

A
SUMMARY
OF
RESEARCH
ACTIVITIES

by

David Brown

Summary of Research Activities	1
Introduction	1
1. Molecular flexibility and mass and momentum transport	2
1.1 Introduction	2
1.2 Homogeneous shear non-equilibrium molecular dynamics (HSNEMD)	2
1.3 Diatomic fluids	3
1.4 N-alkane fluids	3
1.4.1 Ethane vs. Propane	3
1.4.2 Hexane vs. Flexane	4
1.5 Summary of Ph. D. thesis work	4
1.6 The effect of mass distribution in rigid triatomic molecules	4
1.7 Frozen distribution sampling (FDS)	5
1.8 Conclusions	6
1.8.1 Boundary driven shear	7
1.8.2 Rates of Reaction	7
2. Polymers	10
2.1 Introduction	10
2.2 Mechanical properties	10
2.2.1 Amorphous/crystalline interface	10
2.2.2 Amorphous systems	10
2.2.3 Effect of chain configuration on mechanical properties	13
2.2.4 Conclusions	14
2.3 Constructing samples of amorphous polymers	14
2.3.1 Excluded volume growth (EVG)	14
2.3.2 Phantom Chain Growth (PCG)	15
2.3.3 Pivot Monte Carlo sampling	15
2.3.4 Conclusions	18
2.4 Configurations of chains in the melt	18
2.4.1 Introduction	18
2.4.2 Linear n-alkane-like chains	19
2.4.3 Polyethylene oxide (PEO)	24
2.4.4 Polyvinyl chloride (PVC)....and PEO revisited	25
2.4.5 Conclusions	28
2.5 Polymer electrolytes	29

2.5.1	Introduction	29
2.5.2	Potential Optimization in Crystalline PEO and PEO ₃ NaI	29
2.5.3	Preparation of amorphous relaxed PEO and PEO _x NaI systems	35
2.5.4	Polymer mobility	35
2.5.5	Local structure around cations in PEO _x NaI	36
2.5.6	Ion mobility	38
2.5.7	Phase separation	41
2.5.8	Conclusions	41
3.	Biopolymers	42
3.1	Introduction	42
3.2	Parallel Molecular Dynamics	42
3.2.1	Introduction	42
3.2.2	Domain decomposition: General considerations	44
3.2.3.	Domain decomposition for molecular systems	47
3.2.4	The general purpose parallel MD code ddgmq	47
3.2.5	Coulombic summations	48
3.3	Protein folding	48
3.3.1	Introduction	48
3.3.2	T4-Lysozyme	49
3.3.3	Ubiquitin	50
3.4	Enzymes	51
3.4.1	Introduction	51
3.4.2	Bovine Pancreatic Trypsin	52
3.4.3	Human Hageman Factor XII-a	53
3.4.4	Glyceraldehyde-3 Phosphate Dehydrogenase (GAPDH)	54
3.5	Drug design	55
3.5.1.	Introduction	55
3.5.2	Echistatin	55
3.6	Conclusions	55
References		57

Foreword

In very broad terms, the topic of my research career, thus far, could be said to be an investigation of the physical structure and dynamics of matter at the molecular level. In particular, the way in which the microscopic details of the molecules affect macroscopic properties has often been the goal; be it the relationship between molecular flexibility and the transport properties of the corresponding fluid or the interaction of an ion with the polymer electrolyte matrix and the resulting electrical conductivity, for instance. In all of this work the technique used has been molecular level simulation, mostly molecular dynamics (MD) but also Monte Carlo (MC). All simulations have been carried out within a completely classical, often highly simplified, World. The emphasis in this case being placed on generic behaviour and tests of theory rather than direct comparisons with the real World. After all, many properties of matter don't need to invoke quantum mechanics to be explained and a theory which works for a truly classical system ought to work for a real system which behaves to a high degree classically.

To a large extent classical dynamics is forced upon us in any case as the limited computing power available dictates that studies of the motion of even a small number of molecules requires that we make this approximation. Certainly this was true in the early 1980s, when I began my research career, and despite the impressive rate at which computer technology has developed in the last two decades, this remains true today. Indeed, we are still a long way off having the power to compute the motion of even a modest number of atoms from a numerical solution of the time-dependent Schroedinger equation. Of course, techniques to treat a small part, a reactive site for instance, of an otherwise classical system with some degree of quantum mechanical approximation are emerging and in the immediate future offer the prospect, particularly if coupled to parallel computers, of taking simulation a significant step forward. Anyway, the purpose here is to look back rather than to look forward.

So how to summarize concisely the endeavours of the last 16 years? It's important to state at the outset that this period has encompassed a fairly frenetic period in the development of molecular simulations. From the early days of simulations of purely atomic systems in the NVE ensemble things have come a long way. Increasingly elaborate in the types and size of molecules studied, the ensembles and boundary conditions employed and also the computers used; first scalar, then vector and now parallel architectures. As in many areas, the major driving force behind this rapid development has been the desire to do interesting science on more and more complex systems. This desire pushes at the limits of the discipline and often requires developments in the techniques. Certainly looking back some of these developments will turn out to have been blind alleys or side issues, but this is just part of the normal "evolutionary" process. At the same time, the implementation of the techniques used in simulation has had to adapt itself to the technology available in order to get the most from advances in that area. Thus the science, the methodology and the computer technology are intermingled in a way which might seem difficult to separate. Being actively involved in this developmental period my research career mirrors this to a certain extent. In the following summary I will try to unravel these interwoven threads and instead focus on the science within three broad areas which account for a good deal, though by no means all, of my research to date.

1. Molecular flexibility and mass and momentum transport

1.1 Introduction

A common thread running through my research career has been the connection between molecular variables and the transport properties of fluids. This was originally the subject of my Ph. D. thesis which was partly sponsored by *Shell Research (Thornton, UK)* who have obvious interests in improving the performance of their lubricants. Later it was also the subject of a Shell sponsored post-doctoral contract and one other Ph. D. studentship at UMIST. The objective of the work was to try and understand what properties of a molecule affected most the ability of the corresponding bulk liquid to transfer mass and momentum. By using computer simulation it was envisaged that one would be free from the situation in the laboratory where one is generally forced to compare molecules which differ in many respects. Thus the effect of changing just one parameter on a certain system property could be determined unambiguously. As it turned out, this is not that straightforward and the reasons for this will be discussed. In later work ways round the problem were found which do allow the desired comparisons to be made.

1.2 Homogeneous shear non-equilibrium molecular dynamics (HSNEMD)

With a clear need to obtain information on rheological properties much of the initial work involved implementing homogeneous shear non-equilibrium molecular dynamics (HSNEMD) algorithms to simulate periodic systems undergoing shear flow. Comparisons were then made for model diatomic systems of the results obtained from both steady state and perturbation HSNEMD approaches to those obtained from the equilibrium correlation function (Green-Kubo integrand) route to the shear viscosity. At the time it was concluded that the steady state HSNEMD approach was the most reliable route to the shear viscosity.

In HSNEMD a simple modification of the periodic boundaries, very often used in molecular simulations to reduce system size effects, can be made which leads to a constant shear rate, $\dot{\gamma}$, being set up through the system in a smooth and gradient free manner. An illustration of these Lees-Edwards boundary conditions¹ is shown in Fig. 1.2.1.

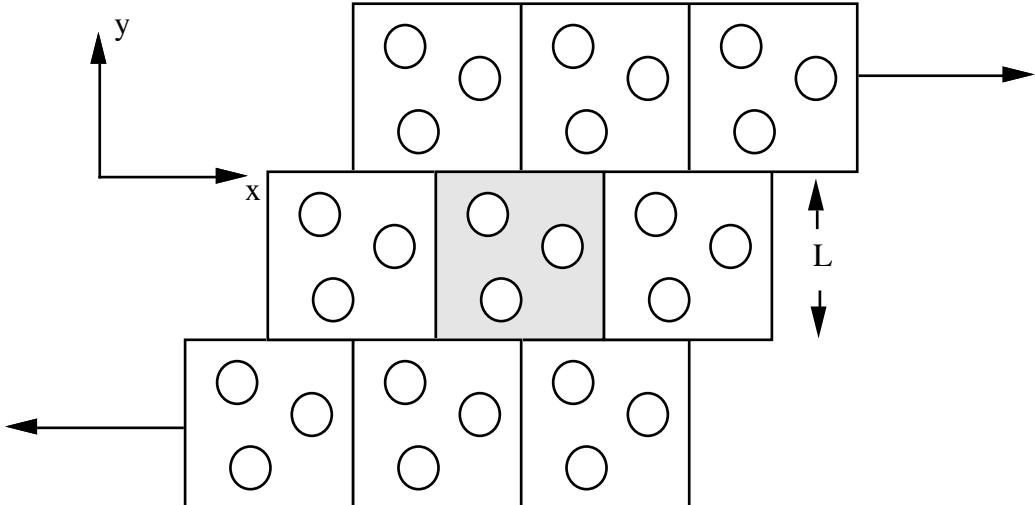


Figure 1.2.1 Illustration of Lees-Edwards boundary conditions in 2-D for shear imposed in the xy plane. Particles in the central shaded cell perceive their images in the layer above moving at a velocity of $\dot{\gamma}L$ and in the layer below at a velocity of $-\dot{\gamma}L$.

In a relative way, all particles see their images drifting past them in the same way. The infinite system is thus completely homogeneous and free from the temperature and density gradients which complicate results obtained from small systems by boundary driven methods. Once a steady state has been achieved the viscosity can be calculated simply by averaging the relevant off-diagonal component of the pressure tensor and then using Newton's Law; for the geometry of Fig. 1.2.1 this gives

$$\eta(\dot{\gamma}) = -\langle P_{xy} \rangle / \dot{\gamma} \quad (1)$$

There is, however, one rather serious problem with HSNEMD in that to measure a viscosity with any precision requires that the signal, $\langle P_{xy} \rangle$, be resolvable from the spontaneous fluctuations. For the size of system generally used in MD ($\sim 10^3$ atoms) a typical error on $\langle P_{xy} \rangle$ would be ~ 1 bar ($= 10^5$ Pa) even for a 1 ns simulation. This implies that for good precision (1%) a shear rate of $\sim 10^{10}$ s $^{-1}$ has to be used for a fluid with a viscosity of ~ 1 mPa s. This order of shear rate is far in excess of that accessible in the laboratory experiments where 10 6 s $^{-1}$ is considered as high. Moreover, the heat generated by the work done on the system would accumulate in a periodic system, where there are no walls to conduct away the heat, unless removed. In HSNEMD the heat can only be removed in a fictitious way by coupling some form of "thermostat" to the particle momenta. Several variants of HSNEMD appeared in the literature following the original work of Lees and Edwards but, thermostats aside, all correct ones reduce to integrating Newton's equations of motion in a frame of reference given by Lees-Edwards boundary conditions.

1.3 Diatomic fluids

Initial attempts to elucidate the influence of molecular parameters on the transport properties of liquids first involved comparing the behaviour of systems comprising two-centre model rigid diatomic molecules which differed only in the bond length. In terms of the σ parameter of the Lennard-Jones 12-6

interatomic potential, the bond lengths were 0.608 and 1; the former value being that used in a study of liquid chlorine from which the parameters were taken. This was more than sufficient to cause a significant difference in the phase diagram of the two liquids and immediately led to the problem of how to make meaningful comparisons. Clearly diffusion and viscosity are functions of the state variables so the key question is how can one possibly separate the effect of the change in bond length from that of the change in phase diagram? In the absence of any corresponding states criterion for diatomic fluids comparisons were decided to be made under conditions of the same temperature and pressure. Two state points were chosen, both at a temperature of 260 K, one at ambient pressure and one at a raised pressure of 1 GPa. Comparisons between the fluids were then made not only for the equilibrium and shear rate dependent viscosities but also the shear rate dependent alignment and thermodynamic properties and structure.

At ambient pressures the initially surprising result obtained was that the short bond length fluid was more viscous. This could be rationalized by the fact that to maintain the same pressure the density of the long bond length liquid had to be decreased substantially. At the higher pressure the situation was reversed and although no correlation between the pressure dependence of viscosity was found, interestingly, the density dependencies were quite similar suggesting that the lubrication industry's pre-occupation with pressure coefficients of viscosity perhaps led to some useful correlations being overlooked.

1.4 N-alkane fluids

1.4.1 Ethane vs. Propane

A study was then made of the model alkane-like liquids ethane and propane using two centre (diatomic) and three centre (triatomic) united-atom representations, *i.e.* where one interaction site represents a CH₂ or CH₃ group. Again no attempt was made to parametrize the models to comply with experimental data. The models simply used the same mass and interaction parameters for each site so as to form an ideal homologous series. Calculations of the viscosity at a temperature of 200 K and ambient pressure were made using HSNEMD, both in steady state and perturbation mode, and also from the stress correlation function. Conclusions regarding the efficacy of the different methods were the same as for the previous case. The addition of the extra interaction site causes quite marked differences in the rheological properties with propane being significantly more viscous and showing a greater dependence of viscosity with shear rate. This study really just underlined the importance of the density of interaction sites in determining the properties of the liquids.

1.4.2 Hexane vs. Flexane

The study on alkanes were then extended to a six-centre model of n-hexane. Up to that point in time HSNEMD calculations had only been carried out on rigid molecules so the extension to systems of semi-flexible molecules was novel in itself². In the n-hexane study a comparison between more similar liquids was achieved by simulating a second liquid, christened *flexane*, which was in every way identical to the model of n-hexane except that the torsional potential, that hindering internal rotations, was set to zero. This was the first attempt to look at the influence of molecular flexibility on transport properties.

From equilibrium simulations carried out at 200 K and 300 K there was clear evidence that these liquids could be said to be much more in corresponding states than the previous comparisons. The diffusion

coefficients at 200 K differed though significantly with a 70% enhancement of the diffusivity of flexane over that of hexane³. In terms of the centre-of-mass velocity correlation function this difference could be largely explained by the much shallower first minima, *i.e.* less of a cage effect, in the flexane case as its enhanced intramolecular flexibility allowed it to respond less as a rigid body. This difference largely disappeared at the higher temperature as the diffusivity of hexane increased markedly quicker. It was concluded that this increased correlation between intramolecular flexibility and translational diffusion came about not because of any increase in conformational transitions, still relatively uncommon at 300 K, but because of the increased amplitudes of torsional motions.

The non-equilibrium simulations also showed that the molecular flexibility had a strong influence; at 200 K the extrapolated zero shear rate viscosity of n-hexane being twice that of flexane⁴. Again this difference largely disappeared at 300 K, as it did also for the pressure, energy and molecular extension as functions of the shear rate, and so the conclusions were much the same as for the case of mass transport.

1.5 Summary of Ph. D. thesis work

The greatest problem with the line of research summarized above is in separating the interesting molecular effects from the associated changes in the phase diagram which are normally associated with any change in molecular structure or interaction potential. This was particularly the case in the work on diatomics and for the ethane/propane comparison. Although some progress was made in the hexane/flexane case, the pressures for example differed by ~300 bar at the same temperature and density and, of course, there is absolutely no guarantee that their phase behaviour would be similar at all state points. Thus, subsequent work has avoided these complications by ensuring that the fluids compared do have *exactly* the same phase diagram and only differ thus in their dynamic properties.

1.6 The effect of mass distribution in rigid triatomic molecules

To the above end, the effect of mass distribution on the viscosity and diffusion of model triatomic molecules was studied⁵ as part of a Shell sponsored post-doctoral research assistantship. In this case equivalent states can be guaranteed for all the liquids so the comparison was free from the ambiguities of earlier work. In the study the molecules consisted of three interaction sites separated by rigid bonds and with a rigid tetrahedral bond angle. The total mass of the molecules was kept constant and three mass distributions were considered initially for which the points of interaction coincided with the points of mass. For these liquids the mass ratios were 6-6-6 (A), 1-16-1 (B) and 8.5-1-8.5 (C). A fourth liquid (A^+) was also generated by arbitrarily increasing the momentum of inertia of molecules of A by a factor of 100; as the equations of motion for the rotational and translational degrees of freedom were integrated separately this is not a technical problem. All four liquids necessarily have the same equation of state. Simulations were performed at 300 K both under equilibrium conditions and also using HSNEMD.

The results showed that, of the three realistic mass distributions, fluid C had the lowest diffusion coefficient whereas fluid A was the most viscous with differences in general being of the order of about 20%, well outside of the statistical errors. As expected the artificially enhanced moment of inertia fluid, A^+ ,

showed much greater differences being ~10 times more viscous and having a diffusion coefficient ~1/5 of that of A.

It was thus demonstrated that simply changing the mass distribution can alter the coupling between rotational and translational motions in a way which is sufficient to cause differences in the transport of mass and momentum. Rationalizing these changes remains a formidable challenge to liquid state theory.

1.7 Frozen distribution sampling (FDS)

More recently the approach of comparing fluids with the same equation of state has taken a further step forward with the development of the frozen distribution sampling (FDS) technique^{6,7}. The principal of FDS is quite straightforward. If we were to take a macroscopic sample of a system of flexible molecules and at some instant in time render all molecules completely rigid then the configurational probability densities are the same in the flexible and frozen systems and thus all the configurational properties, energy, pressure *etc.* must be the same. The theoretical justification of the technique is detailed in⁶ as too is a demonstration of the method for the case of n-butane. In practice the FDS procedure for the relatively small systems used in simulations is slightly different. One begins with an equilibrated system of flexible molecules and at some point in time the molecules are rendered completely rigid. For a certain amount of time the system is followed before allowing the molecules to regain their flexibility. By alternating periods of flexibility with those of rigidity a thorough sampling is made of the available configurational space. The time variation of the torsional energy and the pressure during such a sequence of alternating frozen and flexible intervals is shown in Fig. 1.7.1. for a system of n-butane⁶.

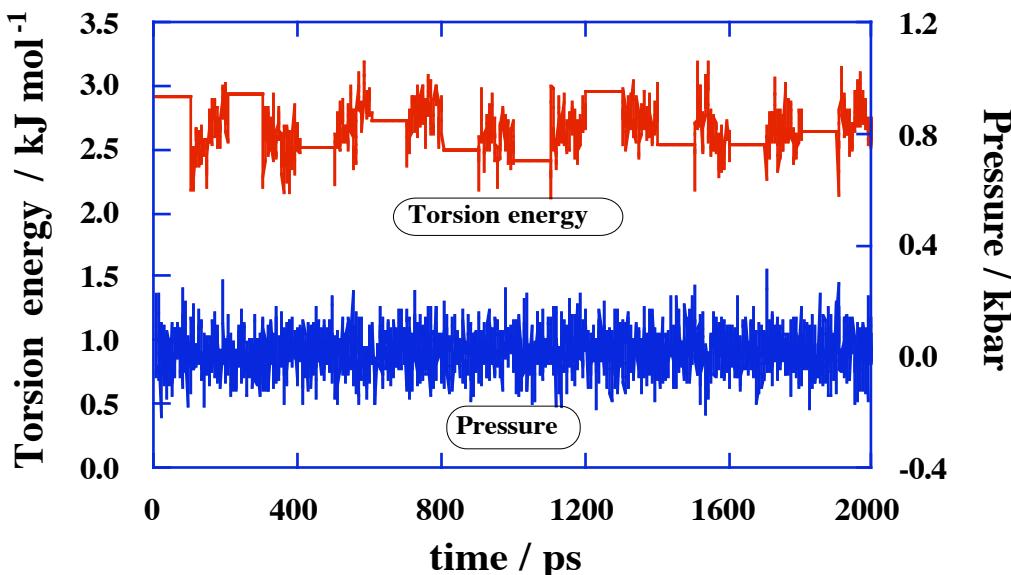


Figure 1.7.1 The time variation of the torsional energy and pressure during a 2 ns simulation of butane during which molecules were rigidified and then allowed to flex alternately for durations of 100 ps using the frozen distribution sampling method.

In Fig. 1.7.1 it can be seen that the frozen intervals, those of constant torsional energy, don't appear to cause any perturbation in the behaviour of the pressure. Indeed, examinations of both the means and the underlying distributions of configurational properties confirms the basis of FDS that the two systems are thermodynamically the same⁶.

In this study of n-butane⁶ it was found that, at the state point chosen, there was little coupling between the flexibility of the molecule and the mechanisms of mass and momentum transport; the diffusion coefficient and the viscosity were the same, within errors, for both the flexible and rigid cases. Extension of the study to other state points confirmed this conclusion⁸.

In a subsequent paper FDS was used to re-examine the hexane vs. flexane comparison³. Samples of both hexane and flexane were subjected to FDS at a range of temperatures between 200 K and 1000 K⁷. The results obtained from equilibrium simulations for the diffusion coefficients of the four different fluids are displayed in Fig. 1.7.2.

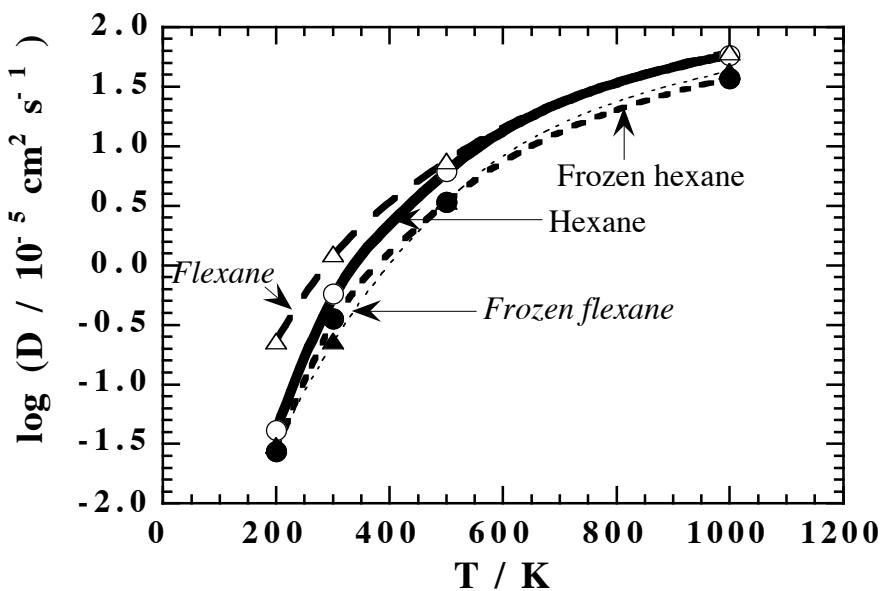


Figure 1.7.2 The natural logarithm of the diffusion coefficient plotted as a function of temperature for hexane, flexane, frozen hexane, and frozen flexane. A smooth curve has been fitted to the data as a visual aid.

Throughout the entire temperature range the coupling with the torsional degrees of freedom leads to an enhancement of the diffusion by about 20% in the case of hexane. Flexane shows considerably more coupling at the lower temperatures and so it can be said that, in comparison, the coupling of the flexibility to diffusion in hexane at low temperatures is rather weak. As the temperature is increased, and the torsional energy barriers in hexane become smaller relative to $k_B T$, there is a crossover to more flexane-like behaviour. Analysis of velocity correlation functions and their power spectra led to the conclusion that outer dihedral angles were mainly responsible for the coupling process. The results also confirmed the conclusion of the

previous study in that it is not the differences in static configurational properties that leads to the different diffusional behaviour of hexane and flexane but rather the differences in dynamical coupling.

1.8 Conclusions

Certainly the advent of the FDS technique represents a significant breakthrough in attempts to resolve the effect of dynamic coupling on transport processes. It removes all the ambiguity which plagued the early work in this area and opens the way to numerous possible applications in any area where it is interesting to distinguish between dynamic and static contributions. The ultimate goal of designing molecules to optimise certain aspects of their rheological properties still lies some way off but from such, and further, studies some useful pointers can be drawn.

Before leaving this topic it is worth drawing attention to a couple of associated pieces of work. The first is a detailed assessment of the HSNEMD technique and the second concerns a technique for measuring rates of reaction in MD.

