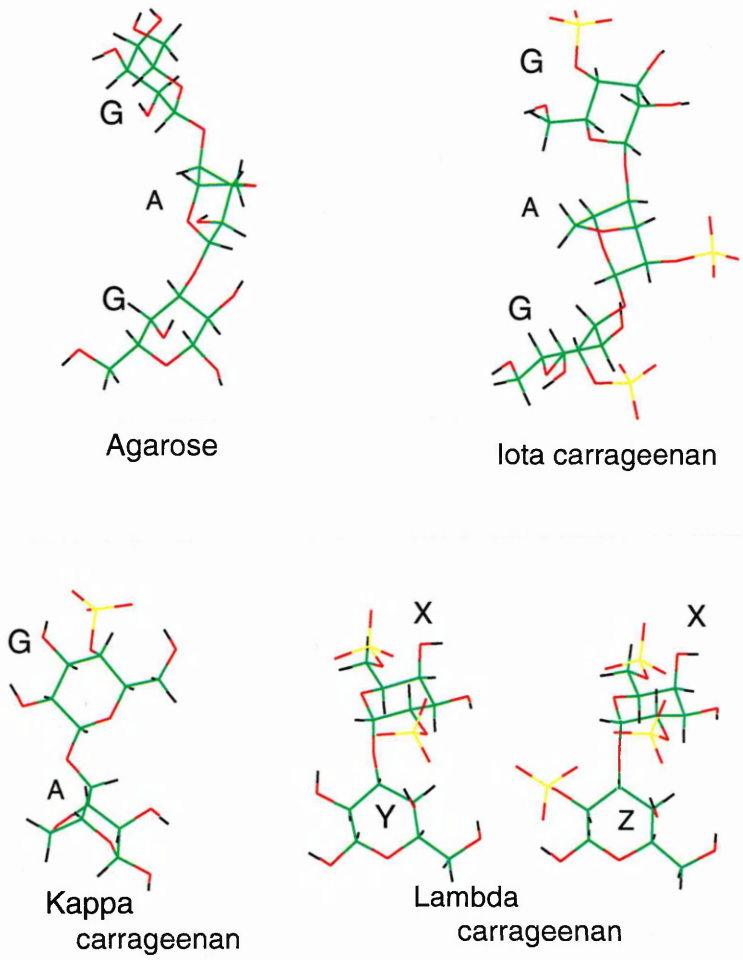


Figure 3.2: Sugar residues found in agarose, and in the three commonly occurring carrageenans. Agarose (top left) has two sugar residues and two glycosidic linkages, **A–G** and **G–A**. Iota and kappa carrageenan also have two different residues, but lambda carrageenan (bottom right) has three different ones, denoted X,Y and Z. (See the text for a description of how they are distributed in the polysaccharide chain.) The major difference between these polysaccharides is the presence or not of sulphate groups, and where they occur.

so two Ramachandran plots are needed. The conditions used were energy minimisation to a derivative of 0.01kcal/A using steepest descents and conjugate gradients, and grid size of 12 by 12 data points.

The results of the grid searches are plotted in figure 3.3. Using the 12×12 array of minimised energies a potential energy surface is generated, which is represented here using contour lines joining points of equal energy.

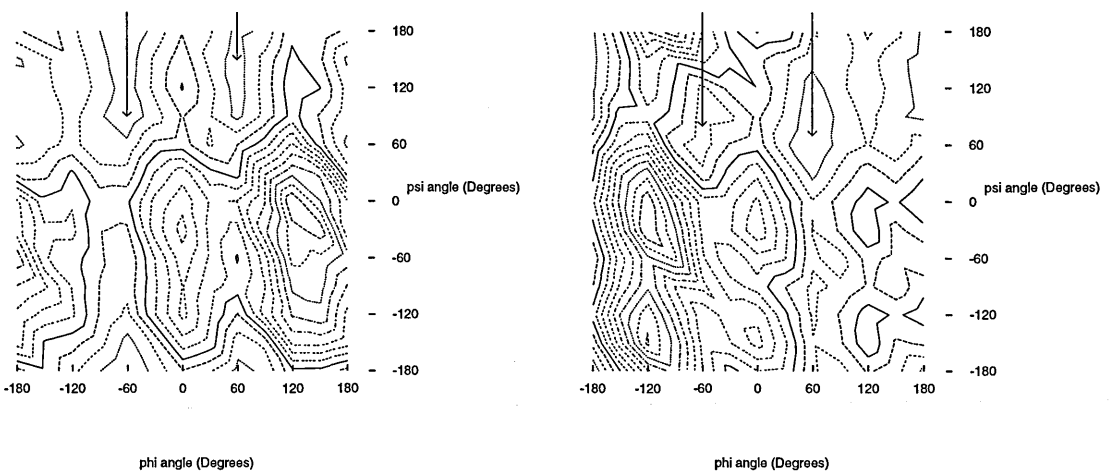


Figure 3.3: Two dimensional contour plots for kappa carrageenan:on the left is the result of a search around the **A–G** linkage and on the right is the **G–A**. The contours are printed every 3.2 kcal, and the arrows indicate the primary minima.

For λ–carrageenan there are four glycosidic linkages, (the three sugar rings are shown in figure 3.2). λ– carrageenan contains three different sugar rings:

- α(1-3) Galactose with no sulphate (Z)
- α(1-3) Galactose with sulphate on O[2] (Y)
- β(1-4) Galactose with sulphate on O[2] and O[6] (X)

The repeating unit always contains ring (X), 30% of the time it contains ring (Z) and 70% of the time ring (Y) [73–75]

The four possible linkages are:

(X)–(Z), (X)–(Y)

(Z)–(X), (Y)–(X)

The conditions used for the conformational search were the same as before. The resulting contour maps are given in figure 3.4.

3.4.2 Iota carrageenan and agarose

For the second pair of polysaccharides a more detailed grid search over (30 × 30) data points was performed, and the minimisation was restricted this time to a derivative of 0.1kcal/A in order to save computer time. The new criteria do not affect the positions of energy minima.

ι–carrageenan contains two sugar rings: (1-4) linked anhydro-galactose with sulphate on O2 and (1-3) linked galactose with sulphate on O4 so there are two glycosidic linkages as with kappa, these being denoted **A–G** and **G–A** for consistency. The resulting Ramachandran plots are reproduced in figure 3.5.

For agar there are also two glycosidic linkages, but no sulphate groups to contend with. The anhydro sugar is a different form to that in the carrageenans: α-L-anhydro-galactose instead of the α-D form. The two links are denoted **A–G** and **G–A** as above.

The results of the conformational searching for the **A–G** and **G–A** linkages in agarose can be seen in figure 3.6. These will be discussed in more detail because most of the thesis deals with agarose, not carrageenans.

The principal low energy conformations for both linkages are given in table 3.1. Note that although values for the dihedral angles are quoted to within one or two degrees, these simply correspond to the point of absolute minimum energy in each "well". The angles can deviate from this by $\sim \pm 15$ degrees without raising the potential energy of the structure by more than 1kcal/A (the lowest rotation barrier separating one energy well from another is about 3kcal/A).

3.4.3 Discussion

The following discussion aims to clarify several issues:

- Justification for using the AMBER forcefield.

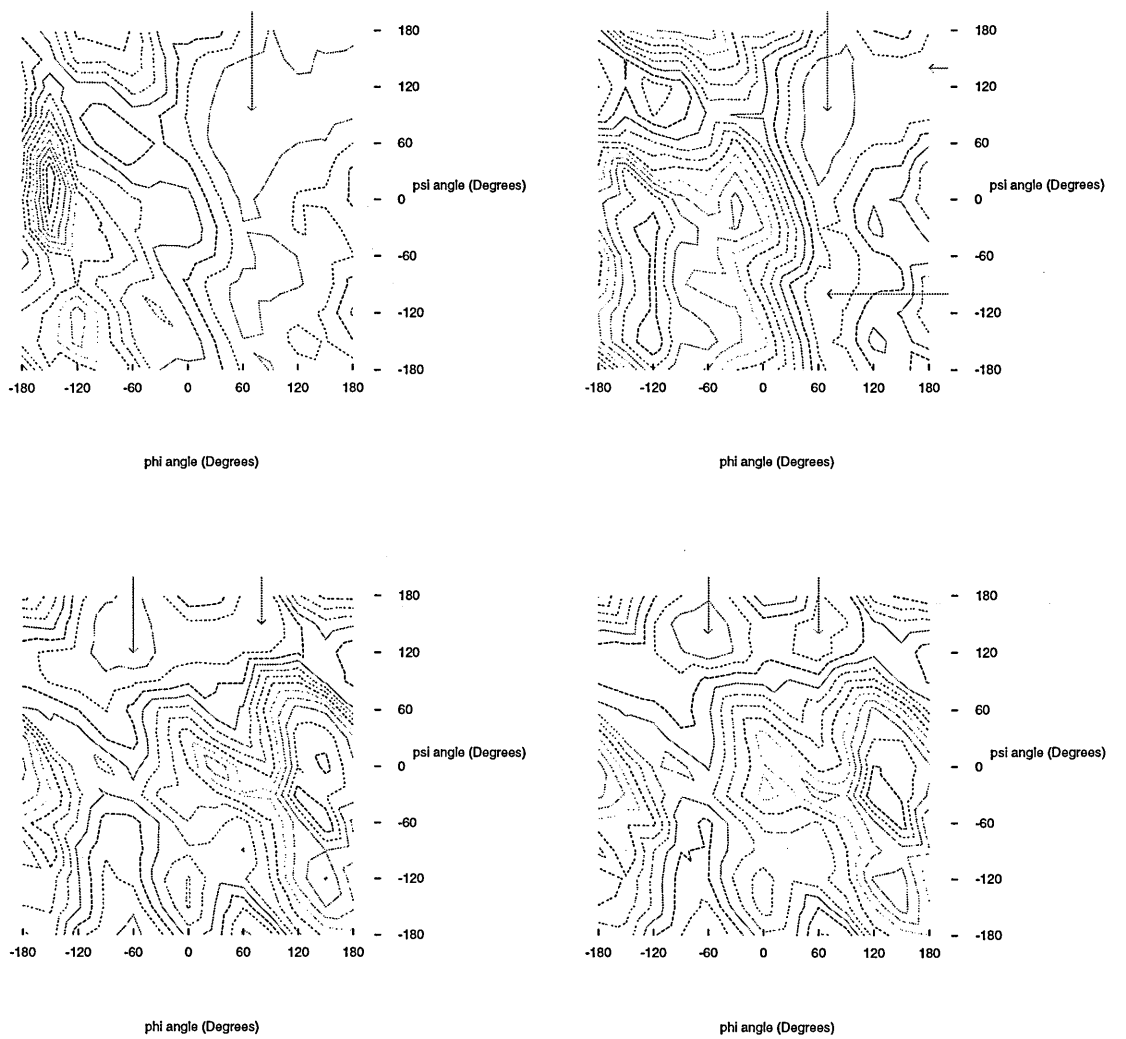


Figure 3.4: Two dimensional contour plots for lambda carrageenan. Top: (X–Y) and (X–Z) linkages, and bottom: (Y–X) and (Z–X) linkages. Note that the extra sulphate groups make the peak values of energy (where groups are clashing) very high. Here, contour intervals are every 3.5, 1.9, 2.6 and 2.6kcal respectively (this variation is a result of the auto-contour function used).

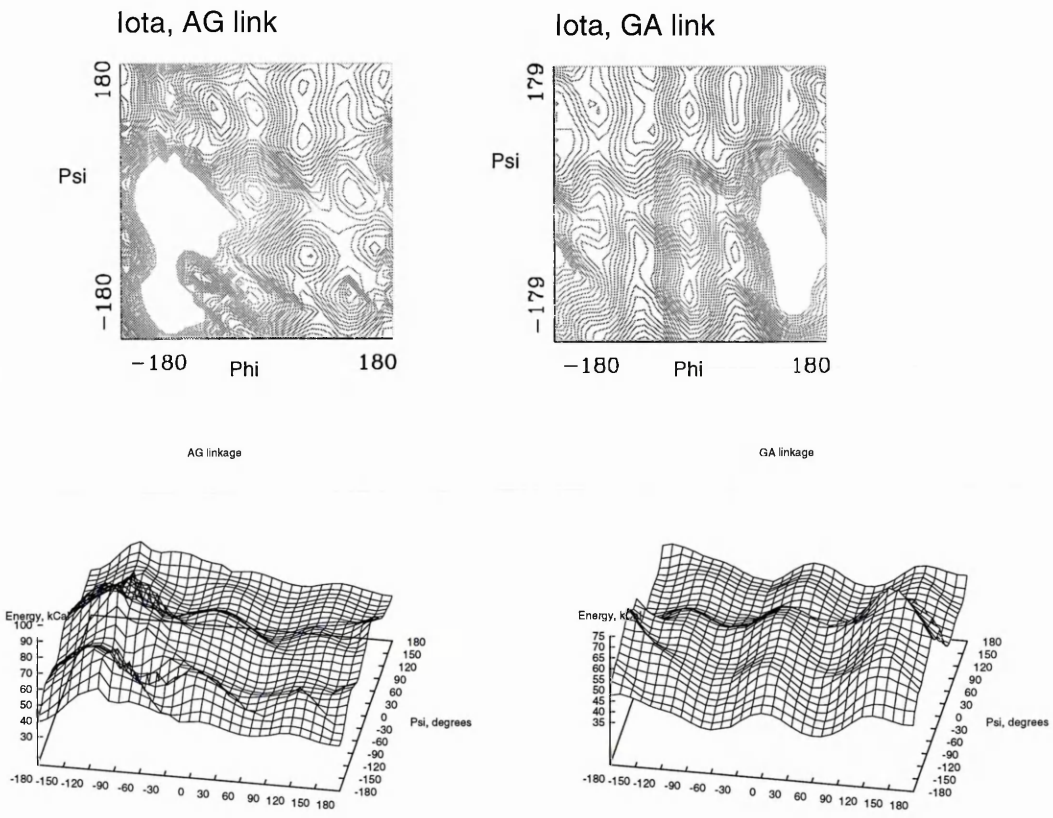


Figure 3.5: Iota carrageenan, **A–G** link (left) and **G–A** (right). The lower pictures are a three dimensional representation of the upper ones, but they have been rotated about the z-axis for clarity. The z-axis represents energy, in units of kcal/A. In the two dimensional plots, contours are printed every 1kcal. Notice that the presence of the bulky sulphate group means that even the lowest 'energy well' is considerably higher in energy than similar wells in the plots for agarose (reflected in the yellow-ish colours).

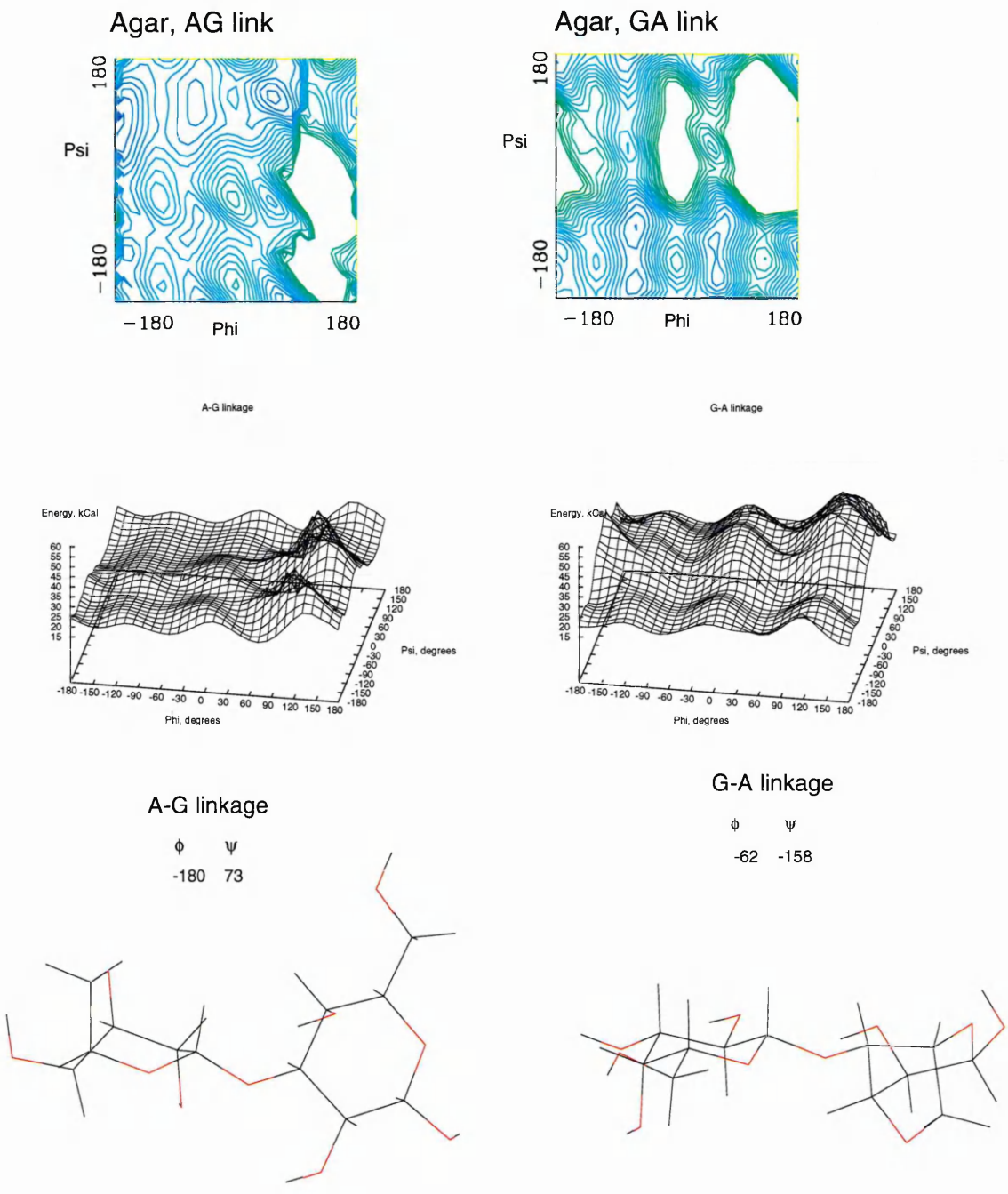


Figure 3.6: Conformational energy maps for agarose glycosidic linkages, with contours printed at 1kcal intervals. On the left is the **A–G** link and on the right is **G–A**. These energy maps characterise the linkages uniquely. Below are reproductions of the minimum energy conformation for each disaccharide.

Linkage	Agarose		Iota Carrageenan	
Primary Minima				
	Phi	Psi	Phi	Psi
A–G link	-180	73	61	87
G–A link	-62	-158	-53	-180
Secondary Minima				
	Phi	Psi	Phi	Psi
A–G link	59	127	60	-70
	-60	115	175	170
	-60	-75	–	–
G–A link	-62	-83	60	140
	60	-95	-70	-180
	60	-150	–	–
Values in oriented fibres				
	Phi	Psi	Phi	Psi
A–G link	-52.2	156.8	77.0	79.0
G–A link	-123.9	-113.2	-87.8	81.0

Table 3.1: Dihedral angle values, in degrees, for which the corresponding disaccharide molecule has the lowest possible energy (the global minimum, here called the primary minimum) and other low energies accessible at room temperature (secondary minima). Note that the list of secondary minima is not exhaustive. The values for the oriented fibres are a result of x-ray determination.

- Positions of energy minima with respect to disaccharide conformation.
- Differences between dihedral angle values for x-ray derived conformations and calculated minimum energy conformations.
- Effect of larger data array (grid size).

Forcefields

The most obvious reason for using the AMBER forcefield is that it has support for hydrogen bonding and carbon atoms in carbohydrate rings. In this respect the forcefield has already been tested, as detailed in chapter 2. In addition however the above Ramachandran plots were compared to identical calculations done using the Universal and Dreiding forcefields [70].

It was found that the Universal forcefield was generally in poorer agreement with the Dreiding and AMBER results, giving very large energies (typically four times greater) and having larger energy barriers between minima. In addition the positions of energy peaks and minima differed slightly. These observations are consistent with the fact that the Universal forcefield is designed for use with a wide range of molecules, so high accuracy cannot be expected. It was therefore quickly dismissed as being of possible use for carbohydrates.

The Dreiding and AMBER forcefields were in much better agreement in that the Ramachandran plots were similar in terms of the range of energies and appearance of the contour map. For a better comparison between them a list of the relevant rotational energy barriers determined by experiment would be useful (nuclear magnetic resonance is a method for finding these). However, no data of this kind have yet come to light after searching in the National scientific literature databases.

Slight differences in the shape of contours may reflect variations in efficiency of convergence during minimisation, since the results for Dreiding were determined using the Cerius modelling package. Here, the energy minima tend to occur when dihedral angles are in the *trans-trans* or *trans-gauche* positions as would be found in a classic propane backbone. This may be due to the incorrect treatment of anomeric oxygens and carbons, and to the partial charges assigned to sulphate groups, which may be too small.

The overall decision therefore is that the AMBER forcefield should be used because it has been tested specifically for carbohydrates.

Disaccharide conformation and energy minima

The potential energy surfaces for the glycosidic linkages display several possible minimum energy conformations.

By inspection of ball and stick Orbit models (Cochranes of Oxford) it can be seen that the minima on the surface plots correspond to positions where groups on the rings, especially those surrounding the glycosidic linkage, are least hindered. In general one can find one or two minima which are lowest of all, and then two or three others which are no more than five kcal/A higher in energy (these could be accessed at room temperature by overcoming the intervening energy barriers). For agarose the global minimum energy conformations are drawn in figure 3.6. For ι–carrageenan the rotation about the **G–A** linkage is less restricted than that about the **A–G** linkage due to the proximity of the sulphate group. Therefore the **A–G** linkage may be expected to be "stiffer".

The minimum energy conformations did not seem to be stabilised by hydrogen bonding between the sugar rings. The biggest effect was steric hindrance, especially for the carrageenans with their bulky sulphate groups.

X-ray derived conformations

It is interesting to compare the values of glycosidic linkage dihedral angles found in agarose and iota carrageenan double helices (x-ray determination of oriented fibres [26, 76]) with those in the minimum energy conformations found by calculation. None of the values coincide exactly, but nevertheless only a slight change in energy is associated with the conversion of the molecular conformation from one to the other. This is reasonable, since in the calculation there are no crystal packing effects. Therefore, in dried fibres the conformations may be a little different from the calculated energy minima.

For the agarose **A–G** link, the closest minimum to the x-ray conformation has a phi/psi value of $-65, 100°$. For the **G–A** link, the closest minimum is a phi/psi of $-60, -72°$. These values appear to be very different from the x-ray derived ones (see table 3.2). The difference however lies in terms of conformation. In terms of energy the difference is less dramatic (table 3.2). For the **A–G** link the x-ray value is a mere 0.8kcal/A higher than the local minimum, while the **G–A** link is 7.9kcal/A higher, which is still not unattainable at room temperature, and other factors which become relevant during formation of fibres (termed 'crystal packing forces') may be sufficient to make one conformation preferable to another within this energy range. When the Ramachandran maps are observed it is plain that the minimum energy conformation is only one point inside a fairly large region ($\pm 40°$) of

	Nearest point to X-ray conformation	Nearest minimum to X-ray conformation
A–G	(-48°;156°) 19.28kcal/A	(-65°;100°) 18.5kcal/A
G–A	(-120°;-108°) 25.78kcal/A	(-60°;-72°) 17.9kcal/A

Table 3.2: Data from the grid search for agarose glycosidic linkages. The degrees given in brackets are the nearest data points to the conformation of interest. From this it can be seen how close in energy are the fibre x-ray conformations to those in relaxed glycosidic linkages.

phi/psi space having a similar low energy, and therefore many conformations can be accessed at only a small energy cost.

These observations have two possible interpretations. Firstly it appears as if the packing forces operative in agarose oriented fibres are sufficient to force the glycosidic linkage conformation away from the most favourable minimum energy positions, at a small energy cost.

Secondly however the forcefield may be reproducing the energy map incorrectly, but examination of 3-dimensional models seems to confirm that the energy map is at least roughly as expected.

Grid Sizes

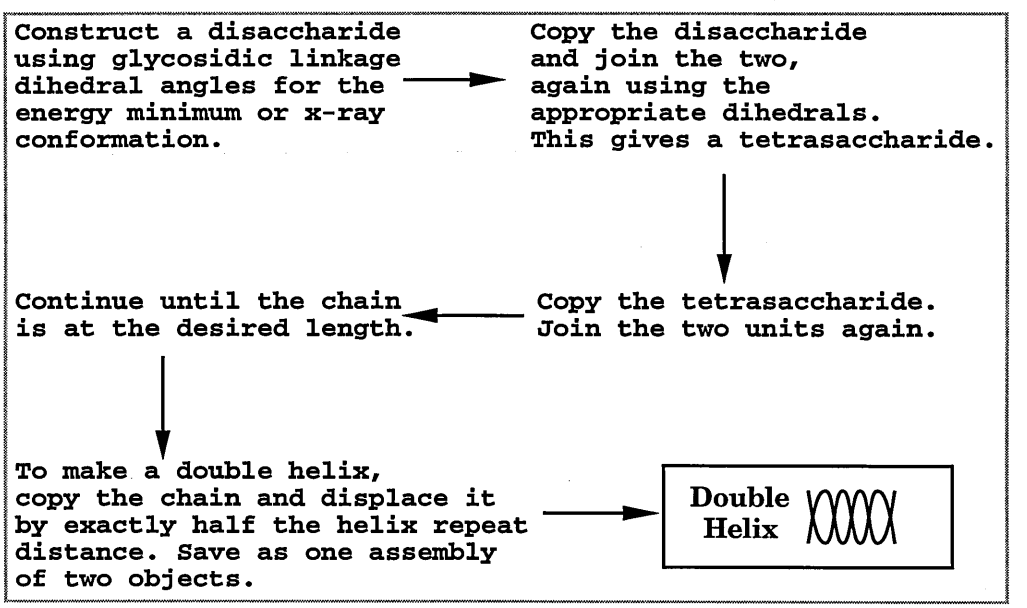
Use of a finer grid results in a smoother and more detailed energy surface (compare the graphs for iota and kappa carrageenan). Compared to the coarse grid, the minima can be identified more accurately because there is more information on the surface texture. The small rotation steps of the fine grid search allow more opportunity for the conformation to adjust as the calculation proceeds. Minimum energy regions are therefore explored more carefully. Hence it is always preferable to employ fine grids, computer time permitting. The 30 × 30 grids shown here took about an hour to produce, depending on system loads, which is quite acceptable. The minima given by finer grids are in non-trivial positions (see table above), reflecting more accurately the influence of bulky and/or charged groups nearby. For coarse grids some detail is inevitably missed out. The actual accuracy level specified for minimisation seems to have relatively little effect on the surface obtained. The positions of minima are the same, although actual energy values obtained

are different if better minimisation criteria are specified.

3.5 Construction of helices

In order to gain further insight into the meanings of the low – energy conformations calculated above, various helices were constructed for agarose using dihedral angles corresponding to energy minima and to experimentally determined fibre conformations. These latter angles should of course produce the same helix as predicted by x-ray fibre diffraction.

The method used for building helices is simple, and is outlined below, starting from readily available sugar units and specific dihedral angles at the glycosidic linkages:

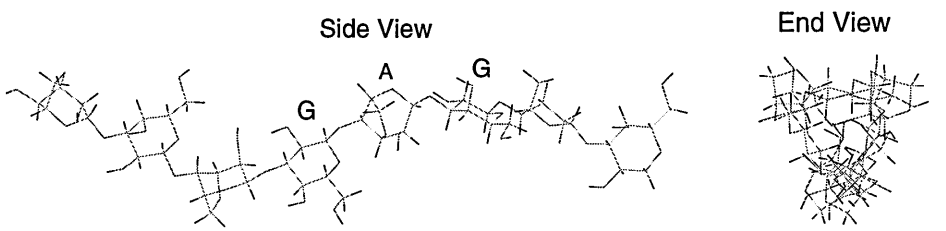


The sugar rings are restored using the INSIGHTII commands `fragment_get` or `molecule_get`, and the torsions can be specified using the `torsion_modify` command. After construction the helices were saved in Biosym archive format; they can then be loaded into the molecular mechanics and dynamics modules. The helices were also converted to `.pdb` format files, which can be read into the molecular viewing programs "RASMOL" and "MOLDEN", to enable further observation.

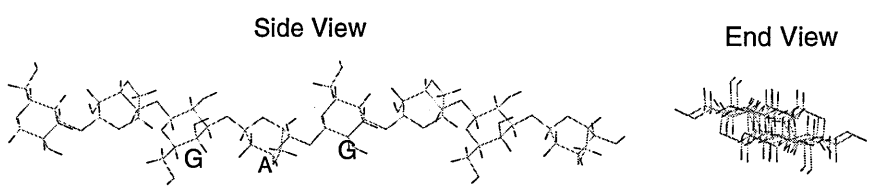
The resulting chain shapes were noted and compared to the known x-ray structure, and are described below.

Agarose : four combinations were made using the calculated minimum energy conformations at glycosidic linkages, and one was made which reproduces the x-ray prediction for the helix.

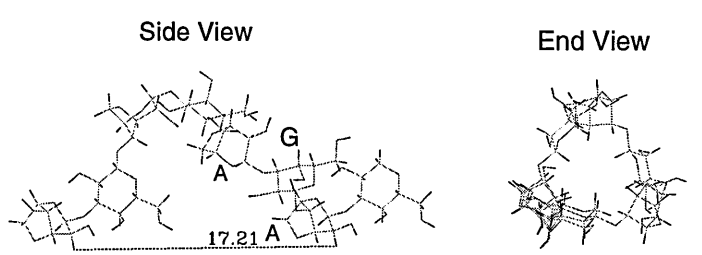
1. **A–G** (-180,73)/ **G–A** (-62,-158)
   This produced a very thin helix with no core and a non-integral repeat:



2. **A–G** (-180,73)/ **G–A** (-62,-83)
   A very flat helix, more like a zig-zag (repeat of two units). It is not possible to form a double helix with structures like these, and in any case the repeat distance is not compatible with the x-ray data.



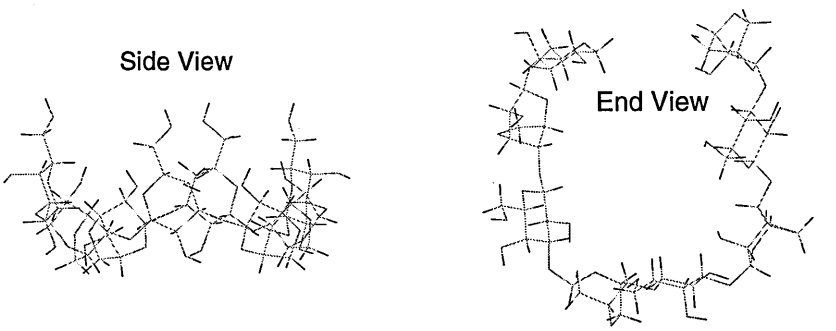
3. **A–G** (59,127)/ **G–A** (-62,-83)
   A wider helix with the repeat distance (one complete turn) being about 17Å:



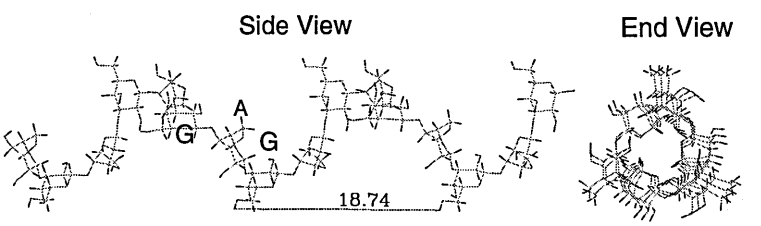
   This helix is a possible alternative interpretation of the x-ray data, as it also has a three-fold axis.