1.8.1 Boundary driven shear

In a project designed to examine the fundamental assumptions implicit in the homogeneous shear non-equilibrium MD method, a study was made of boundary driven shear for a large system between parallel planes⁹. The model fluid was simply a system of spherical atoms interacting through a purely repulsive potential and the walls were made up of layers of the same types of atoms held close to lattice positions by harmonic potentials. To avoid having the results dominated by the boundary effects the simulations used over 40000 atoms. Such system sizes weren't feasible when boundary methods were first attempted and abandoned in favour of HSNEMD; only recently has this kind of simulation been feasible with the introduction of parallel computers. By translating the walls in opposite directions a shear flow is set up within the fluid between them. The heat generated is removed in this case by rescaling the momenta of just the wall particles. This means that within the fluid only the natural processes of conduction are in operation. Consequently gradients of density and temperature exist in the system. These local average density and local temperature at different points between the walls in the system were then used as inputs to HSNEMD calculations, using much smaller samples of 1000 atoms, to see if these gave the same results. Excellent agreement was found under those conditions where comparisons could be made. However, what could not be justified was the extension of the HSNEMD technique to arbitrarily high shear rates where heat is being generated at far higher rates than it could ever be conducted away by natural physical processes. The use of artificial thermostats in HSNEMD means that heat can be removed at an arbitrary, and ultimately unphysical, rate. It is not insignificant that the shear rates at which such phenomena as shear thickening, normal pressure differences and non-equipartition of energy start to occur in simple atomic fluids¹⁰ is well into this unphysical regime. The now rather numerous HSNEMD studies which have reported such "phenomena" must all be called into question. HSNEMD can be a useful way to extract the equilibrium shear viscosity but not too much relevance can be attached to the results of ultra-high shear rate studies.

1.8.2 Rates of Reaction

An important development in the methodology of characterizing rates of conformational isomerization has been made and its use demonstrated in the case of n-butane¹¹. Previously it had been

thought that the correlation function relating the rate constant for the dynamic equilibrium between the *gauche* and *trans* states was a collective property¹². However, a ‘thought experiment’ led directly to the realisation that it is in fact the auto-correlation function for the state which is required. In this experiment we identify a sub-system of molecules which at some particular time are all in the same state *e.g.* *trans*. Instantaneously the distribution in this sub-system corresponds to a non-equilibrium state; it will be driven subsequently towards an equilibrium distribution according to spontaneous fluctuations in the system. A relaxation function can then be obtained by averaging the data obtained from many such sub-systems. This function contains all the information required to define the relaxation of the system to equilibrium. It must be emphasised that this involves no external disturbance of the system and all the analysis can be carried out by post-processing the data from the equilibrium simulation.

Taking the bulk n-butane system as an example consider a system of N molecules. Let $H_T(\alpha_i(t))$ be the characteristic function of the *trans* state where $\alpha_i(t)$ is the dihedral angle of molecule i at time t and

$$H_T(\alpha_i(t)) = 1 \quad \text{if } -60^\circ < \alpha_i(t) < 60^\circ$$

$$\text{otherwise} \quad H_T(\alpha_i(t)) = 0$$

The number of molecules in the *trans* state at a particular time t_0 can then be written as

$$N_T(t_0) = \sum_{i=1}^N H_T(\alpha_i(t_0))$$

At a later time t_0+t the number that are in the *trans* state of those that were originally *trans* is simply

$$N_T(t_0, t_0+t) = \sum_{i=1}^N H_T(\alpha_i(t_0)) H_T(\alpha_i(t_0+t))$$

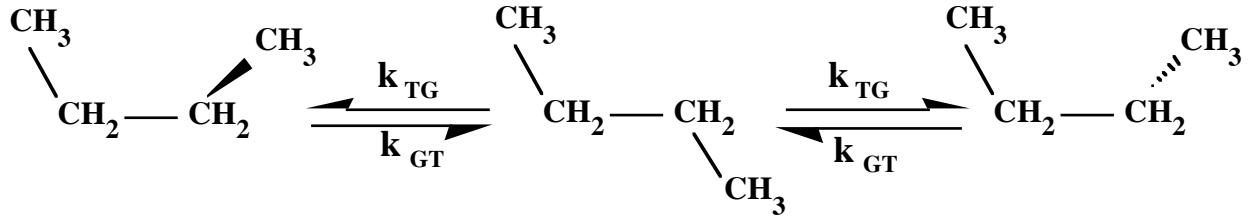
as only those terms where both H functions are equal to one contribute to the sum. The relaxation function, $R_{TT}(t)$, in this case can be defined as

$$R_{TT}(t) = \frac{\langle N_T(t_0, t_0+t) \rangle}{N} = \langle H_T(\alpha_i(0)) H_T(\alpha_i(t)) \rangle$$

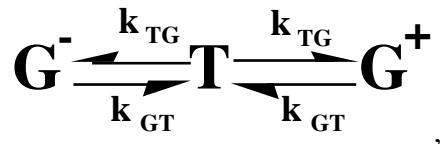
where the angle bracket denotes an average over time origins, t_0 . Clearly $R_{TT}(0) = \langle X_T(t) \rangle = \langle X_T \rangle$, *i.e.* the mean fraction of molecules in the *trans* state, and at long times $R_{TT}(t)$ will tend to $\langle X_T \rangle^2$.

$R_{TT}(t)$ has exactly the time dependence required to describe the way in which on average conformational equilibrium is established in a sub-system. By selecting that subset of particles which are *trans* at any particular point in time we perform an analysis similar to the non-equilibrium experiment of Edberg *et al*¹³ without disturbing the system at all. The comparison with their approach can be obtained simply by dividing $R_{TT}(t)$ by $\langle X_T \rangle$ which is tantamount to starting with a configuration of all *trans* molecules.

A demonstration of the power of this technique was given for the case of n-butane¹¹ where it was found that the long accepted two-rate-constant mechanism,



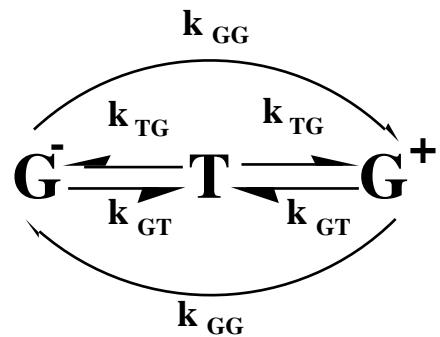
or more succinctly



was inadequate. In particular it did not explain the observed behaviour of the following relaxation functions

$$R_{G^\pm G^\mp}(t) = \frac{\left\langle \sum_{i=1}^N H_{G^\pm}(\alpha_i(t_0)) H_{G^\mp}(\alpha_i(t_0+t)) \right\rangle}{N}$$

which determine the way in which the population of the opposite, initially unoccupied, *gauche* well changes with time given that the *trans* well is also initially unoccupied. Examination of the results obtained revealed immediately that the rate of change of $R_{G^\pm G^\mp}(t)$ was not consistent with the kinetic equations of the original mechanism which predicted zero slope at $t=0$. It could thus be concluded that the traditional *trans*↔*gauche* interconversion mechanism was not sufficient to explain the observed data. Only by including direct *gauche*-*gauche* interconversions through the *trans* well could the data be explained, *i.e.* a mechanism of the form



In this mechanism direct interconversions are postulated between the two *gauche* states and so a third rate constant k_{GG} has been introduced. The third process involves direct $G^- \rightarrow G^+$ transitions via the *trans* well. They do not involve crossings of the very high $G^- \leftrightarrow G^+$ barrier. These are close to adiabatic or resonant transitions in which there is not sufficient energy dissipated for the molecule to settle in the *trans* configuration.

In some further work the method by which transition state theory estimates of the rate constant may be calculated from simulation data have been discussed ^{14,15}. In particular, it was shown that care has to be taken when calculating the average velocity along the reaction coordinate at the barrier. If this is computed just from actual barrier crossings then a systematic overestimation by a factor of exactly $\pi/2$ occurs. Further analysis allowed the connection between the barrier crossing rate and the transition state theory value for the rate constant to be established. Rate constants determined by the "reactive flux" method were compared with values obtained using the relaxation function approach in the case of liquid n-butane. The results of this comparison had implications regarding the identification of the plateau value when using the reactive flux method. A straightforward procedure was described which can establish confidence limits for rate constants calculated in this way.

2. Polymers

2.1 Introduction

The initial interest in polymers stemmed from a short 3 month contract with Du Pont on the mechanical properties of the amorphous/crystalline interface; of which more will be said in the next section. The model polymer used was simply an extension of the united-atom model of n-alkanes used in the thesis work. From this modest beginning has developed a body of work which, for the purposes of this summary, has been separated into four areas: mechanical properties, chain configurations in the melt, constructing samples of amorphous polymers and polymer electrolytes. The boundaries between them are not so clear cut as certain aspects of the work come into more than one area. In particular, a lot of the work required, as a starting point, a sample of an amorphous polymer. This is not a trivial matter as the relaxation times of chains of relatively moderate length far exceeds that which can ever be possibly simulated. Thus, a “good” initial guess of the structure is highly desirable. This then leads to questions about the predictability of polymer chains in the bulk melt. If we know what these are then at least we have some hope of short-circuiting the otherwise prohibitive procedure of waiting for nature to take its course.

2.2 Mechanical properties

2.2.1 Amorphous/crystalline interface

As stated above, the united atom model of n-alkanes, developed during the course of the thesis studies, was used in a study of the amorphous/crystalline interface of polymers¹⁶. The interest being in the tensile properties of fibres and how they are affected by the way in which the amorphous component connects the crystallites.

In the study a lamina model was used in which the chains were suspended between two fixed parallel planes representing the faces of adjacent crystallites. Results were then generated both for an ordered structure, consisting of 16 chains of 25 sites per chain, and an amorphous structure which was generated by an elaborate growth procedure and which contained not only chains attached to both crystallites but also loose ends and loops (chains re-entering into the same face). Both structures were then brought to equilibrium at 300 K and the tension was then measured as a function of the distance apart of the crystallites, thus giving a stress vs strain curve. From these curves it was established that the Young’s moduli in the two systems differed by a factor of 10. For the aligned system the power spectrum of the stress autocorrelation showed distinct peaks attributable to the accordion modes of the chains (angle bending deformations). The associated dispersion plot was used to obtain the Young’s modulus via the speed of sound route. This value was found to be in good agreement with that from the stress vs. strain curve.

The great difficulty with this kind of study is the creation of a realistic model of the amorphous/crystalline interface. With very little evidence to go on the possible ways in which the amorphous region can be generated are numerous. At the time it was too costly to even begin to explore the dynamic

behaviour of a relatively simple model interface with respect to the distance between the planes and the various other growth control factors employed.

2.2.2 Amorphous systems

In principal, purely amorphous polymers are a simpler problem to tackle as one has at least a reasonable starting point for the structure in that of the corresponding bulk melt. Although, as mentioned above, it requires techniques to predict this structure in order to short-circuit the long equilibration times of ensembles of polymer chains, there are good reasons to suggest that this is a tractable problem; more about this will be said in subsequent sections.

As well as the problem of generating relaxed samples of amorphous polymers there is also the problem of how to determine their mechanical properties. In the initial work on the amorphous/crystalline interface there were boundaries which could be moved in order to apply strain to the system. In the purely amorphous case it is more sensible to use periodic boundaries in order to have a homogeneous system. A method was thus developed in order to be able to apply external pressure fields to the relaxed amorphous samples¹⁷. The method chosen was based on constant-temperature constant-pressure loose-coupling ideas of Berendsen *et al*¹⁸. Adaptations were made to the original method though to ensure that initially symmetric MD boxes remained symmetric. The choice of the loose-coupling method in preference to the more rigorous Rahman-Parrinello (RP) technique¹⁹ was made for two reasons. First, the inclusion of rigid constraints into the Lagrangian, required for the RP technique, is a difficult, and at the time unsolved, problem. Second, the pressure differential between the externally applied and internally measured pressure tensor is coupled to the first derivative of the box basis vectors in loose-coupling, rather than the second derivative in the RP technique. Thus the motion of the box is over damped and has little tendency to oscillate as can be the case with RP, this was seen as an important advantage when it comes to performing mechanical experiments with time-dependent pressure fields.

With this technique the properties of a linear polymer model, resembling polyethylene, have been studied over a wide range of temperature²⁰. Five independent samples containing just one long chain of 1000 sites each were generated and equilibrated at 500 K. Samples at different temperatures were then obtained either by heating or cooling. The average density as a function of temperature is shown in Fig. 2.2.1 for these samples.

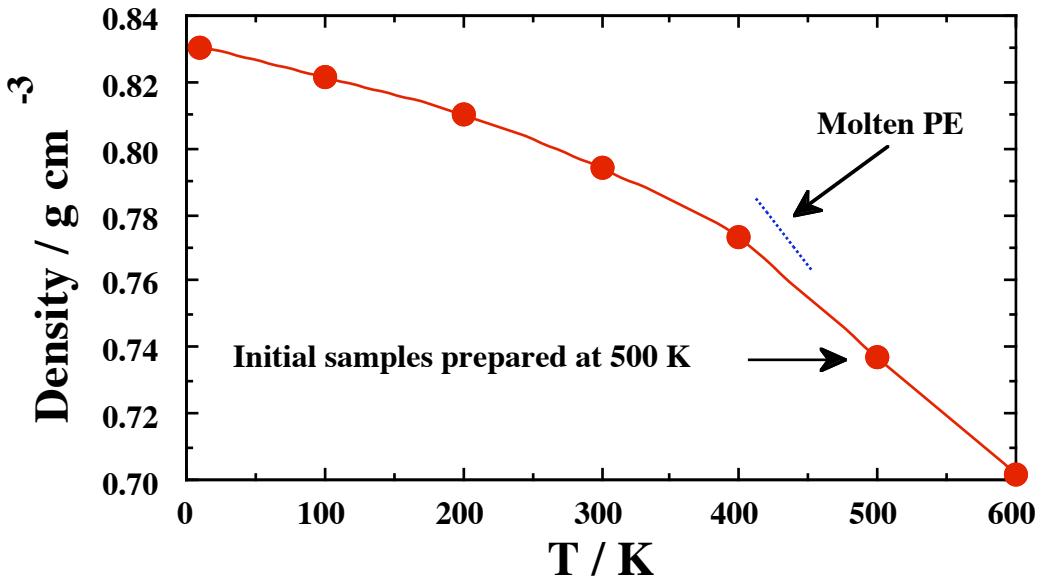


Figure 2.2.1 The average density of the polymer model at a pressure of 1 bar as a function of temperature. The initial samples were grown and relaxed at 500 K and samples at other temperatures were prepared by heating or cooling at a rate of 1 K ps^{-1} . The interaction potential was adjusted to give a density at 500 K which corresponds approximately to the extrapolation of experimental data for molten polyethylene (dotted line). The spread in the results from the five independent samples is less than the size of the symbols.

To investigate the mechanical properties the sets of relaxed samples at each of the five temperatures were subjected to a gradually increasing uniaxial tension by changing the y component of the *applied* pressure tensor; two different values for the tension application rate were used, 5 bar ps^{-1} and 1 bar ps^{-1} , in order to gauge the extent to which the measured properties were rate dependent. The primary interest in these experiments is the strain induced in the sample. The applied tension is then best considered as a control variable which produces a change in the strain the response to which is the *measured* tension. In these kind of experiments both the strain and the tension are thus dependent variables. The method is preferable to direct control of the strain since there is no way *a priori* of predicting how the shape or density of the sample will respond to a change in the external conditions. Studies which maintain a constant volume impose a value of 0.5 for Poisson's ratio and are equivalent to applying a complex pressure field which makes interpretation and comparison with real systems more difficult.

By applying uniaxial tension the elasticity, yield and plastic flow of low temperature samples and the viscoelasticity of higher temperature samples was characterized. The average response to tension of the model polymer samples is shown in Fig. 2.2.2.

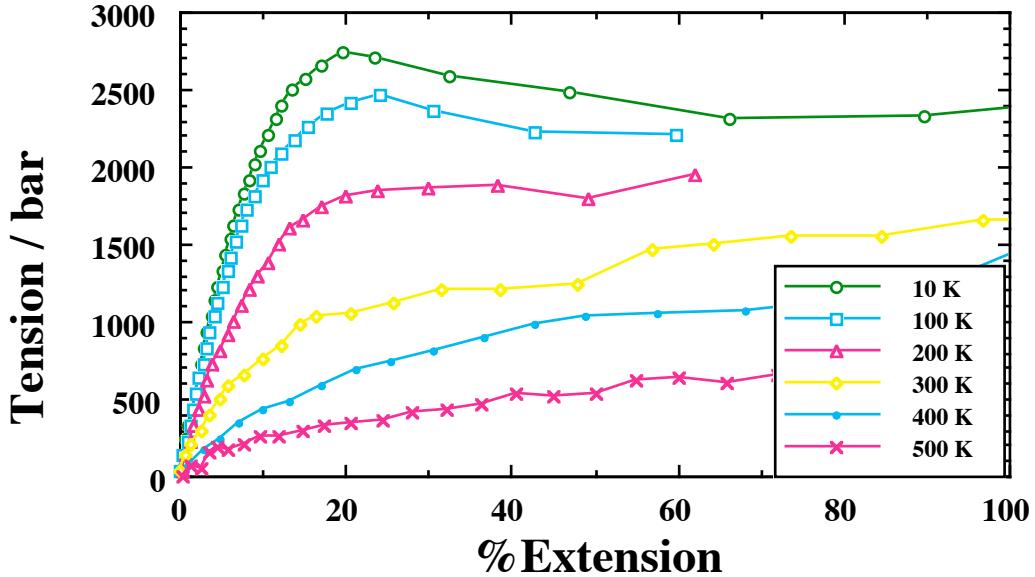


Figure 2.2.2 The measured tension as a function of the percentage extension for tension applied at a rate of 5 bar ps^{-1} . The data at each temperature represent the average behaviour over 5 independent samples.

Qualitative comparisons showed that there were strong similarities between these short time ($\sim 1 \text{ ns}$) simulations and laboratory measurements despite the large difference in time scales. The extensional (Young's) modulus derived from the initial slopes of the plots shown in Fig. 2.2.2 are shown in Fig. 2.2.3. They have values typical of a glassy solid for $T \leq 100 \text{ K}$ and show a significant decrease as the temperature is raised.

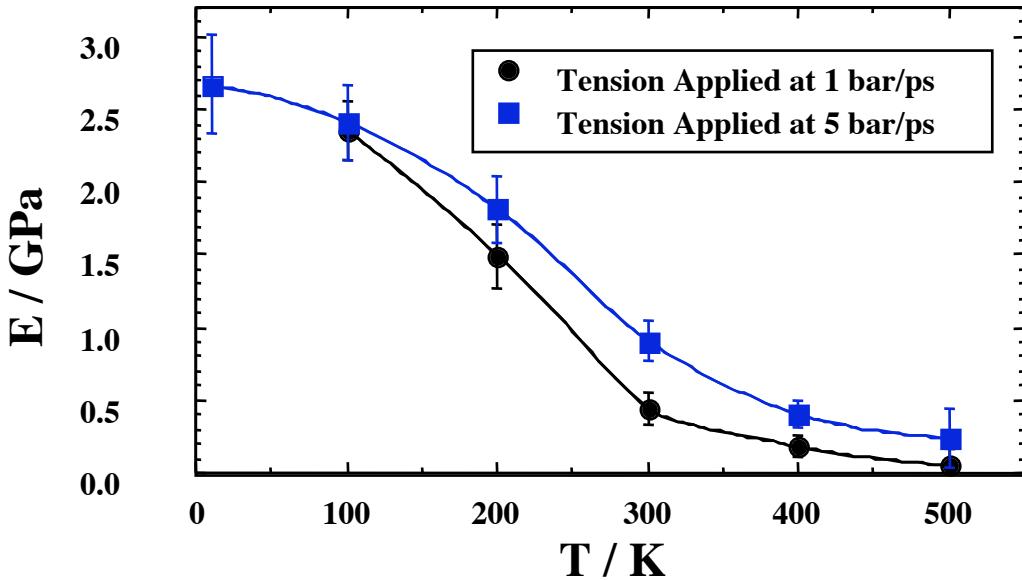


Figure 2.2.3 Extensional (Young's) modulus obtained from small ($<5\%$) strain behaviour as a function of temperature. The squares and circles refer to the data obtained at tension application rates of 5 bar ps^{-1} and 1 bar ps^{-1} respectively.

The high temperature of the glass transformation, as characterized by the disappearance of a yield stress, and the breadth of the transformation region, as characterized by density and modulus, can both be explained in terms of the ultra short time scale of the computer experiments. First of all, the simulated samples are cooled at very high rates compared to experiment thus the structural relaxation rates begin to lag behind the rate of perturbation of the system at quite high temperatures so the transformation starts well above the "normal" glass transition temperature. Secondly, these types of materials show a marked temperature dependence of the activation energies (E_a) for structural relaxation processes. At low temperatures E_a becomes extremely large and relaxation times may change by one or two orders of magnitude over a few degrees - thus producing a very sharp transition. At high temperatures a much smaller value of E_a means that the same change occurs over a much wider temperature range - thus producing a broad transformation.

Despite the lack of resolution regarding the glass transition, the fact that a model as simple as a 1000 site chain filling space via the replicative properties of periodic boundaries displays many of the characteristics of real systems was quite remarkable. In some further work the same model was used to investigate the configurational dependence of the mechanical properties.

2.2.3 Effect of chain configuration on mechanical properties

An interesting question in polymer physics concerns the configuration of the chains in a glassy polymer and the effect this has on mechanical properties. In the laboratory this is a difficult question to answer unambiguously as it is difficult to alter chain topology without changing the chemistry of the polymer. The control that can be exercised over the sample generation has allowed simulation, however, to make a systematic study of this effect ²¹. In this study, samples of the same model of a linear polymer were prepared in four different ways at a temperature well below the glass transition. Two parameters were used to characterize the resultant chain topologies, a correlation length, a , and the percentage of *trans* conformers. The values of these two parameters for the four preparation procedures (five independent samples in each case) is summarized in the following table:

SAMPLE SETS	% trans	$a / \text{\AA}$
A - Cooled from the melt at 500 K	78	~5
B - Grown at 200 K	82	~16
C - As B but annealed at 1000 K for 100 ps	77	~12
D - Grown at 10^5 K	70	~6

These samples were then subjected to uniaxial extension tests in the same manner as described in the previous section. The resulting stress vs strain plots are shown in Fig. 2.2.4.

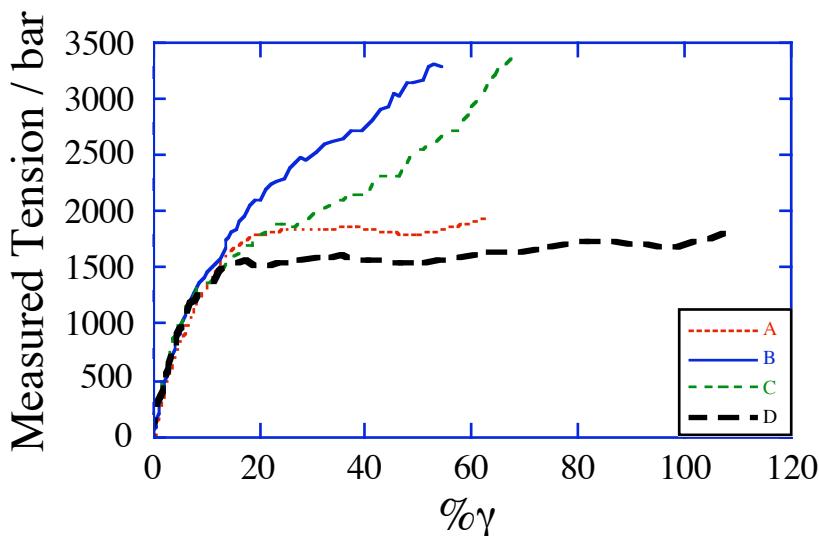


Figure 2.2.4 Measured tension plotted as a function of the percentage extension for the four sets of samples (see table) of the model linear polymer subjected to an applied uniaxial tension increasing at a rate of 5 bar ps⁻¹.

The plot demonstrates very clearly the effect chain configurations can have on the mechanical properties. Up to about 10% strain though the differences are fairly small. At larger strains samples B and C start to show enhanced resistance to extension beyond the point at which the A samples yield (~20% extension). The high *%trans*-high correlation length sample (set B) show the largest extent of strain hardening and in particular produce significantly more stress than the set C which, within the error, has a similar correlation length but lower *trans* fraction. Conversely sample set C has practically the same *%trans* as set A so the difference here must be due to their contrasting configurational structure. Set D shows the lowest resistance of all to the applied tension as was expected from its highly coiled structure with a smaller fraction of *trans* conformers. Thus it can be concluded from this study that both an increase in the persistence length and an increase in the *trans* fraction can *independently* contribute to strain hardening.

2.2.4 Conclusions

For either purely amorphous or purely crystalline polymers the methods outlined above are capable of elucidating the stress vs. strain behaviour for a number of applied perturbations (shear, extension, compression *etc.*). In the amorphous case it is important to remember that the response is likely to be highly time-dependent and so comparisons with experiments made on laboratory time scales have to be made with this in mind. Another way of consider this is in terms of frequency. In the laboratory a dynamic mechanical measurement would be performed typically at frequencies of 1 Hz. In the MD simulation the same kind of low-amplitude oscillatory-strain experiments can be done but only at frequencies in the GHz regime. Nevertheless there are similar problems linking the different frequency regimes probed by dynamic mechanical analysis and, say, dielectric measurements. There is, therefore, some scope in looking into the overlap between these techniques.

For the semi-crystalline polymers the problem still remains as to what is their structure. Although it is clear that structure exists on various scales, at the smallest scale crystal lamellae of thicknesses measured in nanometres are separated by amorphous regions. Since the lamella thickness is generally much less than the stretched out length of the individual chains it is widely thought that chain folding occurs and that chains may span several crystallites and amorphous regions. There is indeed much speculation regarding the structures of chains in these regions but very little in the way of hard evidence. Yet it is the details of the coupling of the crystallite and amorphous regions which is crucial when it comes to the mechanical properties. With the advent of massively parallel computers it is becoming feasible to think in terms of performing a very large scale simulation of a relatively simple polymer model in order to follow the process of cooling from the melt and, at least, the initial formation of crystal nuclei.

2.3 Constructing samples of amorphous polymers

One of the basic problems regarding the simulation of polymers in the amorphous state is how can an initial configuration be generated when it is known that configurational relaxation times of even modest length chains (1000 monomers) are far longer than can ever be simulated using classical atomistic MD. In the previous section it has been demonstrated that the details of the generation procedure are indeed very important so it is clear that this is quite different to the situation in low molecular weight liquids where the method of sample preparation is generally irrelevant.

2.3.1 Excluded volume growth (EVG)

Initial attempts at generating amorphous structures used self-avoiding walks within the confines of the periodic cell^{22,23}. In brief, the procedure involved adding sites sequentially with equilibrium bond lengths and angles and a random torsion angle. A new site was accepted on the basis of the total energy difference, $\Delta\Phi$, caused by the introduction of the extra site by comparing the Boltzmann factor, $\exp(-\Delta\Phi/k_B T)$, to a random number generated on the interval 0 to 1; the new site being accepted if the random number is less than the Boltzmann factor. As the chain is generated the density of sites increases and it becomes increasingly difficult to add sites. To overcome some traps backtracking was incorporated into the method in which sites could also be removed as well as added. Despite these modifications the densities that could be achieved were fairly low and thus a certain amount of volume relaxation had to be done using constant pressure MD. A far more serious problem was that the chain statistics were quite biased. At the start of the growth procedure the density is effectively very low and sites are added very easily. Towards the end the density is relatively high and it becomes increasingly more difficult to add sites. This manifests itself in an increase in the fraction of *gauche* conformations.

2.3.2 Phantom Chain Growth (PCG)

To avoid the problems of growing into an ever increasing density, the phantom chain growth (PCG) method was devised²³. In this method sites are again added sequentially but the range of interaction is limited. In the united-atom polyethylene type chains this meant that only neighbouring torsions were taken into account ("pentane effect") when calculating energy differences. This is in keeping with the spirit of Flory's ideas concerning the configurations of homopolymers in the melt. The problem with this approach is that at the end of the growth procedure there are a large number of high energy overlaps between sites and it

is not possible to begin an MD simulation at the required density with all the non-bonded interactions switched fully on. The most satisfactory way that was found to do this was using a truncated force potential²³. Beyond an inner cutoff the repulsive non-bonded force between two sites is fixed at the value at this cutoff. By doing this and gradually reducing the inner cutoff until the full potential can be switched on allows the excluded volume to be introduced gradually. The advantages of PCG over EVG have been demonstrated in²³. Essentially the chains are a good deal less biased in their configurations and time is saved as systems can be prepared directly at any desired density.

2.3.3 Pivot Monte Carlo sampling

Although PCG represented a big improvement over the EVG technique it nevertheless represents a form of sampling which introduces a bias into the chain statistics. In a further paper details of an improvement have been given²⁴ which completely removes any bias. The method simply involves using a Monte Carlo pivot algorithm to sample efficiently the bond angle and torsion angle phase space of a single isolated chains in which only specific non-bonded interactions are included in the Hamiltonian. A comparison of the PCG sampled chains with those obtained using the pivot MC algorithm²⁴ revealed that the bias increases with chain length to an extent where for 1000 site chains the mean square radius of gyration, $\langle S^2 \rangle$, of PCG chains is ~10% less than the unbiased pivot MC value. Figure 2.3.1 shows a plot of the probability density of S^2 for chains of this length. Similar behaviour is found for the fraction of *trans* conformers.

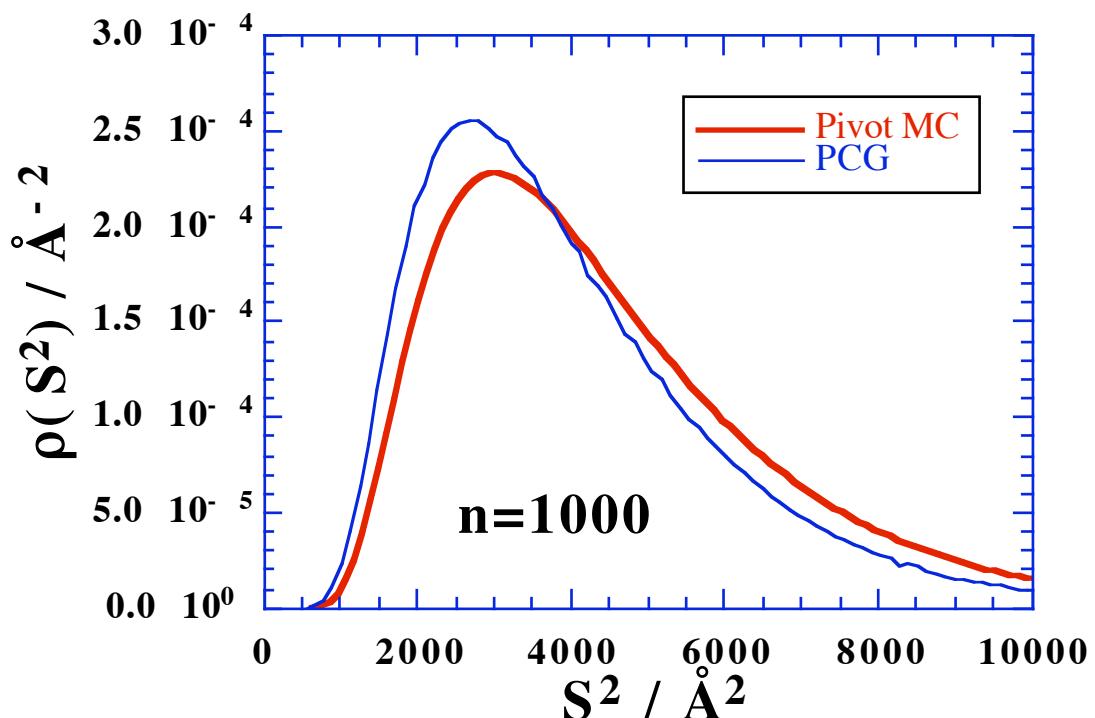


Figure 2.3.1 The probability density of mean square radii of gyration obtained for chains of $n=1000$ sites at 500 K using pivot MC sampling and the PCG method on isolated chains. Note how the inherent bias in PCG distorts the distribution towards shorter values of S^2 .

The bias produced by PCG at the growth stage is not surprisingly retained by the chains when the excluded volume is introduced. A demonstration of this effect is given in Fig. 2.3.2 where results are shown for the variation of $\langle S^2 \rangle$ with time following the introduction of excluded volume for systems containing ten (united-atom n-alkane-like) chains of $n=100$ sites per chain; the results being averaged over 512 independent samples in each case.

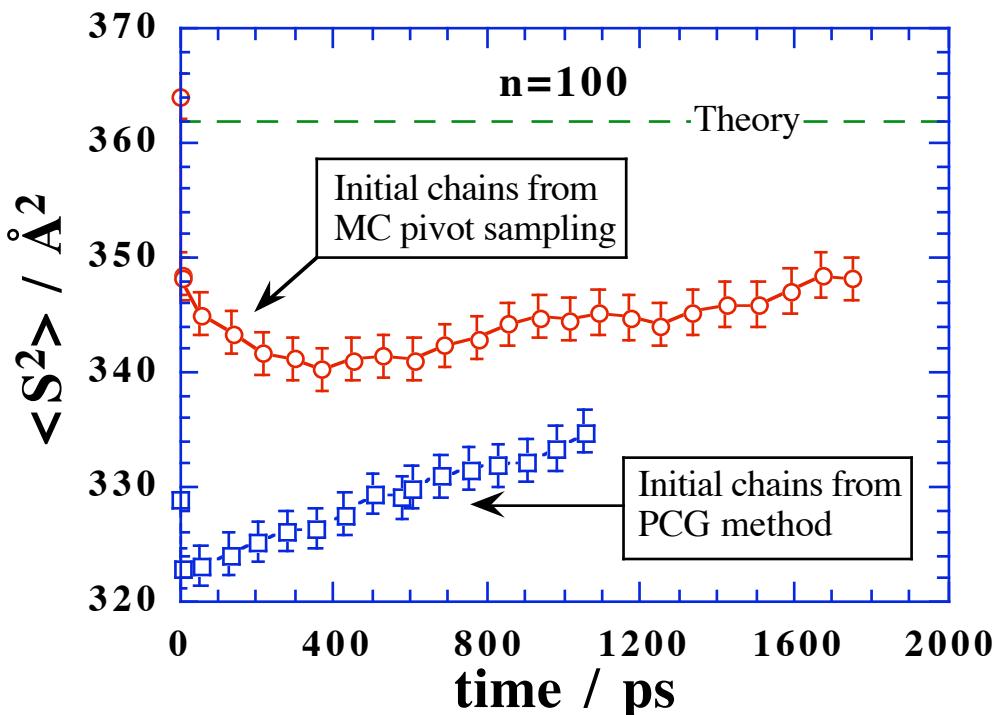


Figure 2.3.2 The behaviour of the mean squared radius of gyration with time for $n=100$ site chains grown using the pivot MC algorithm and the PCG method at 500 K. Both results are averages from 512 independent simulations with each sample containing 10 chains. A rapid initial depression in the value of $\langle S^2 \rangle$ as excluded volume is introduced is followed by a very long time scale relaxation back towards the theoretical result shown by the dashed line.

Apart from demonstrating the effect the bias PCG imparts to chain, Fig. 2.3.2 also illustrates an awkward problem associated with the introduction of excluded volume. Despite the pivot MC sampled chains having the theoretical statistics of chains in the actual bulk melt (in the next section evidence will be given to demonstrate the agreement between theory and simulation for this system) the introduction of the excluded volume perturbs this carefully prepared distribution. On the time scale of Fig. 2.3.2 this is not obvious as the perturbation occurs within the first few picoseconds where all the high energy overlaps are removed; the theoretical value (dashed line on Fig. 2.3.2) and the initial pivot MC value are one and the same

within errors. Once perturbed the distribution only returns to that desired at a rate determined by the natural relaxation processes in the system, *i.e.* exactly the situation such preparation schemes are trying to avoid.

This problem has been studied in greater detail for $n=50$ chains²⁴ where various attempts were made to avoid the initial disruption to the prepared chain statistics. Slowing the rate of introduction of excluded volume just slowed the rate at which the perturbation occurred and did not alleviate its extent and so was counter productive. Similarly, attempts to preserve the chain distribution by using energy minimisation instead of MD to remove the high energy overlaps did not help and just delayed the seemingly inevitable. Suspicions that the real chain statistics in the bulk melt and those obtained by pivot MC on isolated chains were different in some subtle higher order way were scotched by taking samples of fully relaxed chains, randomising the orientations and centre of mass positions of the chains and then re-introducing the excluded volume. Again the same perturbation was observed as is demonstrated in Fig. 2.3.3.

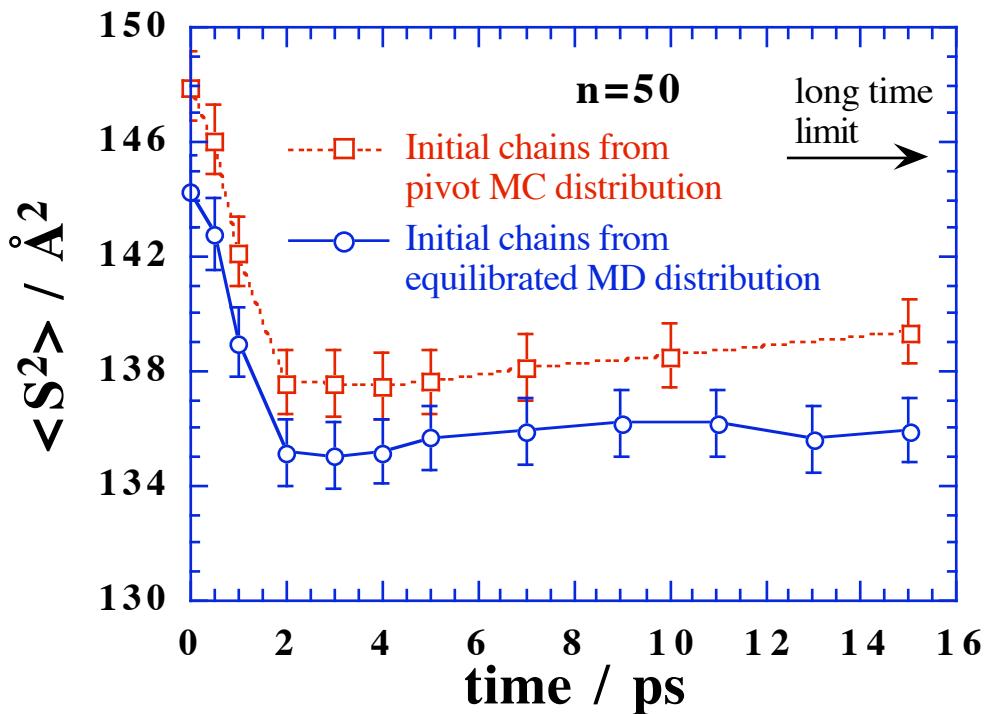


Figure 2.3.3 The effect of the initial distribution of chains on the behaviour of the mean square radius of gyration during the introduction of excluded volume. The results plotted are from averages over 64 independent samples of $N=20$ chains of $n=50$ sites per chain. The results for a set of 64 samples grown using the pivot MC method is compared to a second set of 64 obtained by randomising the orientations and centres-of-mass only of chains taken from MD equilibrated bulk fluid configurations.

2.3.4 Conclusions

Pivot MC sampling has clear advantages over the earlier methods of sampling chain configurations. Not only does it produce a distribution of configurations which is unbiased, it can also produce a distribution close to that *actually found in the bulk melt*; a point which will be expanded on in the next section.

Unfortunately, the rather drastic changes that accompany the introduction of excluded volume render our careful preparation of the initial distribution of chains somewhat in vain. The perturbations in S^2 amount to about 5% which may or may not be significant depending on the property of interest, *e.g.* moduli, penetrant diffusion, *etc.* Such questions remain largely unanswered.

In hindsight it was perhaps naive to expect that a random selection of chains placed in a periodic MD cell will fill the space very uniformly or could ever do so without undergoing some internal distortion. If we take an equilibrium fluid configuration of chains, randomise positions and orientations, as in the experiment described above, would the chances of re-assembling anything close to the original space-filling configuration be anything other than infinitesimally small? This would be a complex 3-D jigsaw problem even if we knew that the "pieces" did fit together.

2.4 Configurations of chains in the melt

2.4.1 Introduction

General issues regarding the predictability of the configurations of chain molecules in the melt were raised during the course of the work, described in the previous section, concerning the preparation of samples of amorphous polymers. In essence, if the configurations of polymer chains in the melt were predictable then we have a sound basis on which to construct models of amorphous polymers. It has thus been seen as of prime importance to establish to what extent, and for which systems, the melt configurations are predictable.

In the examples which follow the approach has been the same. The configurations of chains in the simulated bulk melt are compared to those expected from "theory". The initial work here has again concerned simple linear n-alkane-like chains but then interest has turned to more realistic models of the polymers polyethylene oxide (PEO) and polyvinyl chloride (PVC) which contain explicit hydrogens and, perhaps more importantly, partial charges. The "theory" is based on the ideas of Flory who has argued that there is complete screening of long range interactions in pure melts composed of linear homo-polymers²⁵. The random coils adopted by the molecules are not expanded, as is often the case in a good solvent at high dilution, nor contracted, as they may be in a poor solvent. The dimensions of the chains in the melt can then be determined by considering *isolated* chains with only a certain number of specific near-neighbour interactions. This is the basis of Flory's analysis of chain structure and is a fundamental premise underpinning the rotational isomeric state (RIS) theory approximation²⁵. Although there is some evidence in support of Flory's hypothesis²⁶ precise experimental measurements at high concentrations are difficult to make. Hence simulation is a much more satisfactory way to perform an unambiguous test of these ideas as *exactly* the same model is used both in the theory *and* in the bulk melt simulations.

2.4.2 Linear n-alkane-like chains

In an initial study, molecular dynamics (MD) was used to determine the dimensions of fully interacting chains with $n=4, 5, 8, 20, 50, 100$ interaction sites in the pure liquids which were then compared to the results of Monte Carlo (MC) sampling of chains generated using the Flory model in which only correlations between nearest neighbour torsion angles are considered²⁷. To obtain results of high precision for the $n=50$ and $n=100$ liquids, averages were taken over a number of independent systems. To facilitate this calculation a massively parallel Fujitsu AP1000 computer was programmed in such a way that each of its processors carried out an independent simulation simultaneously. This "cloning" approach to performing MD simulations is a simple and very efficient alternative to other strategies which spread one simulation across all processors.

A comparison of the relaxational behaviour obtained by MD for the $n=100$, $n=50$, $n=20$ and $n=4$ bulk melt systems is shown in Fig. 2.4.1 where the mean square radii of gyration, $\langle S^2 \rangle$, have been plotted in a normalized form, *i.e.* divided by the "theoretical" values derived from the MC calculation, as a function of time.

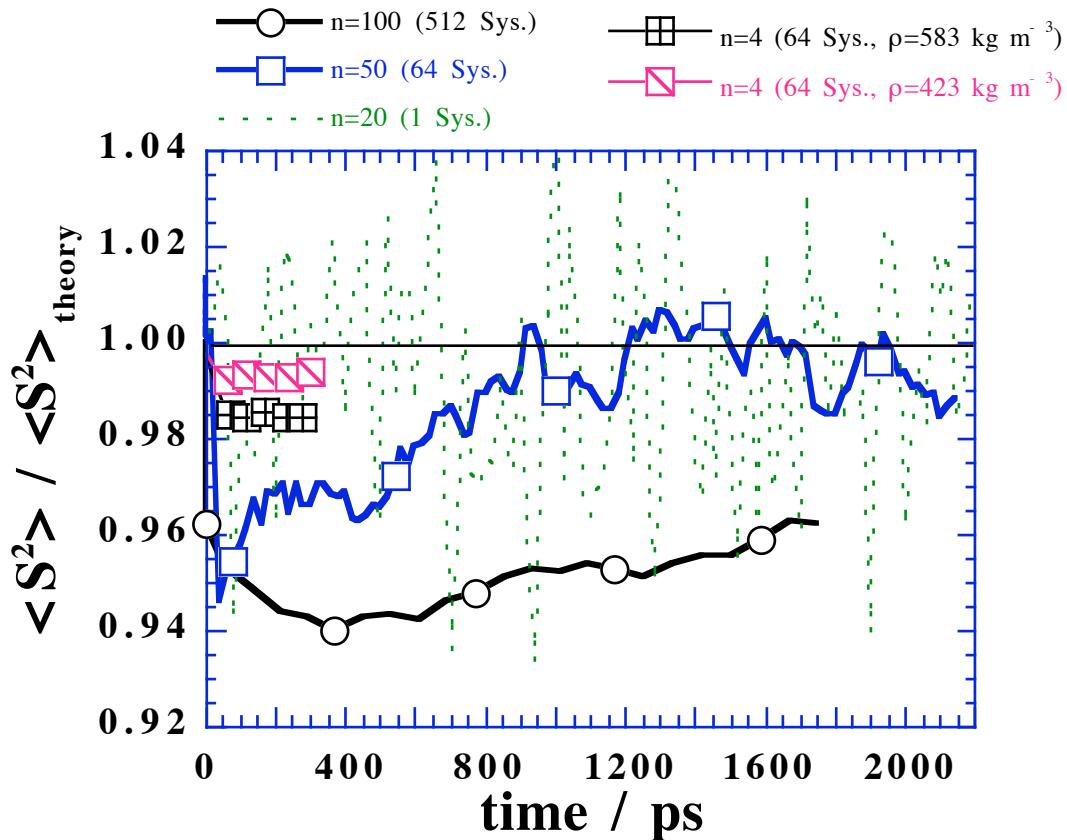


Figure 2.4.1 The relaxation of the mean proportion of the mean squared radius of gyration from the MD simulations of systems of $n=100$, $n=50$, $n=20$ and $n=4$ site chains at 500 K. The normalizing theoretical values are those generated using the pivot MC algorithm for the Flory type chains.

As was pointed out in the previous section, the introduction of excluded volume during the first few picoseconds causes significant perturbation of the $\langle S^2 \rangle$. The subsequent rate of relaxation of $\langle S^2 \rangle$ for n=50 and particularly n=100 is clearly slow on the time scale of the MD simulations. Although less precise the n=50 results appear to have relaxed within 1 ns to a value very close to that predicted by the theory. The $\langle S^2 \rangle$ values for the n=4 systems show some small but significant differences which vary according to the density. Such "solvent" effects are not taken into account by the theory; further discussion of this point can be found in ²⁷.

More convincing evidence that the theory does indeed work very well can be gained by examining the underlying distribution of square radii of gyration. Such a distribution is plotted in Fig. 2.4.2 for the relaxed n=50 chains.

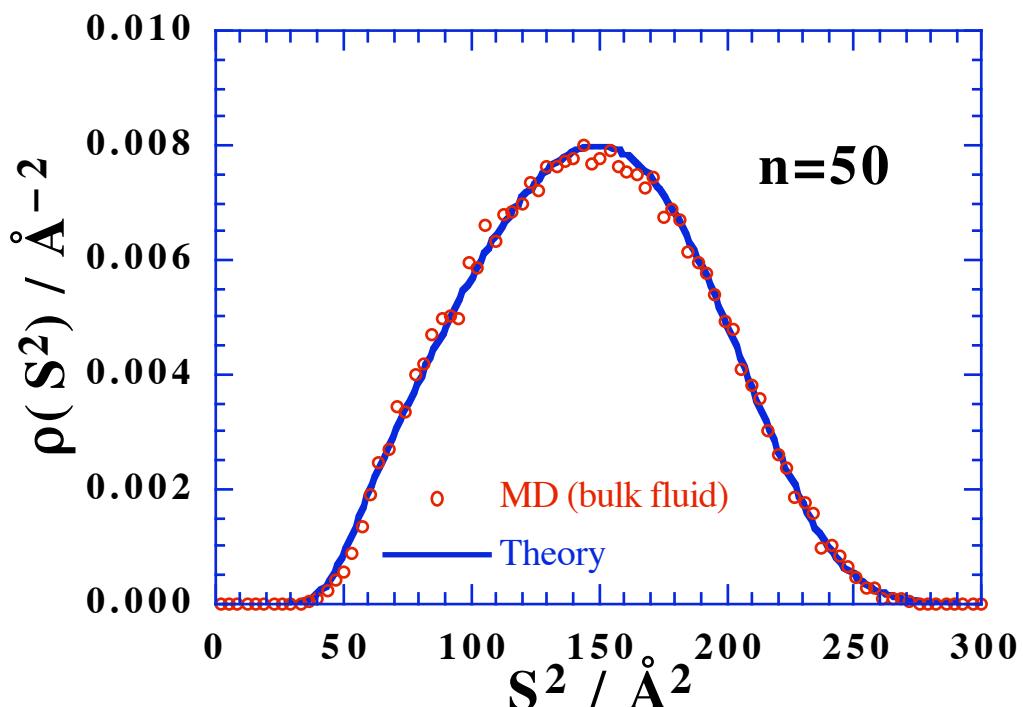


Figure 2.4.2 The probability density for the mean square radius of gyration obtained from the period 1000→2140 ps of the MD simulation of 64 systems each containing 20 n=50 site chains. The theoretical distribution obtained from the MC pivot algorithm sampling of isolated chains (thick line) is also shown.