4. **A–G** (59,127)/ **G–A** (-62,-158)

This made a flat tetrasaccharide ring! Obviously any chain longer than eight sugar residues would not be able to adopt this shape without having some kind of bend in the chain, otherwise the ends would clash.



5. The values of dihedral angles most compatible with the x-ray data are: **A–G** (-52,156)/ **G–A** (-123,-113). These are not at minimum energy conformations, although as mentioned above they are close. They produce a helix with a repeat of 18.8Å and there are three disaccharide units per turn:



Iota : Two combinations were made, and then one using the values predicted by x-ray diffraction.

1. **A–G** (61,87)/ **G–A** (-53,-180)
   This produced an extended chain.
2. **A–G** (61,87)/ **G–A** (60,123)
   This made a left handed helix (the actual one is accepted as being right handed [76]) with just under three disaccharide units per turn.
3. X-ray values: **A–G** (77,79)/ **G–A** (-87,81). The helix made using these has a repeat distance of 21.2Å and there are three disaccharide units per turn. (The x-ray data for this polysaccharide is very good.)

The building exercises show that it may be risky to assume that helices built using only minimum energy conformations found by grid searches, even detailed ones, are the be-all and end-all conformations which exist in nature. However it is an obvious approach to take when faced with such a complex problem, *i.e.* the prediction of the structure of carbohydrates. Many workers have taken this route [43, 44, 75, 77].

The method does not always hit on regular helices, because the calculated results invariably produce a helix shape having a non-integral number of building units per repeat distance. For agarose however, one helix discovered in the above calculations has a repeat distance of 17Å with three disaccharide units per helical turn (see number 3 above). There is no reason to assume that this is not compatible with the fibre x-ray data. Given its similar dimensions (axis repeat) and symmetry it could well be another interpretation for the diffraction pattern of agarose helices; in fact it is conceivable that there is more than one helix anyway. However since there seems to be more evidence for the currently accepted x-ray structure, this will be taken to adequately represent 'the agarose helix' and will be the basis of future calculations in this thesis.

Depicted in figure 3.7 is a view of the double helix, thought to be most compatible with x-ray diffraction patterns of oriented agarose fibres. These structures are thought to exist in the gelatinous phase also [26].

3.6 Summary

Two questions form the essence of the work in this chapter.

1. Is the model used to interpret the x-ray data correct? Are the packing forces in agarose and iota carrageenan fibres really sufficient to force the glycosidic linkage conformations away from the predicted energy minima?

2. Is the forcefield accurate enough? Possibly the energy maps are incorrect.

Either or both of these may apply. It is almost certain that there are packing forces in agarose fibres, so the conformations of glycosidic linkages in them will be different to those found in isolated disaccharides, as seems to be confirmed by the calculation results.

So the forcefield is probably reproducing the energy map correctly. Indeed, examination of 3-dimensional models seems to confirm that the energy

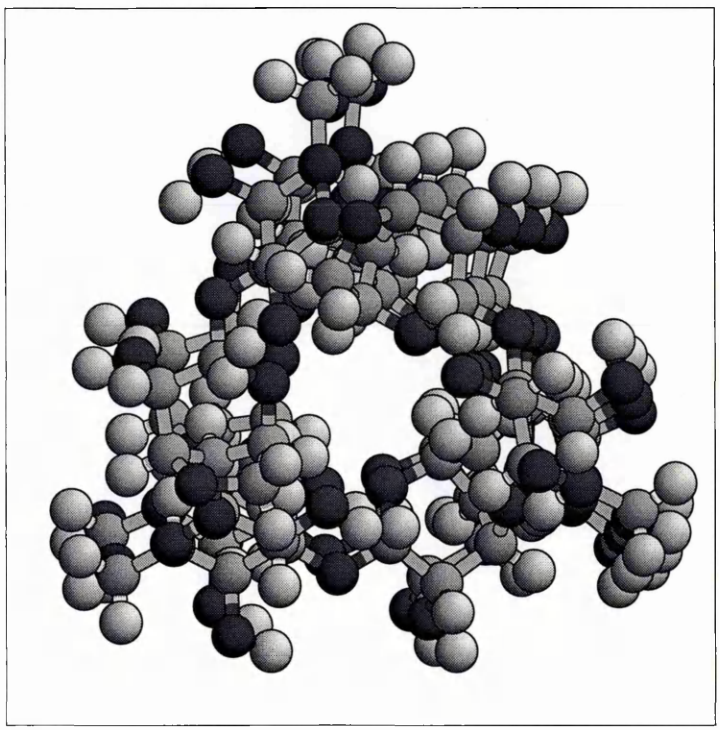


Figure 3.7: The agarose double helix, viewed from the end. The internal core can clearly be seen.

map is at least roughly as expected. The forcefield should reproduce relaxed conformations well enough because it has been parameterised and tested for carbohydrates. It is more likely to be inaccurate in representing strained molecules, such as carrageenans with bulky sulphate groups which are clashing as the glycosidic linkages are rotated, so we should be more cautious about the "mountainous" parts of the Ramachandran maps. Fortunately these are of little importance in the present investigation.

From this work it can be said that environmental factors (solvent effects and inter-chain charge-induced forces) in packed arrays of helices clearly have the potential to favour alternative glycosidic linkage conformations when helices form in nature. It should be noted that *more than one* helix may exist in nature. The helix building exercises described above indicate that another three-fold helix can be built using different glycosidic dihedral angles to those assumed in the fibre x-ray structure.

These factors are still to be investigated by research groups working in the area, and most of this thesis is aimed at exactly this problem. At any rate no data are yet available for agarose in solution and so what conformations are favourable in this case remain to be seen in chapter 4, using calculation methods.

Chapter 4

Modelling of solvent

Contents

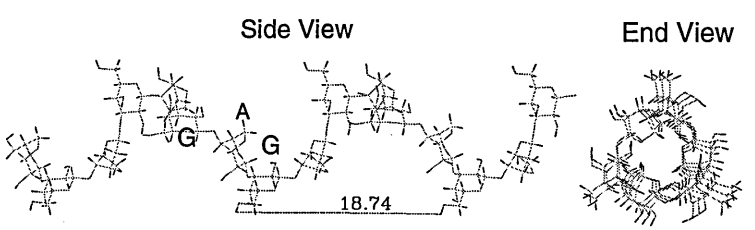
4.1 Introduction

The need to understand more about agarose in the solution state has already been outlined. The only structural data available so far has been obtained using x-ray diffraction of agarose double helices, and these were in the form of dried fibres drawn from the gel phase. This chapter aims to investigate firstly how aqueous solvent should be represented when modelling agarose by computer, and secondly how the carbohydrate itself behaves when in solvent, as opposed to a dried gel.

Molecular modelling techniques using computers are unique in that molecules can be examined directly and spatial relationships between groups observed at first hand. The best alternative approaches available are nuclear magnetic resonance (NMR), where chemical shifts and coupling constants permit information to be obtained about the structure [78], and x-ray diffraction, which gives a mathematical description of the atomic coordinates [79][1].

Molecular modelling allows conformational preferences to be elucidated and processes which normally occur almost instantaneously to be observed. Most investigations to date using this method have been concerned with very small parts of the agarose system; disaccharides or tri-saccharides at most [69]. Here, some calculations on agarose hexa–saccharides are presented.

As detailed earlier, agarose is composed of the alternating sugar units 4-linked β-D-galactopyranosyl (**G**) and 3-linked 3,6 anhydro-α-L-galactopyranosyl (**A**) [80]; see below:



It has been postulated that the double helices of agarose can contain water molecules inside the helical core (*i.e.* parallel to the helix axis) which may assist in holding the carbohydrate chains in their helical shape [26, 81]. In this chapter we propose some doubt regarding the likelihood of this.

[1]Note that only heavy atoms can be detected with any certainty, since the technique relys on the positions and number of electrons in the molecule.

In addition the simulations add to a general understanding of the characteristics of solution behaviour for short chains of agarose, and these characteristics may have implications for models of extended agarose polymers.

4.2 Simulations on agarose solutions

In molecular modelling one is able to represent the solvent using either explicit water molecules, which have their own masses and partial charges, or a representative homogeneous bulk dielectric constant which replaces the solvent, simply representing its relative permittivity. The two methods are compared here.

The starting conformations for dynamics calculations of the agarose double helical fragments were taken from x-ray diffraction data of pure agarose double helices in the form of drawn fibres. Sections of the required size were pruned out of a long molecule corresponding to these data.

Resulting dynamics trajectory data were analysed for the behaviour of dihedral angles, and these were compared to the conformational searches performed on the same angles in chapter 3. The results show that when explicit water molecules are used the oligosaccharide chains are "stiffer" and have a reduced mobility.

4.2.1 Conditions for mechanics and dynamics

For the molecular dynamics (MD) experiments the initial agarose helical fragments were relaxed in two stages, with a final convergence criterion of less than 0.1kcal/molÅ.

```
Minimize: steepest descents for 500 iterations
          conjugate gradients for 500 iterations
```

The minimisation was followed by dynamics calculations (1 iteration is 1 femtosecond):

```
Dynamics: warm to 350K in twelve stages of 200 iterations each;
          calculate a further 100,000 iterations at 350K
          write history file every 100 iterations

Cutoff: 16 Angstroms (switching over a distance of 1.5)

Relative Permittivities: 1.0 (explicit water)
                         80.0 (no specific water molecules).
```

The two simulations were run using exactly the same initial assembly of molecules: two strands of carbohydrate polymer 'pruned' from the double helix conformation, each strand having three anhydro-galactose rings and three galactose rings. This makes a total of 240 atoms.

The first simulation employs a cell containing water molecules at a density corresponding to water at room temperature and pressure (the TIP3P water model [82]), with the double-stranded carbohydrate (the solute) placed at the centre. In this way the water molecules contribute a further 200 atoms to this simulation, as well as all of the associated bonding interactions.

The second simulation employs a cell containing the same carbohydrate molecules but a dielectric of 80.0 instead of specific water molecules was used. This value has been used by other workers when representation of the electrostatic screening effect of water molecules is required for modelling of neutral polysaccharides in solution [83–85].

4.2.2 Periodic boundary conditions

In both cases periodic boundary conditions (PBC) are applied to simulate a bulk solution case. Strong, elastic gels can be formed from agarose solutions at a concentration of just one or two percent by weight [37]. If all the chains were six residues long, this concentration would mean that one molecule would occupy on average a volume of 1 million cubic Angstroms!

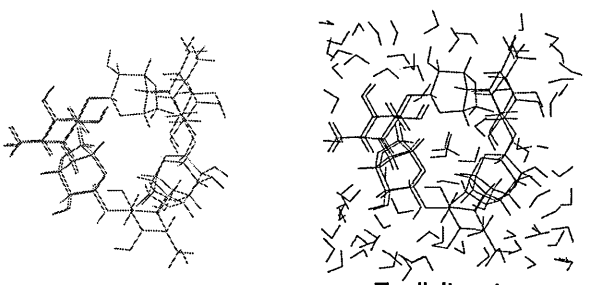
A cell of $1 \times 10^6 \text{Å}^3$ is far too large for simulation purposes, so the cell dimensions need to be reduced to a manageable scale. The choice was made as follows:

Ideally we require as many explicit water molecules as reasonably practical (with regard to the computing resources available) to be fitted in. Therefore the cell will consequently be as large as possible given these restraints.

For the case when specific water molecules are specified, the cell is made large enough to enclose the solute plus two layers of water molecules. Hence, sufficient room between the solutes of adjacent cells is created for water molecules to diffuse during the simulation. The dimensions in this case are $21 \times 15 \times 15$ Å^3, and the system occupies a large proportion of the computing resources on the workstation when calculations are proceeding. The number of water molecules increases according to the cube of the increase in cell side length, so the system can easily be too large to perform calculations on. The size chosen is a good compromise. The same dimensions were used for the implicit water simulation for consistency, even though the absence of explicit water molecules would actually enable a much larger cell to be used.

It should be noted here that the ratio of carbohydrate to water in such a cell as detailed above is in fact equivalent to an extremely concentrated solution, or to a region within the solution in which molecules are very densely packed. This mimics the composition of *aggregated particles* found in solutions and *junction zones* in gels, where the density of molecular packing has been observed by birefringence measurements to be very large at certain points [27].

The two completed assemblies, ready for simulation, are depicted below:



4.3 Results and discussion

It was anticipated that in the implicit water simulation the lack of specific hydrogen bonds between solute and solvent and the absence of the mass which accompanies specific solvent molecules, would make a significant difference to the behaviour of the carbohydrate molecules.

The AMBER forcefield represents hydrogen bonds (in the non-bonding energy expression

$$\sum \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]$$

involving –OH or –NH groups [47, 55]) as covered in chapter 2, and this has the effect of stabilising interactions involving hydroxy groups which are within 2 or 3Å of one another. Therefore a 'hydrogen bond' is created which can persist for one or two pico-seconds. The average vibration period of a real hydrogen bond is 0.2 picoseconds (ps) [86].

Several observations can be made concerning the properties of solvent and solute in the explicit water simulation. These will be discussed in the next few subsections. There then follows a section which attempts to put some values on a property which is quite difficult to measure: stiffness of chains.

4.3.1 Solvent and solute hydrogen bonding

In the explicit simulation, water molecules are seen to interact with hydroxy-groups on the chain. All of the hydroxy-groups on the chain start in close proximity to water molecules except for O(2) on the galactose ring which points inside the double helix cavity (although this group becomes exposed to solvent later on). All of these groups are able to form hydrogen bonds to water molecules at any time during the simulation (see Figure 4.1 for some images), as they are constantly in close proximity to them. Sometimes during conformational movements of the carbohydrate chain the associated water molecules were "dragged" along for a picosecond or so of time.

In other situations a water molecule was able to bind for a period of the order of picoseconds (ps) with two -OH groups around an **A–G** glycosidic linkage: this 'bridging' behaviour could be expected to slightly reduce the freedom of the linkage, making the carbohydrate chain as a whole "stiffer".

Obviously, in the implicit simulation most of these features were absent. In particular, no bridging of glycosidic linkages could occur and the conformational movements of the chain were not hindered by bound water molecules.

4.3.2 Solvent structure

Water molecules (when present) are constantly making and breaking structures within their bulk, this being a consequence of their strongly hydrogen–bonding nature. Such structures are observed to take the form of water molecules arranged in clusters where three or four molecules are able to hydrogen bond to each other simultaneously, with the arrangement persisting for a few picoseconds while non interacting water molecules nearby diffused away. This clustering is a short – range effect only involving a few molecules. Although the modelling software does not permit water molecules within the bulk assembly to be individually highlighted, figure 4.1 shows snapshots of the entire assembly during the simulation. The animated sequences (not given here due to lack of space) demonstrate small pockets of water molecules retaining their positions relative to each other for one or two picoseconds, then diffusing apart.

Again, this cannot happen in the implicit simulation.

4.3.3 End effects

Ends of carbohydrate chains have extra -OH groups and have more opportunity for interacting with other moieties. The extra -OH group on a terminal

sugar residue means that this residue can be better solvated. This fact, and the fact that the end residue has more conformational space available than one inside a chain, may explain why ends of chains are the starting point for unravelling of double helices.

This is most noticeable in the implicit water simulation, where the two chains unravelled in just 10 picoseconds, and the ends of the chain moved to and fro much more than the inner portion.

4.3.4 Water molecules in the helix cavity

Apart from the bulk solvent generated by the solvation program, extra water molecules were placed inside the helix cavity (*internal waters*) at the start of the calculation. These were observed to escape during "breathing modes" displayed by the helix loops. These breathing modes, when helix loops move alternately apart and then back together, have a periodicity of the order of tens of picoseconds. During this cycle, two adjacent helix loops separate enough to permit a water molecule to pass out of the helix cavity. It is statistically much less likely that the reverse process of a water molecule becoming trapped inside the helix will occur, because there are many fewer ways of achieving it compared to the great freedom of an external water. Another way of saying this is that an internal water has less entropy than an external one. Therefore if water molecules are to move to the inside of agarose helices there must be a sufficiently favourable change in *enthalpy* to keep the water molecule inside once it gets there. However, inside the agar double helix there are very few -OH groups, making the cavity hydrophobic in character. It is therefore difficult to see how such a favourable enthalpic change could arise. So, although there may be internal water molecules trapped in crystalline or fibrous forms of agarose, they do not seem to be part of double helices in solution. For this reason, internal waters are not included in future simulations.

4.3.5 Carbohydrate hydrogen bonding

Intra-molecular hydrogen bonding is observed intermittently during both simulations. Common arrangements are hydroxy-groups interacting on the same ring, for example O(6)H$\cdots$O(4) on galactose rings. Hydroxy-groups on rings can also interact with glycosidic bridging oxygen atoms, and these arrangements were observed to persist for up to one picosecond. Hydrogen bonds were not observed from one chain to the other while the double helical arrangement persisted. In order to achieve inter-strand hydrogen bonding

the helix has to be disrupted. This is because most of the –OH groups are positioned on the external surface of the helix.

Therefore in the explicit water simulation, where the helix remained essentially complete until near the end, no inter-chain hydrogen bonding occurred. In the other simulation however, the helix was disrupted after only 10 picoseconds. After 70 picoseconds different non-helical arrangements had established themselves, and here inter-chain hydrogen bonding could occur. Compare the images from both simulations in figure 4.1.

4.3.6 Solvent acting as lubricant

No lubrication of the flexible chains was obvious when observing the animated dynamics trajectory for the explicit water simulation. For this to happen, water molecules need to hold the Van der Waals surfaces of the carbohydrate apart, and so relieve the roughness and reduce the electrostatic forces between them. In this simulation only the ends of the chains were separated enough for water molecules to move between the strands (see figure 4.1). Lubrication may occur when chains are fully solvated at temperatures of over 80°C (in the random chain configuration), or it may be that the simulation needs to be continued for a much greater length of time before the chains become fully solvated. As indicated in chapter six, this latter situation is likely to be true.

4.4 Discussion: stiffness of chains

One of the primary differences between the two simulations is that of chain stiffness. In the current context, a stiff chain is one with greater restriction in its movements compared to a flexible chain. This means that stiff chains show less variation in their glycosidic dihedral angles.

Here, stiffness is described qualitatively by observing movements in the trajectory replay, and quantitatively by extracting dihedral angle information.

4.4.1 Observations

Graphs of total energy variation with time (figure 4.2) for both simulations show that evolution of the systems has settled into a stable state, with respect to energy, after five picoseconds. To compare the differing behaviour of the carbohydrate chains in each simulation, refer to figure 4.1. The simulation without explicit water displayed the greater structural variation and movement.

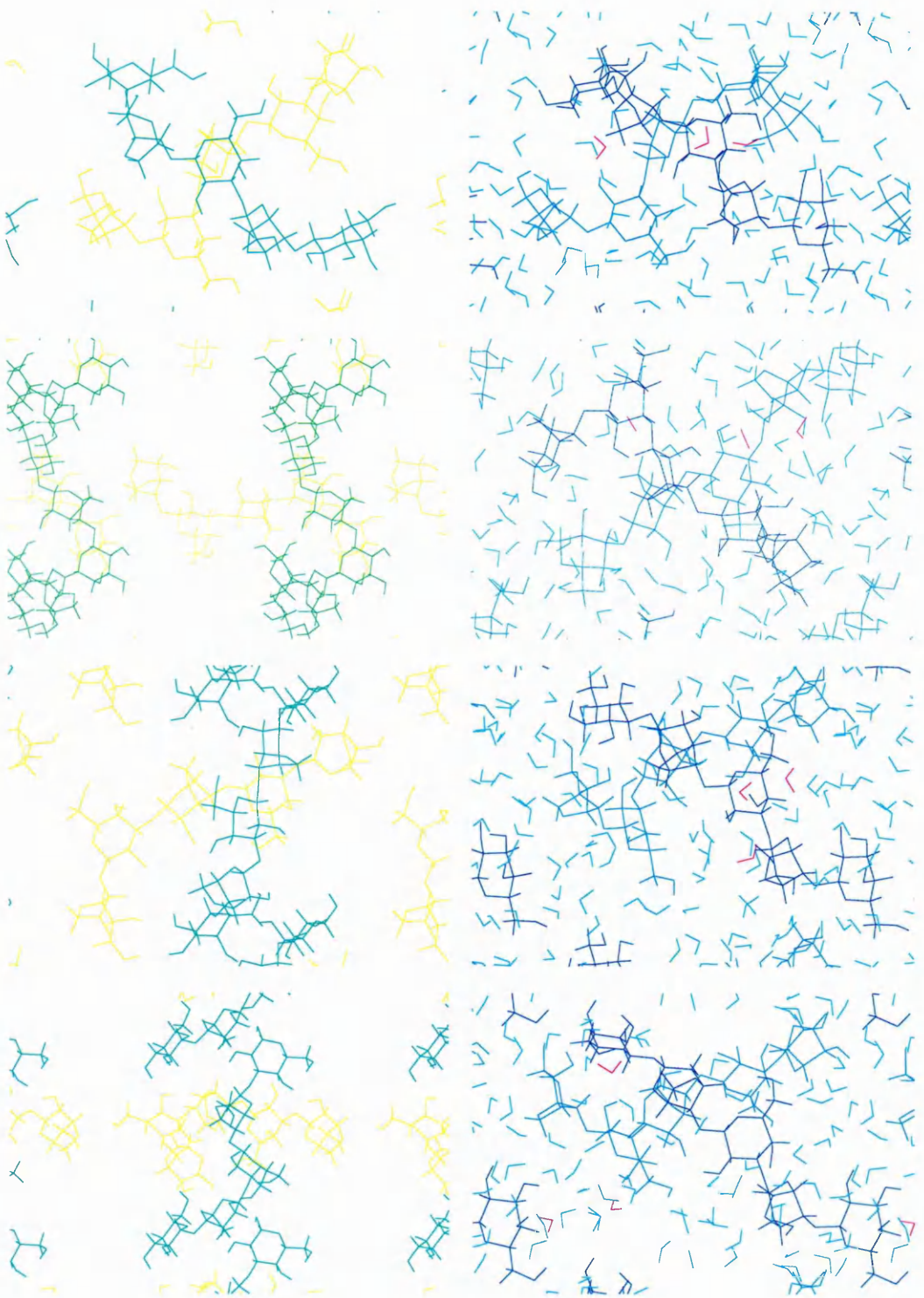


Figure 4.1: Trajectory snapshots from both simulations. On the left is the implicit water simulation at 0, 20, 50 and 100 picoseconds. On the right are the corresponding explicit water snapshots.

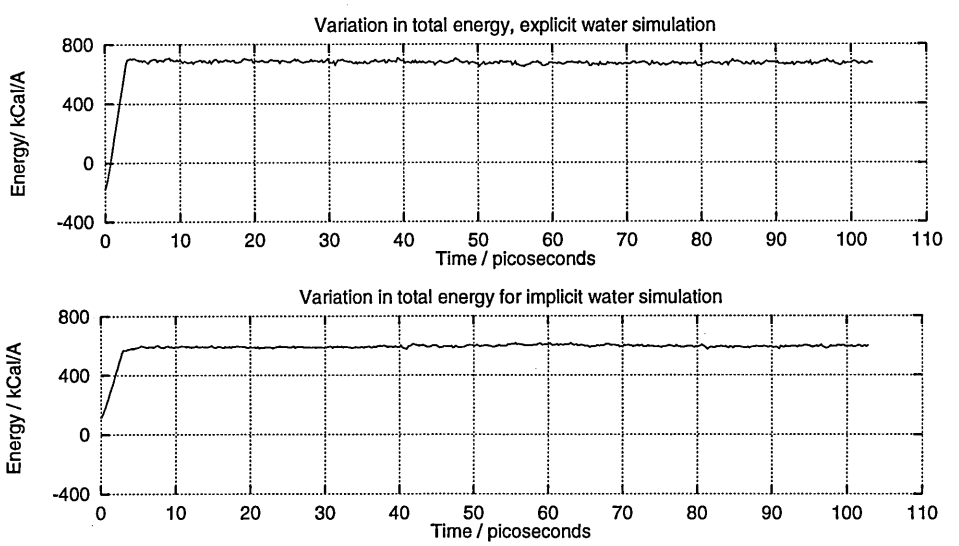


Figure 4.2: Variation in total energy during simulations. Top: simulation with explicit water and dielectric of 1.0. Bottom: simulation with no water molecules and dielectric of 80.0. Note that the total energy for the explicit water simulation is greater, due to the extra atoms present. Extra atoms mean extra atomic interactions and so more contributions to the total energy. However the overall, or mean, energy is stable.

Implicit simulation

The total energy of the carbohydrate molecules in the implicit water simulation varies between 570 and 610kcal/mol (see figure 4.2). Most of this variation is due to changes in non-bonded interactions and torsional energies as the molecules move.

Taking the lower value of total energy as the case where all the dihedral angles are relaxed at once (in the bottom of their potential wells), we can postulate that the variation in energy for one dihedral angle must be approximately $\frac{1}{10}$ of the energy range (40kcal), since there are a total of ten glycosidic linkage dihedral angle pairs in the system. This gives a variation of 4kcal/mol per dihedral angle pair, at 350K. This is necessarily something of a simplification, but gives "ball–park" figures. Occasional variations in the energy of any particular dihedral angle pair of more than 4kcal/mol are bound to occur every so often. These occasional large variations in energy can allow a change from one potential well to another or an exploration significantly away from the minimum position. Indeed it can be seen from the plots in figure 4.3 that the barrier between minimum potential energy wells is traversed a few times during the 100ps simulation.

In the implicit simulation we can observe unwinding of the two chains followed by complete dissociation and some diffusion of the resulting isolated species.

After minimisation the original regular helix–like conformation has already been able to relax significantly, even though the dihedral angles have only rotated slightly. Upon warming up to 350K the two molecules are seen to unwind over a period of less than 10ps ($\frac{1}{10}$ of the simulation); this is clearly a rapid process. The result is that the velocity of molecule 1 is roughly equal in size and opposite in direction to that of molecule 2.

Following the disruption of the double helix there is a drift of molecule number 1's against number 2's in the array (this being a consequence of using periodic boundary conditions), with minor changes in molecular shape constantly occurring.

After 70ps of dynamics, molecule number 1's collide and begin to interact with number 2's by partially enveloping each other (see figure 4.1). Of course, due to the PBC applied, this is repeated in every cell. This interaction continues until the end (another 30ps) and a longer calculation may show that it can last for even greater periods of time.

The chains are constantly altering their shape between extended, arc-shaped and coiled forms. The final situation is one of pairs of molecules, but not in a double helical conformation as at the start. Nevertheless it can be

seen that on the 0.1 nanosecond time-scale the preference is for stable dimers of molecules rather than a completely dispersed solution.

Explicit simulation

For the simulation involving explicit water molecules the variation in total energy is greater than that seen above due to the large number of extra atoms. Unfortunately it cannot be analysed readily because of the random contributions from the many interactions involving solvent, so the dihedral angle information is lost in amongst the general 'noise'.

Both chains remain together for the entire calculation; main trend to be observed is a slow loosening of the helical loops, and the escape of water molecules from inside the helix loops (these are shown in purple in figure 4.1). The conformational freedom of the glycosidic linkages can be seen to be reduced compared to the case when no water is included.

Comparing the two simulations it is readily apparent that the lack of specific water molecules has a large effect on the diffusion rates of the two carbohydrate chains and hence the time taken for the double helical fragment to be disrupted: compare the corresponding images in figure 4.1.

4.4.2 Population of energy wells

Using statistical thermodynamics the favoured conformations at glycosidic linkages can be better understood. More flexible chains should be able to populate different energy wells on the Ramachandran energy maps more often, while a stiff chain can be said to be effectively at a lower temperature, so could not populate higher energy wells so easily.

In the present context there are two energy maps to consider: the one for the **A–G** linkage and the one for the **G–A** linkage. Both of these have a point of global minimum energy plus two or three other low energy points. These points are separated by energy barriers, and will be referred to as *energy states* hereafter. The *Boltzmann distribution* for i different energy states gives the number of molecules n_i in each state as:

$$\frac{n_i}{N} = \frac{e^{-\frac{1}{kT}\varepsilon_i}}{\sum_j e^{-\frac{1}{kT}\varepsilon_j}},$$

where N is the total number of molecules, k is the Boltzmann Constant, T is the absolute temperature and ε_i is the energy of state i relative to the lowest energy state. The sum is over all states j.

The fraction n_i/N therefore gives the population in state i.

Linkage	Energy state (J)	Population at 350K
A–G	0	(two states) 49.670%
	2.082×10^{-20}	0.650%
	4.16×10^{-20}	0.001%
	5.55×10^{-20}	*negligible*
G–A	0	74.000%
	6.94×10^{-21}	17.600%
	1.39×10^{-20}	(two states) 4.10%
	4.16×10^{-20}	0.001%

Table 4.1: Statistically most likely populations (by percentage) to be found in the low energy states of each glycosidic linkage at 350K. Of course these are not discrete states; there can be conformational (and energetic) fluctuations about these positions.

Taking each energy map in turn:

- The **A–G** map has five energy states with relative energies of 0, 0, 3, 6 and 8 kcal/mol respectively.
- The **G–A** map also has five energy states with relative energies of 0, 1, 2, 2 and 6 kcal/mol respectively.

All of these states can be populated at room temperature, but to varying degrees: see table 4.1.

Comparing this information with the dihedral angles actually observed in a dynamics calculation (figure 4.3), it is immediately obvious that such a short calculation does not represent a good statistical sample of the relative populations in each conformational energy well. For this to be done, thousands of molecules would have to be sampled for several nanoseconds at least, in order to allow chances for energy barriers between wells to be traversed. Therefore stiffness of chains cannot be easily ascertained just by using dynamics calculations to predict how different conformations are populated. Instead, the amount of movement within energy wells can be compared for different simulations, and this is the approach taken in the next section.

4.4.3 Dihedral angles and correlation functions

In order to examine stiffness in detail some method of quantification is required. As mentioned before, the configurations of glycosidic linkages (dihedral angles) determine chain shape and stiffness. So, the dihedral angles must be monitored. Dihedral angle information was therefore extracted from

the trajectories and plotted onto Ramachandran energy maps (generated in chapter 3), so that the most visited parts could be identified. Here though, instead of looking at relative populations of energy wells as in the previous section, the actual amount of phi/psi space explored will be directly compared for both experiments. These overlay plots will hereafter be referred to as scatter plots.

Scatter plots of dihedral angles visited for the explicit water case show a more restricted behaviour (less phi/psi space explored) than that seen with implicit water: see figure 4.3. The variation in total energy is however similar[2] as shown in figure 4.2. This can be interpreted as a difference in stiffness of the chains.

To provide some measure of this "stiffness factor", normalised correlation functions have been calculated [66, 87] for dihedral angles in both cases. The statistical relation for the correlation coefficient between two variables $\mathcal{A}$ and $\mathcal{B}$ is (the subscript 'ens' denotes a statistical operation over the whole ensemble):

$$c_{\mathcal{A},\mathcal{B}} = \frac{\langle \delta\mathcal{A}\,\delta\mathcal{B} \rangle}{\sigma(\mathcal{A})\,\sigma(\mathcal{B})}$$

where

$$\sigma(\mathcal{A}) = \left[\langle \mathcal{A}^2 \rangle_{\text{ens}} - \langle \mathcal{A} \rangle_{\text{ens}}^2 \right]^{\frac{1}{2}}$$

and

$$\delta\mathcal{A} = \mathcal{A} - \langle \mathcal{A} \rangle_{\text{ens}}.$$

These formulae were evaluated for each data point using a short FORTRAN program. The results for each simulation (two glycosidic linkages in each case) are given in figure 4.4. The same trends were displayed by all other dihedral angles.