The good agreement between the theory and MD results shown by the relaxed n=50 site chains compares starkly with that of the clearly unrelaxed n=100 site chains; see Fig. 2.4.3.

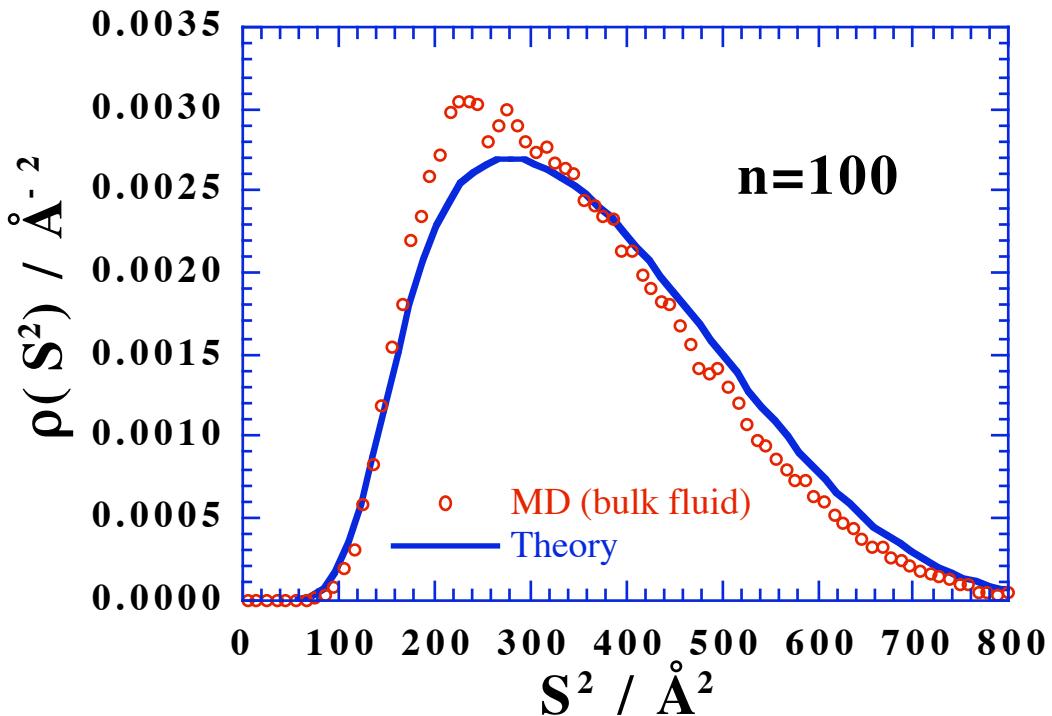


Figure 2.4.3 As Fig. 2.4.2 for the period 1085→1745 ps of the MD simulation of 512 systems containing 10 $n=100$ site chains. Note that the MD distribution has not reached equilibrium.

In the case of the $n=100$ site chains the "cloning" approach to parallel MD is not particularly suitable as the relaxation times are too long given the limited speed of the CPUs used. For this reason a domain decomposition (DD) approach was used instead so as to spread a large scale simulation over a number of processors. In principle the DD approach is not limited by the speed of individual processors as we can always use more processors on the problem. The code developed utilised the same minimum communication scheme as previously developed for evaluating efficiently the non-bonded pair interactions in purely atomic systems²⁸ but could also handle the three- and four-body potentials, required for the linear chain models, and also included a parallel version of the SHAKE routine for maintaining the rigid bond constraints²⁹.

With this program running on a Fujitsu AP1000 with 1024 processors it was possible to perform a much longer simulation of ~8 ns on just one large system containing $N=640$ $n=100$ site chains³⁰. The relaxation of the mean squared radius of gyration in this case is shown in Fig. 2.4.4 where it is compared to the previous cloning result obtained from 512 samples of $N=10$ chains per sample.

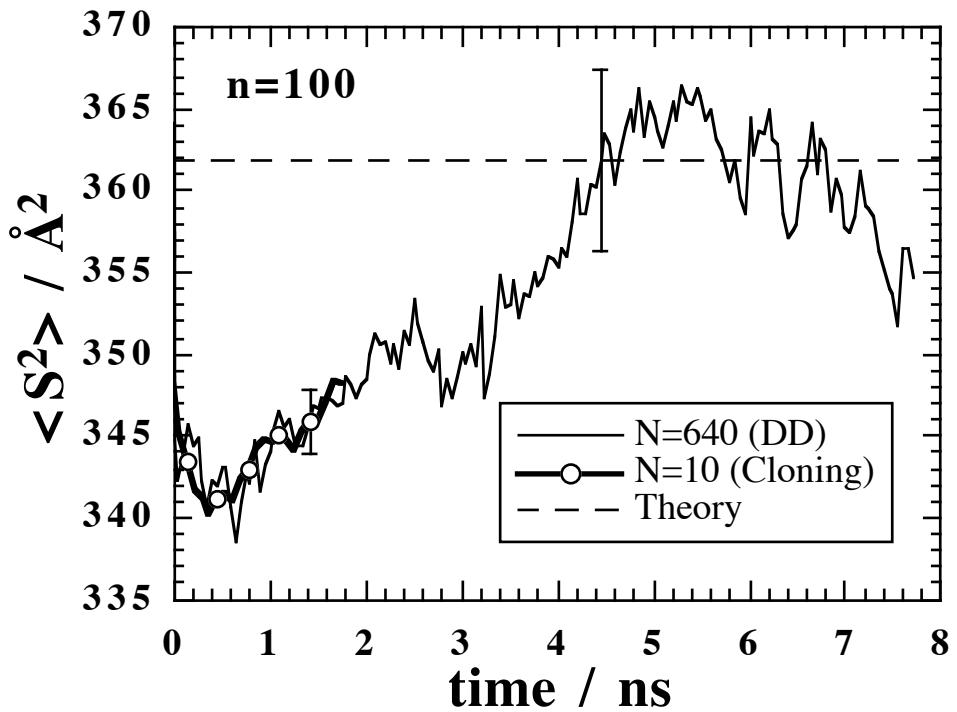


Figure 2.4.4 The relaxation of the mean square radius of gyration. The result from the large system containing $N=640$ chains is compared to that obtained previously for systems containing just 10 chains. When relaxed fully it can be seen that $\langle S^2 \rangle$ tends to the value predicted by theory. The two error bars shown indicate the relative precision of the two simulations.

Although it is true that the equilibrium value of $\langle S^2 \rangle$ is as yet unknown for the $N=10$ system, the similarity of the initial relaxation suggests that the value of $\langle S^2 \rangle$ is apparently insensitive to a 64-fold increase in the size of the sample. For the large sample containing 640 chains the ratio of the root-mean-square end-to-end distance to the unit cell length, $\langle R^2 \rangle^{1/2}/L = 0.39$, whereas for the smaller samples $\langle R^2 \rangle^{1/2}/L = 1.56$, implying that sites at one end of a chain will be interacting with images of sites at the other end of the same chain. That this has apparently no noticeable effect on the configurational properties can be explained by the short correlation length along alkane chains. Further comparisons of the two system sizes are made in ³⁰. It is shown that the auto-correlation function for the square end-to-end distance is also largely number-independent. Where the N -dependence does show up though is in the diffusional behaviour of the chains. The $N=10$ samples show a clear depression of the diffusivity with respect to the $N=640$ sample.

The evidence shown in Fig. 2.4.4, that when the $n=100$ chains can be fully relaxed the results agree well with the theory, is backed up the underlying distribution of square radii of gyration shown in Fig. 2.4.5.

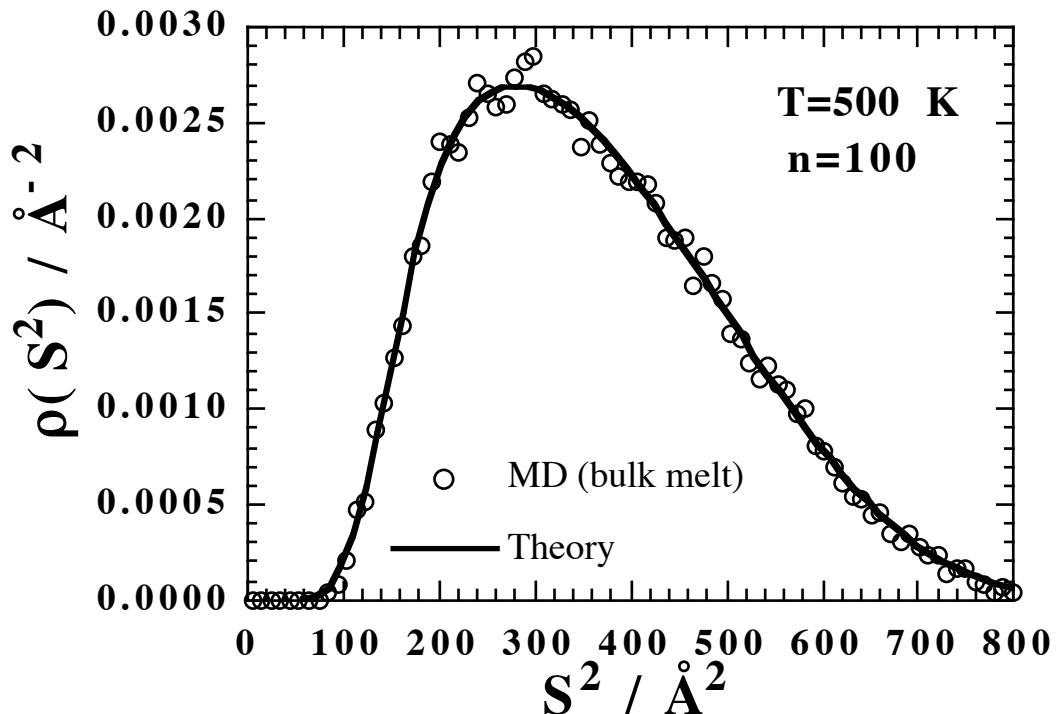


Figure 2.4.5 The probability density for the mean square radius of gyration obtained from the period $5 \rightarrow 7.7$ ns of the MD simulation of $N=640$ $n=100$ -site chains. The theoretical distribution (thick line) is also shown.

Thus in conclusion it can be said that for the n-alkane-like chains Flory's notion leads to predictions of the bulk melt configurations which are in exceptionally good agreement to that actually found. There is, thus, a sound basis in this particular case for growth procedures based on the sampling of isolated chains subject to near-neighbour interactions only. The question that naturally follows is can this be extended to other models of polymers, in particular those which are more "realistic" containing all atoms and in particular have less uniform non-bonded interactions, *i.e.* Coulomb charge-charge interactions. In the following sections is described attempts to answer this question.

2.4.3 Polyethylene oxide (PEO)

In a study designed to test the extent to which the chain dimensions of polymers could be predicted the configurations of a series of diethyl ether chain molecules with the general formula $C_2H_5-O-(CH_2-CH_2-O)_m-C_2H_5$, *i.e.* homologues of poly(ethylene oxide), PEO, were determined from NpT molecular dynamics simulations of the pure melts at 400 K³¹. As in the previous section, the results were compared to those predicted by a pivot Monte Carlo sampling method based on the assumption that chain configurations in the melt are largely determined by highly localized intramolecular near-neighbour interactions. In this case, though, the model polymer contained explicit hydrogens and, in addition, each atom

carried a partial charge. A comparison of the theory and bulk melt results is shown in Fig. 2.4.6 using in this case the characteristic ratio C_{n_b} , defined as

$$C_{n_b} = \frac{\langle R^2 \rangle}{(n_b - 1) \langle l^2 \rangle}$$

where $\langle R^2 \rangle$ is the mean square end-to-end distance, $(n_b - 1)$ is the number of backbone bonds and $\langle l^2 \rangle$ the mean of the square of their lengths. C_{n_b} is plotted in Fig. 2.4.6 as a function of chain length for PMC-sampled and MD-relaxed chains.

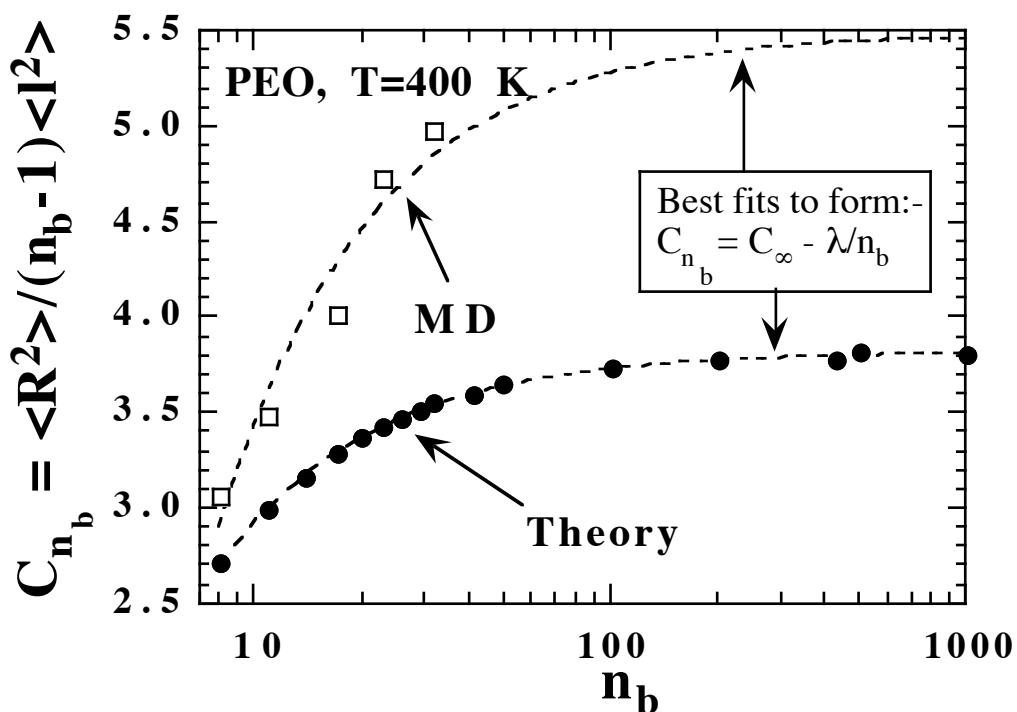


Figure 2.4.6 Variation at 400 K of the characteristic ratio, C_{n_b} , with the number of backbone atoms in a chain, n_b , for the PMC-sampled (filled circles) and MD-relaxed (open squares) chains. The dashed lines are non-linear least squares fits to the form shown.

The discrepancy in this case is clearly quite large. From further simulations carried out on the same chains in the *absence* of partial charges it was evident that the difference was due to the Coulombic interactions³¹; the uncharged chains having dimensions in as good agreement with the theory as in the n-alkane-like case. Various attempts to modify the nature of the local Coulombic interactions, *e.g.* by adjusting the range of the "local" interactions or screening the Coulomb potential, did not lead to any better agreement³¹.

The implication of this study is that packing effects can have a significant influence on the configurations of chains in the bulk melt when charges are present. Sampling under conditions where intermolecular and non-local intramolecular interactions are ignored is thus not a completely valid approximation for these types of chains. Clearly, RIS studies which obtain good agreement with

experimental evidence are taking implicitly into account this effect in their parametrisation procedure. In principle, it would also be possible to use a completely different set of potential parameters in the PMC-sampling procedure than the ones used in the bulk melt simulations in order to ensure much closer agreement between the PMC-generated and the known MD-relaxed structures. Whereas this would undoubtedly be very useful in the preparation of relaxed starting configurations of amorphous polymers, an improved growth procedure based on a different model than the MD simulations is less satisfactory and would have to be considered on a case-by-case basis.

2.4.4 Polyvinyl chloride (PVC).....and PEO revisited

To see if the case of PEO was the general rule or just an exception further investigations were carried out on a second polar polymer, PVC³². The basic formula of the PVC-homologue model used was Me-(CHCl-CH₂)_m-CHCl-Me, with Me representing a united-atom methyl group, otherwise all other atoms were treated explicitly. This definition allows both ends of each molecule to be considered equivalent. PVC is known to occur in either isotactic, syndiotactic or atactic forms. While the latter is by far the most common in practice it is much simpler to use a stereoregular structure to compare single-chain and bulk melt models in the best conditions. In the end the main PVC study was subsequently limited to the isotactic case in order to reduce the required number of simulations to manageable proportions.

As before, theoretical melt chain statistics were generated within the local energy approximation using the pivot MC approach. These were then compared to those found in bulk melt samples of chain sizes with the number of backbone atoms $n_b = 5, 7, 15, 23$ and 31 containing approximately 1000 atoms at 600 K. In this case it was found that good agreement could be achieved but only by extending the range of the local interactions in the pivot MC sampling to include interactions between backbone groups separated by $n_s=5$ backbone bonds, *i.e.* so-called "1…6" interactions. Figure 2.4.7 shows the characteristic ratio as a function of the number of backbone bonds for the bulk melt chains and the theoretical ones generated using both $n_s=4$ and $n_s=5$.

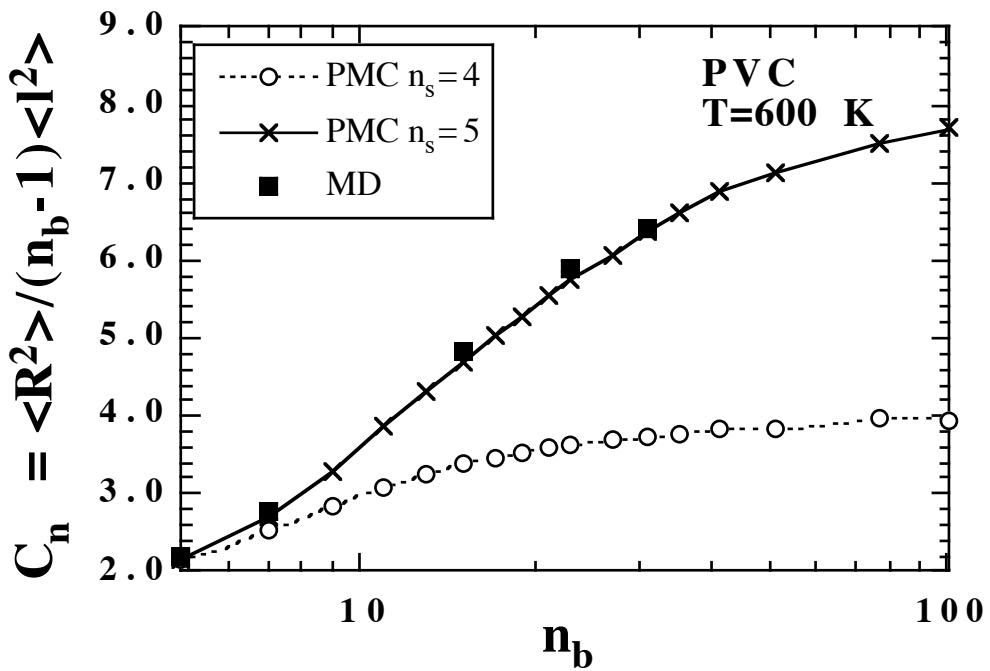


Figure 2.4.7 Variation at 600 K of the PVC characteristic ratio, C_{nb} , with the number of backbone atoms in a chain, n_b , for PMC-sampled configurations with $n_s=4$ (open circles) and $n_s=5$ (crosses), compared with those relaxed by MD (filled squares).

The reason why 1…6 interactions had to be included in PVC was demonstrated in ³². With $n_s=4$ only 1…5 backbone interactions are considered and overlaps between 1…6 groups were found to be brought about by sequences of consecutive *gauche* states of the same sign in these isotactic chains. By including 1…6 interactions in the definition of the local energy these overlaps were removed and, as can be seen from Fig. 2.4.7, the theoretical chain statistics were then in very good agreement with the MD results. Thus it could be concluded that, so long as attention is paid to the definition of the local energy, good agreement can be achieved even in polar polymers.

Further tests made on PVC models without charges and also with more PEO-like charge distributions also showed good agreement with theory and the bulk melt simulations ³². Attention thus switched back to PEO to try and explain the reason for the failure in this one particular case of the theory to predict the melt configurations. Various further trials were made using different charges on the oxygen and hydrogen atoms ³². The evidence gathered pointed to the specific charge distribution in PEO which leads to the "gauche effect", *i.e.* the preference for torsions around C-C bonds to be predominantly in the *gauche* state, being responsible. In all cases where this was removed, by changing the charge distribution, then the melt chain configurations could again be well predicted.

Within the model PEO chains the '1…5' C-H…O intramolecular electrostatic interactions play a dominant role in the *gauche* effect. In order to clarify the obvious differences between models with and

without a *gauche* effect, probability density distributions functions were calculated for the key ‘1…5’ $C_{(i)}\cdots O_{(i+4)}$ distances. For the standard model of PEO these functions are given in Fig. 2.4.8.

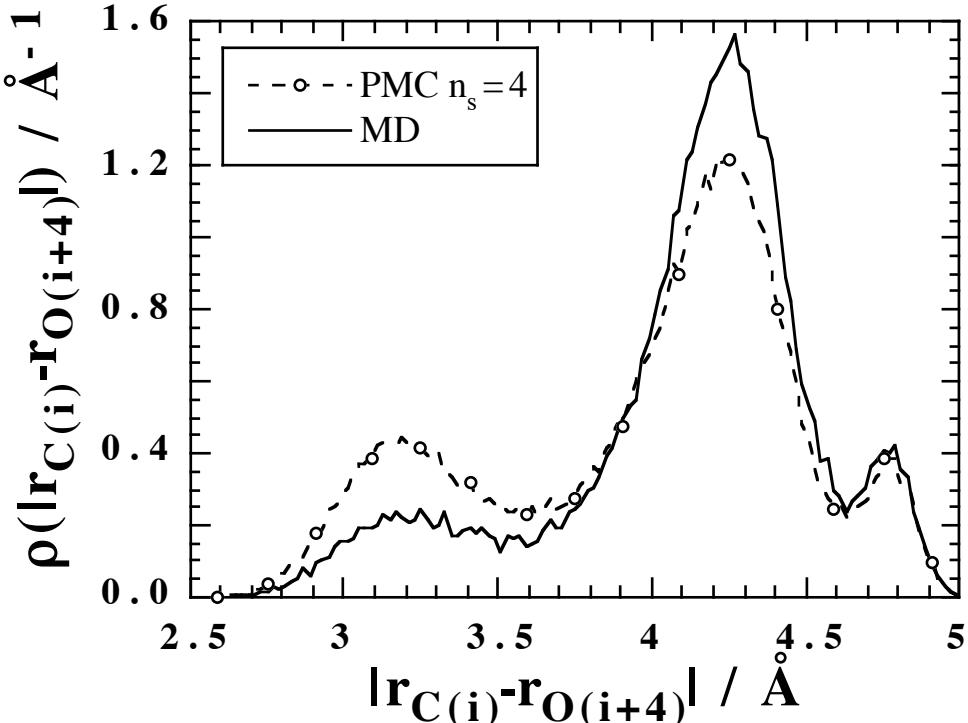


Figure 2.4.8 Probability densities for intramolecular ‘1…5’ C…O distances for the standard model of PEO with $n_b=17$ at 400 K. The PMC-sampled theoretical configurations with $n_s=4$ (broken line with open circles) are compared to those relaxed by MD (solid line).

Both the PMC-sampled and MD-relaxed configurations exhibit peaks at the same location. However, the intensity of these peaks clearly differs in Fig. 2.4.8 for the standard model of PEO, whereas in all non-*gauche*-effect cases (plots not shown) they do. The first peak at 2.6–3.6 Å corresponds to the closest ‘1…5’ $C_{(i)}\cdots O_{(i+4)}$ distances and consequently directly to the *gauche* effect. It is much stronger for isolated chains than for molecules in the bulk melt, and suggests in the latter a competition between intramolecular and intermolecular interactions. This hypothesis was supported by the evidence presented in Fig. 2.4.9, where the intramolecular $g_{\text{intra}}(\text{C}\cdots\text{O})$ and the intermolecular $g_{\text{inter}}(\text{C}\cdots\text{O})$ contributions to the C…O radial distribution function are shown in the same range of distances than Fig. 2.4.8.

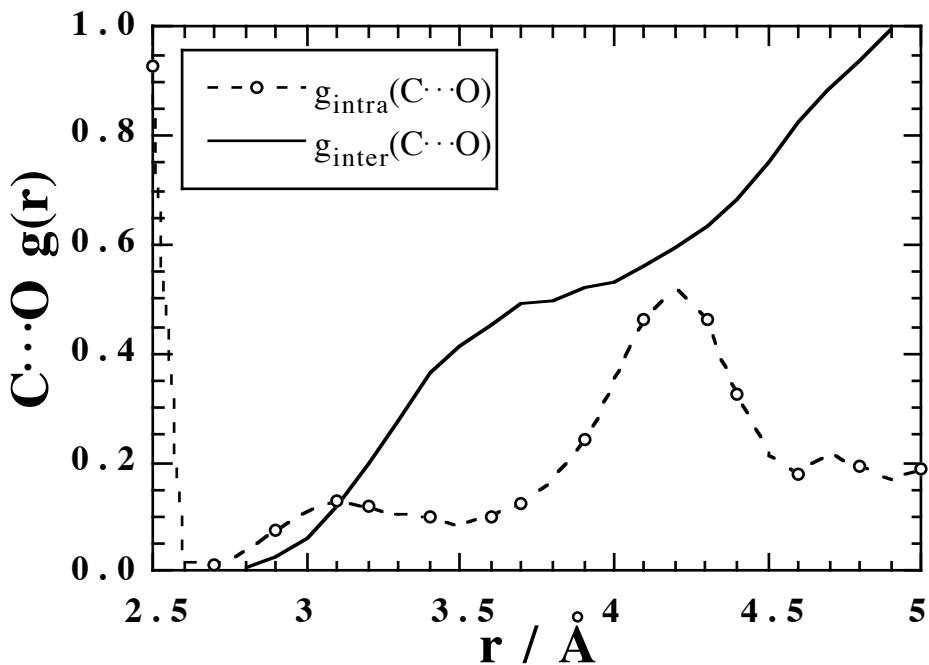


Figure 2.4.9 Radial distribution functions for the C···O interactions in the MD-relaxed standard PEO model with $n_b=17$ at 400 K. Both intramolecular $g_{\text{intra}}(O\cdots C)$ (broken line with open circles) and intermolecular $g_{\text{inter}}(O\cdots C)$ (solid line) contributions to the total radial distribution function are shown.

A similar peak at 2.6–3.6 Å is still visible in the intramolecular structure, but $g_{\text{inter}}(C\cdots O)$ becomes increasingly predominant from approximately 3.1 Å. An oxygen taking part in a *gauche* effect may obviously interact instead with a CH₂ group belonging to a different molecule, thus allowing some of C-C-O-C torsions to return to their intrinsically favoured *trans* conformations. Indeed, the weaker MD peak at 2.6–3.6 Å is followed in Fig. 2.4.9 by a stronger peak at 3.8–4.5 Å, which indicates a lengthening of some ‘1···5’ intramolecular C_(i)···O_(i+4) distances. Taking into account only local intramolecular interactions to describe the polar *gauche* effect in these molecules clearly cannot account for the competition it encounters in the melt. However, the local energy approximation is valid for PEO model melts as soon as this highly specific polar feature is removed.

2.4.5 Conclusions

Given the simplicity of the theoretical approach its success in predicting the configurations of chains in the corresponding bulk melt is quite impressive. Only when specific strong intramolecular interactions exist does the method fail and these are probably the exception rather than the rule. Without any hard and fast rules regarding the definition of what interactions to use in specifying the “local” energy it is always better to compare results with those obtained in the bulk melt. These of course have to be performed on chains which are capable of being relaxed within a reasonable amount of time, *i.e.* relatively short chains.

However, one thing which this work emphasizes, and something often forgotten with the concentration on long chains, is that it is in fact the shorter chains which show the most change in characteristic ratio and are thus their behaviour is more important in determining the limiting value.

2.5 Polymer electrolytes

2.5.1 Introduction

Polymer electrolytes are ionically conducting systems formed by dissolving salts in a polymer matrix and are technologically interesting particularly with regard to battery production. Poly(ethylene oxide), PEO, with the monomer unit -(CH₂-CH₂-O)-, is especially suited as a polymer host in these systems because of its availability, low toxicity, high chemical stability, and its ability to solvate a wide variety of inorganic salts ranging from alkali, alkaline earth, lanthanide to transition metals ³³⁻³⁶. The distance between electron-donor oxygens on the PEO backbone is optimal to efficiently multi-coordinate with the cations and compensate for the lattice energy of the salts ³⁵.

In addition to the good salt complexation capabilities of PEO, the lability of the cation-oxygen bond and the low barriers to rotation around the backbone bonds make for potentially significant ionic transport ³⁶. The situation is, however, complicated by the tendency of PEO to crystallize in helical structures ³⁷; at room temperature, as much as ~70-85% of pure PEO will be crystalline ³⁸. It is now well established that diffusion occurs in the amorphous regions ³⁹, and thus crystallinity is usually seen as a hindrance to efficient ionic conductivity. Much of the experimental work in polymer electrolytes has been aimed at 'amorphizing' PEO-based or related systems, *e.g.* by including plasticizers ⁴⁰, but then new problems, such as poor mechanical strength, have then appeared. To date, PEO-based systems are still the most widely studied polymer electrolytes.

Despite the extensive experimental work carried out on these systems the mechanisms of ionic transport remain unclear in many respects as do certain aspects of the structure like the formation of ionic pairs. It is now well accepted that cation transport in high molecular-weight polymers above the glass transition temperature, T_g, is associated with the dynamical relaxation modes of the polymer chain. Anions are not as strongly solvated by the polymer host as the cations, and tend to move into available holes in the structure.

In this context, computer simulation methods appear promising as a complement to experimental approaches. A fully atomistic model was thus developed for molecular dynamics simulations of both pure PEO and PEO/NaI. The approach taken was a systematic step-by-step one in which parameters were first tested and fine-tuned using the crystalline structures of PEO ^{41,42} and PEO₃NaI ⁴³ before the model was used to study the charge conducting amorphous phases of PEO_xNaI ⁴⁴⁻⁴⁶.

2.5.2 Potential Optimization in Crystalline PEO and PEO₃NaI

Since there is still little structural information available on the amorphous phases of PEO-based systems, successful simulations of the known structures of PEO and PEO₃NaI crystals were considered as an essential first step towards the development of a realistic potential for PEO/NaI systems. Both crystals contain close-packed helical coils of PEO which were modelled in the simulation as effectively infinite chains by bonding one end of the chain to the other end of the nearest periodic image. A number of

parameter trials were then made, firstly on pure PEO and then afterwards on PEO_3NaI . Schematic representations of the two structures are given in Figs. 2.5.1 and 2.5.2.

In determining the degree to which a certain parameter set was reproducing the properties of the real systems emphasis was placed on the extent to which the models preserved the X-ray structures and the behaviour of the pressure tensor. If a system is in mechanical equilibrium then all three on-diagonal components of the pressure should be the same and equal to the ambient hydrostatic pressure and all off-diagonal components should be zero. In the model systems being simulated an immediate problem arose in that conventional methods of measuring the pressure tensor could not cope with systems which used "infinite" chains, rigid constraints and Ewald summations. For this reason a new hybrid scheme of calculating the pressure tensor had to be first developed ⁴². Essentially the mechanical definition of the pressure tensor was mostly used, but the thermodynamic definition was also resorted to for those forces (*i.e.* the Fourier space part of the Ewald sum) where the former is not applicable.

A large number of parameter combinations were tried for crystalline PEO. Short 20 ps NVT simulations were carried out at 300 K with an MD box containing $2 \times 2 \times 2$ monoclinic PEO crystallographic unit cells, *i.e.* 1568 atoms in 16 chains. The sensitivity of the pressure tensor to small details of the potential provided an effective criterion for assessing possible parameter sets and this was illustrated in ⁴². With the optimized parameter set a longer simulation of the crystal was carried out ⁴¹.

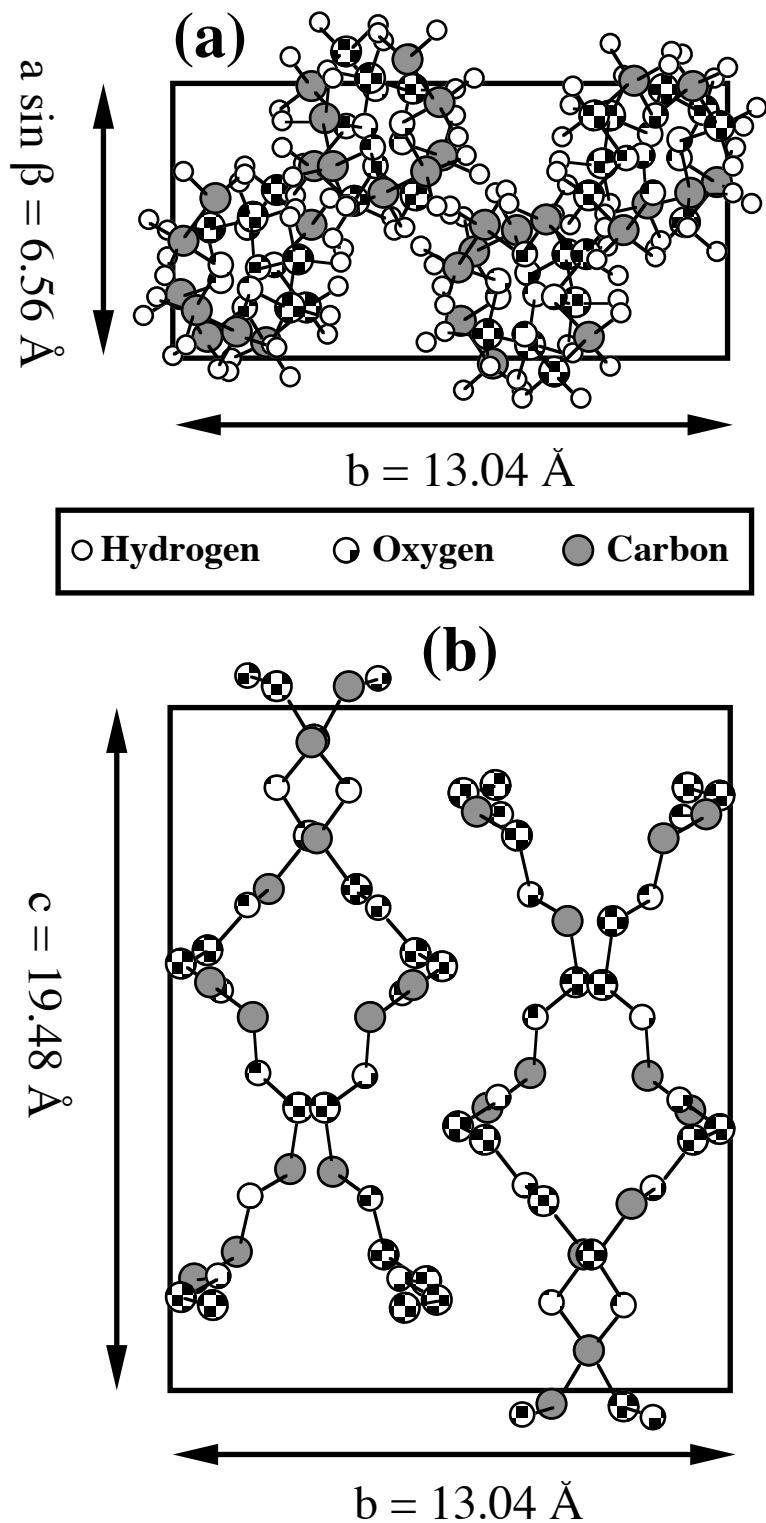


Figure 2.5.1 A schematic representation of the published crystal structure of PEO³⁷, viewed (a) along and (b) perpendicular to the chain axis. For clarity only backbone atoms are shown in (b).

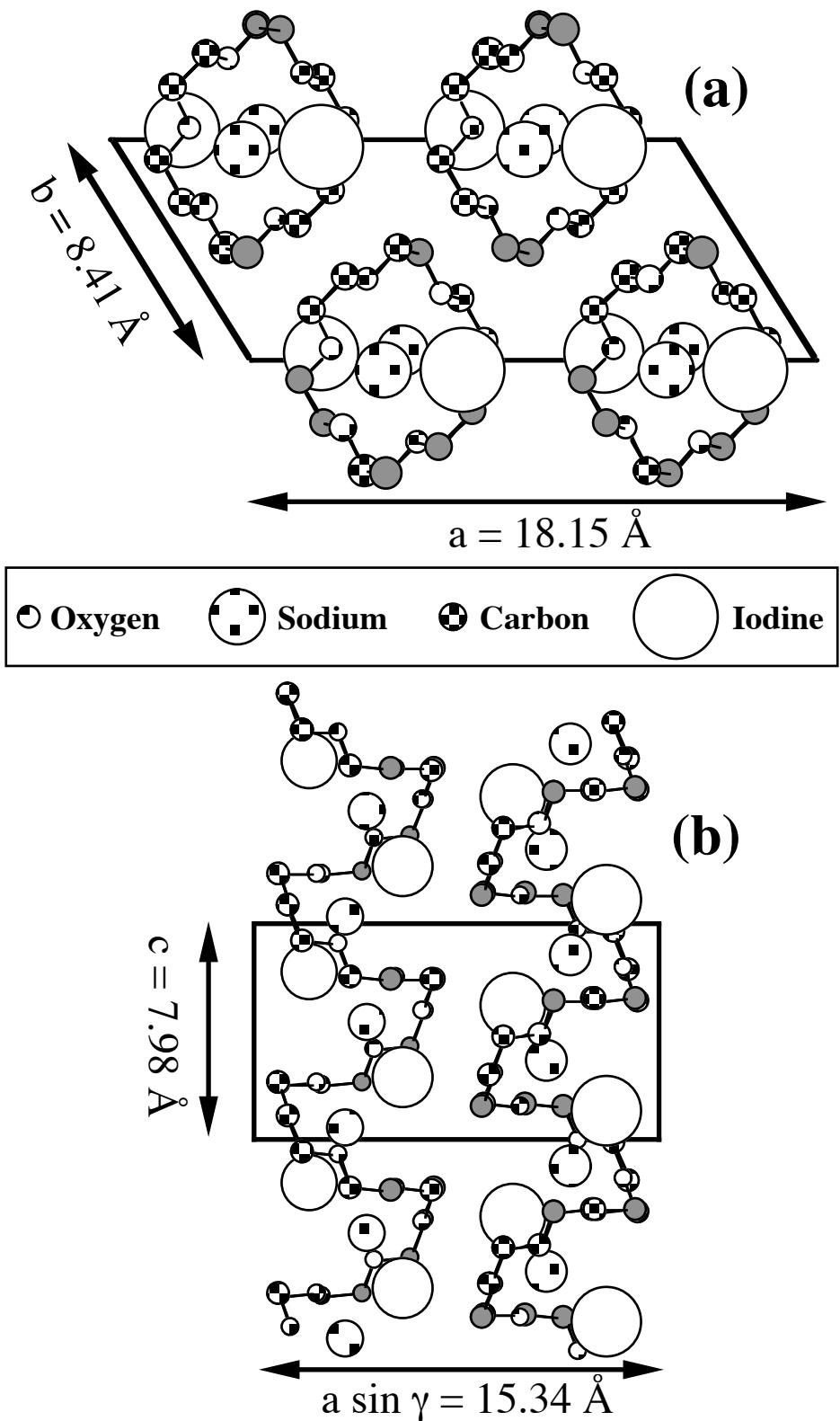


Figure 2.5.2 A schematic representation of the published crystal structure of PEO₃NaI⁴⁷, viewed (a) down the *c* axis (fibre axis) and (b) along the *b* axis. Hydrogen atoms are omitted.

For PEO₃NaI all the parameters for the polymer-polymer interactions were directly transferred from the crystalline PEO work. However, the ion-ion and ion-polymer interactions had to be added to this set so a

new series of short 20 ps NVT simulations was carried out at 300 K with an MD box containing $1 \times 3 \times 3$ monoclinic PEO₃NaI crystallographic unit cells, *i.e.* 828 atoms in total and 6 polymer chains. Longer 100 ps production runs were then carried out with the potential selected from the trials⁴³. For these PEO₃NaI simulations, though, an empirical correlation was found between P_{xy} and the γ angle of the cell. It was found that P_{xy} could be minimized by increasing γ from 122.3° to 123.8°, whilst scaling the b parameter from 8.41 Å to 8.56 Å to maintain a constant density. PEO₃NaI production runs were thus undertaken for two MD cell shapes; the first used the published crystallographic cell, and the second used the cell modified in the manner described above.

The magnitudes of the displacement vector between the experimental crystallographic position and the MD-averaged position are displayed in Figs. 2.5.3a and 2.5.3b for PEO and PEO₃NaI, respectively.

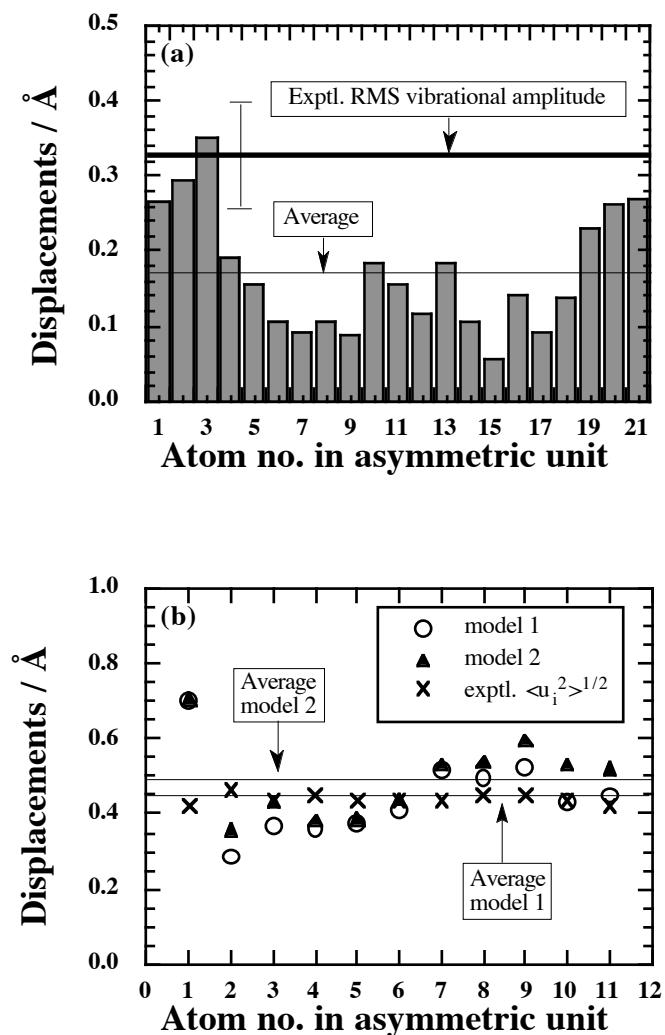


Figure 2.5.3 Magnitudes of the mean vector displacements between the experimental crystallographic and the MD-averaged positions for all atoms in the asymmetric unit of (a) PEO and (b) PEO_3NaI . Atoms are ordered according to their position in the respective asymmetric units, i.e. in (a) the seven ($\text{O}-\text{C}-\text{C}$) monomers and in (b) iodine as no. 1, sodium as no. 2, and then the three ($\text{O}-\text{C}-\text{C}$) monomers. In (a), the thick horizontal line indicates the expected rms displacement of atoms based on the experimental isotropic temperature factor, and the thin line shows the average displacement for all atoms in the PEO asymmetric unit over the MD simulation. In (b), data are shown for both PEO_3NaI models studied, and are compared to the experimental rms displacements, $\langle u_i^2 \rangle^{1/2}$, for each atom. Thin lines indicate the average displacement for all atoms in the PEO_3NaI asymmetric unit over the MD simulations.

All atoms clearly lie very close to those positions expected from experiment and the standard deviations, although not shown, in the magnitudes of the displacement vectors are all less than 0.55 Å. This excellent

agreement was confirmed for pure PEO by calculating an X-ray ‘powder’ diffraction profile from MD-derived coordinates, and comparing it with an experimental profile⁴¹.

Another interesting comparison with the X-ray data was the sequence of average torsional angles along the PEO fibre, Fig. 2.5.4. Despite artificially imposing symmetry through the use of infinite chains and periodic boundary conditions, the systems naturally adopt the experimentally determined (7/2) and (6/1) helices in PEO and PEO_3NaI , respectively, with the sequence of average torsion angles repeating itself every asymmetric unit. The spread in average torsional angles from the MD simulation is much smaller than that predicted by the X-ray data for crystalline PEO³⁷. As it would require an extremely strong intermolecular force to maintain a torsional angle 20° from its minimum, the MD model might actually be a better starting structure for further position refinement than the distorted fibre X-ray diffraction-refined study.

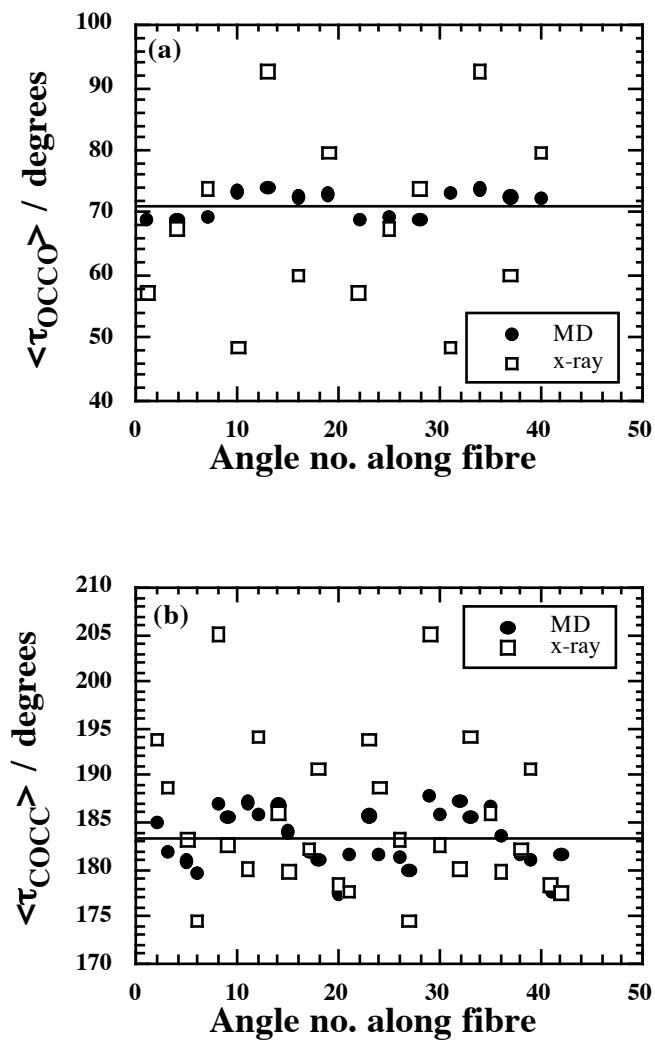


Figure 2.5.4 Average values for the OCCO (a) and COCC (b) dihedral angles from the MD simulation (filled circles) and from the X-ray determined coordinates (open squares) as a function of the dihedral position along a polymer fibre in crystalline PEO. In (a), the gauche⁺ and gauche⁻ wells have been symmetrized. The horizontal lines represent the overall average MD values for each type of dihedral angle.

With the ultimate purpose of the work in mind, the degree of agreement between the MD models and the X-ray structures was considered sufficient. In PEO there are good reasons to treat the refined X-ray data with some circumspection and so there is little to choose between them. In PEO_3NaI the model has some problems to reproduce in exactly the same way the zigzag chains of alternating cations and anions⁴³. These ionic chains rearrange themselves in a slightly "flatter" structure than that reported in the diffraction study⁴⁷. Whether this is due to the use of point charges for the ions in the model is an issue which remains to be resolved. Discrepancies are, however, small in MD terms.