All the plots begin with large fluctuations in the time correlation function and tend towards zero later on. The increased stiffness of the carbohydrates in the explicit water calculation is reflected in a faster decay of the correlation function and a smoother trace value as time progresses towards infinity. In the plots from the implicit simulation variations in the time correlation function continue to manifest themselves up to the end of the data.

4.5 Summary

From the above dynamics calculations it is seen that whatever method of solvent representation is used, the helical nature of an agarose chain is toler-

[2]Of course, the total energy itself is different since different numbers of atoms are involved.

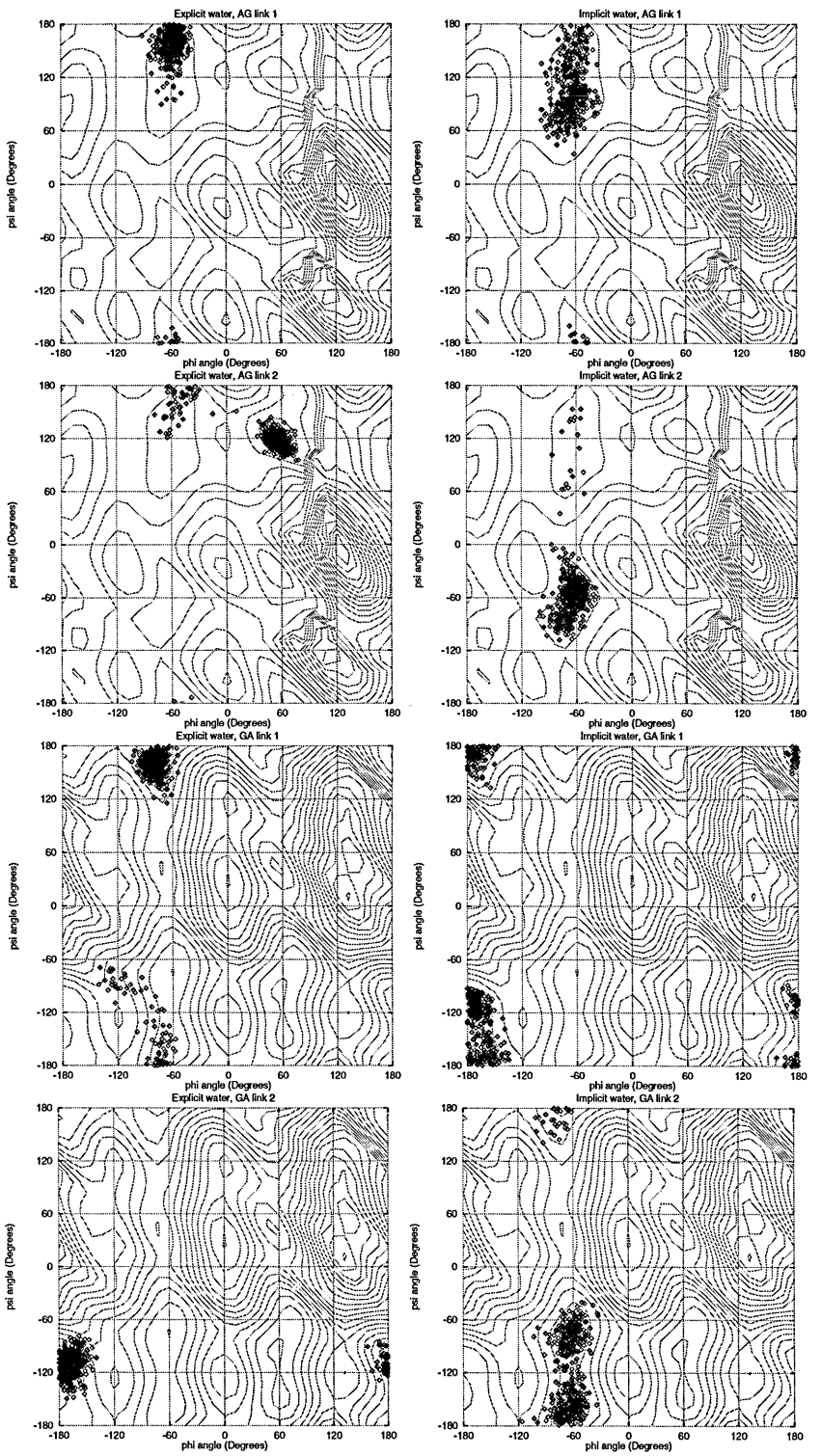


Figure 4.3: Scatter plots of dihedral angles visited (black points) superimposed onto Ramachandran energy maps for the isolated glycosidic linkages. On the left are two examples for each linkage type for the explicit water simulation, and on the right are the corresponding implicit water results. Contour intervals are 1.3 and 1.2kcal for the **A–G** and **G–A** linkages respectively. The data for the implicit water simulation is more scattered, indicating that more conformational space has been explored.

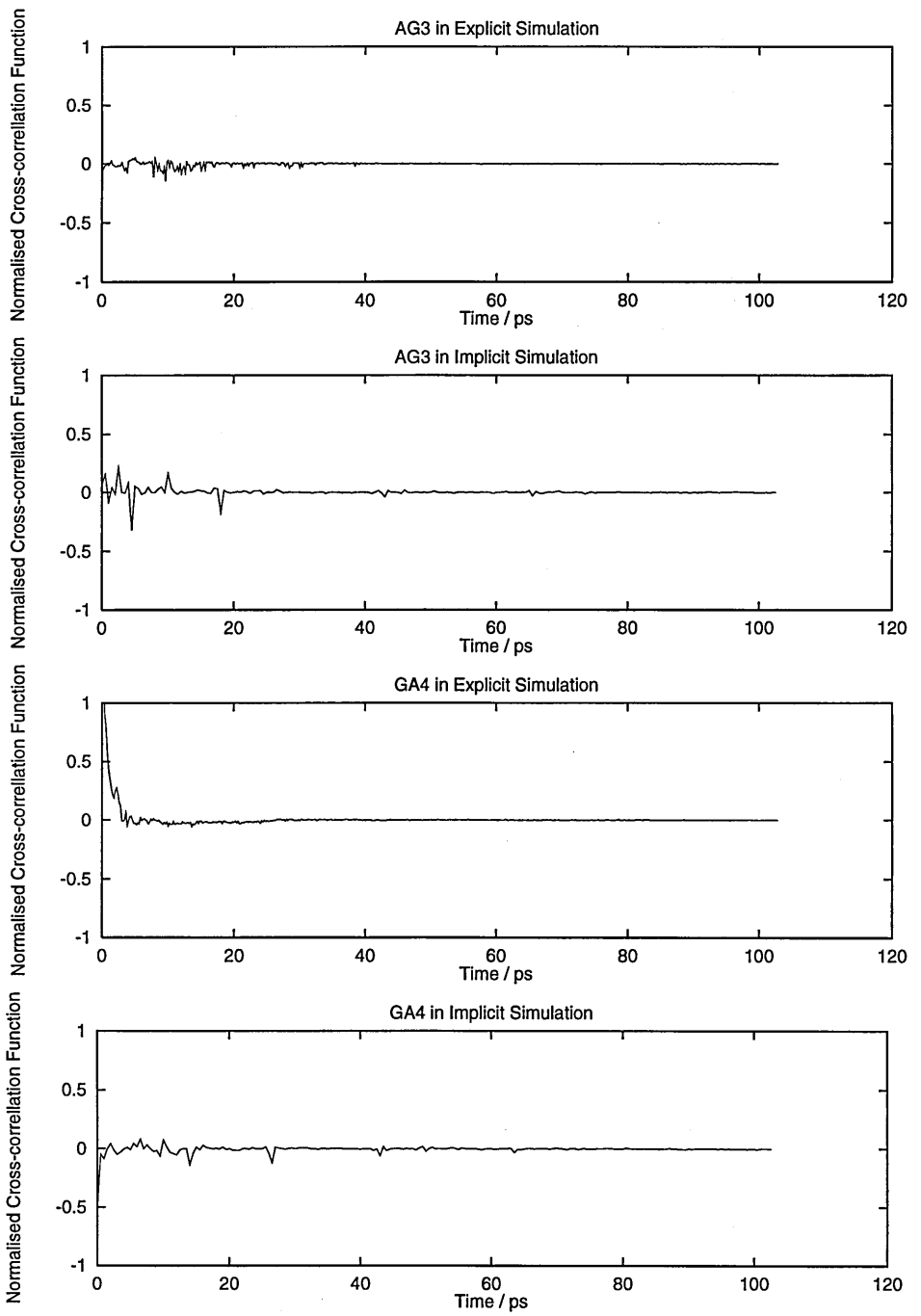


Figure 4.4: Normalised correlation functions for dihedral angles. Top two graphs: for **A–G**; bottom two graphs: a **G–A** linkage. For method of calculation, see text. The feature of interest is that of visible smoothness of the two plots. Those for the explicit water calculation are smoother and converge towards zero more quickly indicating less variation in the dihedral angles and a stiffer chain.

ant to small changes in the glycosidic linkage dihedral angles without major disruption to the main chain shape.

For the short chains simulated here helices are definitely not stable in the long term (for more than 100ps). It is simply too improbable that all the torsions will be in the correct double helix configurations at the same time.

During dynamics calculations of small chains using the implicit water representation the "perfect" double helical conformation is soon lost as all the dihedral angles are independently varying and indeed moving from one minimum energy configuration to another. In addition the rings themselves can flex slightly. In explicit water where the presence of solvent molecules hinders chain movement the same process takes much longer. It has been demonstrated that the chains are stiffer when explicit water is used, and diffusion rates are reduced.

For long agarose chains, helices may be stable due to other factors:

- Overall, as the chain length increases the enthalpy of dimerisation will eventually be enough to favour association of separate strands, possibly into the helices proposed for strong gels and oriented fibres.
- For long chains there is a much lower proportion of terminal sugar rings, termed "ends", in relation to the number of internal sugar rings. Ends are more mobile, so short helices are more likely to be disrupted.
- Water molecules assist by binding to the outside of helices, encouraging their association and restricting their flexibility to some degree. For longer chains this could be the origin of a *hydrophobic effect*, which encourages the association of two chains.
- Although the conformational space for the dihedral links is centered mainly about one or two minima, significant flexibility of the chains is possible by twisting the linkage by 10 or 20 degrees away from the minimum energy conformation. This allows parts of each chain to adjust their conformation independently without having to unwind large parts of the assembly.

The difference between the full explicit water treatment and the dielectric of 80 manifests itself in reduced conformational flexibility (less space explored), and slower relaxation and tumbling of molecules when the solute of interest is surrounded by water molecules. In other words the system takes

much longer to evolve if the full solvation treatment is specified when modelling. Increased mobility of chains due to lubrication by solvent molecules appears to be very much smaller than the loss of mobility due to reduced diffusion rates, when the two dynamics calculations are compared. Although the distribution of conformations about energy wells during the explicit water calculation are slightly different to the distribution about the same energy wells for the implicit simulation, the difference in energy terms is so small that the effect on the overall outcome is minimal.

Given sufficient time both assemblies are likely to evolve to the same stage reached by the implicit water simulation; that of full helix disruption. However this cannot be simulated with current computing resources because of the time-scales involved: it takes a minimum of several milliseconds for real agarose helices to dissolve even at 100°C (see chapters 5 & 6).

If longer chains were to be used we should expect that the balance of enthalpy, ΔH, which favours dimerisation *versus* entropy, ΔS, which favours dispersion into solution would shifted to higher temperatures. With small molecules the entropy term favours total dissolution, while for a long chain the total enthalpy of the chains would favour dimers in a helical form. The present simulation confirms the small molecule case. Ueda *et al.* [83] found that longer chains of the related polysaccharide β- carrageenan formed relatively stable dimers on a time-scale of 550 picoseconds. We will return to this argument at several points in the following chapters.

Natural agars have considerable amounts of sulphate and methyl substitution on O(2) of the anhydro-galactose ring and/or O(6) of the galactose ring: these groups can be expected to affect dimerisation enthalpies by virtue of their bulk and charge. Some highly methylated agars are known to have elevated melting temperatures [88] indicating a large enthalpy of dimerisation or bigger junction zones. However small random amounts of substitution are more likely to disrupt junction zones. The general argument concerning enthalpy *versus* chain length is still valid, but predicting the effect of charged groups on water structure and molecular conformation is not trivial, and presents a considerable challenge for all biopolymer modelling scientists.

Chapter 5

Laboratory support work

Contents

5.1 Introduction

As previously indicated, the investigation of agarose gel structure by molecular modelling methods presents formidable demands upon computational resources. Molecular dynamics in fact cannot be executed for systems which adequately represent all long range structural features in gels.

Some molecular properties may however be probed by the use of laboratory techniques. The primary feature to be investigated in this respect is the agarose temperature dependent *random coil to aggregated double helix* transformation, which we shall call ordering.

The work was carried out at UNILEVER Research, Colworth House, Sharnbrook, Bedfordshire and is described in this chapter.

The first part of the work involves preparation of small agarose molecules. Starting with agarose polymers (the pure "ideal" form of agar[1] [80, 88]), the preparation of oligosaccharides in a range of lengths by acid hydrolysis is described. Hereafter in this text, a length of agarose chain will be referred to as being of a certain *degree of polymerisation*, or Dp. A Dp.17 chain for instance is 17 sugar residues long. The results of high performance anion exchange chromatography (HPAEC) and mass-spectrometry (Ms) reveal the distribution of chain sizes (Dp.'s) obtained. Specific samples containing known lengths of chain (monodispersed samples) were then prepared using gel permeation chromatography (GPC).

Secondly, using these monodispersed samples, tests were done in order to investigate correlations between chain size and ordering. The first approach is to simply put agarose oligomers through a freeze/thaw cycle to see which chain sizes precipitate.

A more quantitative physical test on individual agarose chain sizes then follows, and is described in the second half of this chapter. Using differential scanning calorimetry (DSC), agarose ordering transformations are followed directly and enthalpy changes measured. The same transformations may be part of gelation of the agarose polysaccharide system itself.

The results are finally used to estimate energies of association for different length chains, and to propose some conclusions about agarose itself.

[1] "Native" agar contains an amount of sulphate and also random substitutions of methyl groups

Run	Conditions	Results
1	80°C; 0.2% agarose; 0.5M acid	The polymer was completely hydrolysed in less than one minute
2	80°C; 0.2% agarose; 0.05M acid	The polymer was completely hydrolysed in under 25 minutes.
3	80°C; 0.2% agarose; 0.01M acid	After 20 minutes the sample was found to contain a selection of oligomers
4	80°C; 1.0% agarose; 0.01M acid	The optimum range of oligomers was found in the sample after about one hour of hydrolysis.

Table 5.1: Summary of testing to find optimum agarose hydrolysis conditions.

5.2 Preparation of agarose oligomers

In the first part of this investigation a mixture of oligomers was obtained from agarose using hydrolysis under acidic conditions. It was expected that after hydrolysis the reaction mixture would contain a huge range of chain lengths, from mono-saccharides right up to non-hydrolysed agarose polymer, given the right conditions.

5.2.1 The hydrolysis of agarose

The objectives for the best hydrolysis were a reasonable time for reaction (about 30 minutes), with the production of as many chain sizes as possible. Ideally oligomers in the size range expected to be around the minimum needed for aggregation were required. Samples were periodically withdrawn from the hydrolysis mixture and analysed by HPAEC after neutralisation and cooling, to monitor the progress of hydrolysis. The full procedure is outlined below, and a summary of the development process is given in table 5.1.

5.2.2 Optimisation of agarose hydrolysis

Agarose powder (1.0g, SIGMA chemicals) was dissolved in 50mℓ water at 100°C. 50mℓ of 0.02M sulphuric acid at 80°C was added and the mixture shaken gently in a stoppered flask. The reaction was then incubated for one hour at 75°C in a water bath.

After this time about 10mℓ of 0.1M NaOH was added and the pH adjusted to be close to 8. The solution was allowed to cool to room temperature, when cloudy precipitates formed[2]. The resultant liquid was stirred with de-ionising resin beads (about 5g), frozen in an acetone/dry-ice (−80°C) mixture and dried under vacuum overnight. Yields were in excess of 95%.

5.2.3 Anion exchange chromatography (HPAEC) of agarose

HPAEC uses two eluents which are mixed to form a continually changing concentration of acetate ions ([-$OCOCH_3$]). More details on the operation of the HPAEC are given in Appendix B. The acetate gradient is able to sequentially elute oligomers of increasing size.

The two eluents were in the machine were:

- 100mM NaOH (Eluent A)
- 100mM NaOH + 500mM NaOAc (Eluent B)

A simple gradient of 4.4mM/minute made by combining these was found to be effective at separating agarose oligomers. A single analytical run took 35 minutes, with the sample being injected at $t = 0.1min$. See figure 5.1 for a typical output graph.

The mass of unresolved peaks before $t = 3$ minutes represents the short chain oligomers. Thereafter the peaks were clearly resolved. Samples corresponding to two peaks were collected and sent away for Ms analysis. Once these two had been identified, the assignment of all other peaks in figure 5.1 could be done.

5.2.4 Assignment of HPAEC analysis

Mass spectrometric characterisation of material from two well - resolved peaks was used to assign the HPAEC analysis. The masses of potential fragments were calculated as shown in table 5.2. ('Gal' denotes the β-D-galactopyranose ring, and AGal the α-L-anhydro-galactopyranose ring.)

[2]All samples were filtered through 45μm filter tips prior to HPAEC analysis to prevent damage to the chromatography column by colloidal particles.

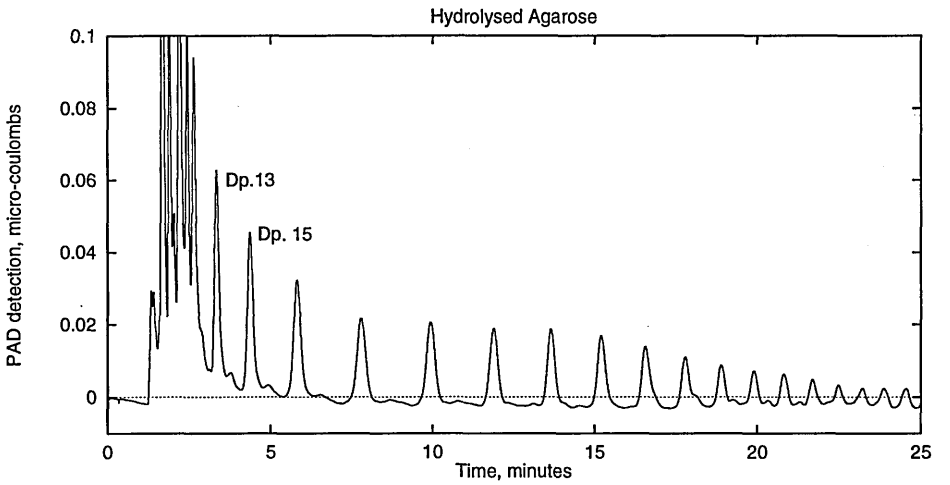


Figure 5.1: HPAEC analysis of hydrolysed agarose, where "PAD" refers to the pulsed amperometric detector output on the chromatography instrument. All hydrolysis preparations were checked to ensure that they gave this distribution, and therefore contained the same range of chain lengths.

Fragments	**Masses (Daltons)**
Sodium	23
Potassium	39
AGal sugar ring (inc. H_2O)	162
Gal. sugar ring	180
Disaccharide unit (2 rings and 2 glycosidic links):	306

Peaks observed in first sample (estimated to be Dp. 12)	Peaks observed for second sample (estimated to be Dp. 16)
2040	2348 (7%)
	2653 (75%)
	2959 (11%)
	3268 (7%)

Table 5.2: Some agarose fragments and their masses. Top: potential fragments and calculated masses; bottom: peaks actually observed in the mass spectra. All masses are given in Daltons.

Analysis

There are three possible structures for products of the hydrolysis, if it is assumed that either glycosidic linkage can equally well be hydrolysed.

1. $(\text{Dimer})_n$: A series of even sugar unit numbers.

   Mass: $[310]_n + 18$

2. $(\text{Dimer})_n$ + Gal: A series of odd sugar unit numbers where the extra unit is Gal.

   Mass: $[310]_n + 162 + 18$

3. $(\text{Dimer})_n$ + AGal: A series of odd sugar unit numbers where the extra unit is AGal.

   Mass: $[310]_n + 144 + 18$.

A study of the numbers obtained shows that the oligosaccharides are a series with odd numbers of sugar units, where the extra sugar is Gal. (scheme 2 above). For example, taking the second sample in table 5.2:

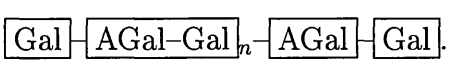
$$\text{Expected mass (Dp.16)} = (306 \times 8) + 18 = 2466\,\text{Daltons}$$

Actual value is 2653 Daltons, according to the spectrum. Thus the discrepancy is

$$\begin{aligned} 2653 - 2466 &= 187 \\ &= 162 + 23\,\text{Daltons} \end{aligned}$$

This is the mass of Gal (less one water) plus sodium. The fragment is therefore Dp.17 plus a sodium ion. The sodium ion could have been picked up during the neutralisation of sulphuric acid with sodium hydroxide after hydrolysis.

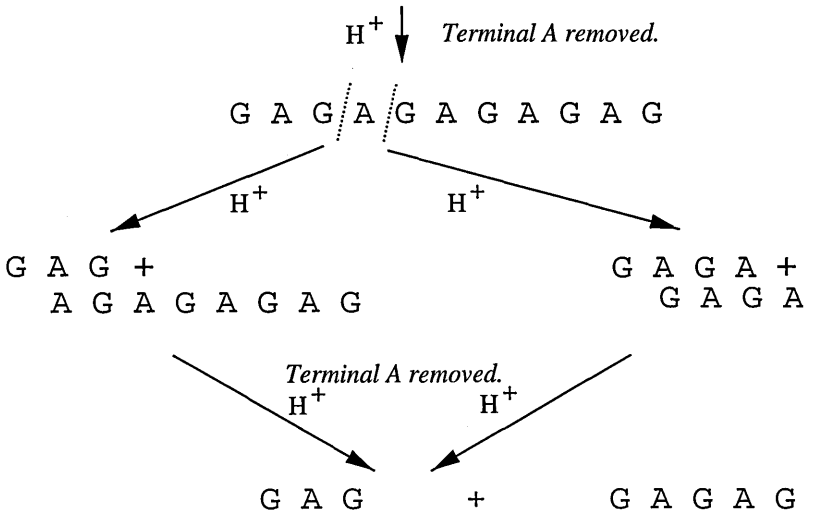
Thus any oligosaccharide fits into the general form



The hydrolysis therefore clearly has two parts, since we have the cleavage of an agarose polymer at any glycosidic linkage followed by hydrolysis of any terminal anhydro-galactose sugar rings. Hydrolysis of a chain such as

G-A-G-A-G-A-G-A-G-A-G

may proceed *via* the following scheme (here, A represents the anhydro galactose ring and G the galactose ring). Note that two-step hydrolysis at any point on a chain gives two odd–numbered chains where the extra unit is galactose:



Observations

- Ms is so sensitive that even very small amounts of potassium salt impurities can be seen. Therefore some "doubling" of peaks is seen, with the separation being 16 Daltons.
- In the second GPC sample extra peaks separated by 306 Daltons can be seen. This difference corresponds to a single dimer unit and the extra peaks are impurities from neighbouring GPC fractions. Therefore the Ms data is consistent with a series of oligosaccharides in which every successive member has an odd number of sugar residues and differs by one dimer unit from the previous member.
- Ms confirms observations made using HPAEC analysis on fractions from the GPC column (described in section 5.3); *i.e.*: as longer oligosaccharides are eluted they tend to be less well resolved. In Ms, this results in more peaks being detected in the long chain sample.

5.3 Gel permeation chromatography (GPC)

Although hydrolysed agarose can be resolved very well using HPAEC, this technique is unable to deal with large volumes of material. To obtain sig-

nificant amounts of resolved (monodispersed) oligomers, mixtures of agarose oligomers were separated using GPC.

Background to the technique

Over the past 20 to 30 years the importance of GPC has greatly increased. The technique and theory has been examined and reviewed extensively [89, 90].

GPC is a method for separating molecules on the basis of their hydrodynamic volume and other molecular parameters play only minor roles in the process [89]. Also, GPC can operate in very mild conditions as well as a range of organic solvents. In the field of macromolecular study, separations can be performed on an analytical or preparative scale.

GPC columns are packed as evenly as possible with fine beads of gel which have been previously swelled with solvent. Figure 5.2 illustrates the internal structure of a GPC column. Small molecules can penetrate the beads and thus have more solvent volume available than large ones do. At a constant flow rate, the small molecules will therefore take longer to elute through the column.

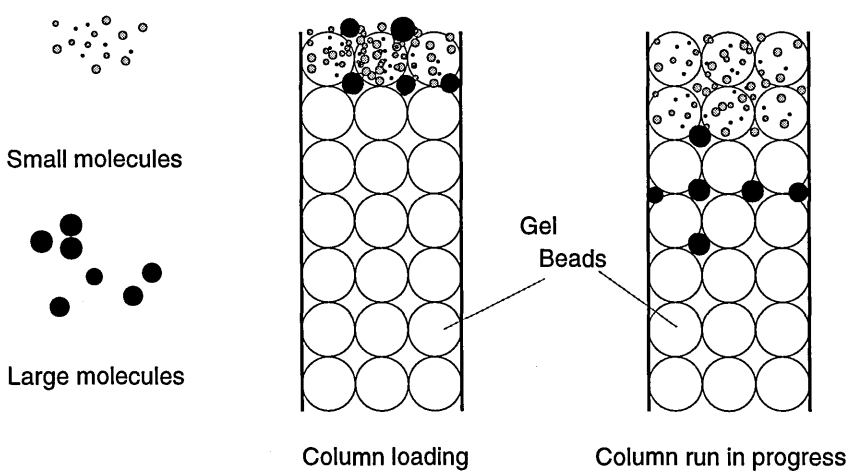


Figure 5.2: Diagram showing passage of large and small molecules through a gel column.

Commercial gels are available which can separate molecules of weights anywhere from 100 to several million Daltons [89]; it is a matter of choosing the one appropriate for the desired separation.

Column hardware

Figure 5.3 shows the equipment required for the process, and figure 5.4 is a photograph of the same setup, in situ.

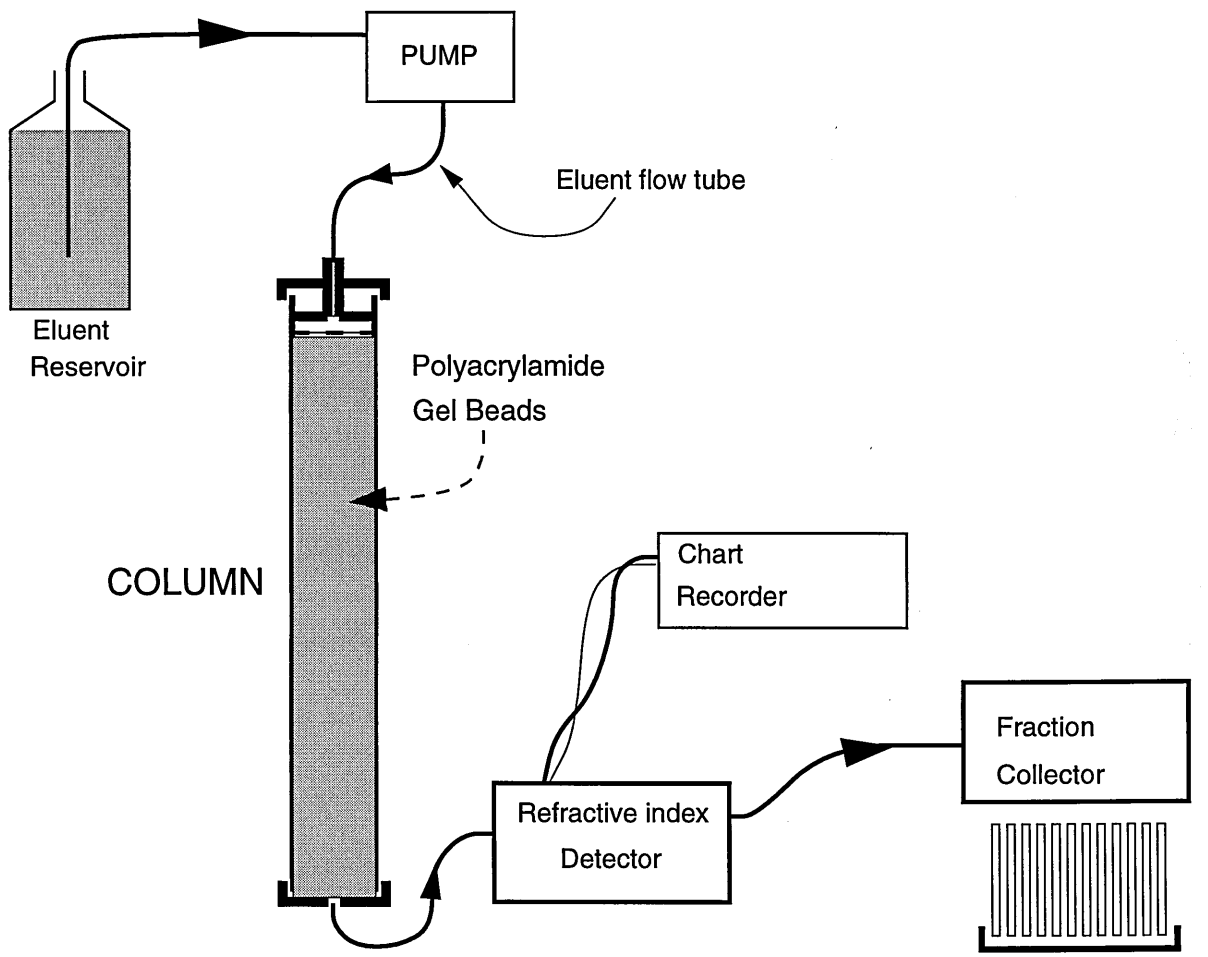


Figure 5.3: Gel Permeation Chromatography column, eluent reservoir, refractive index detector and fraction collector, shown schematically.