2.5.3 Preparation of amorphous relaxed PEO and PEO_xNaI systems

An amorphous relaxed sample of PEO was generated by growing a single chain of 142 EO monomers (431 backbone atoms and 1009 atoms in total) at 400 K using pivot MC-sampling, and then allowing the system to relax to mechanical and conformational equilibrium using MD. The average density, $\langle\rho\rangle$, of this system after relaxation is plotted with those of the smaller chain bulk melts discussed in Section 2.4.3., as a function of the inverse polymer chain length in Fig. 2.5.5. The extrapolation of the mean densities is clearly very consistent with experiment and provides good evidence of the transferability of the potential, developed for crystalline PEO, to amorphous PEO.

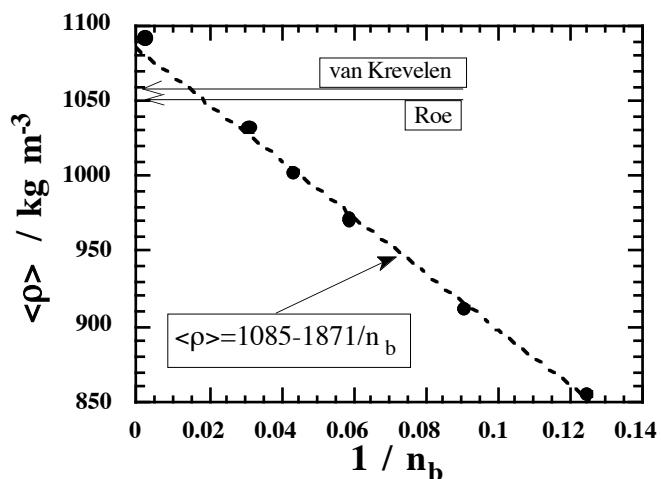


Figure 2.5.5 Relaxed average densities, $\langle\rho\rangle$, of PEO bulk melt systems at 400 K, plotted as a function of the inverse chain length of the polymer, $1/n_b$ (filled circles). Error bars are contained within the symbols. An estimate of the density of molten PEO obtained from the work of van Krevelen⁴⁸ and the experimental value of Roe⁴⁹ are indicated by the arrows.

Despite the misgivings detailed in the Section 2.4 about the pivot MC growth procedure when applied to PEO, the lack of any better alternative methods meant that there was little choice but to persevere. The fact that the density of the amorphous model agreed reasonably with experimental data and extrapolated linearly from the equilibrated shorter chain length systems was certainly reassuring. In any case, as the main interest lay in the local structure and dynamics of the ions within the polymer matrix, details of the large

scale structure of the polymer chains were probably less important although it was probably worth bearing in mind.

In the final stage of preparation amorphous PEO_xNaI systems, $x = 48, 20$ and 3 , were generated by distributing randomly the appropriate numbers of Na^+ and I^- ions in the periodic MD box of the relaxed PEO-sample at 400 K ⁴⁴. MD simulations were then carried out for each of the three salt concentrations both at 400 K and at 500 K . A comparison of the specific volumes and the energies for pure and mixed systems supported the visual evidence that all PEO_xNaI systems were single-phase under the conditions specified.

2.5.4 Polymer mobility

The conformation of the polymer chain upon introduction of ions remained on the whole relatively similar to that of pure PEO. However, salt concentration and temperature both influenced its mobility, as can be discerned from the normalized autocorrelation functions for the *gauche* state for the polymer backbone OCCO angles shown in Fig. 2.5.6. The CCOC angles in the *trans* state show the same tendency towards lower polymer chain flexibility with increased ionic concentration, *i.e.* an anti-plasticiser effect, and decreased temperature. Relaxation times exhibited by our models are in good agreement with the value of $\sim 10^{-10} \text{ s}$ reported for the segmental motions of the polymer chain³⁵.

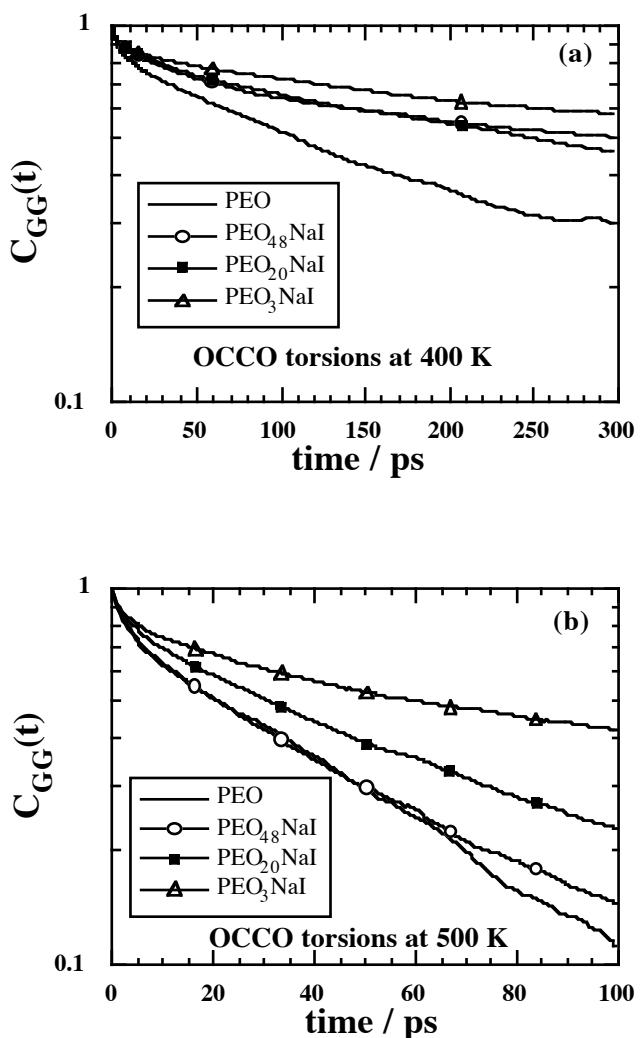


Figure 2.5.6 The normalized autocorrelation functions calculated for the polymer backbone OCCO angles in the gauche state at (a) 400 K and (b) 500 K. The results for different PEO_xNaI systems are compared to those for pure PEO.

2.5.5 Local structure around cations in PEO_xNaI

The structure around the sodium cations within the PEO matrix has been analysed in great detail^{44,45}. With regard to ion conduction, the presence or absence of ion clusters is important and so these were monitored⁴⁴. At the highest salt concentrations all the ions were effectively in one "cluster", or percolating network rather, so the picture that emerges is one of an extended salt network intertwined with the polymer chain. At lower concentrations clusters certainly exist but the lack of statistics prevents any definitive statements to be made about their distribution of sizes.

At low concentrations the cations are preferentially coordinated by continuous oxygens on the same strand of polymer. It is common for as many as 6 continuous oxygens to pack around a single cation. Such polydentate loops are similar to the coordination of cations by crown ethers. Anions compete for coordination with the oxygens and the probabilities of so-called 'NaO_mI_n complexes' has been given in⁴⁵.

Many combinations of (m,n) values are seen although some, particular the NaO_3I_2 complex, appear more stable than others. At the lowest salt concentrations, iodines tend to be replaced by two oxygens to give NaO_5I_1 and NaO_7I_0 complexes.

The geometry of the NaO_mI_n complexes have also been studied in detail⁴⁵. The angles subtended by successive oxygens on the same PEO strand forming ligands with the same sodium cation are on average around 60°. The strands of successive oxygens also tend also to form into planar loops. Any coordinated iodines tend to be perpendicular to these planar loops. Examples of the coordination around a sodium ion are shown in Fig. 2.5.7.

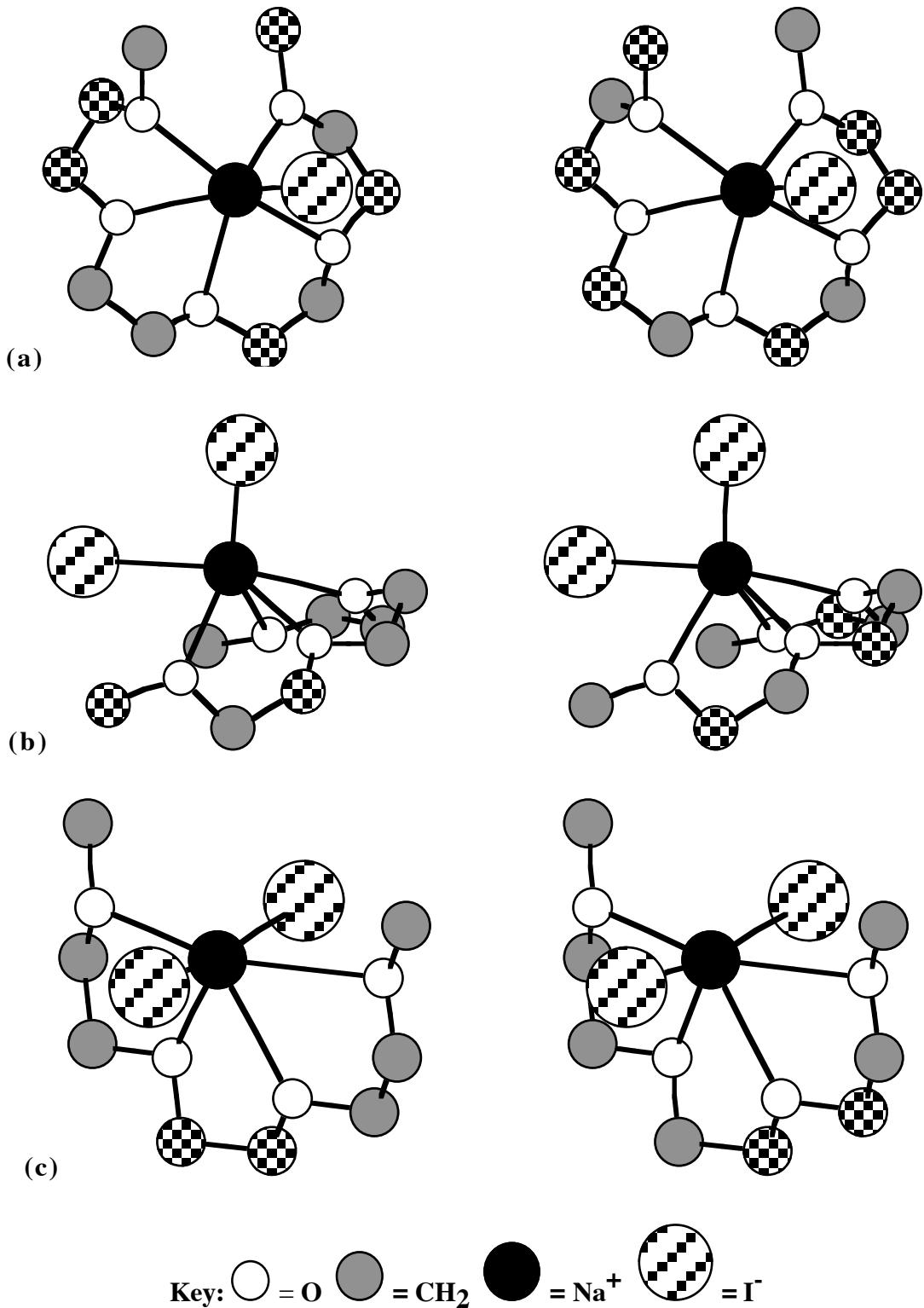


Fig. 2.5.7

Stereo views of NaO_mI_n complexes with (a) one iodine ligand in the $\text{PEO}_{20}\text{NaI}$ system at 400 K, (b) two iodines ligands in the $\text{PEO}_{20}\text{NaI}$ system at 500 K and (c) the X-ray refined PEO_3NaI structure⁴⁷. ‘Bonds’ have been drawn between the cation and its ligands. In the latter case, the radius of coordination to the cation was set to 3.8 Å to be consistent with our simulations of the bulk melt.

2.5.6 Ion mobility

From a technological point of view, one of the most important properties of solid polymer electrolytes is their ability to conduct electricity. In these materials the conductivity is related to the diffusion of ions in the polymer matrix. Experimentally the diffusion is known to steadily fall as viscosity increases, due to the stiffening effect of the added salt and the increase in ion association^{36,50}. Reported diffusion coefficients of ions in PEO are typically in the range 10^{-11} to $10^{-10} \text{ m}^2 \text{ s}^{-1}$ ^{50,51}. This order of resolution is not accessible to MD simulations carried out on a nanosecond time-scale on nanometre-sized samples with very few ions. In agreement with simulations of similar systems⁵², our own calculations for the ions in PEO_xNaI systems showed that the mean square displacements were very small on the nanosecond time-scale and amounted to average displacements of only a few Ångstroms. In addition, they were prone to erratic behaviour and differed substantially depending on which part of the simulation was analyzed. This was hardly surprising, given the expected diffusion coefficients and the small numbers of ions over which averages were taken.

The erratic behaviour of the mean squared displacements could be traced in fact to distinct diffusional steps that were occurring. On closer inspection a rich array of various jump events could be observed in the vicinity of the cation coordination spheres, including hopping between neighbouring sites on the polymer chain, replacement of an anion by oxygen atoms and replacement of oxygen ligands by an anion. A number of these are illustrated in Fig. 2.5.8.

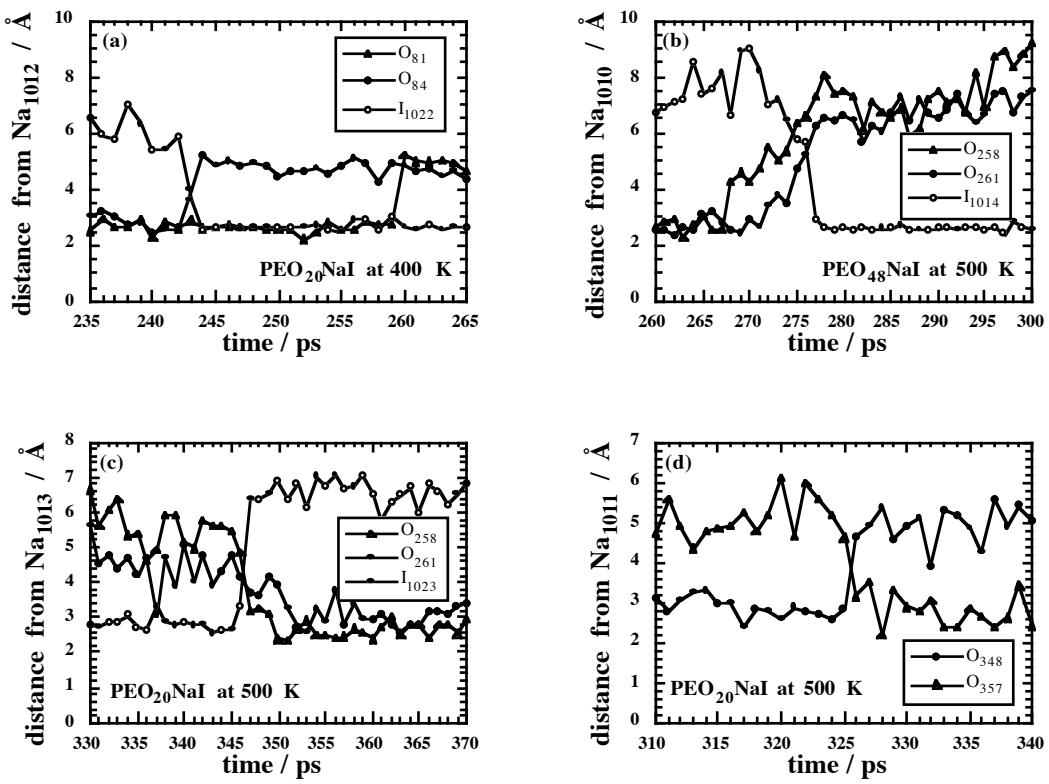
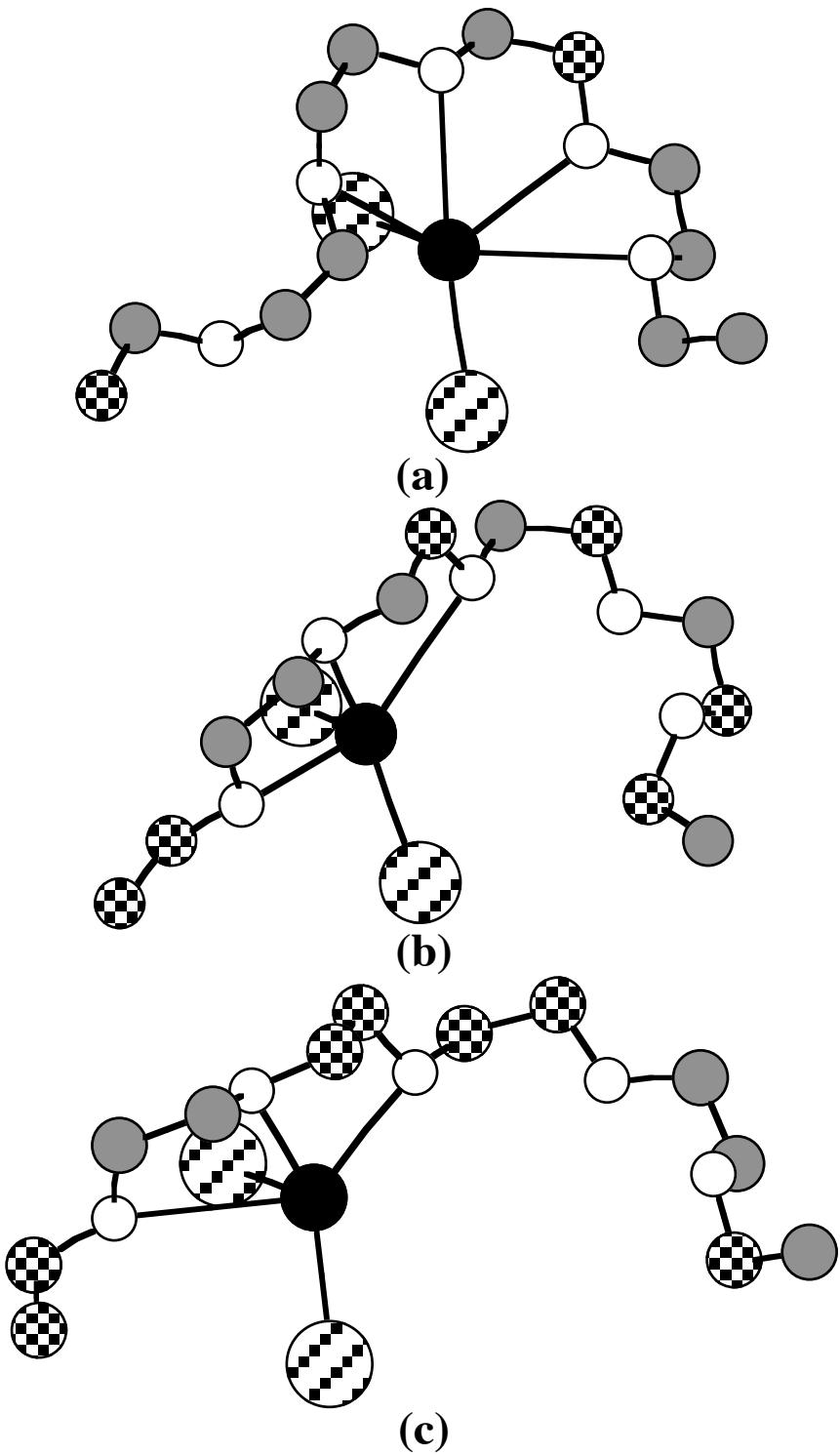


Figure 2.5.8 Examples of dynamical changes in the coordination shell around cations featuring (a) the replacement of two oxygens by an iodine in two steps, (b) the replacement of two oxygens by an iodine, (c) the replacement of an iodine by two oxygens, and (d) the replacement of an oxygen by another oxygen situated a bit further along the polymer chain. Note that in (a), (b), and (c), both oxygens involved in the events are only separated by two carbons.

A particularly interesting pattern of motion involves cations moving along the PEO chain in distinct steps. Such motions have been characterized in detail in Refns.^{44,45} and an illustration is given in Fig. 2.5.9.



Key: $\circ = \text{O}$ $\bullet = \text{CH}_2$ $\blacksquare = \text{Na}^+$ $\square = \text{I}^-$

Figure 2.5.9 Cation jump between neighbouring sites on the PEO chain. ‘Bonds’ to the cation indicate those oxygen and iodine ligands that are found within 3.8 Å of the sodium. Please note that in all snap-shots, exactly the same atoms are displayed and the viewing angles are identical. The sequence is taken from the local environment of Na_{1011} in the simulation of the $\text{PEO}_{20}\text{NaI}$ system at 500 K at (a) 325 ps, (b) 326 ps and (c) 327 ps.

2.5.7 Phase separation

In a further paper the response of the three amorphous PEO_xNaI model systems to a temperature increase from 500 K to 1000 K and a cooling back down to 500 K was investigated. Although such a computational procedure did not allow for even quasi-equilibrium to be reached at a given temperature, interesting trends were obtained by monitoring the evolution of a number of properties. Thermodynamic analyses showed that the systems underwent a net modification upon heating and that this was not fully reversible when they were returned to their initial temperature. The polymer conformations were rather insensitive to the temperature cycle but did indicate some restructuring at the highest salt concentration. Ionic association was clearly very affected for all systems under study by the rapid increase in temperature, and both Na^+ and I^- ions settled into single clusters containing all the ions in the simulations. These final phase-separated states were very stable, irreversible on the time-scale of the simulations, and in accordance with experimental findings on the so-called ‘salting-out’ effect.

2.5.8 Conclusions

There is no doubt that the results of the amorphous polymer electrolyte simulations outlined above give a good qualitative picture of the structure and short time dynamics of these types of systems. Quantitative measurements of diffusion coefficients require a substantial investment of computer time but with the parallel computers now available this is possible. It would be instructive, for instance, to see if the diffusion coefficients followed the same trend with increasing concentration as seen experimentally. Actual quantitative comparisons with experimental data would probably show up some of the fundamental weaknesses of the model; in particular the complete disregard of polarisability. In the near future it is conceivable that such effects could be incorporated at least in an approximate manner.

3. Biopolymers

3.1 Introduction

The time and length scales that can be reasonably simulated using classical atomistic simulations are still major limitations for these techniques. Indeed, many technologically and biochemically interesting systems which extend over distances greater than 10 nm and/or relax on time scales longer than a nanosecond are too time consuming to simulate. In the area of biochemical research, for example, membranes containing typically many thousands of atoms⁵³ are increasingly popular subjects of simulation studies. While tens to hundreds of picoseconds are considered adequate for the study of their structural properties⁵⁴⁻⁵⁷, transport of molecules through membranes will require far longer simulations. Diffusion through membranes of relatively small molecules such as helium and methane⁵⁸, monovalent ions⁵⁹ and benzene^{60,61} have been reported, but, significantly, long trajectories of ~1 ns were necessary in the latter case. Clearly, extending these calculations to the case of the membrane permeation of even a relatively small antibiotic molecule such as valinomycin⁶² will require even longer trajectories. The same remarks also apply to the study of membrane permeation through pores formed by proteins⁶³⁻⁶⁶, as well as to reactivity in enzymes⁶⁷.

In the previous chapter, similar problems with long relaxation times were encountered in the study of polymer melts. In that case, a domain decomposition algorithm was developed for the highly specific linear homopolymer system in question and it was clearly demonstrated that parallel computers can deliver impressive increases in performance if they are efficiently programmed. In general, though, the biochemical and technological systems are not limited to molecules with relatively straightforward connectivities so we were faced with the problem of *extending the domain decomposition method to molecules of arbitrary connectivity*. Solving this problem opens up the prospect of simulating systems of a much larger scale than was ever previously possible and thus the opportunity of advancing our understanding of many processes at a fundamental level. In this final chapter, it is first shown that it is indeed possible to generalize the domain decomposition approach to systems of arbitrary connectivity. After implementing this new algorithm, we also start tackling problems previously considered too time consuming.

3.2 Parallel Molecular Dynamics

3.2.1 Introduction

Early attempts at parallelizing MD simulations were largely limited to atomic systems and to highly specific problems, *e.g.*^{9,68-71}. During this time, a number of strategies were proposed for mapping MD onto multiple-instruction multiple-data (MIMD) machines based on either particle decomposition, *i.e.* assigning particles to home processors using their indices, or domain decomposition, *i.e.* assigning particles to processors according to their position in space^{68,69,72-76}. Although there have been more recent developments involving force decomposition^{77,78} and attempts to combine strategies⁷⁹, the problem has remained that the best parallel solution for a particular simulation depends on factors like the average number of particles per

processor and the speed of the individual processors. However, for large scale simulations where there are many more particles than processors, there is little doubt that domain decomposition (DD) is the optimum approach.

In the DD strategy, each processor element (PE) is responsible for particles residing in a particular sub-volume of the MD box and each processor accumulates the total forces on the particles in its domain and integrates their equations of motion. Although DD is relatively straightforward to implement for atomic systems in which only pair interactions exist (and there are now numerous examples^{28,71,74,80-85}), the connectivity of molecular systems introduces a number of new difficulties^{29,86}. These can be summarized as follows:-

- (1) Multiparticle potentials and specialized constraints require that the group of n atoms involved in the calculation be simultaneously known by one PE which is assigned to carry out the necessary computations.
- (2) Particles subject to rigid constraints may lie in different domains. The procedure SHAKE, generally used to impose rigid constraints, implies a communication between processors at each of several iterations.
- (3) Not all pairs of atoms interact via non-bonded pair potentials.

As a result of the above, most attempts to parallelize general purpose MD codes have relied on simpler to implement particle decomposition techniques⁸⁷⁻⁹³ which are unsuitable for large-scale simulations as each processor requires a knowledge of all atomic positions. Some improvement is possible using force decomposition techniques⁷⁷ but, as yet, there are only a few examples in the literature of DD MD programs being used for molecular systems^{29,86,94-96}.

Chynoweth *et al*⁹⁴ have described a DD method in which space is divided in just two dimensions (2-D) and each PE is responsible for a column of the system, but they give few details of how they handled the connectivity problem. Esselink and Hilbers⁸⁶ have also used a 2-D partitioning of space with a method which guarantees that all n atoms of a multi-particle system will be found by just one PE. In partitioning space in only two dimensions, both these latter methods incur some communication penalties for not using the optimum 3-D decomposition²⁸. Furthermore, they also perform more communications than is required for a search of all interacting pairs.

Clark *et al*⁹⁵ have recently presented a program called EULERGROMOS, which is a parallel version of the general purpose MD program GROMOS. They use domain decomposition to calculate the non-bonded forces but incur some redundancy by evaluating the same pair forces in different domains. Their method has the advantages that it is not limited to domains of size greater than the potential truncation distance and that it can also be operated in a self load-balancing mode. However, to evaluate the bonded forces, each processor has to scan the *global* connectivity table to extract those terms involving atoms in its own domain, a process which probably also involves some redundant calculations and which they admit could limit the scalability of

the algorithm. This may be a marginal issue in the case where CPU costs are dominated by the calculation of the non-bonded forces but it is likely to be fairly crucial in the cases where they do not.

Plimpton *et al*⁹⁶ have also recently described a fairly comprehensive general purpose parallel MD program, called LAMMPS, which incorporates multiple time step methods and efficient Coulombic sums. For the bonded forces, however, they either scan a global connectivity table or use a 'hash table', but few details are given of the latter approach.

We have developed a method which allows the optimum 3-D decomposition of space to be combined with an efficient minimum communication scheme and non-redundant search for all interacting pairs²⁸. We have then shown how this can be retained in a DD code capable of handling model linear homo-polymers²⁹. This algorithm for molecular DD incorporated a parallel SHAKE algorithm to constrain bond lengths, and a crude, but effective, method to handle the three- and four-body forces which were used to restrict the flexibility of the linear molecules. The feasibility of being able to use a parallel SHAKE algorithm was seen as an important part of this work. The generally quoted reason for choosing rigidly constrained bonds as opposed to flexible ones in molecular dynamics simulations is that of a means of saving CPU time; the relatively high frequency of most realistic chemical bond vibrations implies that short time steps have to be used for stable integration of the equations of motion. However, a second and much more fundamental, although less well appreciated, problem with the dynamics of classical systems where there are modes of motion with widely separated frequencies present, is that of equipartition of kinetic energy. It is possible for the coupling between modes to be so poor as to make equilibration times extremely long⁹⁷. For this latter reason alone it is important to retain the constraints routine in the DD framework. It is worth pointing out that although multiple time step methods, *e.g.*⁹⁸, can overcome the problem of short time steps, they cannot address the problem of poor coupling. Our method was then extended to molecular systems of arbitrary connectivity⁹⁹.

In the following section, some general remarks are made about our implementation of domain decomposition. We then extend this approach to molecular systems of arbitrary connectivity. Many of the features used in homo-polymers²⁹ are retained, but a totally new method is implemented for searching for all n atoms which constitute a group (henceforward referred to as an 'n-plet') required for a multi-particle potential or a rigid constraint. Further details can be found in a recent article⁹⁹ and in the on-line manual for the *gmq* and *ddgmq* programs¹⁰⁰.

3.2.2 Domain decomposition: General considerations

In the DD scheme used, the atoms are initially assigned to the $N_{px} \times N_{py} \times N_{pz} = N_p$ processors on the basis of the x, y and z coordinates, *i.e.* 3-D decomposition of space. The exact shape of the MD box does not have to be orthorhombic and assignments are actually carried out using the set of scaled coordinates $s_i = \mathbf{h}^{-1} \mathbf{r}_i$ where the matrix $\mathbf{h} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is defined from the basis vectors of the MD box. The scaled space is cubic and facilitates easy assignment of particles to processors. The scaled space also facilitates the assignment of atoms to the $N_{cx} \times N_{cy} \times N_{cz} = N_c$ link-cells within each domain. The link-cell strategy¹⁰¹ is useful both as a way to speed up the creation of a neighbour table¹⁰¹ and as a means of determining which atoms lie close to domain boundaries and thus have to be communicated to neighbouring domains.

At this point, each PE is only aware of particles in its own domain. As atoms are assigned to processors, it is inevitable that some molecules will be distributed among two or more of them. Decompositions based on molecular centres-of-mass can avoid this problem¹⁰² but they are not an option in the general case where molecules span many domains.

Once the assignments to domains and link-cells have been made, the first step is a search for all interacting pairs of atoms. This requires the communication of those coordinates required by other domains to complete their "working" environment. In order to minimize communications at this stage, the conventional link-cell search pattern is modified²⁸. A schematic representation of the form of the modification in 2-D is shown in Fig. 3.2.1; the extension to 3-D follows straightforwardly.

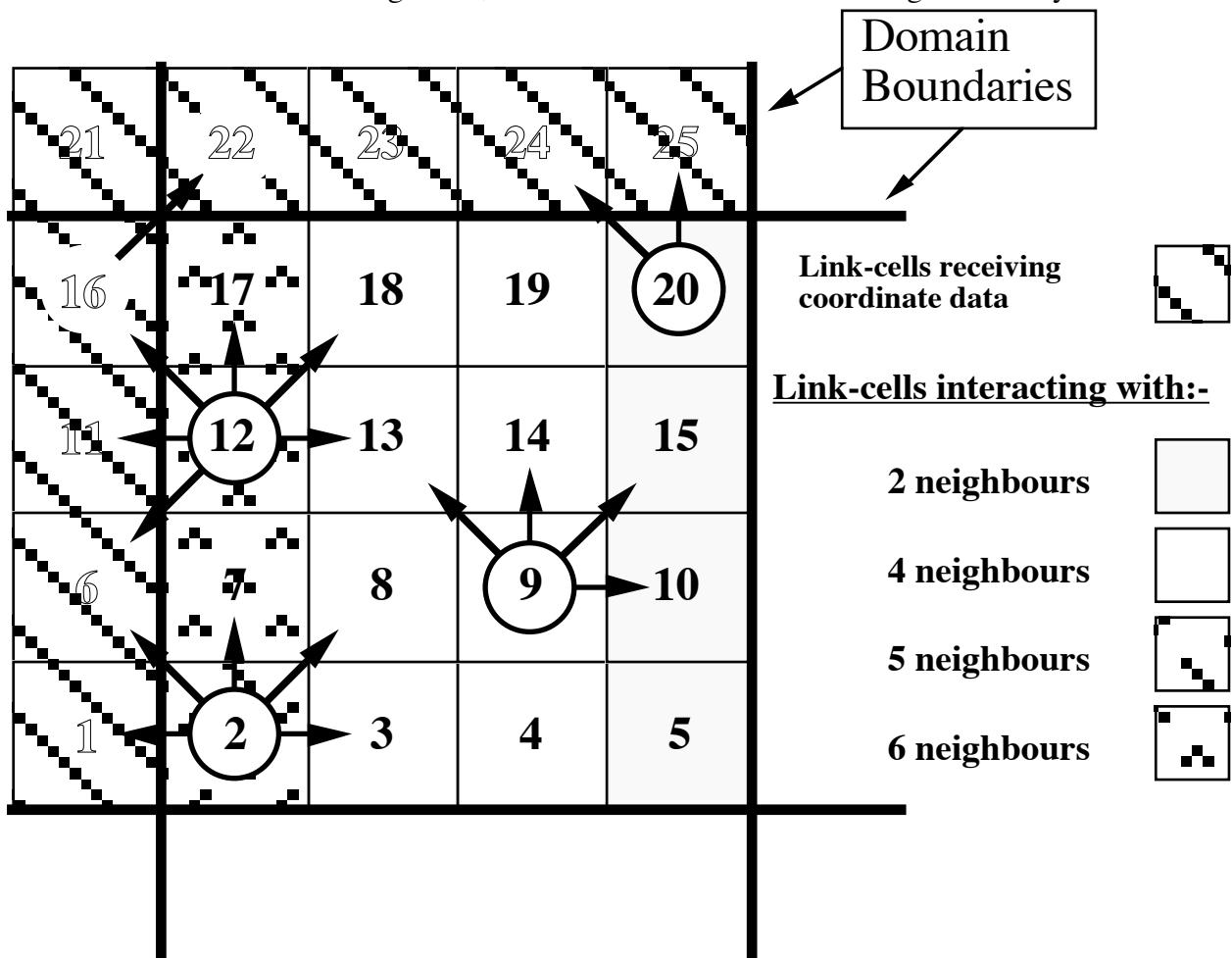


Figure 3.2.1 A schematic 2-D representation of the link-cell neighbour search pattern used in the domain decomposition program. The thick black lines mark the boundaries between regions which different processors are responsible for. In this example all the 16 link-cells that are resident on the central processor (numbered in **bold** type) are searched using the pattern shown. In addition all the particles in the received cell **16** are interacted with those in **22**.

A link-cell search for interacting pairs normally treats each link-cell in the same way. Particles within a link-cell are interacted amongst themselves and then with the contents of half the 26 (8 in 2-D) nearest neighbour

link-cells¹⁰¹. If this search method was to be applied in a domain decomposition code, it would require direct or indirect communications with 17 (5 in 2-D) of the nearest neighbour processors. By modifying the search pattern of link-cells according to their position within a domain, direct and indirect communications are only required with 7 (3 in 2-D) of the nearest neighbour domains and can be accomplished efficiently with just three (two in 2-D) communication steps. The flow of information during these three communication steps²⁹ is demonstrated in Fig. 3.2.2.

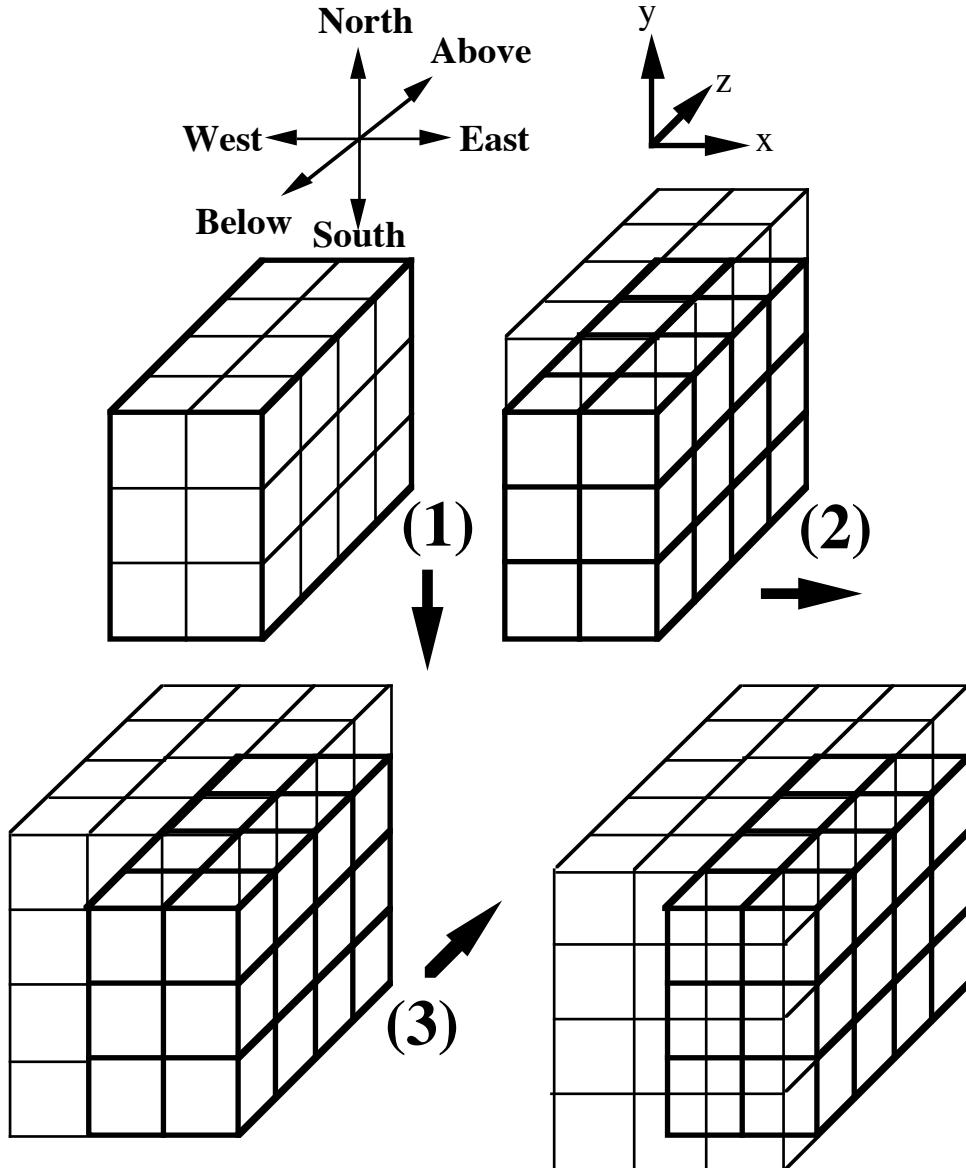


Figure 3.2.2 The flow of information in the communication stage prior to the all-pair search. The figure shows an initial set of $2 \times 3 \times 4$ link-cells in a domain (top left). At the first communication step all PEs pass to their southern neighbour all the coordinates of sites in the link-cells on their southern periphery. This results in each PE now having access to coordinates in an extra layer of link-cells (top right). Two further communication stages follow, to the east and to the above neighbours, in an analogous manner to give the final working environment (lower right).

As can be seen from Fig. 3.2.2, a subject domain's working environment of link-cells is built up by communications in each of the three perpendicular directions, *i.e.* with respect to the aforementioned scaled space used to assign particles initially to PEs. The nomenclature used is that the neighbouring PE in the -y (“southward”) direction is called South, that in the +x (“eastward”) direction is called East and that in the +z “upward” direction is called Above. In the first stage of communication, each PE simultaneously sends information for all particles found on the southern layer of link-cells to its southern neighbour. At the same time, each PE receives some information from its northern neighbour. This increases the working environment of a PE by one layer of link-cells. In the second and third stages similar exchanges occur in the other two directions, as shown in Fig. 3.2.2.

Following this initial communication stage, each processor has access to enough of the coordinate information to perform its share of the all-interacting-pairs search. This initial pair search not only allows the interaction of all non-bonded pairs but also the formation of neighbour tables of non-bonded pairs. These neighbour tables are used in subsequent steps until they become redundant. In the worst possible case, this is when any site has diffused a distance greater than or equal to half the shell width (a parameter used in the construction of the neighbour table²⁹) since the last rebuild. Only at this stage is it also necessary for sites to get reassigned to different PEs, should they have diffused out of their previous home PEs, so as to ensure that sites appear in their correct link-cells for the next rebuild. Subsequent steps are handled by the neighbour tables and other lists formed so it is irrelevant if a site wanders out of the region that its home PE is nominally responsible for.

3.2.3. Domain decomposition for molecular systems

The main problem encountered in applying domain decomposition to molecular systems is the unique identification of the n-plet groups required for multi-particle potentials and constraints. In their work on a domain decomposition algorithm for a molecular system, Esselink and Hilbers⁸⁶ show that, *as long as the distance between any two atoms in an n-plet is no greater than that of the truncation distance of the non-bonded pair potential, it is certain that at least one processor will have all the n atoms in its working environment*. As a result, no extra communication costs need to be incurred over those of the search for all-interacting-pairs. The same result was also found for linear chain systems²⁹, despite the fact that a different communication environment was used.

Given the above guiding principle, the problem thus reduces to one of *identifying* that processor which will be temporarily responsible for a particular n-plet and *signalling* to it that this is the case. In the work on linear homo-polymers, advantage was taken of the fact that all atoms were of the same type and that atoms bonded together were identifiable by their global indices²⁹. Despite this relatively simplified connectivity, the unique identification of the triplets and quadruplets proved quite a difficult problem. In the most general case, atom type and connectivity, although both fully specified, are completely arbitrary. For this reason, the previous approach had to be abandoned in favour of a more general one applicable to n-plets of any size. Extensive details of the solution are given in⁹⁹; in summary it involves the following novel features:-

- (i) In addition to its coordinates and velocities, each atom carries around with it details of its atom type, global atomic index, global molecular index and information concerning the neighbours with which it is has ‘bonded’ interactions.
- (ii) All bends, torsions and groups requiring special constraints, *e.g.* those for CH₂ groups as in ¹⁰³, are assigned to a particular atom and are also carried by this atom whichever processor it moves to.
- (iii) At the first step, and other steps where the neighbour lists *etc.* have to be rebuilt, a table is required indicating the local index and processor identity of all atoms known to each PE, *i.e.* both those a processor is directly responsible for and all those it receives during the communication stages prior to the search for all interacting pairs.
- (iv) Using the tables formed in (iii), any two atoms found during the pair search to be involved in a ‘bonded’ interaction make further lists which contain each others indices and processor identities in all the PEs they are known in. This information is communicated back to each atom’s home PE at the end of the pair search. By doing this, each atom knows for each of its ‘bonded’ neighbours all PEs in which they both appear in and their local indices therein. It is then a simple matter of finding for each n-plet the processor which is aware of all n atoms and signalling to it that it is responsible for that n-plet. This procedure allows each PE to build lists of all the n-plets it will handle for this and following steps until the lists are deemed redundant.

Although this procedure increases some of the communication costs, much of the information is invariant for the period that each processor works with the same set of atoms, *i.e.* between neighbour table rebuilds, and so the effect of the extra overhead is ameliorated.

3.2.4 The general purpose parallel MD code *ddgmq*

Using the techniques summarized above, an existing general purpose scalar code, *gmq*, has been successfully parallelized. The working code, *ddgmq*, has been implemented on Silicon Graphics Power Challenge and Origin 2000 machines with communications between processors being handled by calls to the MPI message passing library. The force-field utilised is fairly standard containing harmonic bond stretching and angle bending potentials, a torsion potential (expressed as a sixth order polynomial in the cosine of the dihedral angle) and also a potential to limit the out of plane motions of certain planar groups. Non-bonded interactions take the form of a Lennard-Jones 12-6 potential in tandem with a Coulombic potential operating between charged species; more about the Coulombic potential is said in the next section. The program also contains a parallel SHAKE algorithm to handle both rigid bond constraints and special CH₂ constraints.

Simulations are performed using 3-D periodic boundary conditions within a primary MD box defined by three arbitrary basis vectors, *i.e.* even triclinic cells can be simulated. Techniques are implemented for allowing control of the temperature and the pressure tensor in the system. Full details of the functionality of the *gmq* and *ddgmq* codes and their mode of operation can be found in the manual ¹⁰⁰.

3.2.5 Coulombic summations

A very important issue within domain decomposition, and MD in general, is the treatment of the Coulombic potential. From several studies already conducted ¹⁰⁴⁻¹⁰⁸, it is very clear that approximate

techniques which either simply truncate the $1/r$ potential or force it to zero using a switching function lead to drastic unphysical effects which only disappear when the significant long-range part of the Coulomb potential is handled in a proper manner. In three dimensionally periodic systems, lattice summation techniques such as the Ewald method¹⁰⁹ can be used. The Ewald summation nominally fits well into a domain decomposition strategy, and is currently used in *ddgmq*, as the sum over reciprocal lattice vectors, \mathbf{k} (defined with respect to the matrix of basis vectors \mathbf{h}) as

$$\mathbf{k} = 2\pi (\mathbf{h}^{-1})^t (L, M, N)^t,$$

with L , M and N integers) requires just the atom coordinates whilst the real space calculation is short-ranged and can be simply incorporated with the non-bonded pair force calculation. However, problems do exist with the Ewald summation when systems become large in extent, in that the number of \mathbf{k} with a modulus (the determining factor in the convergence of the reciprocal space sum) less than a certain value increases for a fixed cut-off in real space. The scaling of the method is thus non-linear and ultimately the cost becomes prohibitive.

Alternative methods with better scaling characteristics do exist and current opinion favours particle mesh approaches¹¹⁰. These involve solving Poisson's equation on a regular grid and three variants are currently under investigation for use in *ddgmq*. The first is based on the original particle-particle particle-mesh (P³M) description of Hockney and Eastwood¹¹¹ and uses fast Fourier transform (FFT) techniques. The second is the particle mesh Ewald approach of Darden *et al*^{112,113} which also uses FFT methods. The third approach uses an iterative scheme to solve Poisson's equation on the grid¹¹⁴ which, although not normally competitive with FFT techniques, has two possible advantages. First, it requires only near-neighbour communications and thus does not have the problems related to distributing the 3-D FFTs, and second, the method is, according to Beckers *et al*¹¹⁴, adaptable to only partially-periodic or non-periodic systems. This is not possible for the FFT methods which necessarily impose periodicity. The ability to be able to treat systems with limited periodicity is important with regard to projects involving interfaces like membranes simulations mentioned above.

Despite the obvious need for more efficient Coulombic summations, the code as it stands is being used at the moment in various projects to perform ~1 ns simulations involving up to ~100000 atoms with full account taken of the electrostatic interactions. The new parallelized general MD code can perform such computations in reasonable times (2 or 3 weeks). Such simulations were unfeasible to do within a reasonable time scale using scalar codes. In the following sections these various projects are described.

3.3 Protein folding

3.3.1 Introduction

Even though it has long been known that the amino acid sequence of a protein encodes all the information necessary for that protein to fold into its 3-D form¹¹⁵, predicting the tertiary structure of the protein given this sequence alone still remains a very difficult task. Some recent work, however, has shown that certain amino acids in the primary sequence are more important than others concerning the folding

process¹¹⁶. Ultimately, an understanding of the molecular basis for the sequence-structure relationship will allow rational modifications of the structure and stability of existing proteins and the design of new proteins able to perform specified functions¹¹⁷.

Although it is still unfeasible to follow the entire folding process using explicit MD simulations, the problem can be split into a number of tractable related parts which are open to investigation. In the following two sections, two different complementary approaches to the folding problem are summarized for molecules for which there is experimental data already available.

3.3.2 T4-Lysozyme

Bacteriophage T4-lysozyme is a single-chain globular polypeptide containing 164 amino acid residues and having a molecular weight of 18700. Its carboxyl-terminal domain consists mainly of α -helices while its amino-terminal domain includes both β -sheets and α -helices. These two domains are connected by a long α -helix. The active site is located at the junction of the two domains, opposite to the helix linking the two domains, and it might be expected that the positioning of one domain relative to the other will be critical for catalytic activity¹¹⁸. The crystal structure of the native form of the protein has been accurately determined¹¹⁹ and numerous amino acid replacements have been produced by appropriate mutagenesis and cloning¹²⁰. For several of these mutants, crystal structures are also available from the Protein Data Bank, giving precise information about the influence of a given mutation on the stability of the protein. From experimental studies, it has been established that several of these mutants display large hinge-bending displacements compared to the native structure.