The column consists of a $(100 \times 4.9)\text{cm}^3$ glass tube and plastic water jacket from Pharmacia, at the bottom of which is fitted a retaining membrane and output tube. Hot de-gassed water (to prevent bubble formation) is fed into the top of the column (see figure 5.3) using a Bio-Rad EP - 1 Econo-Pump through a flow adaptor positioned just above the gel surface. The water

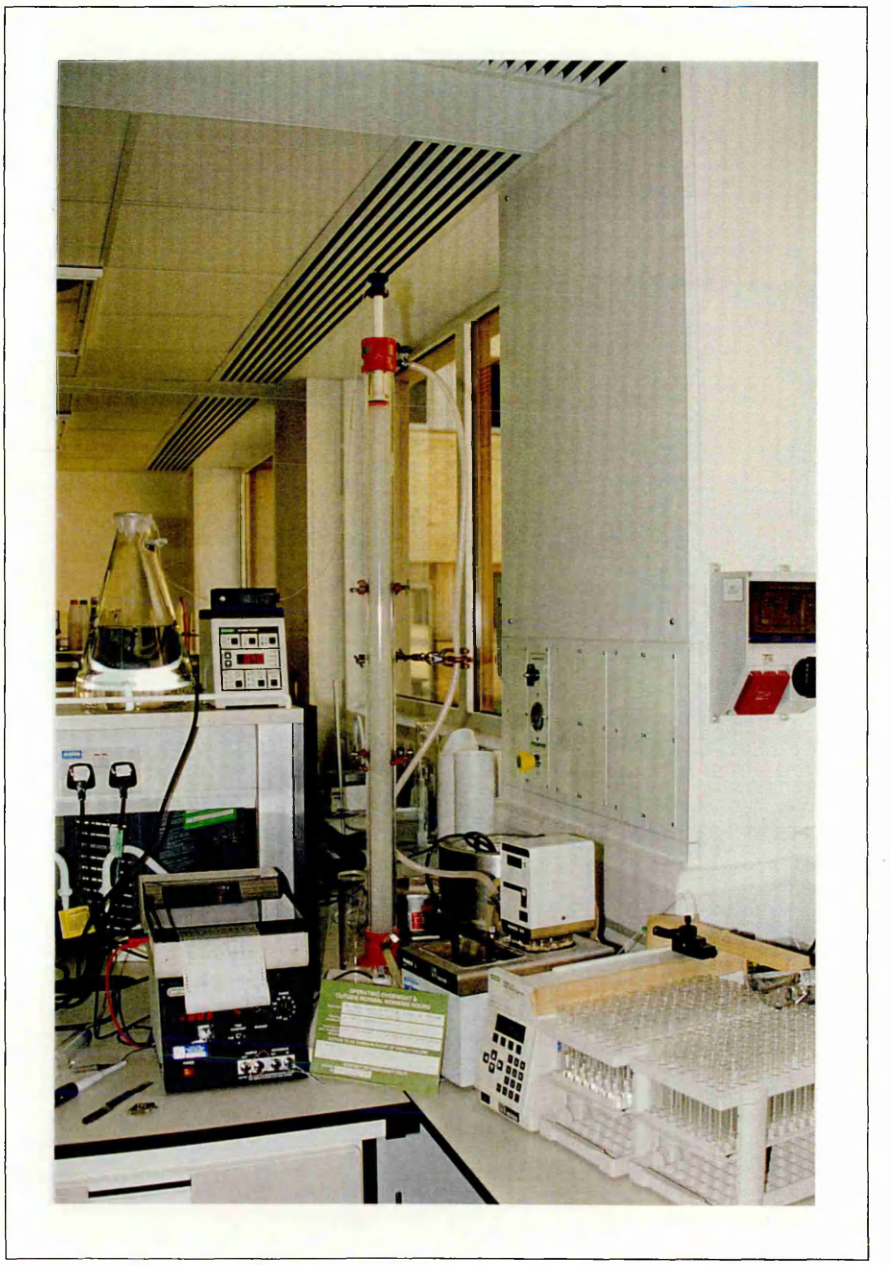


Figure 5.4: Gel Permeation Chromatography column, eluent reservoir, refractive index detector and fraction collector: actual setup.

jacket permits a constant temperature of 55°C to be maintained. Detection of carbohydrates is by a Bio-Rad refractive index detector (model 1755) connected to a Bio-Rad 1327 Econo-Recorder chart plotter. Fractions of interest are collected by the Bio-Rad fraction collector (model 2128), and the rest is drained to waste. By correlating the output from the chart plotter to the fractions collected, the contents of every fraction can be ascertained.

5.3.1 Column characterisation and sample application

The packed column permitted a throughput rate of solution of 0.5cm^3 per minute, giving an average run time for each batch of 19 hours. The bed dimensions were 80cm long by 4.9cm^2 cross sectional area.

Bio-Rad P6 fine polyacrylamide gel beads were used: these have a typical diameter of 45 to 90μm and will fractionate molecules in the range 1000 to 6000 Daltons.

Characterisation

The properties of the column were determined by running Blue Dextran 2000 (SIGMA). This substance has a well characterised molecular mass and uniform molecular size, and is also visible so that its progress on the column can readily be observed. The sample was made up to 0.4%*wt/vol* in water, since it is essentially monodispersed and is therefore narrowly distributed within the column when eluting. A flow rate of 0.5cm^3 per minute was again used. The following was noted:

- Peak broadening in the column; i.e. as the band moves down it gets wider: this needs to be minimised by good packing.
- The volume of eluent collected when the dextran emerges. Dextran 2000 is a large molecule (the molecular weight is of the order of millions); thus its elution volume is equivalent to the *excluded volume* of the column. For agarose separations no fractions need be collected before this time because no molecules can elute sooner than this.
- Signal amplitude and sign (plus or minus) on the chart recorder: allow correct settings for the agarose separations to be applied.

The column *plate count* and *peak resolvability* were also calculated; these are essentially a measure of how well the gel is packed. Uniformity of gel bead size and ratio of gel bead volume to void volume will affect performance (*i.e.* resolution of the column).

Plate counts N are found using any peak, numbered from $1 - m - n$, and are determined by examining the ratio of retention volume V_R (volume eluted at the time peak m appears) to peak base width W_m (volume eluted for the duration of the eluted peak) [90]. All peaks should give the same result, so later peaks are therefore wider than earlier ones.

$$N = m \times \left(\frac{V_R}{W_m}\right)$$

Theoretical number of peaks [90]:

$$n = 1 + 0.2N^{\frac{1}{2}}$$

The calculation is based on diffusion into pores in the gel matrix. Since the system is not at equilibrium during chromatography these results may be approximate.

Resolution **R** of peaks can also be described according to the following formula [90]:

$$\mathbf{R} = \frac{2(V_{R1}V_{R2})}{W1 + W2}$$

The plate count, N, was found to be 2,500 units. This is a relatively high value, being partly due to the height of the column (nearly a metre) and partly due to the fine bead size employed. Essentially this means that the column should resolve well: the theoretical highest number of peaks resolvable [90] should be eleven. In the present case more peaks than this are being resolved (up to 20), but the quality of resolution is correspondingly reduced. To get better separation a longer column would be needed: this would multiply the disadvantageous factors.

Sample application

All samples (volume $2cm^3$) were made up to 2% (*wt/vol*) and heated to 100°C before application, to ensure that the solute was completely dissolved. Agarose samples had to be separated at temperatures above 50°C, to prevent aggregation from taking place. Hence, the samples were applied directly to the surface of the gel column while still hot.

It can be noted that the column performed remarkably well given the complexity of the mixture (see chromatogram, figure 5.5).

Once the column had been packed one batch run of agarose mixture gave a result much the same as another. A typical output trace is given in figure 5.5.

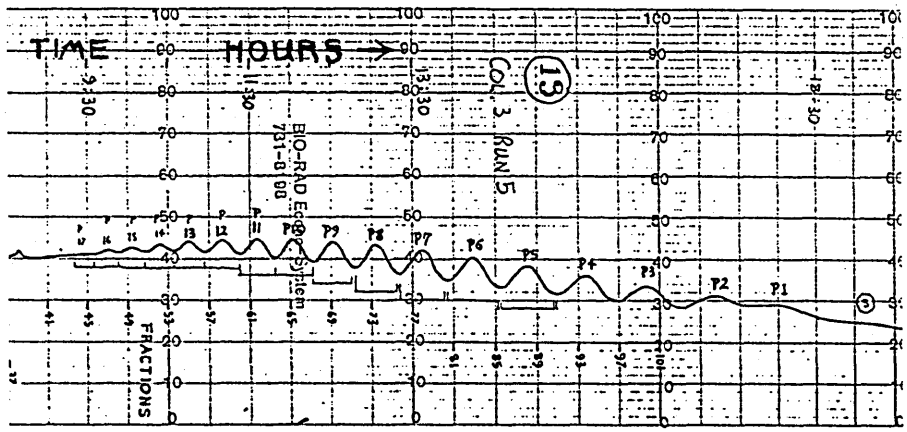


Figure 5.5: A GPC trace. Note that the shorter, smaller chains are resolved more distinctly, and they are the last to emerge from the column.

Peaks in the GPC trace were assigned unambiguously by use of known agarose oligomer sizes (measured using Ms; see above). To label the output trace extra agarose of known size was added to the mixture, thus augmenting one of the output peaks and permitting full assignment, since it has been shown that the oligomers differ in size by one disaccharide unit.

The average batch size of 0.04 grams of hydrolysed agarose produced two or three milligrams of each oligomer after chromatography. The fractions were pooled after a number of runs and freeze-dried carefully.

5.3.2 Analysis using HPAEC

By this stage it was hoped to have prepared monodispersed samples of each oligomer size. Purity of samples can be analysed using HPAEC by observing the size of the expected peaks compared to the "rogue" peaks. HPAEC is extremely sensitive with regard to detection of carbohydrates, so the slightest quantity of oligomers of different sizes in a particular sample are readily observed.

Most of the samples produced by the GPC column were of a good standard of purity, and two examples of HPAEC analysis can be seen in figure 5.6. HPAEC showed that chains bigger than Dp.35 could not be separated very well due to overlap of their retention times in GPC.

The data in figure 5.6 show that, compared to the raw hydrolysed form, a considerable degree of monodispersity in the samples of agarose oligomers has been achieved. The best separations give a monodisperse excess of about 60%.

5.4 Solution characteristics of agarose oligomers

Agarose undergoes a reversible coil to helix transition in aqueous solution [15, 91] (see chapter six). At the boiling point of water agarose chains are fully dissociated and show a random free coil configuration. During cooling the coils begin to wind up into regular helices, reaching a peak rate of molecular conformational change at the *ordering temperature*. The transition is not at all sharp [92]. The ordered helices then aggregate together. Ordering, aggregation and remnants of free coil are all thought to play a part in the formation of gels.

The aim of the present work is to obtain values for enthalpies and ordering temperatures associated with different sizes of agarose oligosaccharides, when cycled through a heating/cooling regime.

There is still debate over whether agarose helices are double helices or just single ones. Some workers have interpreted x-ray diffraction patterns and small angle neutron scattering of agarose as being more consistent with single helices [93, 94]. This view is however in the minority. Most workers accept the double helix model for agarose and it is supported by a variety of evidence which, though not conclusive, is consistent with the double helix. Examples include x-ray diffraction [26], small angle x-ray scattering [37] and differential scanning calorimetry concentration dependencies [92].

If the chain length of agarose is systematically increased the number of residues involved in ordering increases, so the ordering temperature and ΔH per mole of molecules should increase also. This assumes that there is a certain "stability" per residue for helix formation. Work of this kind has not been previously reported, as monodispersed samples have not previously been prepared.

5.4.1 Freeze–thaw

A freeze–thaw experiment was conducted on a sample of hydrolysed agarose. In freeze–thawing the sample is cycled between being frozen at -18°C and slowly melted in a refrigerator; here three cycles were executed. The agarose solution was 3% by mass.

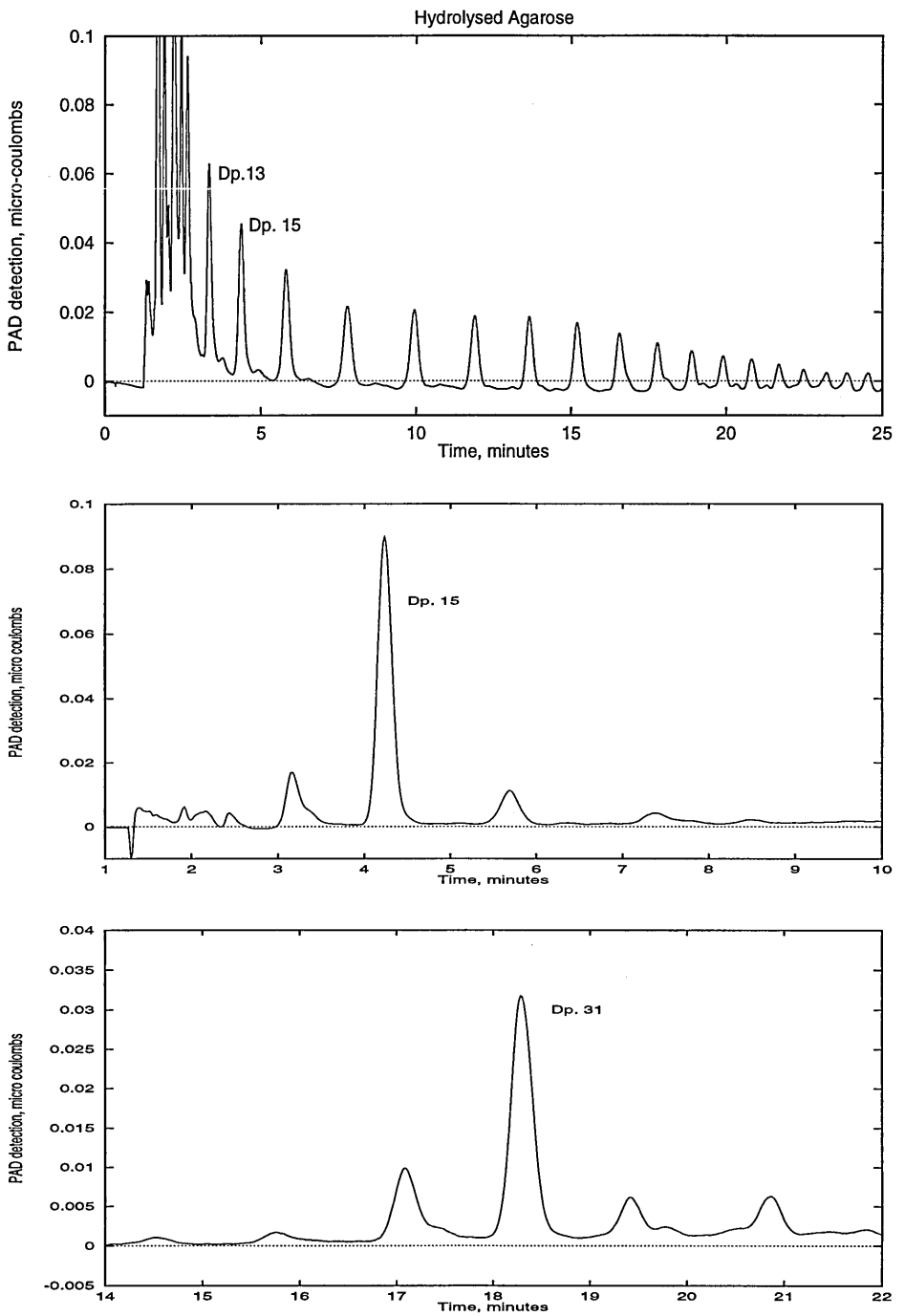


Figure 5.6: Examples of monodispersed agarose oligomers. Top: hydrolysed agarose mixture; middle: Dp.15 and bottom: Dp.31 from the GPC column.

The result is a precipitate of opaque material, which re-melts when warmed to 100°C. The precipitated material was centrifuged out of the sample, leaving just the short chain oligomers which did not precipitate. HPAEC analysis was done on this solution and the result is given in figure 5.7.

From this simple test it can be determined that chains shorter than seventeen residues cannot be induced to form precipitates, and in order to be sure of inducing precipitation the chains must be longer than nineteen residues. This in itself is a valuable guide. The caveat however is that freeze - thawing forces precipitation of agarose by two methods:

- the temperature is reduced
- the formation of ice crystals forces solute molecules together (freeze–concentration).

In addition, we are observing precipitation only, and no information on chain ordering is obtained. A more accurate method of observing molecular transformations is differential scanning calorimetry, and this is described below.

5.4.2 Differential scanning calorimetry (DSC)

DSC is able to accurately measure enthalpic events. In the experiments presented here, the temperature is gradually changed and the heat flow between a reference and sample cell recorded. If an enthalpic event occurs, such as melting in the sample cell, the rate of heat flow changes and a peak is recorded. Integrating a peak gives the total enthalpy of the event responsible for it (ΔH). To obtain a value of ΔH per residue for agarose the length of the chains and the exact mass of material used are required. Both of these are known in each case.

The small sample size of just a few milligrams of each pure agarose oligomer means that very sensitive (and costly) instruments are needed to measure energy changes.

The machine used was a Setaram Micro DSC II with a nitrogen atmosphere and a cold water bath, enabling temperatures in the range -20 to 100°C to be explored. The output was fed to a personal computer for data analysis and printing.

The temperature profile consists of a heat to 90°C, where the temperature was held for five minutes in order that the sample should be fully dissolved, followed by rapid cooling down to 35°C (no events were observed in this period). The sample is then cooled slowly down to 1°C and held there for

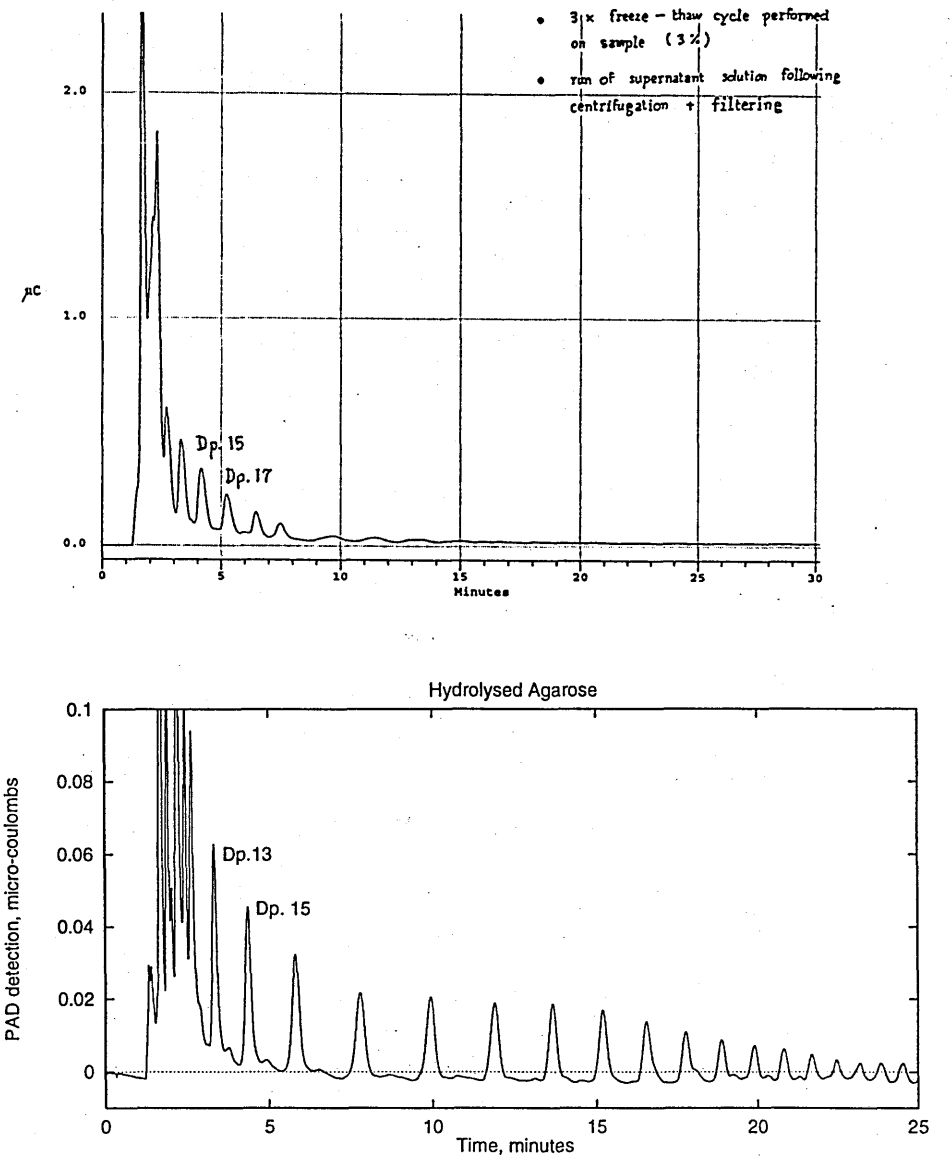
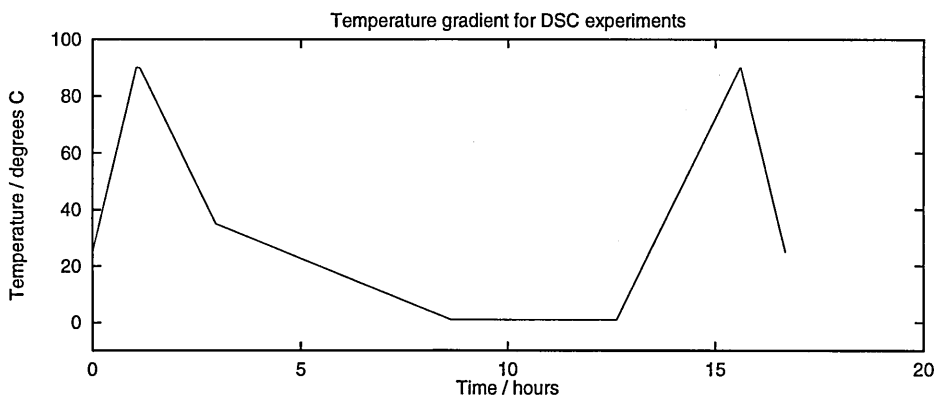


Figure 5.7: Freeze–thawing of hydrolysed agarose. The lower plot shows the full range of agarose oligomers before freeze-thawing. The upper plot shows the oligomers which remain in solution after freeze–thawing.

four hours to permit full ordering, except for Dp. 15 and 17. In these cases cooling had to be continued down to -20°C in order to observe the ordering process. The cooling gradient during the period of interest is 0.1°C per minute.

The melting process is observed during a 0.5°C per minute heating, and finally the sample is cooled to 25°C.



The sample cells were washed thoroughly with water and acetone, and dried using compressed air. All the samples were made up to about 0.8cm^3, at 1% concentration of agarose oligomer by weight; the solid was accurately weighed to ±2%. Ultra pure de-ionised water was used. The reference cells were made up to within 0.001 grams of the sample cells. Both cells were allowed to equilibrate inside the machine at the starting temperature of 25°C. Experiments were not initiated until the heat flow rate had stabilised to within 0.01mW (milli-watts).

5.4.3 DSC: results and discussion.

In order to describe the DSC results adequately, typical features of an arbitrary DSC cooling curve are shown in figure 5.8. These can be found on all of the DSC results given in this work.

Typical ordering peaks are shown in figure 5.9. The important features are:

- A sharp leading edge to each peak, and a long tail (not visible in every one),
- Longer chains (higher "Dp.") start ordering at a higher temperature.

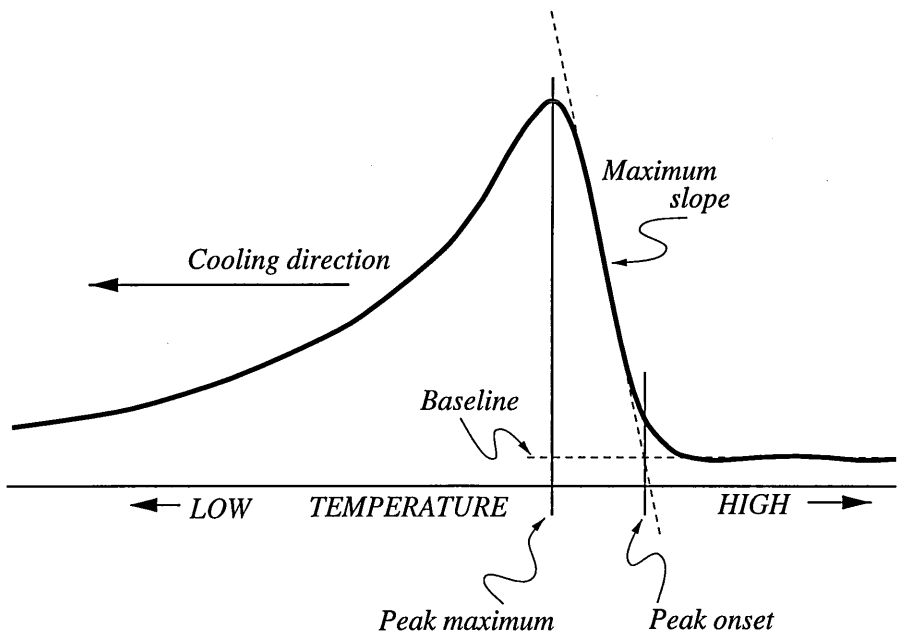


Figure 5.8: Definitions of peak onset and peak maximum in relation to a typical DSC cooling curve. The same features can be seen on heating curves.

- The size of the peak is proportional to the total enthalpy of the transition, and not simply the size of the chains used. The amount of material is the most significant factor here, as it is equivalent to the amount of ordering which takes place.
- The peak maximum represents the point at which the net conversion rate of coils to helices is greatest. The process is an equilibrium, so during cooling there will be a particular point at which the forward (coil to helix) reaction is faster than the reverse reaction by the largest margin.

Several aspects of agarose chain behaviour may be probed by analysis of features of the DSC data. Presented in figure 5.10 are plots of ordering and melting peak onset temperature values; these are discussed later in this section. The ordering peak on the DSC trace is always exothermic, while the melting event is endothermic. Values of temperature were measured to the nearest 0.5 degrees.

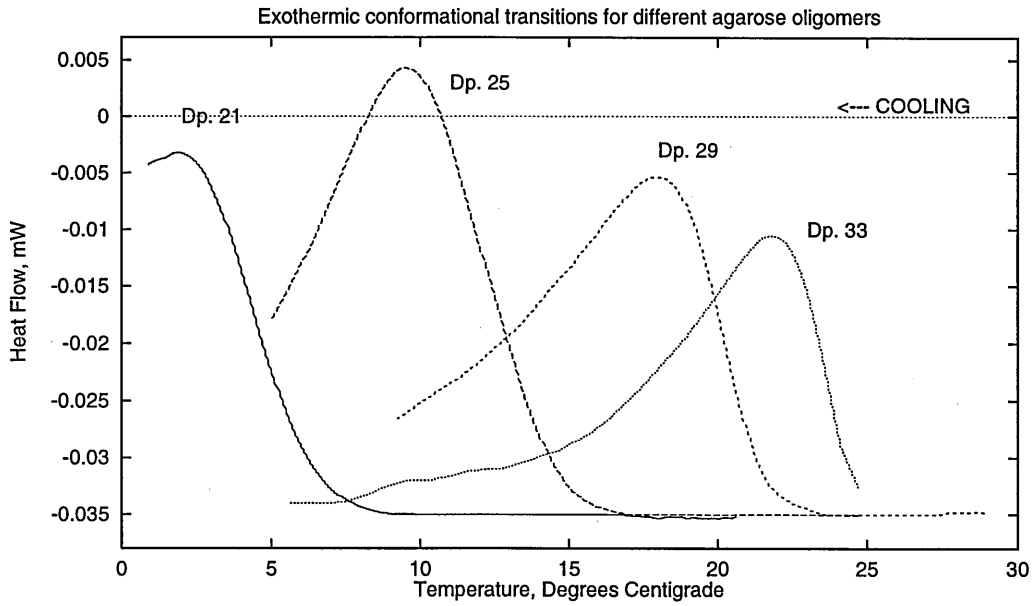


Figure 5.9: Double helix ordering for a range of agarose chain sizes. Note that the graphs are for cooling, so time would progress from right to left.

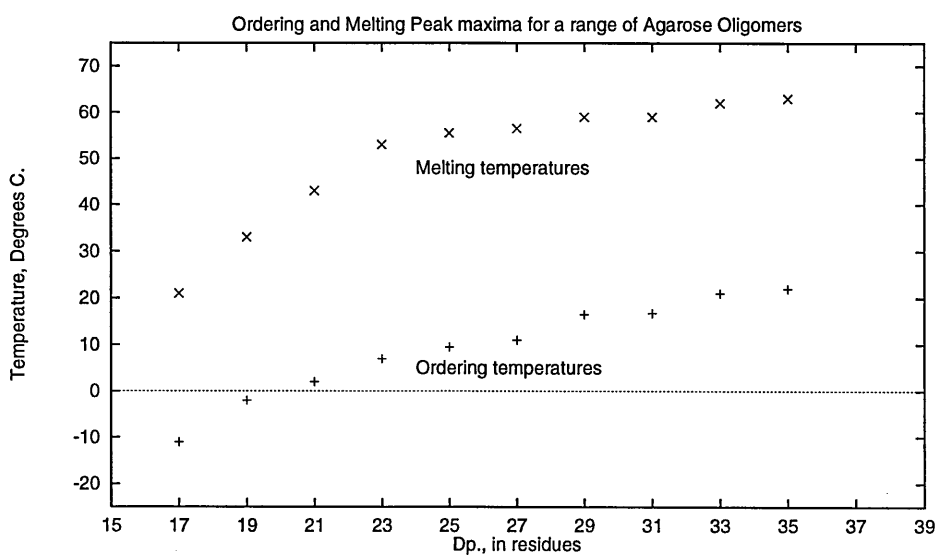


Figure 5.10: Ordering and melting temperatures for a range of agarose oligomers, determined by DSC. The points for melting form two approximately straight lines intersecting at Dp. 23. The points for ordering display a curve which is abruptly cut off below Dp. 17. No oligomers below this size could be induced to order.

Minimum size for double helix formation.

It can be seen from figure 5.10 that the temperatures at which helices form and melt are determined largely by the length of the carbohydrate chains present. This fits the notion that there is a certain ΔH and ΔS associated with the change from coil to helical conformation of any particular residue in the chain. Therefore if the chain is very long the overall process is more favourable, so occurs at a higher temperature.

From figure 5.10 the shortest possible double helix for agarose can be ascertained. The ordering (lower) curve terminates after chains smaller than 17 sugar rings because the Dp.15 oligomer could not be induced to order, even when cooling down to −20°C. If the Dp.15 oligomer was to order, the peak onset point would be around −10°C with a maximum at about −20°C[3]. It was not observed. In addition, even following several hours at −20°C to allow ordering to occur, no melting peak was observed during heating.

This also indicates that a significant portion of the chain at each end never takes part in a helix (the so-called *end effect*). Other gelling polysaccharides are known to form helices with only a few residues and have much smaller end effects. Agarose helices are therefore not particularly stable, although aggregation may help to push the reaction over towards the helix side of the equilibrium.

In the limit of polysaccharide length represented by agarose itself the ordering transition begins at about 33 to 36°C and re-melts again at well over 70°C. If the ordering curve in figure 5.10 is extrapolated to 35°C the average size of helices would be about 40 to 45 residues. From this the average length of a junction zone can even be estimated. Using the helical parameters of agarose [26] (height of one helical turn is 18Å containing six residues) then 40 residues would be about 120Å long. These facts have not been reported before.

Kinetic and thermodynamic effects.

The observation of hysteresis in the agarose system indicates that some part of the coil to helix equilibrium is kinetically controlled. So, although the system would like to melt into coils when it is warmed, some process with a large activation energy is preventing it.

Firstly we have coil to helix ordering, which is accepted as being a reversible process [95, 96], and secondly we have aggregation, which is thought to be kinetically controlled and to have only a small enthalpic contribution [97]. During melting, aggregates mush be disrupted before helices, so

[3]Ice only forms inside the DSC sample vial (constant volume) at -20°C.

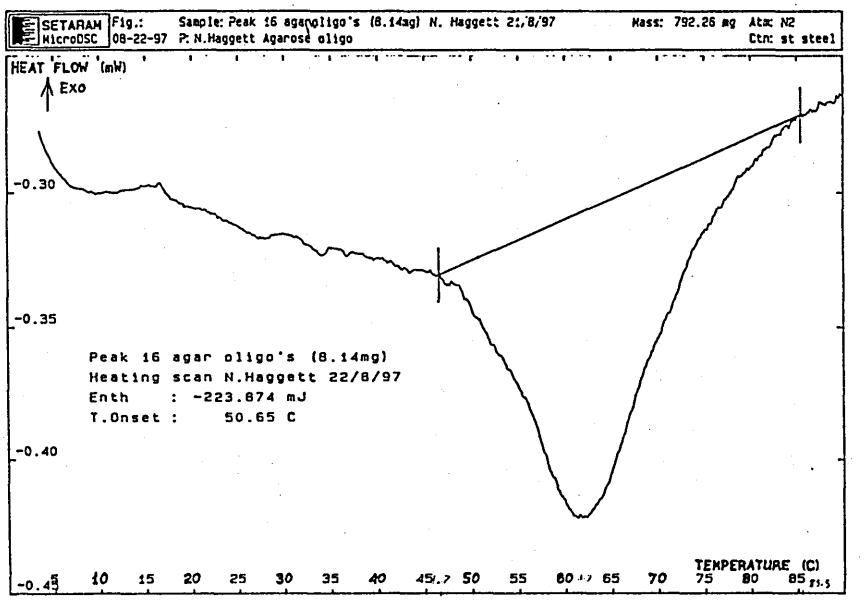


Figure 5.11: Melting of aggregated material for Dp.33 agarose oligomers, measured by DSC.

aggregation means that the entire melting process is kinetically controlled, and also since this is the rate determining step the melting peaks are fairly symmetrical (helices are disrupted easily) [97]. A melting curve for Dp.33 is given in figure 5.11.

The ordering process is therefore responsible for the major part of the exotherm seen during cooling. It involves extensive molecular conformational change, and is essentially a thermodynamically controlled process.

The cooling curve displays a long "tail" region which indicates some residual process taking place for some time after the peak maximum has passed. Since agarose is thought to form double helices:

$$\text{COIL} + \text{COIL} \rightleftharpoons \text{DOUBLE HELIX}.$$

it is expected that the concentration of the sample will affect the rate of ordering and the shape of the peaks. This is consistent with the long tail of the DSC cooling curve. There is an initial rapid rise in the rate of ordering, followed by a long decay as the concentration of free coils in solution is reduced. These observations do support the assumption made in this work: that agarose forms double helices, not single ones. This is the main reason for using the same amount of material in each sample, and the same cooling

rate. For an in-depth analysis several measurements should be done with differing sample concentrations, however this work has not been done here.

There are no concentration effects for the melting process, since by this stage the temperature is above the coil to helix equilibrium point. During melting aggregates are disrupted, releasing helices which themselves melt into coils.

Optical rotation and light scattering characteristics during the cooling of the hot solutions are of considerable interest at this point. Other workers have frequently applied these techniques in their efforts to learn more about agarose sol–gel transitions [37, 98]. In the present case however insufficient material could be prepared for either of these techniques to be applied.

Optical rotation is especially appropriate, since it gives direct information on the proportion of residues within the sample which have been rotated into the ordered phase [37]. Light scattering responds to particles in the solution, hence the onset of aggregation could be monitored by this method [90].

Double helix aggregation.

Aggregation of agarose double helices has been seen to be unavoidable and very favourable. The extensive hysteresis observed with agaroses longer than 17 residues is evidence of this. The two plots in figure 5.10 show that the melting temperature (of aggregated helices) is always 20 to 40°C higher than the setting temperature (of random coils).

It would be interesting to know if the ordering/aggregation processes can be separated in the DSC experiments, but tests where samples were cooled to just below the ordering onset temperature, and then re-melted quickly to try and avoid aggregation still displayed hysteresis. It seems therefore that as soon as molecules order into helices, they immediately form aggregates which are then difficult to melt.

The melting peaks are quite broad, covering a range as large as 40°C (see figure 5.11). It is likely that the internal structure and distribution of sizes of aggregate particles is reponsible for this.

The plots of melting points of different lengths of agarose chain (figure 5.10) display two distinct regions; firstly for small chains shorter than Dp.23 where the plot gradient is steep, and secondly for chains larger than Dp.23 where the gradient is less than half as steep. This observation must reflect the ease of melting of aggregate particles. One hypothesis is that there are two possible modes for packing together of double helices into aggregated particles. One mode is favoured by short chains, while the other becomes dominant if long chains are used.

It may be that helices are allowed to overlap by an amount denoted here as "mismatch". If the helix is shorter than 23 residues, the mismatch is too small for extended aggregates to form. If longer ones are used then three dimensional aggregates can form (see figure 5.12).

Note that this phenomenon can only be observed for monodispersed samples where packing must take place between identical units, as in the present case. It is to be expected that the number cannot be generalised to any agarose sample but may vary by a few residues.

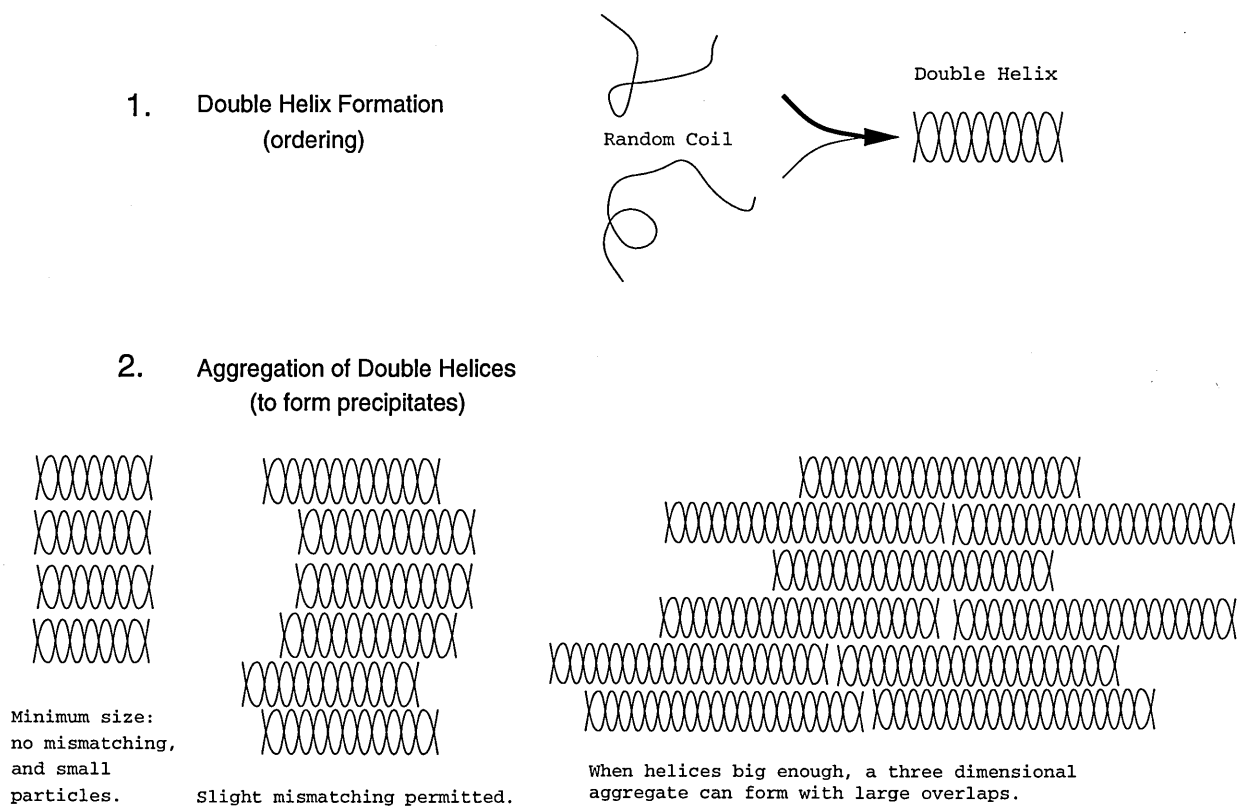


Figure 5.12: Agarose double helices form when hot solutions of random coils are cooled (top scheme). These double helices then form aggregates. The shortest helices must be lined up perfectly in order to form an aggregate, with no mismatching. With large helices the mismatch is large enough to permit three dimensional particles to form (lower scheme).

Mismatching is the most likely reason for broadness in melting peaks. Indeed tests where random–coil samples were cooled to just below the ordering onset temperature and then re-melted after a time interval of some minutes

did display narrower peaks, this being evidence that only well - matched aggregates had been allowed to form. However for long chains (Dp. 25 or more) the effect could not be isolated.

5.4.4 Enthalpies from DSC

This part of the analysis deals with enthalpies measured from DSC results. It should be noted here that there are several errors involved:

- Errors in weighing samples (0.25%)
- Unavoidable microscopic fragments of solid or droplets of liquid which may become stuck around the top of the DSC sample vial.
- Water contained in the dry agarose samples (this can be up to 10% by weight).
- Errors in the analysis of graphs: noisy baselines and problems with sample / reference mass difference.

These errors are all the more significant due to the small sample sizes used. Enthalpies are calculated directly from the DSC curves by measuring the area under the curve of heat flow *versus* time. They are in milli-joules (mJ), and sample masses are in milli-grams. The results were corrected to the joules per unit number of molecules by normalising to the same concentration. The values are presented in table 5.3 and figure 5.13.

'Internal' and 'external' agarose chain residues

The formation of a double helix from two agarose chains is an equilibrium process (although perturbed by aggregation). On the one hand there must be a source of stabilisation enthalpy to favour the helix forming process, and on the other hand there is an entropy which favours free coils. For short chains (Dp. 10 and below) the computer modelling work clearly indicates that it is unlikely that enough stabilisation is available (see chapters four and six). In addition, Dp. 17 chains are not seen to form helices by DSC; this indicates that there is an *end effect*. In an agarose double helix of arbitrary size there are "end regions" (external residues) which do not participate in helices, and internal residues which do, and which have an associated (exothermic) enthalpy of helix formation.

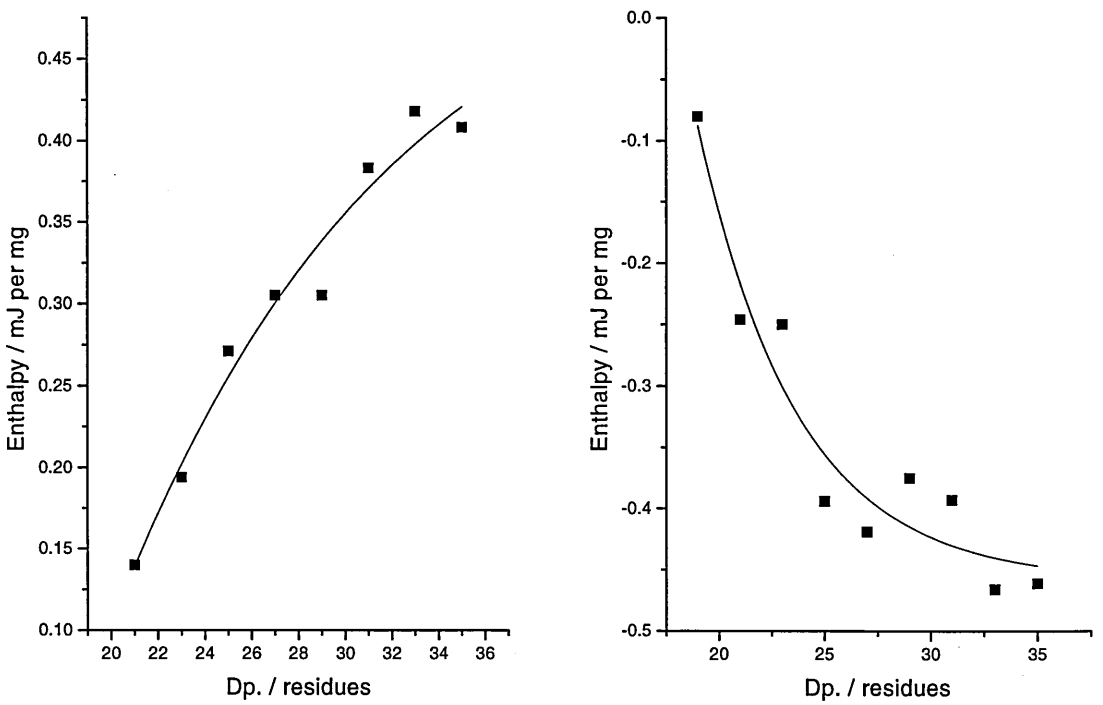
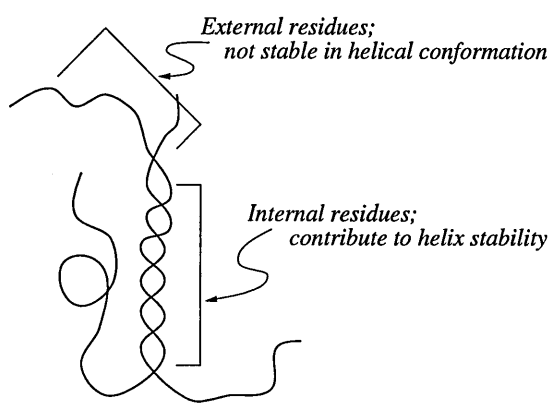


Figure 5.13: Integrals from ordering and melting curves measured by DSC, for agarose chains ranging from 19 to 35 residues in length.

Dp. (residues)	Ordering		Melting	
	Total peak integral (mJ/mg)	Integral adjusted to conc. 3.4 mM (in mJ/mg)	Total peak integral (mJ/mg)	Integral adjusted to conc. 3.4 mM (in mJ/mg)
19	–	–	0.080	0.080
21	-0.126	-0.140	0.222	0.246
23	-0.159	-0.194	0.206	0.250
25	-0.208	-0.271	0.302	0.394
27	-0.204	-0.305	0.280	0.419
29	-0.198	-0.305	0.261	0.375
31	-0.231	-0.383	0.236	0.393
33	-0.236	-0.418	0.263	0.466
35	-0.210	-0.408	0.206	0.250

Table 5.3: Peak integrals for cooling (ordering) and heating (melting) curves measured by DSC.

In this instance, we should observe that (above a minimum Dp.) the enthalpy of ordering increases regularly with chain size as each extra residue contributes a small $\Delta H_{res.}$ to the overall helix stability.



Helix forming ability for residues in an agarose chain (simple model).

A more complex model would introduce a third residue type intermediate between the fully helix forming ones and the end ones. It really depends on how cooperative the coil to helix equilibrium is. If fully cooperative then residues will "see" each other and the chain will tend to be all coil or all

helix, since once a helix nucleates it will grow quickly. If the process is not cooperative small helical regions could, in principle, form independently at any point on the chain. Unfortunately the plots in figure 5.13 are not really accurate or extensive enough to permit an idea of the relative numbers of end, near-end or internal residues to be estimated. The fact that small chains shorter than seventeen residues do not form double helices under normal conditions quantifies the "end" region as being about eight residues long. In other words, only residues which are eight residues along from the end of a chain will participate in ordering.

Enthalpy per residue for double helix formation

Although few statements can be made about numbers of internal and external residues, nevertheless an average figure for the binding energy per residue can be derived from the DSC results. The most accurate data are from the two largest chains: Dp. 33 and 35. Here, the entire peak can be observed and readily integrated.

The data required for the calculations are:

- the relative molecular mass of the agarose oligomer under study
- the exact mass taken for DSC
- enthalpy of the transformation for ordering, corrected for different concentrations
- and finally an estimate for the number of residues not taking part in the double helix (external residues).

Assuming the only two types of residue exist, we can say that when the chain size is seventeen residues, the external residues balance the internal residues exactly, so that a stable helix is only just prevented. When the chain is nineteen residues long the helix can form. In other words only residues which have at least eight additional residues on both sides can initialise growth of a stable helix. From these results and discussion, a reasonable value for the number of residues not taking part in the helix is 17, per chain.

1. Calculate the relative molecular mass (RMM) for each chain size used:

$$\text{RMM} = n(306) + 162 + 18 \text{ g/mol}$$

where n is the number of disaccharide units of mass 306 Daltons, 162 is the mass of the extra galactose unit, and 18 is the mass of the extra water molecule, formed from -H at one end and -OH at the other end of the chain.

Dp. (*residues*)	**Conc. (M)**	**Enthalpy**/mJ (*Measured from DSC results*)	**Enthalpy**/mJ per mol of chains	**Enthalpy**/J per mole of residues
33 (21 binding) RMM: 5076g/Mol	1.77×10^{-3}	155	106	6.3
35 (23 binding) RMM: 5382g/Mol	1.27×10^{-3}	139	111	5.9

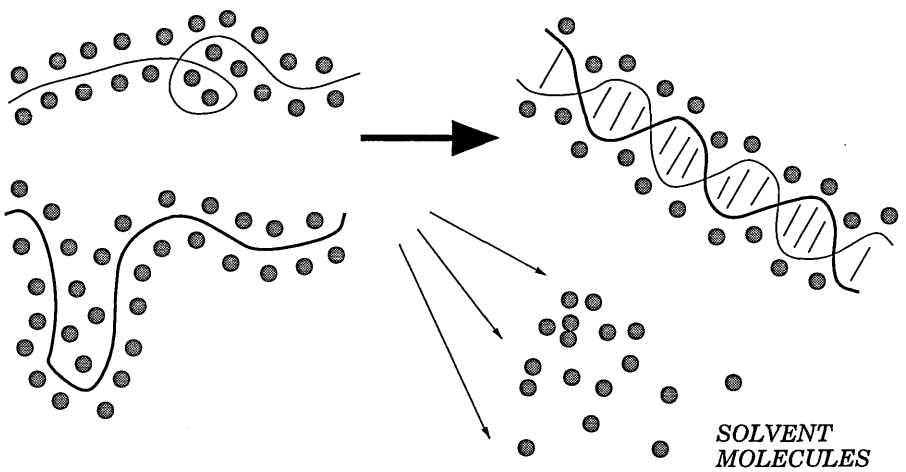
Table 5.4: Values of enthalpy per mole of residues calculated from DSC results; these are for the *ordering transformation*. For the method of calculation, see text.

2. Find the number of moles of monodispersed agarose chain used in each DSC test using the actual mass taken and the RMM.
3. Find the actual enthalpy change for the ordering transformation (random coil to helix) for each DSC test by integrating the area under the DSC peaks observed during cooling.
4. Compute the enthalpy per mole of chains.
5. Compute the enthalpy per binding residue using the number of moles of chains present and the estimated number of binding residues per chain.

This yields an average value for binding energy per residue of **-6.1kJ per mol**. Since the peak is an exotherm, this value is negative.

The role of solvent

Solvent molecules (in the current case these are water) play a part in molecular dimerisation reactions. In order for random solvated coils to meet, solvent must be removed from their surfaces. This involves disruption of `solvent:solute` complexes, requiring energy, but there is in addition a favourable increase in entropy due to the extra freedom imparted to released solvent.



The breakdown of total free energy into these specific terms cannot be attempted here because the techniques used only give an overall measurement. In observing disruption of helices by molecular modelling methods, the solvent is omitted due to computational limits. This is likely to favour helix disruption since there is no entropic cost in making random coils: they have no solvent bound to their surfaces.

5.5 Summary

This chapter has described the hydrolysis of agarose using dilute acid, to produce all lengths of chain from a few sugar units up to hundreds. The smaller chains, which are of most interest in the present work, have been successfully characterised and separated using mass spectrometry, high performance anion exchange chromatography and gel permeation chromatography. These techniques show that the hydrolysis can produce any size of chain in the form $(GAL–AnGAL)_n–GAL$, and that there is a limit of chain size below which the well known conformational ordering properties of agarose itself are not displayed.

DSC has been used to probe the changes in molecular configuration occurring when agarose oligomers are cycled through a 1 to 90°C temperature regime.

It is shown that the temperature onset of ordering and of melting is indeed dependent on molecular size, as expected, and graphs displaying this relationship have been constructed.

Some insight into the two processes of *ordering* and *aggregation* has been obtained: these processes cannot be easily separated. The relationship be-

tween ordering temperature and chain size (during cooling) displays an abrupt cutoff for chains smaller than seventeen residues. It may therefore be postulated that only chains longer than this can ever form stable double helices. The relatively large size of this minimum length, compared to other helix–forming biopolymers, indicates a significant "end effect".

In the melting of ordered agarose oligomers it is noted that two different packing mechanisms for aggregation of double helices are likely to be in operation. For very short helices only small, two dimensional plates can form. For large helices extended bundles with internal overlapping of double helix units can occur. Scanning electron microscopy may be able to show these types of particle. These two modes of packing give rise to a discontinuity in the melting temperature *versus* chain size plot.

Some values of enthalpies of ordering and aggregation are obtained for the oligosaccharides. These confirm the prediction that longer chains will have larger enthalpies of ordering. From the results a value of binding energy per residue was calculated at -6.1kJmol^{-1}.

Possibilities exist for further work to be done in this area, particularly using the techniques of light scattering and optical rotation. The results of these tests would improve the present understanding of agarose ordering reactions. In addition electron microscopy of agarose aggregated particles, prepared from monodispersed samples, would confirm the hypothesis regarding the packing of double helices themselves.

In the next chapter the DSC work is studied more closely in the light of some modelling work which has been done on the stability of agarose helices.

Chapter 6

Agarose coil to helix equilibria

Contents

6.1 Review of previous chapter

The previous chapter has shown that, on cooling down agarose solutions containing molecules in random coil conformations, a distinct transition point can be determined with respect to temperature where molecular reorganisations are taking place. This transition point depends on chain size, and has an associated enthalpy change.

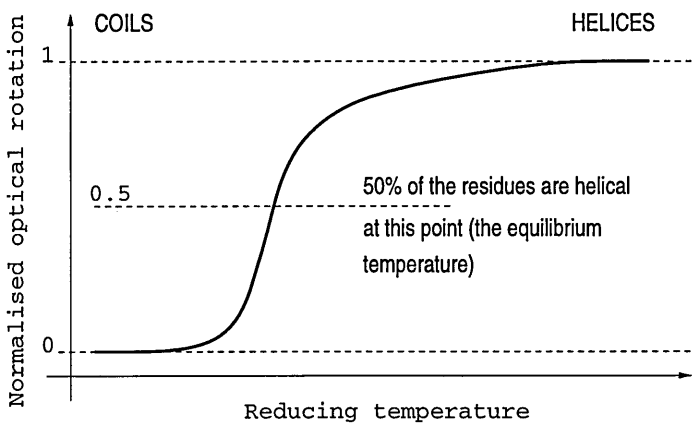
In this chapter a theoretical model for random coil to helical transformations is presented and applied to agarose. In addition the reverse process of agarose double helices melting, has been modelled for significantly large chain sizes using a supercomputer.

6.2 The Zimm–Bragg model: application to agarose.

6.2.1 Background

It has been well known for some years that many biopolymers which exist in solution as a random coil form at high temperatures will undergo a conformational change to an ordered helical form if the temperature of the solution is reduced [99, 100]. A change in the quality of the solvent (for example by adding salts or other solutes) can also effect the transformation. For some polymers, for instance *polybenzyl*-L-*glutamate*, this transition is totally reversible with a specific equilibrium point where 50% of the residues are in a helical state [99].

The ordering process may be followed using various techniques, such as ultraviolet absorption [101] and optical rotation [102]. The principle relies on a change in some physical property from the value corresponding to all–random–coil to that for all–helical conformation. This change can then be normalised to represent the transition itself. Experiments of this nature result in a typical 'S' shaped curve, shown below:



This curve is effectively a progress meter which shows how much material has ordered as the transition proceeds. For this reason it is usually normalised to represent values from 0 to 1.

Such a molecular transformation has been firmly established as a property of the agarose system. For agarose however ordered helices have a strong tendency to *aggregate*. The formation of ordered helices and their subsequent aggregation is thought to be a precursor to the network forming process (*i.e.* gelation). Aggregation should favour the transition from coils to helices, by removing the products of the transition. This is the origin of the hysteresis between transition points during cooling and heating in agarose.

Many people have contributed ideas for a model to describe coil to helix reversible transformations, but there have been differences according to which system was used for the development. The main principles were brought together by *B.H.Zimm & J.K.Bragg* [39], and models for various different systems are excellently presented in a volume by *D.Poland & H.A.Scheraga* [95].

The Zimm-Bragg theory does not account for aggregation and the results will be discussed with this in mind.

6.2.2 Theory

The theory is an application of statistical mechanics [103]. Some general points are given first, followed by the development specific to short double helices (20 to 40 residues in length). Agarose oligomers such as these were studied in chapter five.

Statistical aspects of coil to helix transformations

In the context of coil to helix transformations, a polymer chain can be considered to be made up of two different types of residue[1]:

- helical residues h (of the conformation found in the perfect helix)
- non-helical residues c (any other conformation)

The likelihood of a particular residue being in the coil state or the helical state is determined by what state neighbouring residues are in, and by the temperature of the system.

Before a model can be developed, the partition function for the chain must be written down. To do this, all possible chains are considered. Each chain can be thought of as sequences of h's and c's such as

$$\cdots cchhhcchhhhccchhhhhcc\cdots.$$

Using the free energy of any particular chain G , the partition function $Z(N)$ is

$$Z(N) = \sum_{\{h,c\}} \exp\left[\frac{-G\{h,c\}}{RT}\right] \tag{6.1}$$

where $G\{h,c\}$ is the free energy of the chain sequence $\{h,c\}$, and N is the chain length in residues.

The probability of one particular sequence $\{h,c\}$ is:

$$\text{Probability of } \{h,c\} = \frac{\exp[-G\{h,c\}/RT]}{Z(N)}$$

In order to make any real progress in evaluating the probability of occurrence of particular chain shapes it is necessary to assume that neighbouring sequences (of h's or c's) in a chain are *independent*; that is, their free energies are unaffected by each other[2]. The free energy of a particular chain then depends only upon the sum of free energies of the constituent sequences, which in turn are related to how many h or c residues are present in each sequence. This is really all that is required. The main problem is to compute the sum in equation 6.1, and it is here that differences between the various biomolecular coil to helix systems manifest themselves.

[1]In the most general case there are a multitude of states for each residue.

[2]Note that residues within sequences are not independent.

Since a coil sequence can only be followed by a helical sequence, the entire chain can be expressed as a series of [helix–coil] sequence pairs. Representing u_i as the statistical weight of a sequence of i coil type residues, and v_j as the weight of j helical residues, we now arrive at the partition function for a chain of N residues:

$$Z(N) = \sum_{i,j} \prod u_i v_j. \tag{6.2}$$

This is a product of partition functions over all sequence pairs.

Relation to Zimm–Bragg parameters

Zimm and Bragg [39] used the so-called s, σ notation to define their helix-coil transformations; this has now become the standard. Its main application was in α-helix models, which have specific hydrogen bonding between residues i and $i + 4$ in helical sequences. Its relation to the u, v and w (Lifson and Roig) notation is as follows.

Coil state	u/u	=	1
Helical state	w/u	=	s
Helical state; no hydrogen bond	v/u	=	$\sigma^{1/2}$

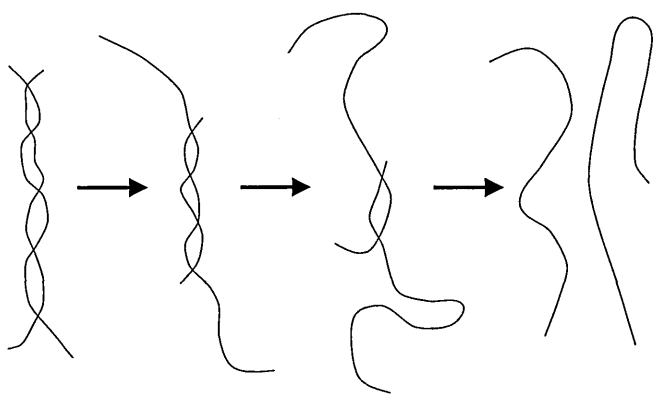
Physically, this means that if s is decreased, for example by raising the temperature, then the chain as a whole becomes less helix–like and more coil–like.

Another complication arises when helical sequences are forming. How long do we expect a helical sequence to be? On the one hand, helical sequences will want to be as long as possible in order to maximise the stability of the sequence. On the other hand if helical sequences are short, they can be placed in many different ways in the chain and there is an entropy gain. This leads to the proposition that for chains of limited length where a helical sequence cannot be placed in many ways, the *one-sequence* ("all or nothing") rule will apply. That is, at high temperature we have one coil sequence only, and at low temperatures we have one helical sequence.

Application to agarose

The agarose oligomers studied in chapter five may be described as short chains, and so it will be assumed that there is one (double) helical sequence per chain pair (the "all or nothing" case). Another name given to the model is *the staggering zipper model* [101], due to the obvious parallel with a zipper.

The definition of h and c states is the same as before. Finally, in agarose, we have the possibility of *imperfect matching* during unwinding of a double helix, which increases the number of ways to select the residues from each chain for helical sequences (see below).



Imperfect matching during unwinding of a helix

The primary aim of the theory is to estimate the fraction of residues bound into a double helix. This can be done by computing the products of all factors for residues in the bound helical state, and presenting it as a fraction of possible number of bound residues. This is given the symbol θ:

$$\theta = \underbrace{\left(1 + \frac{1}{4\gamma\Sigma} - \left[\left\{1 + \frac{1}{4\gamma\Sigma}\right\}^2 - 1\right]^{\frac{1}{2}}\right)}_{1} \underbrace{\left(\frac{\Sigma'}{N\Sigma}\right)}_{2} \tag{6.3}$$

where

$$\Sigma = \sum_{n=1}^{N} (N - n + 1)^2 s^n \tag{6.4}$$

and

$$\Sigma' = \sum_{n=1}^{N} n(N - n + 1)^2 s^n \tag{6.5}$$

Term (1) in equation 6.3 is the fraction molecules existing in double stranded pairs, as opposed to isolated coils, while term (2) is the fraction of residues in the ordered (double) helical form. The symbol γ is used to make the final relation simpler to follow. It is

$$\gamma \equiv \frac{Z_{\text{ext.}}(C_2)C_0}{Z_{\text{ext.}}(C)^2}$$

The terms in C arise from consideration of the equilibrium between isolated coils and double stranded helices:

$$C + C \rightleftharpoons C_2$$

where C is a single coil and C_2 is the associated species. C_0 is the total concentration of C, and $Z_{\text{ext.}}$ is the partition function for the external degrees of freedom of the species.

Modelling the helix to coil transition of agarose

A FORTRAN program (see Appendix A) was used to evaluate the coil/helix model. The fraction of paired molecules and fraction of ordered residues are written out to a file against temperature.

6.2.3 Fitting to experimental data for agarose

From chapter five we have sets of data for some chain lengths which represent rate of ordering with temperature. This is simply the DSC data for cooling (ordering transition). We do not have data pertaining to the *extent of ordering*[3], which is what is really required. The following reasoning, however, leads to an interpretation of the data which can be used for this purpose.

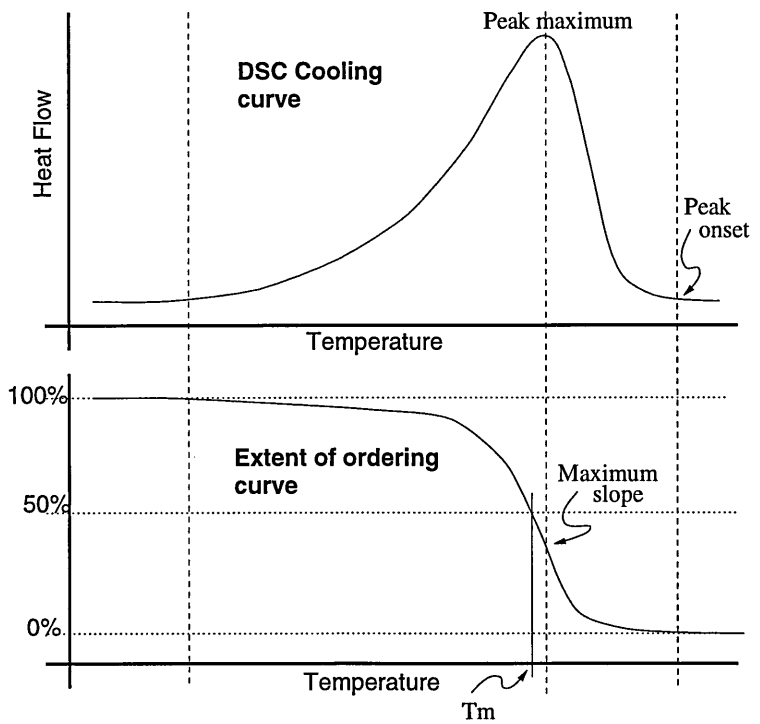
Relation of ordering theory to experimental DSC data

Consider the entire DSC peak measured during cooling of a solution of agarose oligomers. At the start (high temperature) the predominant form for a residue is the coil state; at the end it is the helix, and the total change in enthalpy is the area of the DSC peak. In between are points where the fraction of ordered residues is increasing and a heat flow is recorded. In fact, the DSC enthalpy peak can be considered as a record of extent of ordering with temperature, provided that aggregation does not contribute to the enthalpy[4], and that the system is in equilibrium during cooling. The scheme

[3]The extent of ordering is equivalent to the fraction of residues in helical form.

[4]This is the reason that optical rotation results are preferable.

below illustrates this reasoning (T_m is the temperature at the transition mid-point):



Graphs of extent of ordering with temperature were therefore produced from the DSC data, and are given in figure 6.1.

The transition temperature T_m for the helix to coil transformation in the Zimm–Bragg theory is defined as the point at which 50% of the residues have ordered (see figure 6.1), and ΔG is zero. This is equivalent to the point of maximum slope on the accumulated DSC integral plots in figure 6.1, and these are plotted in figure 6.2.

The model

The model used is more or less that given in section 6.2.2, ("all or nothing", with imperfect matching during unwinding) but it has been slightly adjusted to give more physical meaning.

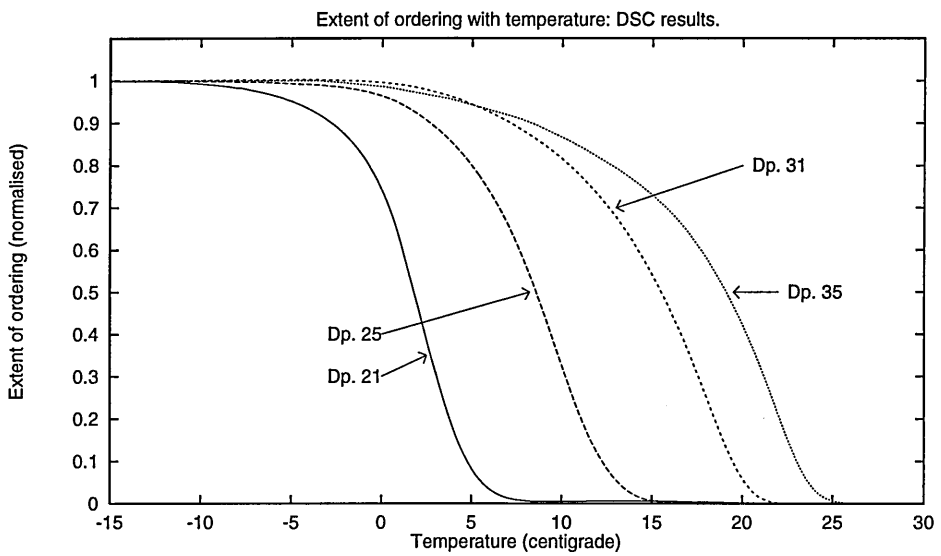


Figure 6.1: Normalised, accumulated DSC integral data for some agarose chain lengths. As the solutions are cooled (right to left) the extent of ordering increases. The transition temperature is defined as the point where 50% of residues are ordered (see arrows).

The physical interpretation of the parameter gamma (γ) in this model is as a 'concentration parameter' [101], of the form

$$\gamma = \beta c$$

where β is an association factor independent of the length of the chain, and c is the total molar concentration of chains including double helices. β therefore gives an indication of the tendency of chains to pair up. It has units of ℓ/mole.

The Zimm-Bragg parameter s (see equations 6.4 and 6.5) is redefined in terms of thermodynamic parameters [101]:

$$\ln s = \frac{\Delta S}{R} - \frac{\Delta H}{RT}$$

ΔS and ΔH are the enthalpy and entropy per residue for the coil – helix molecular re-organisation.

The parameters which can be varied (and are read in by the model) are:

- the size of chains, N

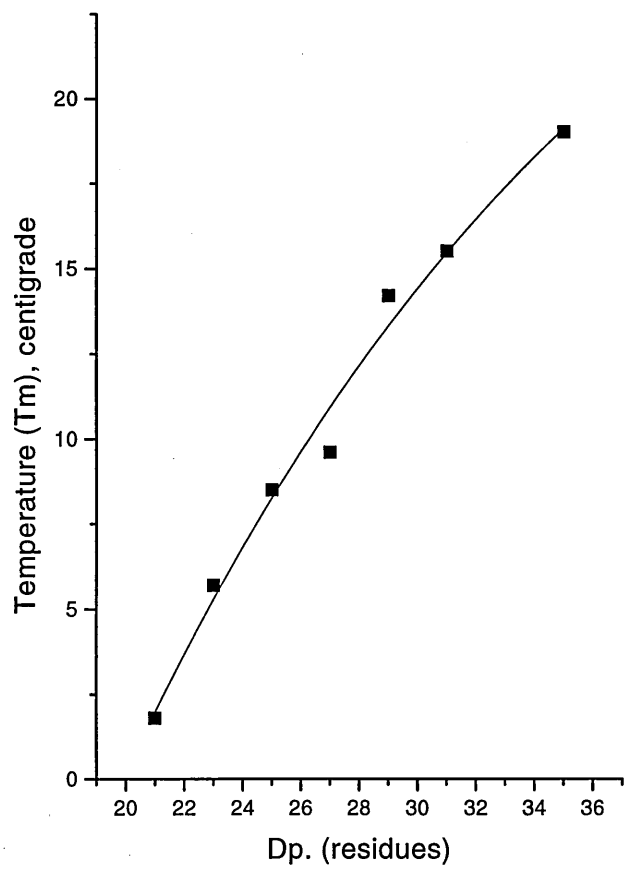


Figure 6.2: The transition temperatures (Tm, black squares) corresponding to 50% ordered material, for a range of agarose oligomers determined by experiment.

- an estimation of the enthalpy and entropy of the transition, in kcal per mole of residues (ΔH, ΔS).
- the parameter γ

The size of the chains was varied between 11 and 35; these are the same sizes which were studied in chapter five. A value of -1.5kcal per mole of residue pairs was taken from chapter five as a starting value for the enthalpy. The error in this value is $\pm$0.5kcal.

The parameter γ and the entropy ΔS are unknown. Therefore the model is used to fit the experimental data from chapter five and find values for them.

The first approach is to reproduce the transition temperatures T_m for each chain size. These were found by experiment (chapter five) and are accurate to within a few degrees.

The value of ΔH calculated by experiment is much lower than that typically expected for these types of transitions [101, 104]. Therefore the experimental error was taken into account and the largest possible value of -2.0kcal used.

At the transition temperature it should be noted that

$$\Delta S = \frac{\Delta H}{T_m};$$

this gives a guideline for choosing an initial value of ΔS. ΔS and γ were then adjusted in order to fit the experimental data curves as closely as possible. The results shown in figure 6.3 are the best fit to the experimental data.

The value of ΔS which best fits all the experimentally determined T_m's is -6.2calK^{-1} per mole of residues. The best value of gamma is equivalent to $\beta = 8 \times 10^{-5} \ell \text{mol}^{-1}$.

Discussion

It was found that an intrinsic property of the model is a trade off between γ and ΔS. On the one hand T_m can be fitted very well, but then the onset of ordering is too high. On the other hand if a poor fit to T_m is accepted, the transition is predicted to occur over a range of about 25°C, as observed by experiment. The experimentally determined T_m is accurate to within a degree or so, therefore the model had to reproduce this as well as possible. The value of enthalpy and shape of the order/disorder curve for DSC measurements are less precise, so these were allowed some flexibility. Overall a

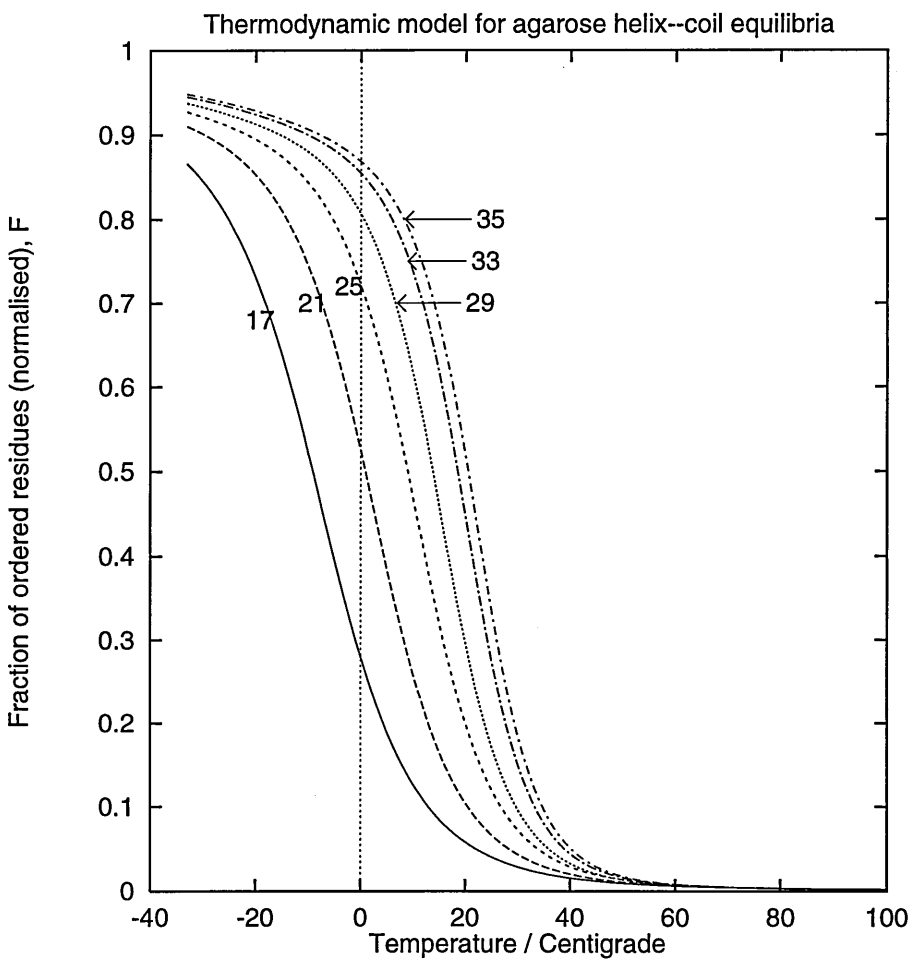


Figure 6.3: The fraction of residues in the ordered (helical) conformation, calculated using the Zimm–Bragg based model for a staggered zipper (see text). The numbers next to the lines indicate the size of chain, in residues.

reasonable fit was obtained, but it is clear that the model is not correct for the experimental data. There may be two reasons for this.

Firstly, aggregation may disturb the equilibrium; this is not accounted for by the model, and may move the transition temperature.

Secondly, aggregation may trap species which are not of the "all helix" conformation, again meaning that the model is not appropriate. In addition, residues trapped in the coil state will not be able to contribute to the enthalpy measured by DSC, so the estimated value used above might be too small.

The model predicts a small value of β: $8 \times 10^{-5}\,\ell/\mathrm{mol}$. This indicates that agarose strands do not pair up readily, and therefore need to be quite long before helices form.

In summary, the model does not reproduce DSC data for the agarose coil to helix transition. This is likely to be due to aggregation and non-cooperativity in the agarose system. (In other words, residues are not encouraged to be in the helical state by neighbouring helical residues.) In DNA type systems such as those studied by Applequist *et al.* [101], residues in paired chains are specifically bonded and short chains are cooperative. In agarose there is probably never a 100% fraction of helical residues, and so coil–like regions of chain may be more important than has been hitherto assumed.

6.3 Modelling of agarose using supercomputers

6.3.1 Introduction

It would be of interest to model agarose chains of varying length using molecular modelling techniques, to see if any of the above points could be reproduced. A start on this task was made in chapter four, where very short agarose chains (five or six residues) were modelled in and out of water. We know however that chains of this size do not form stable helices, but unfortunately the computer workstations available in the laboratory are unable to perform simulations on larger molecules.

100 processing hours was therefore obtained on the DEC supercomputer `Columbus` at the Rutherford Appleton Laboratories for this purpose, and this was all used.

6.3.2 Chain conformation and helix stability

Agarose chain conformations for chain lengths of 13,17 and 19 residues were studied.

In real aqueous systems, the winding up of two coils to form a double helix takes far longer than is possible to simulate. Even unwinding (melting) is still impossible to observe in full; except for extremely short chains, the process simply takes too long.

Therefore the only strategy is to design molecular conformations which may exist at various stages along the helix to coil reaction pathway, and then perform calculations on these. This is illustrated in figure 6.4, which shows the conformations used for the 13 and 19 residue chains.

Simulation conditions

The AMBER forcefield was used for modelling the carbohydrates, which is the same forcefield used previously. The general method for simulation using the `Columbus` x–terminal interface is as follows:

1. Build sugar residues. For agarose we require four residues:

   - galactose
   - anhydrogalactose
   - galactose with a terminal -OH group
   - galactose with a starting HO- group.

   Set the potential types and partial charges for each atom.

2. Use the local workstation to prepare the desired chain conformation, then transfer this to the supercomputer.

3. Create a command file for the dynamics simulation.

4. Put the calculation on a batch queue.

All the starting conformations were relaxed to relieve strained regions before dynamics calculations.

The conditions used for dynamics calculations were (1 iteration $\equiv$ 1fs):

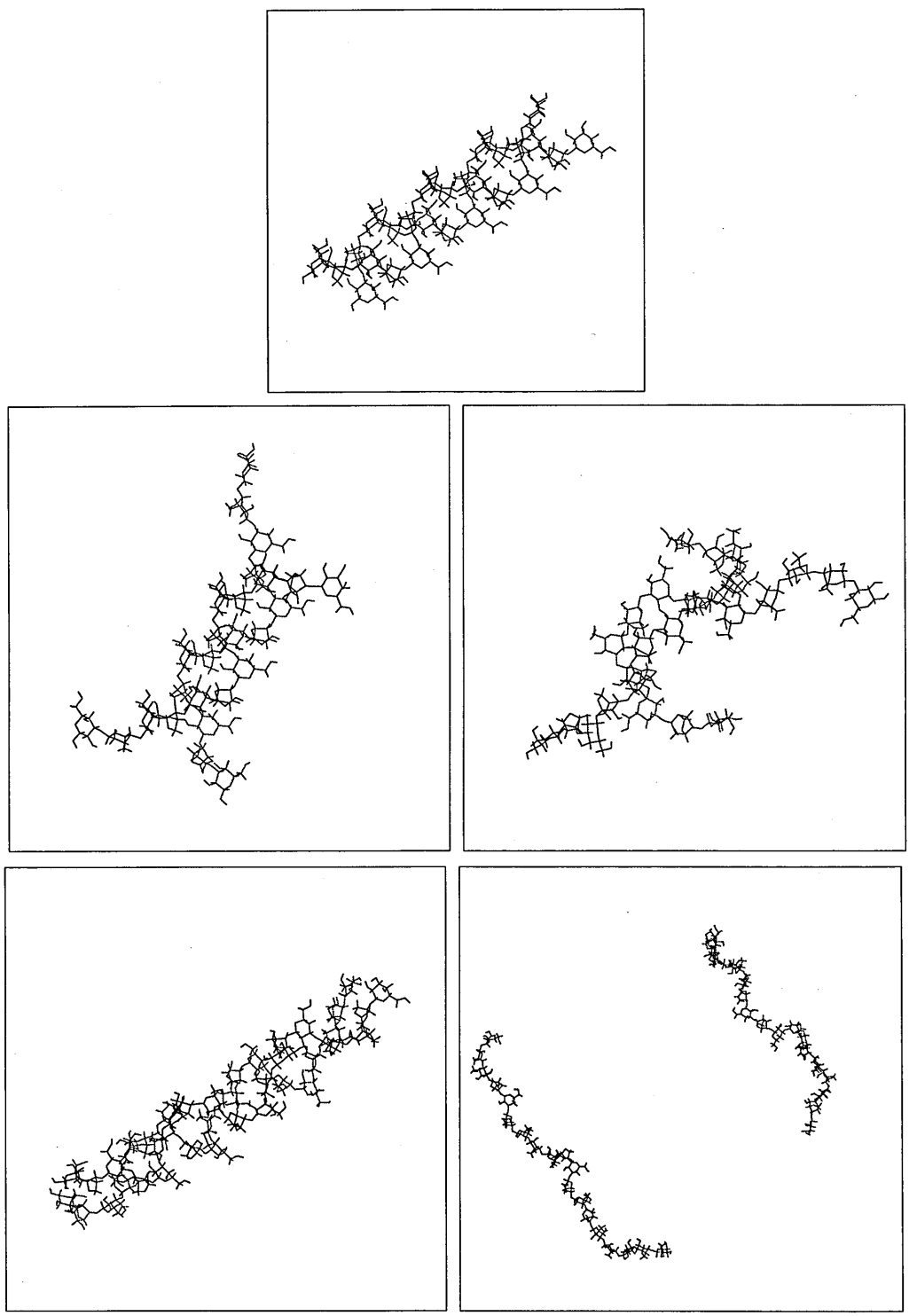


Figure 6.4: Starting conformations for molecular dynamics. The top three pictures are respectively perfect helix, partial helix and fully random coil forms of a 13 residue chain. The lower two pictures are for the 19 residue chain, showing a perfect helix and two isolated random coils.

```
dielectric        =    80
initial T         =     0K
running T         =   300K
cutoff            =    10A
no. iterations    = 50000 or more
data stored       = every 500 iterations
```

A temperature of 300K was used in order to be in the equilibrium region for agarose chains of the sizes chosen.

6.3.3 Results and discussion

Energy difference between coils and helices

The results in figure 6.5 give the total energy of fully helical, partially helical and fully extended conformations of agarose chain pairs calculated for dynamics simulations. Three sizes of chain were studied.

In each case there is an induction period of about 30 picoseconds during which the system settles down to a stable energy after heating from 0K. This period is ignored when energies are being discussed.

Total (average) energies of the coil compared to the helix for each chain size are:

Chain size	**Coil - helix energy difference** (kcal/mol)	**Energy per residue**
13	-31	-1.2
17	-31	-0.9
19	-107	-2.8

The differences reflect insufficient exploration of different random coil conformations. If enough computer time was available repeat experiments would be done in order to produce statistically reliable results. Qualitatively, it is clear that the total energy of two random coils is indeed higher than the total energy of the corresponding double helix on the 50 picosecond time-scale. A very rough estimate of the helix binding energy per residue is -1.6 kcal/mol of residues from the table above. This agrees very well with the value used earlier for the Zimm-Bragg model (2 kcal/mol).

The source of this stabilisation energy is the Van der Waals energy term (see figure 6.7). Therefore the orientation of the surfaces of the two molecules when pairing up will be of importance in determining the probability of ordering.

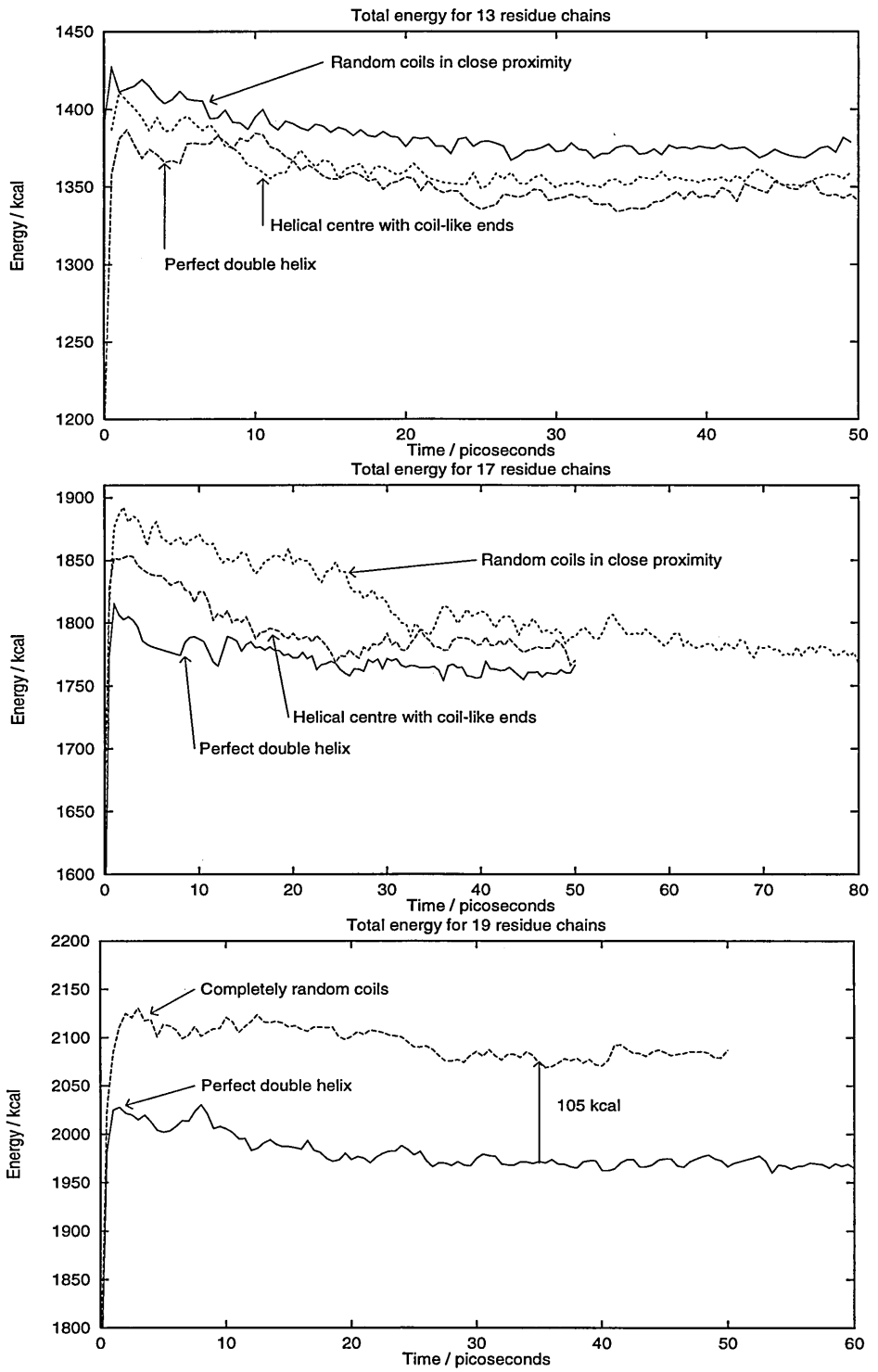


Figure 6.5: Total energy of two strands (lengths indicated on graph titles) for various initial conformations of agarose molecules.

Helix dynamics

Snapshots for the 13 and 19 residue chains, beginning as double helices, are given in figure 6.6. Although the helices relax and bend during the simulation they do not unwind at all. The Zimm–Bragg model applied earlier is based on this kind of system ("all or nothing"). In the discussion in chapter five we assumed that as many as 17 residues did not take part in the double helix; the current simulations seem to dispute this. Again however, the effect of aggregation has not been taken into account here.

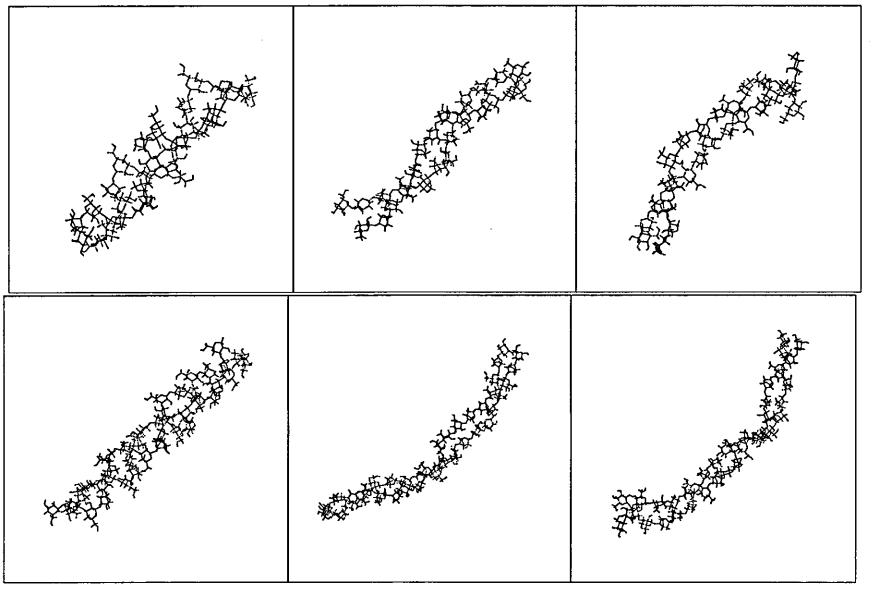


Figure 6.6: Dynamics calculations on agarose double helices: sizes 13 (above) and 19 (below). From left to right the snapshots correspond to 1, 20 and 50 picoseconds respectively. These images can be compared with the relevant points on the energy graphs in figure 6.5.

These images show that the helices actually relax during the simulation, and the partially helical conformation also evolved towards a similar structure (results not shown). Initially, x-ray conformations were specified for the helical sequences. These do not appear to be stable during dynamics calculations, as can be seen from the following table of helix parameters (a 'turn' means one complete revolution about the helix axis):

	Start		End (50ps)	
Chain size	No. residues per turn.	Height of one turn (Å)	No. residues per turn.	Height of one turn (Å)
13	6	19	8	27
19	6	19	7	24

Detailed conformational analysis would help to throw light onto this[5]. The relaxed helices may be an intermediate stage in the complete dissociation of the two strands, or they might be a more stable helix than the initial one. The simulations here are not long enough to provide an answer, but what is certain is that the relaxed helix has a lower energy than the x-ray helix. This difference in energy arises from the dihedral energy term (see figure 6.7. Note that these results are only for the 19 residue chain, which displays the biggest difference.).

As was shown in chapter four, a small change in the dihedral angles at each glycosidic linkage can result in large changes in energy and conformation over the chain as a whole.

6.3.4 Solvent

Two simulations, each modelling two 19-residue agarose chains, were done on the supercomputer using explicit aqueous solvent; the all–helix conformation and the all–coil conformation.

The agarose strands were prepared for simulation using the `alignAxes`, `solvateBox` and `saveamberparm` commands within the `xleap` interface. This sequence creates a periodic box aligned with the molecular axis, builds the solvent around the solute and sets the AMBER potentials. The walls of the periodic box were set to be no closer than 5Å from any part of the solute, generating 4000 water molecules.

Results and discussion

The results of both simulations are shown visually in figure 6.8. The all–helix model relaxed in a similar way to the same model in implicit water (see figure 6.6). Again, the two strands do not unwind. The average height of one turn in the relaxed helix is 27Å (compare to 19Å in the x-ray helix and

[5]Unfortunately the software on the supercomputer does not permit dihedral angles to be analysed, and the output files could not be read into the InsightII analysis package.

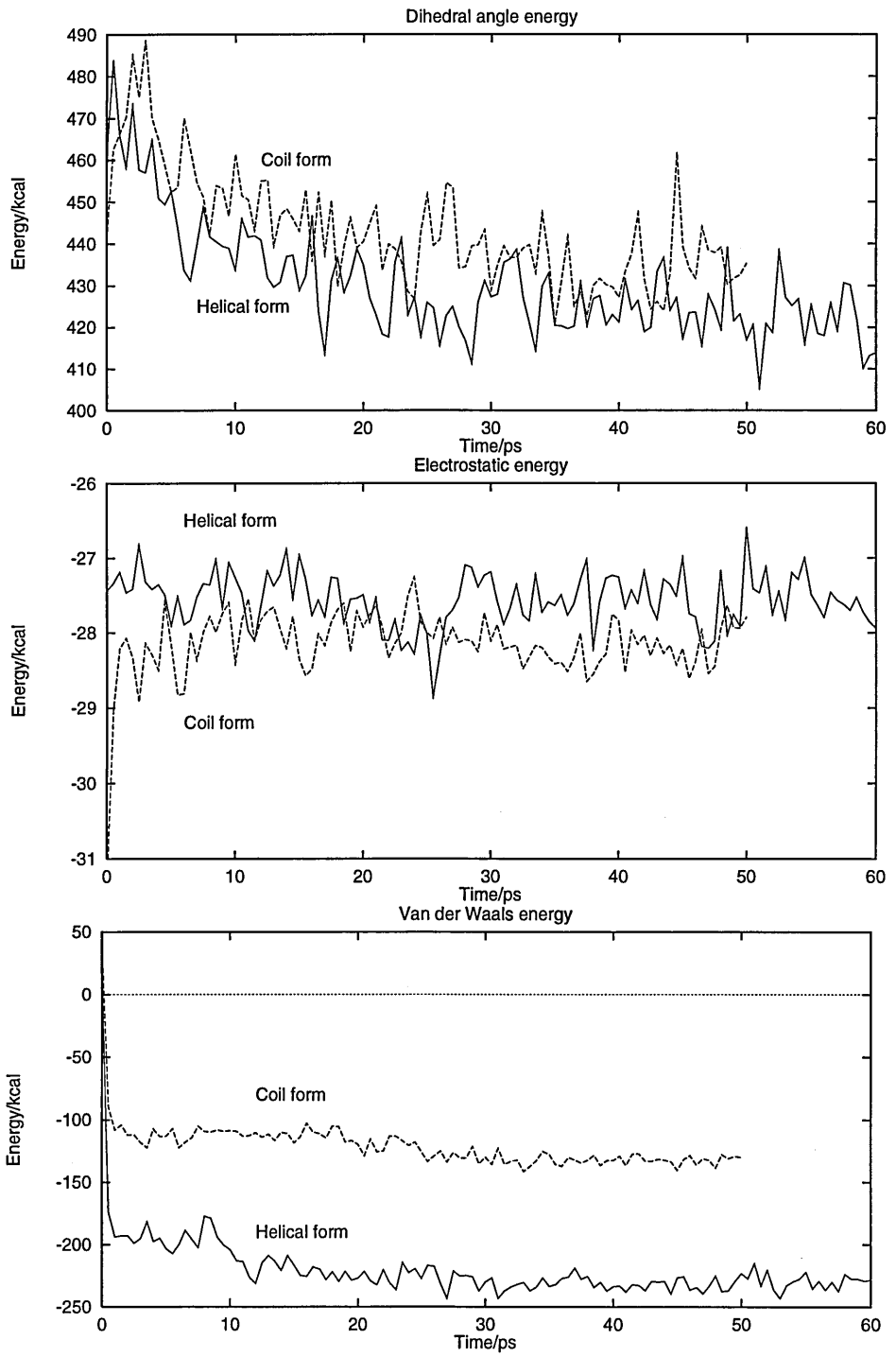


Figure 6.7: Comparison between coil and helix for 19 residue chains: specific energy terms for dihedral angles, electrostatic and Van der Waals interactions.

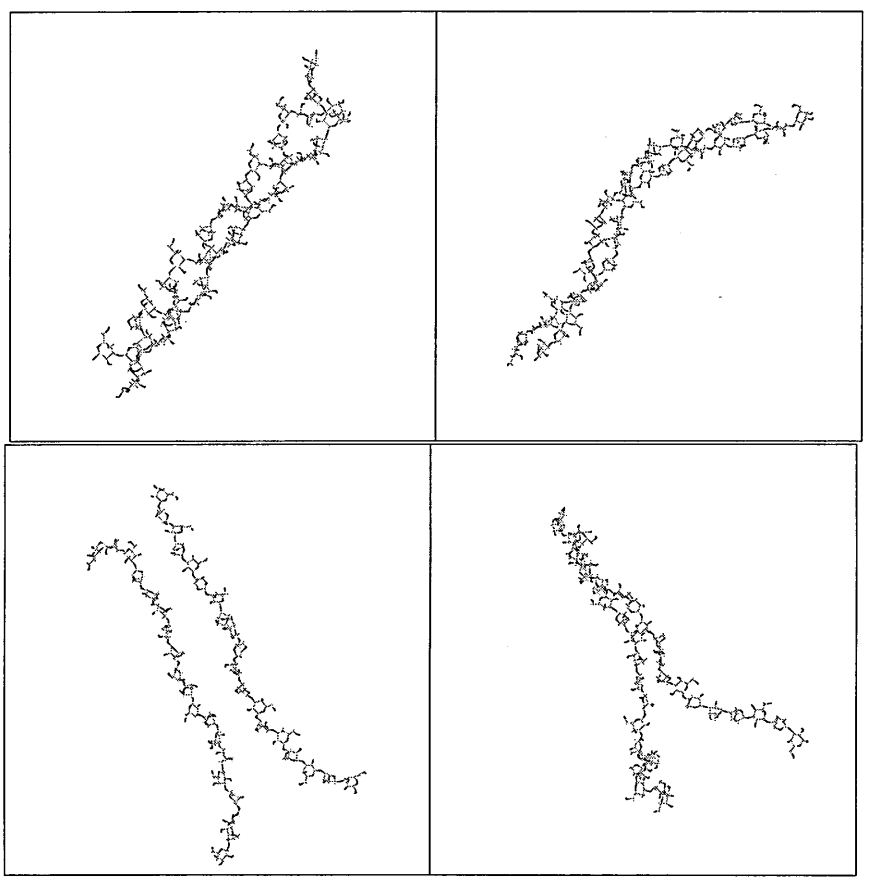


Figure 6.8: Dynamics simulation of 19 residue chains with the inclusion of specific water molecules (to keep this figure simple, these are not shown). Top: the all–helix simulation. Bottom: the all–coil simulation. The right hand images give the result after 50 picoseconds.

24Å in the implicit water simulation) and there are about 7 to 8 residues per turn.

The result for the random coil simulation in explicit water parallels the earlier result in implicit water, with the two carbohydrate molecules meeting up and forming a paired double strand. No helix was formed.

The energy of the simulation cannot be analysed as was done for the results with implicit water, due to the overwhelming contribution from the solvent molecules which masks the carbohydrate itself. To indicate the complexity of the system, a colour snapshot of two random coils surrounded by water molecules is given in figure 6.9.

6.4 Agarose coil interacting with double helices

So far no mention has been made in this chapter of what happens to double helices after they have formed. We know from the previous chapter that the formation of microcrystalline aggregates is a favourable process.

In agar gels the equivalent of the aggregate is the *double helix bundle.* Various workers have looked at the dimensions of these helix bundles, and have estimated their thickness [27, 38, 90, 98]. The results indicate that the bundles probably contain six to eight double helices. It seems likely that bundles will be linked by coil–like chains, either as part of helices within bundles or bound to bundle surfaces.

A hypothetical agarose junction zone containing six double helices was created and packed using a hexagonal cell. The length of side of the cell (a and b sides) determines the size of the voids between junction zones: see figure 6.10 and 6.9. For simulations the free coil (length 14 residues) was then placed into a void in the cell.

For the simulations the actual unit cell used was:

$a = b = 56.8$Å.
$c = 55$Å.
$\alpha = \beta = 90°$; $\gamma = 60°$

Calculations were done on a Silicon Graphics workstation and all double helices were kept fixed; only the coil was allowed to move. The free coil was energy-minimised, then warmed up in steps of 25 kelvin every 400 iterations. Dynamics were executed for 100 picoseconds at four temperatures. The initial assembly is shown in figure 6.9.

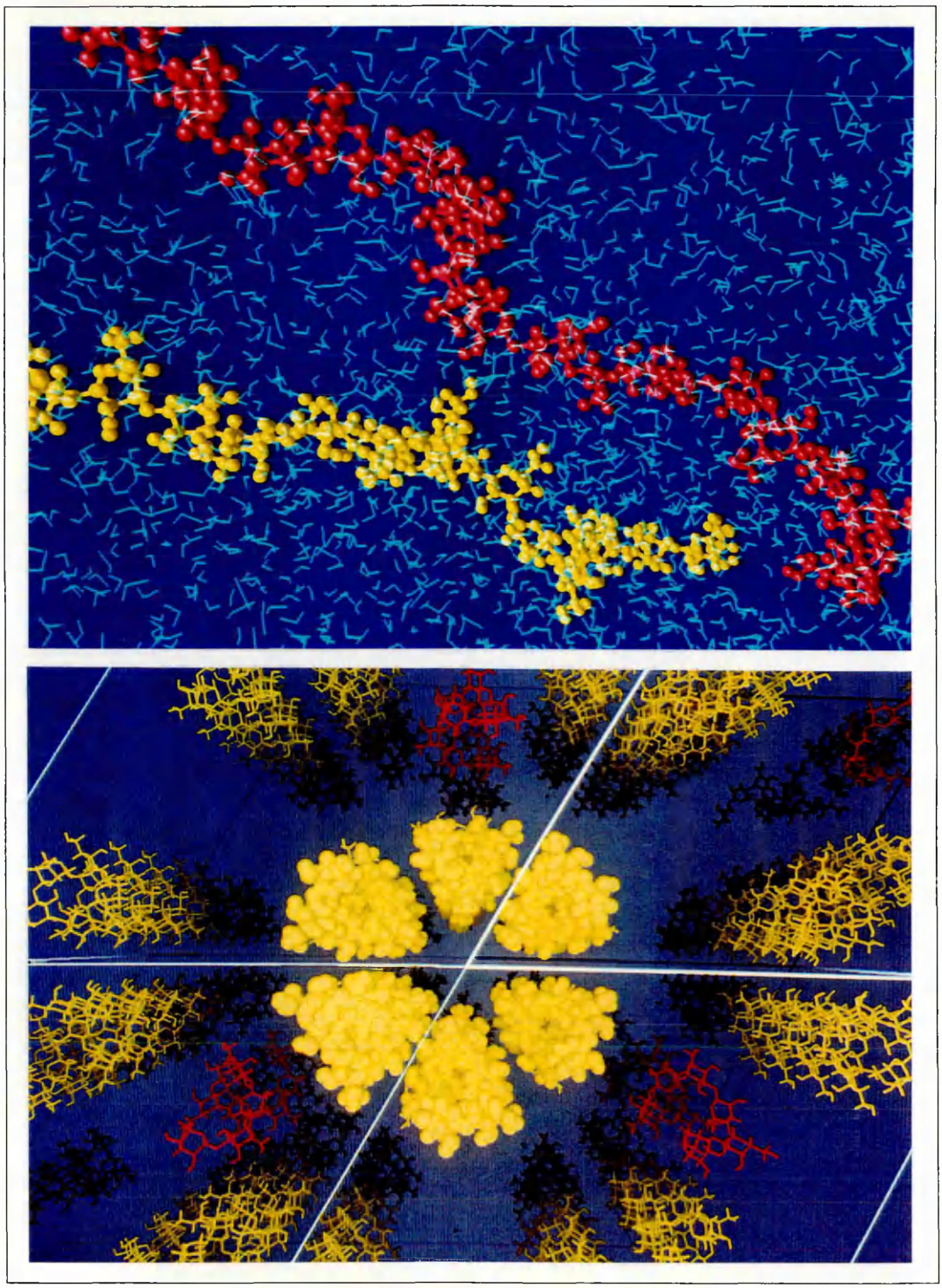


Figure 6.9: Images representing two simulations involving large assemblies of atoms (photographed from computer screen). Top: agarose fully solvated in water; bottom: agarose random coils positioned in between bundles each comprising six double helices.

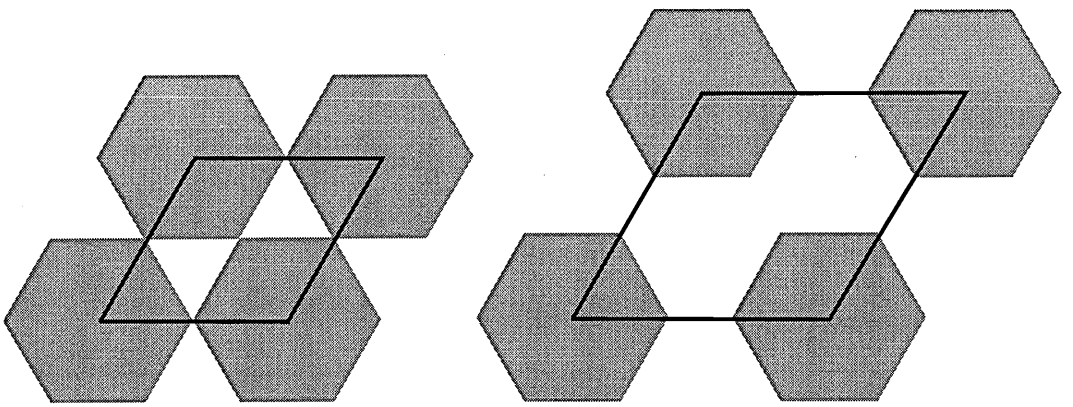
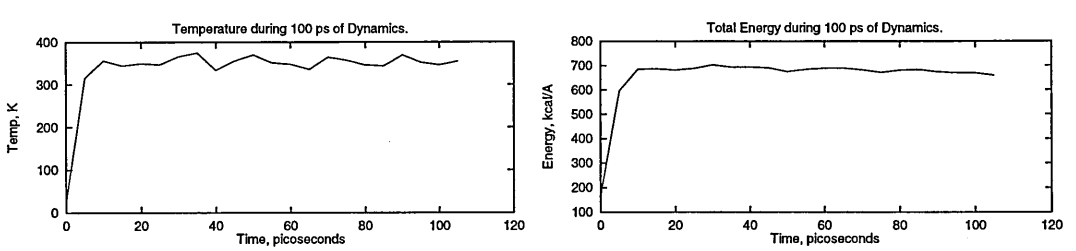


Figure 6.10: Unit cells of different side length, shown as black lines. The grey hexagons represent bundles of six double helices.

6.4.1 Results

During the simulation, temperature and total energy were monitored to confirm that they remained stable:



These graphs were very similar in form for all three cases.

In the descriptions below, "binding" means that the molecules concerned are in contact with regard to the Van der Waals surfaces, and that hydrogen bonding is able to occur between the participating residues' –OH groups. In addition a residue was not considered to be bound unless it remained in one position for several picoseconds.

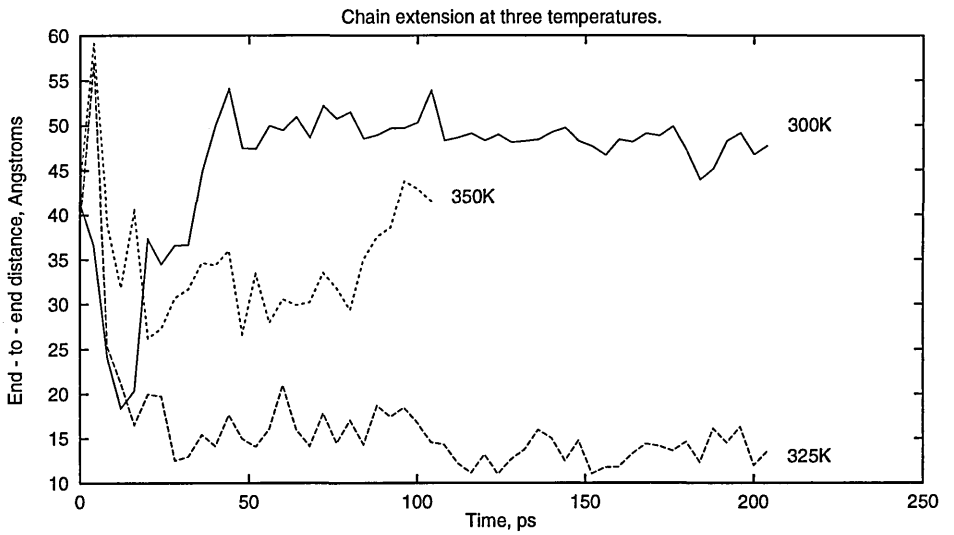
Temperature	Observations
300K	After 16 picoseconds there was an interaction with one bundle. The coil had only drifted about 30Å before this. After 60 picoseconds six residues were firmly bound and this did not change much thereafter. The chain shape was that of a floppy loop (not fully extended).
325K	The coil was again binding after about 17 picoseconds and was firmly bound at about the same stage as in the above case. About five residues were bound. This did not change further.
350K	The free chain was binding with one bundle after 17 picoseconds: this occurred very suddenly within 2 picoseconds. Then between 26 and 30 picoseconds it was observed to unbind and drift away. After 50 picoseconds had passed the coil became bound to two bundles (20 Å apart) and in total six residues were bound.
375K	After 21 picoseconds the free coil was linking two bundles, binding with residues 1 and 2 on one and 8 and 10 on another about 20Å away. This arrangement persisted until the end, with only slight movements of a few Å when the binding position was adjusted.

6.4.2 Discussion

There were only minor observable differences between the three cases given here. Although we are dealing with a short coil here (14 residues), which should melt out into solution at about 300K, it must be remembered that

- there are frozen molecules present which can help to trap the free coil
- the simulations were only 100 picoseconds long.

Chain extension became steady after binding, and showed no discernable pattern with respect to temperature. Binding to the frozen helices greatly reduces the freedom of the chain.



Dihedral angles visited are a good indication of the flexibility of a chain, as demonstrated earlier (see chapter four). However binding reduces freedom and can favour unusual conformations to some degree. This is illustrated below in figure 6.11.

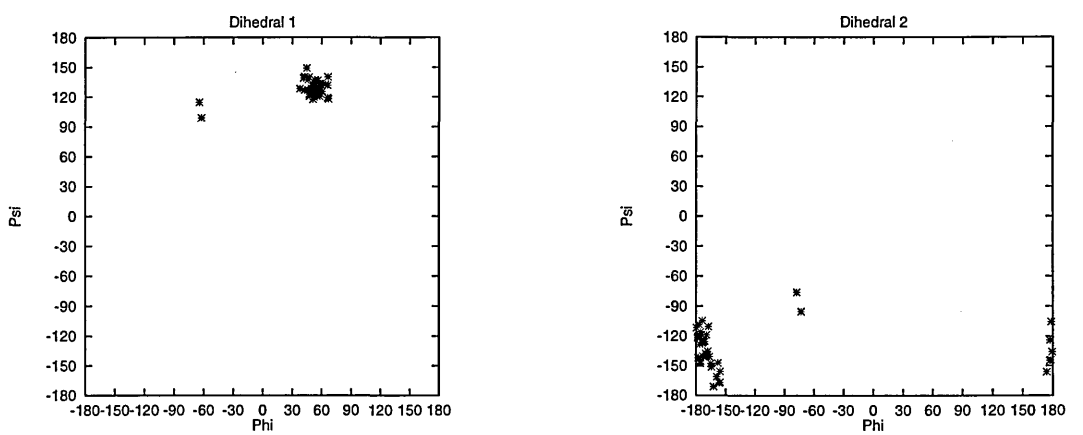


Figure 6.11: Two dihedral angles monitored for the simulation at 325K. The left hand plot corresponds to the linkage between residues 5 and 6; the other to that between residues 4 and 5. The conformations show little variation after binding has been established.

6.5 Summary

A helix–coil theory originally developed for double stranded DNA with mismatching was applied to the agarose system and compared with experimental DSC data. It was assumed that the DSC observed 100% ordering and that equilibrium was established; the comparison indicated that this may not be valid. However some agreement was found. ΔS was found to be -6.2calK^{-1} per mole of residues; β was $8 \times 10^{-5}\ell\text{mol}^{-1}$. This indicates that agarose chains need to be long in order to pair up. This agrees with the minimum size for ordering measured by DSC (17 residues).

Modelling of long agarose chains (in and out of water) using a supercomputer showed that the double helix has a stabilisation energy with respect to random coils of about -2kcal/mol residues. The main source of this is the favourable Van der Waals interactions between closely aligned strands. It was also found that double helices calculated from fibre x-ray diffraction relaxed to an extended version with a repeat distance of 25Å. The reason here is a reduction in dihedral angle energy.

Finally it has been demonstrated that agarose coils can link helical bundles together by binding to their surfaces, and can bridge 20 or 30Å of space in this manner.

The modelling indicates that helices in fibres are not realistic when solutions are being discussed, and single random coils may more important than hitherto suspected in the gel structure.

Chapter 7

Conclusion

Contents

7.1 Outline summary

The main outcome of this thesis has been to show quite vividly that the structure of agarose, a simple polymer of two sugar units, is very complex particularly when its solution in water is considered. A hint of this was given first of all in the introductory chapter when a brief review of other work done on related gelling biopolymers was presented. In order to learn more about these types of molecules, every possible technique available to the chemist needs to be employed. This thesis has looked mainly at the application of computer modelling to the task, with some support also from bench analytical techniques which studied 'the real thing'.

Chapters 3,4,5 & 6 cover the major research topics, and the results from these are summarised below.

7.2 Achievements

Firstly the various structures of agarose and carrageenans which have previously been elucidated were reviewed, using the relevant literature. To enable sulphated carbohydrates to be modelled, forcefield parameters were located, also in the literature.

Energy maps which characterise the glycosidic linkages found in the different carbohydrates were calculated. This revealed that the x-ray derived configurations corresponding to molecular helices were not quite the same as the configurations of lowest energy. Thus some other factors, not represented by simple energy calculations, were operating. We saw that there are several energy minima accessible at room temperature with only a slight energy cost, and maybe in 'crystal field packing' such as found in dried gels these could be accessed.

Some model helices were built in readiness for simulation. At this stage it was noted that a chain would need to be very long in order to favour, in solution, a deviation from the ideal minimum energy conformation seen in energy maps.

The first simulation was a comparison between the two techniques of solvent representation, namely implicit and explicit. Dynamics calculations on six ring double agarose helices were used to investigate favoured conformations at glycosidic linkages and these were compared to minimum energy positions. Correlation plots showed that in explicit water the chains were stiffer. Diffusion rates were also greatly reduced.

However it was also seen that the actual conformations favoured by explicit and implicit simulation were similar to within a few degrees. Therefore in

the long term the use of the (economical) implicit technique seems to be vindicated.

An obvious result from this work was also that a six-ring agarose chain is not going to be stable in the double helix conformation! It was not even close to being stable: both molecules separated very quickly.

Chapter five contained exclusively experimental work, on real life agarose molecules. Agarose itself is too complex to analyse; the molecules are too big and the structure cannot be probed by current techniques. By using monodispersed samples of short chains (oligomers) one process could be observed, namely helix to coil ordering, which was touched upon in previous chapters.

DSC showed without doubt that as agarose chains become longer the temperature (during cooling) at which helices form becomes higher. If the chains are shorter than 17 residues no helix will form in a 1% solution. The melting of the ordered phase showed a similar trend: as the chains become longer the melting of the ordered phase occurs at higher temperature. By extrapolation to the gelation temperature of agarose itself, an estimate of the average size of (helical) junction zones in agarose gels was set at about 40 residues, or 120Å.

By using DSC to study the temperature dependent phase transition, it was shown that at least two processes are observed: ordering and aggregation. These could not be separated by DSC even when special routines such as fast cooling and cooling to just below the transition temperature were employed. The cooling curve showed that following ordering, which gives a large measured enthalpy, there is long period of aggregation. The heating curve demonstrated that melting occurs at a higher temperature than ordering; thus melting is kinetically controlled (by aggregation), just as it is in the agarose polymer itself.

Unlike for agarose however, the enthalpy per residue for the conformational coil to helix equilibrium could be estimated, because the chain sizes were known. It is not clear exactly how much of the measured enthalpy is due to ordering itself. It is generally thought that aggregation only accounts for a small part. An estimate of the enthalpy of ordering per residue was thus found to be -1.5kcal/mol (these units were used so as to be consistent with other workers).

The experimental work gave valuable insight into the properties of agarose oligomers, and helped to put the modelling work into context. With theoretical results only, it would be easy to get carried away and begin to assume that all the problems had been solved and features correctly represented. The DSC results serve to remind one that we are dealing with a real and

complex system whose internal structures are still not fully understood.

The experimental work on agarose of known chain size raises the natural question "can the helix–coil equilibrium temperature be predicted using a theoretical model?". Work has already been published in this area, and was applied specifically to agarose for the final chapter in this thesis. Several models exist; the most appropriate one for agarose is for mismatched double helices. Using an iterative procedure appropriate values for the variables in the model were found which fitted the experimental transition temperatures. Hence, a value of -6.2calK^{-1} per mole of residues was found for the entropy of the coil to helix transformation. The model was unable to predict a minimum size for helix formation, but this is not surprising given that it ignores end-effects. The model indicated that agarose chains do not readily associate into pairs; a pre-requisite for double helix formation and confirmation that agarose chains need to be large before association occurs.

The helix–coil equilibrium model is based on statistical mechanics. However using supercomputers and molecular modelling it is also possible to study the solution behaviour of any particular oligomer directly, taking into account each atom individually. Therefore agarose chains with sizes ranging from 13 to 19 residues in length were simulated on such a computer, and this formed the second major molecular modelling investigation. The double helices were found to be stable on the 100 picosecond time-scale with respect to the same molecules in a random coil configuration. In fact there was a stabilisation energy of about -1.5kcal/mol of residues. However the modelling work also predicted that the agarose dried–fibre x-ray type configurations (height per turn 19Å) used as starting points for the calculation would relax into more extended helices with a height per turn of 25 to 30Å. These longer, thinner helices were more stable and more flexible than the starting configuration. Even when explicit solvent was used the same behaviour was observed. Therefore, in solution, the helical conformation of agarose seems to be different to that observed in a dried fibre.

7.3 Future and related work

The work summarised above only touches on one aspect of a complete description of agarose. Agarose has many applications, and there are other ways of probing its structure. Some related topics and possible avenues for future work will now be briefly mentioned.

Optical rotation

It has already been stated that optical rotation measurements taken during the cooling of agarose random coils would be of great interest, since the conformational changes (ordering) could then be entirely separated from the aggregation event. Thus, a better estimate of the coil to helix equilibrium temperature and its dependence on chain size could be obtained. Optical rotation measurements could be used to characterise the effect of aggregation on the enthalpic transition measured by DSC, and to see how cooperative the coil to helix transformation is.

Computational models

Opportunities exist for more computational models to be applied to agarose. The concept of a random, statistical, "paperclip" chain to represent polymer chains is not new. Modules exist in molecular modelling packages (for example in the MSI Cerius^2 software) which permit dihedral angles in a long polymer chain to be varied while the monomers themselves remain rigid. This results in a saving of computer time, allowing large chains to be simulated. Such calculations on agarose would increase the scope of modelling studies, and could be done in the future.

Gel electrophoresis

Agarose, the pure form of agar, is used not only in medical and food applications as described elsewhere in this thesis, but also in analytical chemistry where its porous properties, in gel form, permit it to be used for chromatographic separation techniques. In this context, its internal structure becomes extremely important.

In gel electrophoresis, particles are separated on the basis of the size of their charge, by inducing them to move through a fixed medium (the gel) under an applied electric field [105]. The essential characteristic is that the fluid medium is held in one place, allowing ions to diffuse according to their charge[1]. The gel medium must not adsorb the ions of interest (these may be proteins, nucleic acids or rare earth metals), and must be inert. Agarose proves to be ideal for many applications [105].

What is interesting is that many aspects of this subject are unexplored, or only accounted for by various "coefficients" in electrophoresis theory [105].

[1]This is different from gel permeation chromatography, where the stationary phase is in the form of beads.

As an example, the friction coefficient (experienced by particles migrating through the gel network) is well known but not fully characterised. Possibilities exist here for molecular modelling. Using the estimates for junction zone size for example, molecular dynamics of a hypothetical gel network with charged particles migrating through could be done. The network itself need not be moved. Similarly molecular sieve effects could be explored.

Nuclear magnetic resonance

Nuclear magnetic resonance, a physical technique of great importance in the study of biopolymers, has not been used in this thesis. However many groups have done work on carbohydrates using this method [106, 107].

Although gels do not remain stable in the NMR sample tube, there is no reason why double helices formed from short agarose chains should not be perfectly able to withstand tumbling motions. Useful structural information about these helices could be extracted from such a study. Removal of the labile hydroxy-hydrogens by deuterated water would greatly clean up proton spectra from such samples.

Infra-red spectroscopy

Another possible way to study the internal structure of agarose double helices is to use infra-red spectroscopy. It may well be possible to monitor the extent of hydrogen bonding between agarose chains during helix to coil transitions. The experiments would have to be done in a non-hydrogen bonding solvent such as CCl_4, meaning that any hydrogen bonding observed must be due to carbohydrate–carbohydrate interactions.

The -OH stretching frequency is broad if hydrogen bonding is able to occur due to the variation in O–H bond strength [51]. Therefore if hydrogen bonding does exist within double helices two -OH frequencies should be visible: a sharp one due to external -OH groups and a broad one due to internal hydrogen bonded groups.

The main problem would be that agarose coil to helix transformations are essentially an aqueous phenomenon, and so would probably not occur in CCl_4, but instead the molecules would simply precipitate. Nevertheless if any hydrogen bonding could be measured it would be of use.

Carrageenans

The huge area of agarose–related polyelectrolytic polysaccharides has not been covered in this thesis to any significant degree. Many (such as iota carrageenan) have very good gel-forming properties, and helices have been

characterised by x-ray diffraction. Here, there is a big potential for the study of different ions and functionality possible on the carbohydrate chain, and NMR can be used to follow the locations of some ions [73, 108].

Finally the study of carrageenan and protein reactions (not seen with agarose) merits attention [109].

7.4 Conclusion

Molecular modelling and sensitive laboratory techniques have allowed the examination of one small part of the agarose gelling system; a picture of the full gel network still remains elusive.

As molecular modelling software becomes more sophisticated, computers become more powerful, and laboratory imaging techniques constantly improve, we can expect that more and more interesting details concerning the intimate processes at work behind apparently simple bio-molecular interactions will be revealed. With the advent of faster personal computers and the growth of operating systems such as LINUX, huge assemblies of atoms such as a bundle of agarose helices no longer have to be studied in the laboratory, but can now be observed at home. Meanwhile laboratories will continue to reveal an even more detailed picture of the internal structure of these bio-molecules.

Appendix A

Computer codes and forcefield data

This appendix contains the DISCOVER macro command file which calculates a two dimensional map of energy for glycosidic linkages, and two short FORTRAN codes used for calculations in the thesis. The first FORTRAN program was originally written by Charles R.Heald [110] in 1993 to obtain correlation functions for various angles in a polymer chain. It was adapted to deal specifically with the dihedral angle pairs found in carbohydrates, for which data was obtained from the Insight modelling package.

The second code predicts the percentage of residues in an agarose chain which will be in a helical conformation at any specified temperature. This is the Zimm-Bragg relation adapted for short double helices [101]. The code was originally composed by Professor A.H.Clark (at UNILEVER), and adapted slightly for its present application (see chapter 6).

DISCOVER macro for analysing glycosidic linkages

```
!  INPUT FILE FOR DISCOVER GENERATED BY INSIGHT
!  AND ADAPTED BY N.M.W.HAGGETT 1996
!
      overlap = 0.01
      cutoff = 15.000000
      cutdis = 14.000000
      swtdis = 1.5

      begin simulation
     *    add-automatic bond torsion valence out-of-plane
      reduce
```

```
!
      set dielectric  = 80.000000
!
!  DEFINE THE TORSION ANGLES
!
!
      Assign the torsion name  cam to
     * residue ALGA    1 atom O
     * residue ALGA    1 atom C1
     * residue ALGA    1 atom O1
     * residue BDGA   1B atom C3


      Assign the torsion name  oxo to
     * residue ALGA    1 atom C1
     * residue ALGA    1 atom O1
     * residue BDGA   1B atom C3
     * residue BDGA   1B atom C4
!
!  BEGIN INCREMENTATION OF ANGLES AND
!  EVALUATION OF MOLECULE ENERGY
!
      loop0 = 1
      angle0 = -180.00
      ifnum = 1
!
!
L0    force  cam in  ALGA 1  to angle0 using 1000.000000
      loop1 = 1
      angle1 = -180.00
!
!
L1    force  oxo in  ALGA 1  to angle1 using 1000.000000
!
!
      Minimize
     *    no cross terms
     *    no morse
     *    for 100 iterations
     *    using steep descents
     *    until the maximum derivative is less than 10.000000000 kcal/A

      Minimize
```

```
     *    no cross terms
     *    no morse
     *    for 900 iterations
     *    using conjugate gradients
     *    until the maximum derivative is less than 0.1000000000 kcal/A
!
!
      archive as file number ifnum
      ifnum = ifnum + 1
!
!
R0    loop1 = loop1 + 1
      angle1 = angle1 + 12.00
      if loop1 .le.   31 then L1
      loop0 = loop0 + 1
      angle0 = angle0 + 12.00
      if loop0 .le.   31 then L0
!
!
      end
```

Program to calculate the time - correlation functions for glycosidic linkages

```
       PROGRAM CORR
C
C  --  PROGRAM TO OBTAIN THE NON-NORMALISED CORRELATION COEFFICIENT
C      OF VARIOUS ANGLES IN THE BACKBONE OF A POLYMER CHAIN
C
C  --  CREATED: 5 DEC 1993, CHARLES R HEALD
C  --  Edited July 1997 to work with different data, N M W Haggett
C
       PARAMETER(JAMOUNT=2500,JCOR=2)
       DIMENSION TORS (JAMOUNT,JCOR),norm(jamount)
       REAL COREL(JAMOUNT),time(jamount)
       CHARACTER*1 CREPLY
C
C  --  READ IN DATA FILE AND PUT DATA INTO PROCESSING ARRAY
C
10     CONTINUE
C
       OPEN(UNIT=1,FILE='dihedral.dat',status='old',ERR=10)
       JTIME=0
```

```
20      CONTINUE
        JTIME=JTIME + 1
        READ(1,*,END=100)(TORS(JTIME,J),J=1,JCOR),time(jtime)
        WRITE(*,*) TORS(JTIME,1)
        GOTO 20
100     CONTINUE
        CLOSE(1)
C
C  --   COSINE ALL ANGLES
C
        DO 200 J = 1,JTIME
           DO 202 J1 = 1,JCOR
              TORS(J,J1)=COS(TORS(J,J1)*(3.14159/180))
202        CONTINUE
200     CONTINUE
C
C  --   PROCESS THE DATA TO GET THE NON-NORMALISED CORRELATION FUNCTION
C
205     CONTINUE
        JCOR1=1
        JCOR2=2
C
C  --   CLEAR OUTPUT ARRAY
C
        DO 210 J = 1,JTIME
           COREL(J) = 0
210     CONTINUE
220     CONTINUE
        OPEN(UNIT=2,FILE='dihedral.res',status='unknown',ERR=220)
C
C  --   LOOP OVER ALL STRUCTURES SO AS TO OBTAIN AN AVERAGED ENSEMBLE
C       OF RESULTS
C
c        DO 230 J = 1,JTIME - 1
c           DO 240 J1 = J + 1,JTIME
c              COREL(J1) = COREL(J1) + TORS(J,JCOR1)*TORS(1,JCOR2)
c240        CONTINUE
c230     CONTINUE
C
        DO 250 J = JTIME,2,-1
           COREL(J) = COREL(J)/J
250     CONTINUE
C
```

```
      do 270 j = 1,jtime
      suma = suma+tors(j,jcor1)
      suma2 = suma2+tors(j,jcor1)**2
      sumb = sumb+tors(j,jcor2)
      sumb2 = sumb2+tors(j,jcor2)**2

      sqm = (suma/j)**2
      msq = suma2/j
      sqmb = (sumb/j)**2
      msqb = sumb2/j
      roota = sqrt(msq-sqm)
      rootb = sqrt(msqb-sqmb)
      norm(j) = roota*rootb
 270  continue
c
      do 280 j=1,jtime
      corel(j)=((tors(j,jcor1) - corel(j))*(tors(1,jcor2)-/norm(j)
 280  continue
      DO 260 J = 1,JTIME-1
         WRITE(2,*) time(j),COREL(J)
260   CONTINUE
      CLOSE(2)
C
      WRITE(*,*)'DO YOU WANT ANOTHER GO (Y/N)'
      READ(*,'(1A)')CREPLY
      IF (CREPLY .EQ. 'Y' .OR. CREPLY .EQ. 'y') GOTO 205
C
      END
```

Program to calculate the percentage of helical residues in an agarose double helix as a function of temperature.

```
C      PROGRAM TO MODEL DOUBLE HELIX FORMATION
C      USING SHORT-CHAIN APPROXIMATION OF
C      POLAND AND SCHERAGA - MISMATCHED CHAIN
C      FORM (this version writes the results to a file)
C
      IMPLICIT DOUBLE PRECISION (A-H)
      IMPLICIT DOUBLE PRECISION (O-Z)
      DIMENSION TEMP(1000),THETT(1000),DERIV(1000)
```

```
    1 FORMAT(5I5)
    2 FORMAT(10F10.5)
    3 FORMAT(10E12.4)
    4 FORMAT(1HO,'GIVE A VALUE FOR THE NUMBER OF CHAIN RESIDUES')
    5 FORMAT(1HO,'GIVE A VALUE FOR THE PARAMETER GAMMA')
    6 FORMAT(1HO,'GIVE VALUES FOR DELH DELS IN KCAL/MOLE RES')
      WRITE(6,4)
      READ(5,*)N
      PN=N
      WRITE(6,5)
      READ(5,*)GAMMA
      WRITE(6,6)
      READ(5,*)DELHO,DELSO
      T=240.16
      JJ=0
 1000 CONTINUE
      JJ=JJ+1
      DELGO=DELHO-T*DELSO
      SS=-DELGO/(0.00198*T)
      S=DEXP(SS)
C     BEGIN SUMMATIONS
      SUM1=0.0
      SUM2=0.0
      DO100I=1,N
      PI=I
      SUM1=SUM1+(PN-PI+1.0D+00)*(PN-PI+1.0D+00)*(S**PI)
      SUM2=SUM2+PI*(PN-PI+1.0D+00)*(PN-PI+1.0D+00)*(S**PI)
  100 CONTINUE
C     BEGIN CALCULATION OF DEGREE OF ORDERING
      TERM1=1.0D+00+1.0D+00/(4.0D+00*GAMMA*SUM1)
      TERM1=TERM1*TERM1-1.0D+00
      TERM1=DSQRT(TERM1)
      TERM2=1.0D+00+1.0D+00/(4.0D+00*GAMMA*SUM1)
      TERM3=TERM2-TERM1
      TERM3=DSQRT(TERM3)
      IF(SUM1.EQ.0.0D+00)TERM4=1.0D+00/PN
      IF(SUM1.EQ.0.0D+00)GOTO101
      TERM4=SUM2/(PN*SUM1)
  101 CONTINUE
  110 CONTINUE
      OPEN(UNIT=2,FILE='cl2hx.res',status='unknown',ERR=110)
      THET=TERM3*TERM4
      TT=T-273.16D+00
```

```
C Write out 's', the temperature, degree of ordering and
C degree of association
      WRITE(2,*)S,TT,TERM3,THET
      TEMP(JJ)=TT
      THETT(JJ)=THET
      KK=JJ-1
      IF(KK.EQ.0)KK=1
      DERIV(JJ)=THETT(JJ)-THETT(KK)
      T=T+1.0D+00
      IF(T.LT.473.16D+00)GOTO1000
      DO300I=1,JJ
  300 CONTINUE
      STOP
      END
```

Appendix B

High performance anion exchange chromatography (HPAEC)

Overview

The HPAEC apparatus used in the present work was made by DIONEX. The aim of the technique, which combines liquid chromatography with a novel method for electrolytic detection of -OH groups, is to separate molecules which are similar in chemical makeup but differing in size.

HPAEC: main parts

The DIONEX HPAEC apparatus [111] (see figure B.1 for a scheme showing the main parts) consists of eluent reservoirs connected *via* a sparge and pressurise unit (using helium) to a pump consisting of a gradient mixer and piston unit. The sample is injected from a $25\mu\ell$ tubing loop and is carried along a 4×250mm DIONEX CarboPac PA-100 pellicular anion exchange resin column by the eluent. Detection of organic molecules in the output is by pulsed amperometric detection (PAD) using a triple pulse scanning sequence and gold electrodes. The shape of the scanning pulse can be altered [111–113]; a suitable shape already exists for use with agarose and other neutral carbohydrates.

Programming of the pulse sequences in the detector and of the mixing in the pump module can be carried out using a supplied control panel or using the Windows interface on a personal computer.

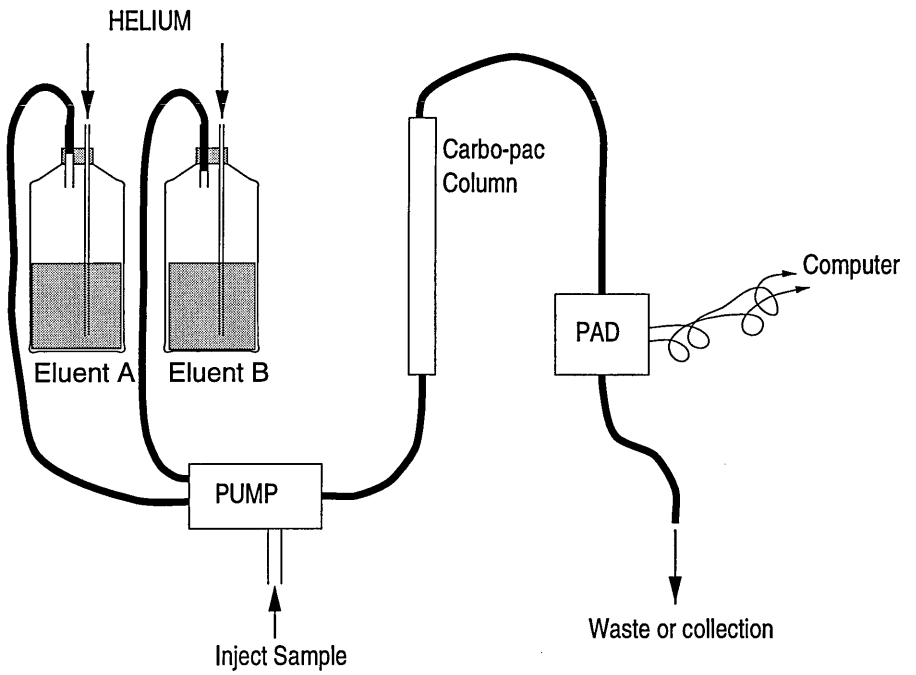


Figure B.1: Main parts of the HPAEC apparatus.

Anion exchange chromatography and pulsed amperometric detection (PAD)

PAD is sensitive enough to detect carbohydrates at the picomolar level in the output from an HPAEC column [112, 114, 115]. This makes it ideal for the analysis of carbohydrate mixtures, especially since other detection methods (such as UV) are much less sensitive. The detection device consists of gold electrodes (for resistance to oxidation) to which are fed rapidly repeating sequences of pulses. Essentially the ease of oxidation of carbohydrates in solution is measured.

Operation

The two eluents used for agarose oligomers are described in chapter five. They are prepared using de-ionised, de-gassed water and analytical purity chemicals. To make a solution 500mM in sodium acetate and 100mM in sodium hydroxide, 68.04g of sodium acetate and 8.5mℓ of 50% sodium hydroxide is used. The solutions must be filtered through 0.45μm Millipore filters and sparged (de-gassed) with helium for five minutes before use. During use the reservoirs are kept pressurised with helium.

The HPAEC machine is started up using a specific cleaning routine, which

gradually increases the flow rate of eluent to the operating level, and cleans the column. After this the sample can be injected, and the elution gradient (see chapter five) initialised.

The column separates carbohydrates out according to pK values (these range from 12 to 14, necessitating highly alkaline eluents), chain length and conformation; and the PAD reports on the contents of the emerging solution. The column is always cleaned with a sodium acetate eluent after each analytical run, to remove residual solutes. The results are down-loaded to a personal computer for later analysis.

Bibliography

[1] L.H.Van Vlack. *Materials for Engineering.* Addison - Wesley, 1982.

[2] N.N.Greenwood and A.Earnshaw. *Chemistry of the Elements.* Pergamon Press, 1990.

[3] J.Kouřimský. *The Illustrated Encyclopedia of Minerals and Rocks.* Sunburst Books, London, 1995.

[4] R.D.Harrison. *Book of data.*, chapter 4, pages 51–52. Longman Group Limited, 5 edition, 1988.

[5] R.T.Morrison and R.N.Boyd. *Organic Chemistry.* Allyn & Bacon Inc., 1987.

[6] R.T.Morrison and R.N.Boyd. *Organic Chemistry*, chapter 39.8, pages 1331–1332. Allyn & Bacon, 1987.

[7] A.Fersht. *Enzyme structure and mechanism.* W.H.Freeman and company, 2 edition, 1985.

[8] W.Pigman and D.Horton. *The Carbohydrates: chemistry and biochemistry.* Academic Press (London), 1972.

[9] G.A.Jeffrey and R.Taylor. The application of molecular mechanics to the structure of carbohydrates. *Journal of Computational Chemistry*, 1:99–109, 1980.

[10] I.Fleming. *Frontier Orbitals and Organic Chemical Reactions.* John Wiley & Sons, 1990.

[11] T.Nogrady. *Medicinal chemistry; a biochemical approach*, chapter 1, pages 3–8. Oxford University Press, 2 edition, 1988.

[12] Eric Dickinson, editor. *Food Polymers Gels and Colloids*, chapter 26, pages 322–327. Royal Society of Chemistry, 1991.

[13] D.A.Rees. Structure, formation and mechanism in the formation of polysaccharide gels and networks. *Advances in Carbohydrates and Biopolymers*, 24:266–292, 1969.

[14] L.Piculell and S.Nilsson. Anion specific salt effects in aqueous agarose systems. 1. effects on the coil-helix transition and gelation of agarose. *Journal of Physical Chemistry*, 93:5596–5601, 1989.

[15] K.Nishinari, M.Watase, K.Kohyama, N.Nishinari, D.Oakenfull, S.Koide, K.Ogino, P.A.Williams, and G.O.Phillips. The effect of sucrose on the thermo-reversible gel-sol transition in agarose and gelatin. *Polymer Journal*, 24:871–877, 1992.

[16] C.Rochas, F.R.Taravel, and T.Turquois. Nmr studies of synergistic kappa-carrageenan-carob galactomannan gels. *International Journal of Biological Macromolecules*, 12:353–358, 1990.

[17] L.Piculell, W.Zhang, T.Turqois, C.Rochas, F-R.Taravel, and P.A.Williams. Effects of added galacto- and gluco-mannans on the nmr spectra of ^{133}Cs ions in kappa-carrageenan gels. *Carbohydrate Research*, 265:281–290, 1994.

[18] G.Stainsby. Proteinaceous gelling systems and their complexes with polysaccharides. *Food Chemistry*, 6:3–14, 1980.

[19] M.G.Lynch and D.M.Mulvihill. The influence of caesins on the rheology of κ–carrageenan gels. *Food Hydrocolloids*, 8:317–329, 1994.

[20] D.G.Dagleish and E.R.Morris. Interactions between carrageenans and casein micelles: Electrophoretic and hydrodynamic properties of the particles. *Food Hydrocolloids*, 2:311–320, 1988.

[21] Eric Dickinson, editor. *Food Polymers, Gels and Colloids*, pages 310–321. Royal Society of Chemistry, 1991.

[22] R.J.Wieme. *Agar gel electrophoresis*, chapter 4, pages 102–110. Elsevier Publishing Company, 1965.

[23] E.Percival and R.H.McDowell. *Chemistry and Enzymology of Marine Algal Polysaccharides.* Academic Press, 1967.

[24] R.P.Millane, R.Chandrasekaran, S.Arnott, and I.C.M.Dea. The molecular structure of kappa-carrageenan and comparison with iota-carrageenan. *Carbohydrate Research*, 182:1–17, 1988.

[25] A.Chiovitti, A Bacic, D.J.Craik, G.T.Kraft, M.L.Liao, R.Falshaw, and R.H.Furneaux. A pyruvated carrageenan from australian specimens of the red algae *Sarconema filiforme*. *Carbohydrate Polymers*, 310:77–83, 1998.

[26] S.Arnott, A.Fulmer, W.E.Scott, I.C.M.Dea, R.Moorhouse, and D.A.Rees. The agarose double helix and it's function in agarose gel structure. *Journal of Molecular Biology*, 90:269–284, 1974.

[27] J.Stellwagen and N.C.Stellwagen. Internal structure of the agarose gel matrix. *Journal of Physical Chemistry*, 99:4247–4251, 1995.

[28] I.Heertje. Structure and function of food products: a review. *Food Structure*, 12:343–364, 1993.

[29] I.J.Miller, R.Falshaw, and R.H.Furneaux. Chemical methylation of agaroid hydroxyl groups. *Carbohydrate Research*, 262:127–135, 1994.

[30] V.L.Larwood. *Modelling Studies of Factors Affecting Gellan Gelation.* PhD thesis, University of Surrey, 1996.

[31] Eric Dickinson, editor. *Food Polymers, Gels and Colloids*, pages 47–64. Royal Society of Chemistry, 1991.

[32] J.R.Mitchell and D.A.Ledward. Gelation of polysaccharides. Journal. P.131.

[33] M.D.Walkinshaw and S.Arnott. Conformations and interactions of pectins i. x-ray diffraction analyses of sodium pectate in natural and acidified forms. *Journal of Molecular Biology*, 153:1055–1073, 1981.

[34] J.R.Mitchell and D.A.Ledward, editors. *Gelation of Polysaccharides*, chapter 3, pages 121–170. Elsevier, 1986.

[35] M.J.Gidley and P.V.Bulpin. Aggregation of amylose in aqueous systems: The effect of chain length on phase behaviour and aggregation kinetics. *Macromolecules*, 22:341–345, 1989.

[36] A.H.Clarke, M.J.Gidley, R.K.Richardson, and S.B.Ross-Murphy. Rheological studies of aqueous amylose gels: The effect of chain length and concentration on gel modulus. *Macromolecules*, 22:346–351, 1989.

[37] M.Djabourov, A.H.Clark, D.W.Rowlands, and S.B.Ross-Murphy. Small angle x-ray scattering characteristics of agarose sols and gels. *Macromolecules*, 22:180–188, 1989.

[38] J.E.Brigham, M.J.Gidley, R.A.Hoffmann, and C.G.Smith. Microscopic imaging of network strands in agar, carrageenan, locust bean gum and κ–carrageenan/locust bean gum gels. *Food Hydrocolloids*, 8:331–344, 1994.

[39] B.H.Zimm and J.K.Bragg. Theory of the phase transition between helix and random coil in polypeptide chains. *The Journal of Chemical Physics*, 31:526–535, 1959.

[40] E.T.Adman, M.W.Mather, and J.A.Fee. Molecular modelling studies on the proposed nacl induced dimerisation of chromatium–vinosum high–potential iron protein. *Biochimica et Biophysica Acta*, 1142(1-2):93–98, 1993.

[41] A.Imberty and S.Perez. Molecular modelling of protein–carbohydrate interactions: understanding the specificities of two legume lectins towards oligosaccharides. *Glycobiology*, 4(3):351–366, 1994.

[42] J.L.Klepeis, M.G.Ierapetritou, and C.A.Floudas. Protein folding and peptide docking: a molecular modelling and global optimisation approach. *Computers and Chemical Engineering*, 22:S3–S10, 1998.

[43] S.Peréz, M.Kouwijzer, K.Mazeau, and S.B.Engelsen. Modelling polysaccharides: Present status and challenges. *Journal of Molecular Graphics*, 14:307–321, 1996.

[44] Claus-Wilhelm von der Leith, T. Kozár, and W.E.Hull. A (critical) survey of modelling protocols used to explore the conformational space of oligosaccharides. *Journal of Molecular Structure*, pages 225–244, 1997.

[45] K.B.Lipkowski and D.B.Boyd, editors. *Reviews in Computational Chemistry*, chapter 8, pages 295–320. VCH Publishers, Inc., 1990.

[46] A.R.Leach. *Molecular Modelling; principles and applications.* Addison Wesley Longman, 1 edition, 1996.

[47] P.K.Weiner and P.A.Kollman. AMBER: Assisted model building with energy refinement. *Journal of Computational Chemistry*, 2:287–303, 1981.

[48] & C.Altona C.J.M.Huige. Force field parameters for sulphates and sulphamates based on *ab initio* calculations: Extensions of amber and charmm fields. *Journal of Computational Chemistry*, 16:56–79, 1995.

[49] J.L.Asensia, M.Martin-Pastor, and J.Jimenez-Barbero. The use of cvff and cff91 forcefields in the conformational analysis of carbohydrate molecules. comparison with amber molecular mechanics and dynamics calculations for methyl α-lactoside.;. *International Journal of Biological Macromolecules*, 17:137–148, 1995.

[50] T.M.Glennon and K.M.Merz. A carbohydrate force field for AMBER and its application to the study of saccharide to surface adsorption. *Journal of Molecular Structure*, 395:157–171, 1997.

[51] D.H.Williams and I.Fleming. *Spectroscopic Methods in Organic Chemistry.* McGraw Hill, 1989.

[52] I.Tvaroska. Computational methods for studying oligo- and polysaccharide conformations. *Pure and Applied Chemistry*, 61:1201–1216, 1989.

[53] A.Poveda, J.L.Asensio, J.F.Espinosa, M.MartinPastor, and J.Canada. Applications of nuclear magnetic resonance spectroscopy and molecular modelling to the study of protein–carbohydrate interactions. *Journal of Molecular Graphics and Modelling*, 15(1):9, 1997.

[54] K.Rasmussen. How to develop force fields: An account of the emergence of potential energy functions for saccharides. *Journal of Molecular Structure*, pages 91–106, 1997.

[55] San Diego Biosym Technologies Inc. *DISCOVER, Theory and Methodology.* Biosym Technologies Inc., San Diego, 1993.

[56] T.Steiner and W.Saenger. Geometry of c-h...o hydrogen bonds in carbohydrate crystal structures. analysis of neutron diffraction data. *Journal of the American Chemical Society*, 114:10146–10154, 1992.

[57] T.Steiner. Unrolling the hydrogen bond properties of c-h...o interactions. *Chemical Communications*, pages 727–734, 1997.

[58] Biosym Technologies INC. Class II forcefield for modelling carbohydrates confidential. *Biosym Manuals*, .:1–end, 1994.

[59] L.Verlet. Computer experiments on classical fluids. i. thermodynamical properties of lennard–jones molecules. *Physics Review*, 159:98–103, 1967.

[60] H.J.C.Berendsen, J.P.M.Postma, W.F.Vangunsteren, A.Dinola, and J.R.Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81:3684, 1984.

[61] S.Homans. A molecular mechanical forcefield for the conformational-analysis of oligosaccharides - comparison of theoretical and crystal structures of MAN-α-1-3MAN-β-1-4GLCNAC. *Biochemistry*, 29:9110, 1990.

[62] D.Lamba, S.Glover, W.Mackie, A.Rashdid, B.Sheldrick, and S.Perez. Insights into stereochemical features of sulphated carbohydrates: X-ray crystallographic and modelling investigations. *Glycobiology*, 4(2):151–162, 1994.

[63] N.Bouchemal-Chibani, I.Braccini, C.Derouet, C.Hervé du Penhoat, and V.Michon. Conformational analysis of disaccharides using molecular mechanics and nmr methods. *International Journal of Biological Macromolecules*, 17:177–182, 1995.

[64] S.B.Engelsen, S.Perez, I.Braccini, and C.Herve du Penhoat. Internal motions of carbohydrates as probed by comparative molecular modelling and nuclear magnetic resonance of ethyl β–lactoside. *Journal of Computational Chemistry*, 16:1096–1119, 1995.

[65] A.M.J.J.Bovin and A.T.Brunger. Conformational variability of solution nuclear magnetic resonance structures. *Journal of Molecular Biology*, 250:80–93, 1995.

[66] G.Lipari and A.Szabo. Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules 1. theory and range of validation. *Journal of the Americal Chemical Society*, 104:4546–4559, 1982.

[67] G.Lipari and A.Szabo. Model free approach to the interpretation of nuclear magnetic resenance relaxation in macromolecules 2. analysis of experimental results. *Journal of the Americal Chemical Society*, 104:4559–4570, 1982.

[68] S.K.Sarkar. *NMR Spectroscopy and its Application to Biomedical Research.* Elsevier, 1996.

[69] J.Jimenez-Barbero, C.Bouffar-Roupe, C.Rochas, and S.Perez. Modelling studies of solvent effects on the conformational stability of agarobiose and neoagarobiose and their relationship to agarose. *International Journal of Biological Macromolecules*, 11:265–272, 1989.

[70] Georg Lautscham. Modelling studies of carrageenans. Master's thesis, University of Surrey, 1996.

[71] Vivienne L.Larwood, Brendan J.Howlin, and Graham G.Webb. Solvation effects on the conformational behaviour of gellan and calcium ion binding to gellan double helices. *Journal of Molecular Modelling*, 2:175–182, 1996.

[72] G.N.Ramachandran and U.Sasisekharan. ,. *Advances in Protein Chemistry*, 1968.

[73] P.S.Belton, V.J.Morris, and S.F.Tanner. Interaction of group i cations with iota, kappa and lambda carrageenans studied by multinuclear nmr. *International Journal of Biological Macromolecules*, 7:53–56, 1985.

[74] H.Grasdalen and O.Smidsrød. ^{133}Cs nmr in the sol-gel states of aqueous carrageenan. selective site binding of caesium and potassium ions in kappa carrgeenan gels. *Macromolecules*, 14:229, 1981.

[75] D.A.Rees. Conformational analysis of polysaccharides part ii. alternating copolymers of the agar-carrageenan-chondroitin type by model building in the computer with calculation of helical parameters. *J. Chem. Soc. Phys. Org.*, B:217–226, 1969.

[76] N.S.Anderson, J.W.Campbell, M.M.Harding, D.A.Rees, and J.W.B.Samuel. X-ray diffraction studies of polysaccharide sulphates: Double helix models for ι and κ–carrageenans. *Journal of Molecular Biology*, 45:85, 1969.

[77] J-Y.LeQuestel, S.Cros, W.Mackie, and S.Perez. Computer modelling of sulphated carbohydrates; applications to carrageenans. *International Journal of Biological Macromolecules*, 17:161–173, 1995.

[78] R.K.Harris. *Nuclear Magnetic Resonance Spectroscopy; a physicochemical view.* Longman, 1991.

[79] D.McKie and C.McKie. *Essentials of Crystallography.* Blackwell Scientific Publications, 1986.

[80] J.S.Craigie and A.Jurgens. Structure of agars from *gracilaria tikvahiae* rhodophyta: Location of 4-o-methyl-L-galactose and sulphate. *Carbohydrate Polymers*, 11:265–278, 1989.

[81] G.Corongiu, S.L.Fornili, and E.Clementi. . *International Journal of Quantum Chemistry*, 10:227, 1973.

[82] W.L.Jorgensen, J.Chandrasekhar, and J.D.Madura. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79(2):926–935, 1983.

[83] K.Ueda, H.Ochiai, A.Imamura, and S.Nakagawa. An investigation of the conformation of β–carrageenan by molecular mechanics and molecular dynamics simulations. *Bulletin of the Chemical Society of Japan.*, 68:95–106, 1995.

[84] B.J.Hardy and A.Sarko. Conformational analysis and molecular dynamics simulation of cellobiose and larger cellooligomers. *Journal of Computational Chemistry*, 14:831–847, 1993.

[85] M.K.Dowd, P.J.Reilly, and A.D.French. Conformational analysis of trehalose disaccharides and analogues using MM3. *Journal of Computational Chemistry*, 13:102–114, 1992.

[86] S.N.Vinogradov and R.H.Linnell. *Hydrogen Bonding.* Van Nostrand, 1971.

[87] M.P.Allen and D.J.Tildesley. *Computer Simulation of Liquids.* Oxford Science Publications, 1996.

[88] K.B.Guiseley. The relationship between methoxyl content and gelling temperature of agarose. *Carbohydrate Research*, 13:247–265, 1970.

[89] M.Watase, K.Kohyama, and K.Nishinari. Effects of sugars and polyols on the gel - sol transition of agarose by differential scanning calorimetry. *Thermochimica Acta*, 206:163–173, 1992.

[90] P.Y.Key and D.B.Sellen. A laser light-scattering study of the structure of agarose gels. *Journal of Polymer Science: Polymer Physics Edition*, 20:659–679, 1982.

[91] E.A.Masimov, V.V.Prudko, and L.I.Khomutov. Phenomenon of hysteresis in the solution-gel transition in the agarose-water system. *Colloid Journal of the USSR*, 45:891–894, 1983.

[92] A.Cesaro, F.Delben, S.Paoletti, and F.Scagnolari. Iso-ionic and isothermal phase transitions in the ionic polysaccharides: a calorimetric study. *Thermochimica Acta*, 85:465–468, 1985.

[93] S.A.Foord and E.D.T.Atkins. New x-ray results from agarose: Extended single helix structures and implications for gelation mechanisms. *Biopolymers*, 28:1345–1365, 1989.

[94] J.M.Guenet, C.Rochas, and A.Brûlet. Molecular structures in biopolymer sols and gels. *Journale de Physique IV*, C8:99–102, 1993.

[95] D.Poland and H.A.Scheraga. *Theory of Helix - Coil Transitions in Biopolymers.* Academic Press, New York and London, 1970.

[96] D.S.Reid, T.A.Bryce, A.H.Clark, and D.A.Rees. Helix–coil transitions in gelling polysaccharides. *Faraday Discussions of the Chemical Society 57: Gels and Gelling Processes*, 57:230–237, 1974.

[97] D.Cooke, M.J.Gidley, and N.D.Hedges. Thermal properties of polysaccharides at low moisture. *Journal of Thermal Analysis*, 47:1485–1498, 1996.

[98] M.Watase, K.Nishinari, A.H.Clark, and S.B.Ross-Murphy. Differential scanning calorimetry, rheology, x-ray and nmr of very concenrated agarose gels. *Macromolecules*, 22:1196–1201, 1989.

[99] P.Doty and J.T.Yang. Polypeptides vii. poly-γ-benzyl-L-glutamate; the helix - coil transition in solution. *Journal of the American Chemical Society*, 78:498–500, 1956.

[100] P.Doty, J.Marmur, J.Eigner, and C.Schildkraut. . *Proceedings of the National Academy of Sciences of the United States of America*, 1960.

[101] J.Applequist and V.Damle. Thermodynamics of the helix–coil equilibrium in oligoadenylic acid from hypochromicity studies. *Journal of the American Chemical Society*, 87:1450–1458, 1965.

[102] I.T.Norton, D.M.Goodall, S.A. Frangou, E.R.Morris, and D.A.Rees. Mechanism and dynamics of conformational ordering in xanthan polysaccharides. *Journal of Molecular Biology*, 175:371–394, 1984.

[103] P.W.Atkins. *Physical Chemistry*, chapter 21, pages 695–702. Oxford University Press, second edition edition, 1984.

[104] A.Cesaro, S.Paoletti, R.Rizzo, and J.C.Benegas. Thermodynamics of polysaccharide polyelectrolytes: Enthalpy changes in the conformational transition of kappa carrageenan. *Pure & Applied Chemistry*, 66(3):461–467, 1994.

[105] R.J.Wieme. *Agar Gel Electrophoresis.* Elsevier Publishing Company, 1965.

[106] P.S.Belton, I.J.Colquhoun, and B.P.Hills. Applications of nmr to food science. *Reviews of NMR*, –:25–51, –.

[107] P.S.Belton and I.J.Colquhoun. Nmr and ftir studies of structure and function in food biopolymers. *Abstracts of papers of the American Chemical Society*, 200(1):208, 1990.

[108] H.Grasdalen and O.Smidsrød. Iodide specific formation of kappa carrageenan single helices. $^{1}27$i nmr spectroscopic evidence for selective site binding of iodide anions in the ordered conformation. *Macromolecules*, 14:1842, 1981.

[109] L.A.Luck. Probing H-bond interactions in carbohydrate - protein systems: applications of ^{19}F nmr to fluoro-deoxy sugars and receptor proteins. *Journal of Cell Biology*, VS2IB SIB:48, 1995.

[110] C.R.Heald. *Synthesis and molecular modelling studies of poly(arylene ether sulphone) dicyanates.* PhD thesis, University of Surrey, March 1995.

[111] DIONEX Corporation. *Dionex Technical Notes*, 1993.

[112] M.R.Hardy and R.R.Townsend. Separation of positional isomers of oligosaccharides and glycopeptides by high performance anion exchange chromatography with pulsed amperometric detection. *Proceedings of the National Academy of Sciences of the United States of America*, 85:3289–3293, 1988.

[113] R.D.Rocklin and C.A.Pohl. Determination of carbohydrates by anion exchange chromatography with pulsed amperometric detection. *Journal of Liquid Chromatography*, 6(9):1577–1590, 1983.

[114] S.Hughes and D.C.Johnson. . *Anal. Chim. Acta*, 132:11–22, 1981.

[115] R.R.Townsend, M.R.Hardy, O.Hindsgaul, and Y.C.Lee. High performance anion exchange chromatography of oligosaccharides using pellicular resins and pulsed amperometric detection. *Analytical Biochemistry*, 174(2):459–470, 1988.

MATERIAL REDACTED AT REQUEST OF UNIVERSITY