As experimental studies of the transient intermediate structures adopted during the protein folding process are extremely difficult to carry out, computer simulation techniques are expected to provide information about the underlying mechanisms. These simulations form part of a combined theoretical/experimental investigation of the folding/unfolding properties of T4-lysozyme with a particular focus on the role of the constitutive α -helices which are known to play an important role in the formation of higher order structures¹²¹.

Rather than performing investigations on the whole protein, joint efforts have first concentrated on the stability of several helical pieces of the complete structure. These studies will increase in complexity towards packing calculations inside each domain. Some insight should thus be gained into the inherent stability of several constitutive structural elements, independent from their protein context first, before adding in more medium-range interactions. Particular attention is devoted to the central α -helix connecting both domains which seems to play a pivotal role with regard to the alignment of the two domains. This step-by-step analysis will enable estimations of the contribution of the different types of molecular interactions (intra- versus inter-domain interactions) to be made to the free energy of protein stabilization, and to better characterize the effect of solvent on each of them. This last aspect is particularly important as it is thought that protein self-organization at physiological temperatures is determined mainly by the entropy gain in the protein plus solvent system. Moreover, several models of protein folding, such as the kernel, the diffusion-collision and the molten globule models, could therefore be tested.

From the above considerations, the folding problem of T4-lysozyme can thus be resolved into four different sub-problems:

- (i) stability of the central helix connecting the two domains,
- (ii) stability and packing properties of helices in the C-terminal domain,
- (iii) stability and packing properties of helices and sheets in the N-terminal domain,
- (iv) stability of the whole protein.

In a first step, computer simulations on all the constitutive structural pieces of T4-lysozyme will be carried out independently of the others. These pieces will then be associated progressively in a steps of increasing complexity. Finally, the whole protein could be considered if necessary. Such a stepwise procedure seems justified from recent experimental studies. First, it has been shown that the constitutive pieces of T4-lysozyme are stable and have structural similarities with the native protein itself¹²². Second, there is evidence¹²³ which suggests that hydrophobic medium-range forces predominantly influence T4-lysozyme stability and thus determine the folding pathway¹²⁴.

For these studies, the logical starting point are the constitutive helices of T4-lysozyme. It is known that the stability of short polypeptide helices in solution probably plays an important role in protein folding¹²⁵. Experimental studies, designed to test the effect on catalytic activity following changes in the alignment of the two domains in T4-lysozyme, have already been carried out¹¹⁸. By substituting various residues in the long α -helix connecting both domains by proline, it was expected that significant distortion would take place and hence change the relative orientations of the upper and lower domains. The residues chosen for substitution were in that case the middle section of the long α -helix (residues 60-80), *i.e.* Gln-69, Val-71, Asp-72 and Ala-74. All mutants could be purified, except for the Val-71 proline substituted one which was either highly unstable or was rapidly degraded by proteolysis. The other three had catalytic activity in the range 50-90% of that of the pure form. The structural changes induced by the mutations are then used to try and rationalise these changes in activity.

In an analogous way, a series of MD calculations have been started using a model of just the interesting middle α -helical section of T4-Lysozyme, *i.e.* roughly from residues 60 to 80. The native structure and the Asp-72 and Gln-69 proline substituted systems have so far been simulated using MD. The fragments were first placed in boxes containing water. Overlapping water molecules were then removed and as many counter-ions as necessary were added in order to maintain charge neutrality.

In all of these, and the following, sets of calculations, the consistent valence force field (CVFF) was the basis of the interaction potential. The potential used in *ddgmq* differs slightly from that used in CVFF in two respects. First, the angle bending term is, for computational reasons, harmonic in the *cosine* of the angle, rather than in the angle itself⁹⁹. Second, the out-of-plane potential is harmonic in the distance of the central (trivalent) atom from the plane of the other three. This form was preferred to the improper dihedral form as the latter is ill-defined; there being three possible definitions of the improper dihedral angle and hence three different energies. In both the angle bending and out-of-plane bending cases we obtain the corresponding force constant for use in *ddgmq* from the CVFF form by equating the curvatures at the minimum of the potentials.

For the native case, the system contained the protein fragment (361 atoms), 3036 water molecules and one Na^+ counterion for a total of 9470 atoms at an initial density of $990.6755 \text{ kg m}^{-3}$. This system was equilibrated at 300 K over a period of 20 ps using loose-coupling to a thermal bath¹⁸ and a coupling constant, τ_T , of 1 ps. It was then allowed to relax under NpT conditions, again using a loose-coupling method¹⁷, at an applied isotropic pressure, p , of 1 bar. After a further 100 ps, the average system pressure had relaxed to that desired and the volume was again fixed giving a final density of $976.6833 \text{ kg m}^{-3}$. A longer constant volume simulation was then carried out with very loose-coupling to the thermal bath using a τ_T of 10 ps, *i.e.* just sufficient to prevent the temperature drifting away from 300 K. The total length of the simulation was 2400 ps including the equilibration. During this entire period the protein fragment retained its helical structure. With the Ewald summation parameters ($\alpha=0.3 \text{ \AA}^{-1}$, $R_c=8.5 \text{ \AA}$, $K_{\max}=14$) optimized to give a less than 10 bar differential between the direct and indirect calculations of the reciprocal space contribution to the pressure⁴² and a tolerance of 10^{-5} used in the SHAKE routine maintaining all bond lengths rigid, the *ddgmq* program required about 0.95 s of CPU time per 1 fs time step (*i.e.* ~ 26 days of CPU time for the 2400 ps) on 12 processors of an SGi Origin 2000. The final configuration of the protein fragment is shown in Fig. 3.3.1

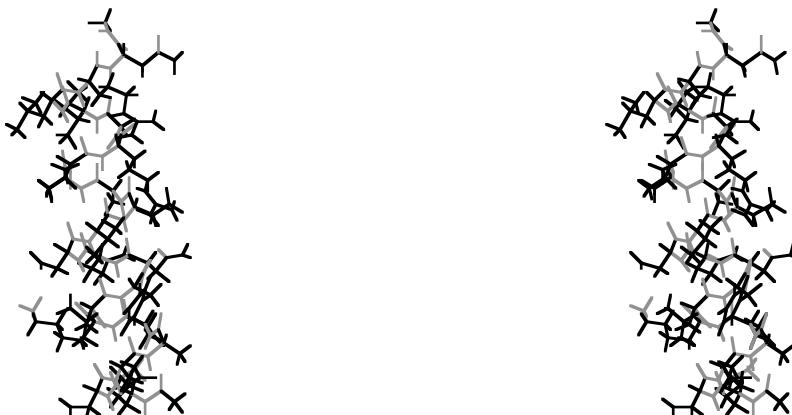


Figure 3.3.1 A stereo image of the native structure of the fragment of the long α -helix linking the two terminal domains in T4-Lysozyme. The plot shows the structure at the end of a 2.4 ns simulation in water at 300 K.

For the Gln-69 proline substituted case, the system contained the protein fragment (358 atoms), 3036 water molecules and one Na^+ counterion for a total of 9467 atoms at an initial density of $990.139 \text{ kg m}^{-3}$. The system was first energy minimized before being heated to 300 K over a period of 10 ps by simply setting the desired temperature to this value and using $\tau_T=1$ ps. However, the protein fragment unfolded within this period. Whether this was due to a real instability of the fragment or resulted from the relatively rapid heating is not fully established yet. Certainly during the heating phase non-equipartition of the kinetic energy was apparent with the temperature evaluated from just the centres-of-mass translational degrees of freedom rising to over 500 K before falling back to 300 K. A second simulation has been carried out with a much slower linear rate of heating of 3 K per ps. Using this approach the fragment has remained stable during the 100 ps heating phase and a further 100 ps equilibration at 300 K.

In the light of the previous result, the system size was increased in the Asp-72 proline case in order to have a larger separation between the periodic images of the protein fragment, particularly if partial unfolding occurred. Thus the full system contained the protein fragment (363 atoms) and 16907 water molecules for a total of 51084 atoms in a cube of side length 80 Å, *i.e.* at an initial density of 996.0785 kg m⁻³. With the larger system size the Ewald parameters were re-optimized and the following set is being used { $\alpha=0.2 \text{ \AA}^{-1}$, $R_c=11.5 \text{ \AA}$, $K_{\max}=10$ }. The real space cut-off chosen allows for a 3×3×3 domain decomposition to be made and on 27 processors of the SGi Origin 2000 the *ddgmq* code runs at about 1.38 s per time step. At the moment, this system is being energy minimized before being equilibrated at 300 K.

Data from these simulations are currently being analysed and a paper is in preparation ¹²⁶.

3.3.3 Ubiquitin

An alternative way of approaching the problem of folding in proteins is to look at the unfolding behaviour while assuming that the process of denaturing (unfolding) follows the same pathway as the folding. The so-called "cold-denaturing" process of globular proteins is a general phenomenon and can be induced in various ways, for example by a jump in temperature, a change in solvent or by varying the pH ¹²⁷. Another way is by increasing the pressure, which does have certain advantages, and a number of experimental studies have been made using this approach ¹²⁸⁻¹³².

Human ubiquitin is a relatively small 76-residue cytoplasmic protein that serves as a protein degradation signal when conjugated to another protein ¹³³. Its structure in the folded native state consists of an extended β-sheet and an α-helix surrounding a pronounced hydrophobic core ¹³⁴; the native structure of ubiquitin is shown in Fig. 3.3.2. It is also known to have unfolded and partially folded structures which have been studied experimentally using NMR techniques, both at ambient ^{133,134} and raised pressures ¹²⁸, and by simulation methods at ambient pressures ¹³⁵. It is thus a good candidate for a simulation study of the pressure induced cold-denatured states.

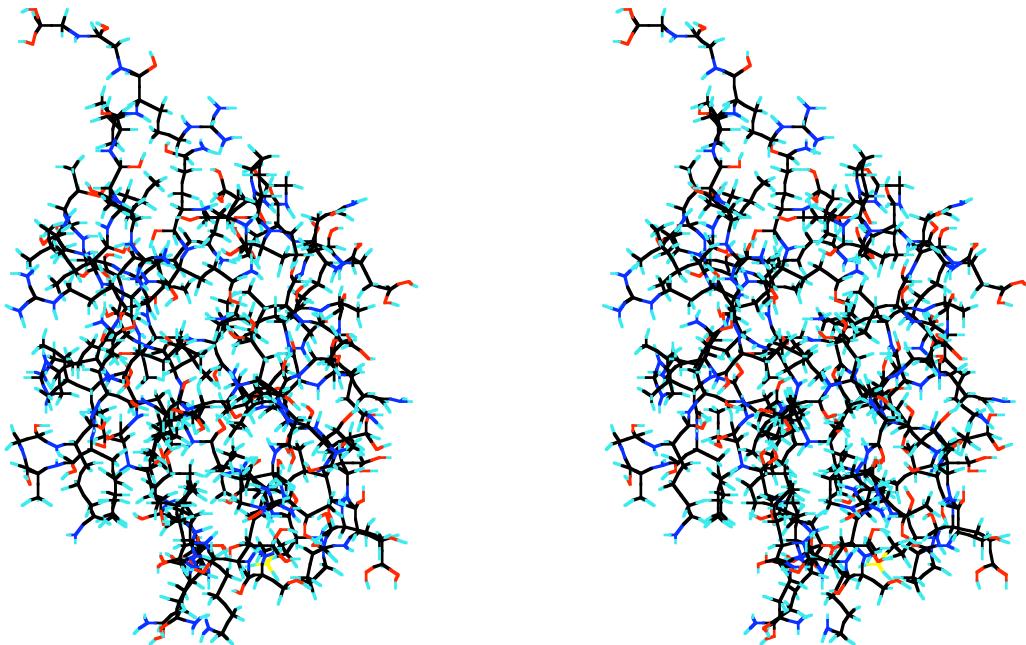


Figure 3.3.2 A stereo image of the structure of ubiquitin.

To this end we have initiated an MD simulation of ubiquitin in aqueous solution. The initial structure was taken from the Brookhaven Protein Data Bank (<http://www.pdb.bnl.gov>) and first solvated in a cubic box of side length 80 Å. 16588 waters were added so that the system contained 50995 atoms altogether at a density of 996.9837 kg m⁻³. This system was first energy minimized before being gradually heated from 0 K to 300 K over a period of 30 ps using loose-coupling to a thermal bath ¹⁸ and a coupling constant, τ_T , of 1 ps. Equilibration was then continued for another 20 ps before the system was allowed to relax under NpT conditions, again using a loose-coupling method ¹⁷, at p=1 bar. After a further 100 ps the system density had smoothly relaxed to a value of 970.3 kg m⁻³. The relaxation of the system density and pressure are illustrated in Fig. 3.3.3.

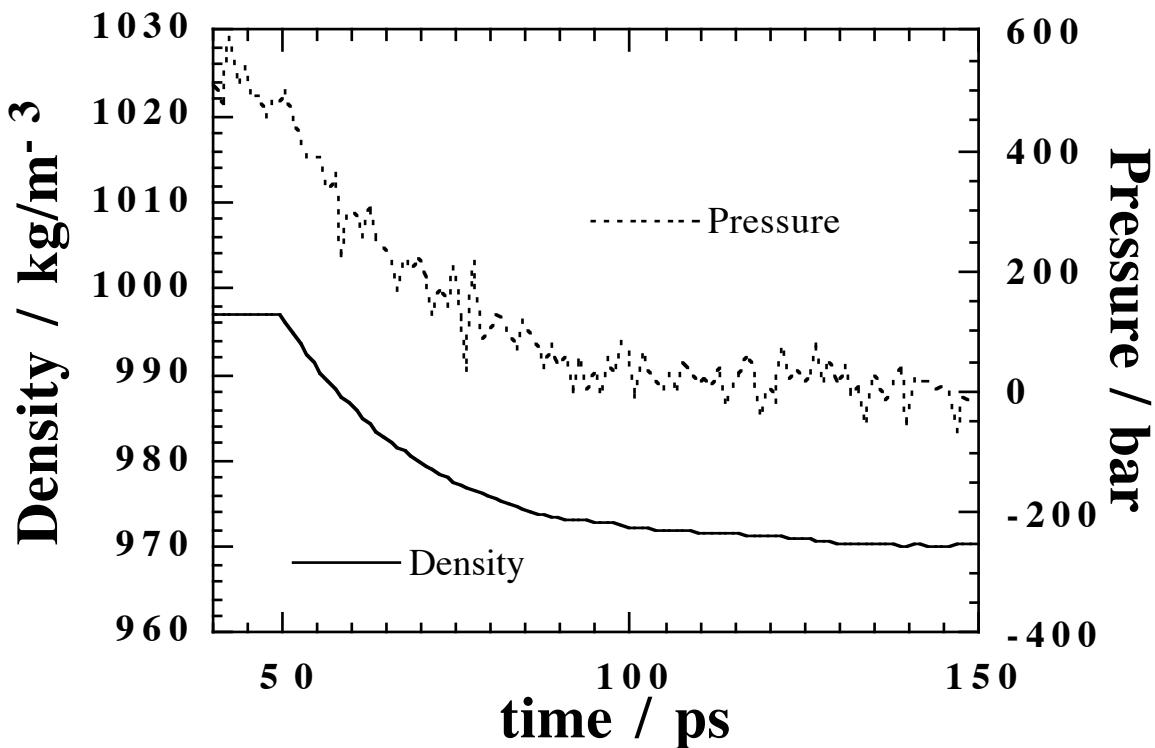


Figure 3.3.3 The behaviour of the sub-averaged (1 ps intervals) density and pressure from the simulation of ubiquitin in water. Up until 50 ps the volume is constant, thereafter the volume is allowed to change in response to the imbalance between the internally measured and externally applied pressure of 1 bar. Note the smooth exponential relaxation accorded by the loose-coupling method.

On 27 processors of an SGi Origin 2000, the *ddgmc* program requires about 1.96 s of CPU time per 1 fs time step. In the next stage, the stability of the model ubiquitin will first be checked by performing a long simulation under ambient conditions. Should this first stage be completed satisfactorily, the system will then be subjected to a gradually increasing pressure and monitored for any changes in configuration.

3.4 Enzymes

3.4.1 Introduction

Of the many different types of enzymes, two important classes involved in many biological functions are the serine ($R=CH_2-OH$) and cysteine ($R=CH_2-SH$) proteases. Both catalyse the breaking of peptidic or esteric bonds by hydrolysis. The catalysis in the case of serine protease involves two other residues, aspartic acid ($R=CH_2-CO_2H$) and histidine ($R=CH_2-C_3N_2H_3$), with the hydroxyl group of serine acting as a nucleophilic agent whilst the imidazole of histidine behaves as an acid/base.

Among the serine proteases, the bovine trypsin, a pancreatic trypsin involved in digestion, has been studied for more than 40 years. Nevertheless, its reaction mechanism is not completely understood. Notably the extraordinary stability of the complex formed between bovine trypsin and BPTI (Bovine Pancreatic Trypsin Inhibitor) is not really explained. The conformation of this complex is often used in the reactivity studies as a model of the Michaelis complex, which generally leads to the formation of the acyl-enzyme and then to the hydrolysis of the protein substrate. The overall aim of this work is to understand why there is no hydrolysis of the BPTI, why this complex does not lead to the cleavage of the peptidic bond of BPTI and why there is no nucleophilic attack of the serine. For this reason, an accurate study of the trypsin-BPTI interaction in the active site of this complex is well under way.

The work on BPT is also important with regard to a second serine protease being studied, human Hageman Factor XII-a. There is a certain homology between the two and so the inhibitors of this second enzyme will first be modelled in complexes with BPT, as a number of the 3-D structures are available, to see if there is a general mechanism of inhibition. Ultimately the aim is to use the findings of the work on BPT to understand the same process in Factor XII-a and possibly suggest better inhibitors. Before this can be done, however, a model of Factor XII-a has to be developed (there being no X-ray structure available) and verified.

In the cysteine proteases there are also two other residues involved in the catalysis; a histidine and an asparagine ($R=CH_2-CONH_2$). It is thought that the active site is the zwitterionic form, $Asn\cdots His^+\cdots Cys^-$, in which the thiolate functions as a nucleophile during the early stages of the catalysis. However, it is not known if this zwitterion is the most stable form or, indeed, whether it is involved in the catalysis.

To try and answer these questions an investigation of the stability of different states of the $Asn\cdots His\cdots Cys$ catalytic triad in Glyceraldehyde-3 Phosphate Dehydrogenase has begun. There are a number of possibilities as the histidine has two tautomeric forms, δ and ϵ , and a charged state, His^+ , and the cysteine can be neutral or charged, Cys^- . The more likely forms will be studied using hybrid quantum mechanics/molecular mechanics (QM/MM) calculations, but in the first stage it will be necessary to optimise the structures using MD and energy minimization.

3.4.2 Bovine Pancreatic Trypsin

The Bovine Pancreatic Trypsin Inhibitor (BPTI) is a naturally occurring protein of 58 amino-acids which inhibits the Bovine Pancreatic Trypsin (BPT) serine protease. It binds to the active site of BPT by a complex network of hydrogen bonds. The native structure of BPTI is stabilised by three disulphide bridges and consists of a 3_{10} (N-terminal part) helix, an anti-parallel β -sheet and an α -helix (C-terminal part).

The portion of BPTI in contact with the active site of BPT resembles that of a bound substrate to such an extent that it is hard to explain why no reaction occurs between them. Instead of reacting, the two form an extremely stable complex with the association constant being one of the highest known (10^{13} mol^{-1}). The structure of the BPT/BPTI complex has been resolved by X-ray crystallography and the associated PDB file was used as the starting point of the modelling calculations.

The catalytic triad is formed by three aminoacids, histidine E57, aspartic acid E102 and serine E195. The lysine I15 of BPTI is the P1 site, its side chain is maintained in the S1 site of the trypsin by hydrogen bonds between the NH_3^+ group and two amino-acids, an aspartic acid (E189) and a serine (E190). We can note the presence of at least three water molecules, which could play a role in the stability of the complex by

forming hydrogen bonds with both amino-acids of BPTI and amino-acids of trypsin. This complex is very interesting because BPTI places a peptidic bond between a Lysine (I15) and an Alanine (Ala I16) in the catalytic triad of trypsin, exactly as a cleavable peptide would do, with the side chain of the Lysine in the binding pocket. It is for this reason that the complex is often used as a model of a transition state for the reaction between trypsin and its substrates. In this case, though, one has to suppose that the peptidic bond between Lys I15 and Ala I16 assumes a tetrahedral conformation. There is, however, no nucleophilic attack from the serine on this bond and consequently no formation of the acyl-enzyme. It is, thus, interesting to try to understand why such a conformation of the active site does not lead to the activation of the enzyme.

As the significant BPT/BPTI interactions are mostly hydrogen bonds, the local self-consistent field (LSCF) hybrid quantum mechanical/molecular mechanical method¹³⁶ has been used to study the strength of these bonds and potential proton transfers; the LSCF treatment allows a small sub-system, here the catalytic site and the "substrate", to be treated quantum mechanically whilst the rest of the system is represented by classical atoms fixed in space. These calculations have, though, proved sensitive to whether neutralising counter-ions were placed close to the charged amino-acid groups. Since these LSCF calculations were done in a vacuum with only two (fixed) layers of water molecules, performing completely classical MD simulations with the system immersed in bulk water and with proper account taken of the long-range part of the Coulombic interactions is very useful. The more representative arrangement of the counter-ions expected from such a simulation can then be used as input to a new set of LSCF calculations.

In these MD simulations, the BPT molecule contained 230 residues (3220 atoms) and the BPTI molecule 58 residues (892 atoms). The system also contained a Ca²⁺ ion, 8 Na⁺ ions and 22 Cl⁻ ions along with 17611 water molecules for a total of 56976 atoms altogether. The potential used was based on the standard CVFF one. Initially this system was contained within a cube of side length 80 Å ($\rho=1003.5 \text{ kg m}^{-3}$). After equilibration at 300 K for 20 ps, the system was allowed to relax under NpT conditions at an applied pressure of 1 bar. After a period of 100 ps, the pressure had relaxed to the desired value and the volume was again fixed; the new density being $\sim 988.6 \text{ kg m}^{-3}$. The simulation was then continued under NVT conditions using loose-coupling to a thermal bath and a coupling constant of 10 ps for another 880 ps for a total simulation duration of 1 ns. With the following Ewald summation parameters ($\alpha=0.2 \text{ \AA}^{-1}$, $R_c=12 \text{ \AA}$, $K_{\max}=10$) and a tolerance of 10^{-5} used in the SHAKE routine maintaining all bond lengths rigid, the program used 1.9 s of CPU per 1 fs time step on 27 processors of the SGi Origin 2000.

The results of this simulation are currently being analysed in terms of the electric field generated by the system around the catalytic site. This analysis will then be used as a start point for further LSCF calculations and findings will be published.

3.4.3 Human Hageman Factor XII-a

Without an X-ray structure for Factor XII-a, other methods have to be used to construct an initial 3-D model. At LCTN, the SWISS-PROT sequence (FA12_HUMAN, P00748) of the human Factor XII precursor (EC 3.4.21.38) has been considered so as to extract from it the sequence of amino acids ranging from Val₃₇₃ to Ser₆₁₅ corresponding to the heavy chain of β Factor XII-a. This sequence of amino acids was then used to search for similarities with other protein sequences using the EMBL databank and the FASTA¹³⁷ program. From the results obtained this way, it appeared that, apart from Factor XII encountered

in other species, the best homology was obtained with three protein types: the hepatocyte growth factor activator (47.7% identity), the tissue plasminogen activator (40.7% identity) and the urokinase-type plasminogen activator (39.4% identity). For the last two of these, X-ray structures were recently made available; PDB codes 1rtf and 1lmw, respectively, for the tissue plasminogen activator and the urokinase-type plasminogen activator. The sequence alignment between the β Factor XII-a and the protease part of these two plasminogen activators was then used next for homology modelling.

The *Homology* program of the BIOSYM modelling package assigned co-ordinates for the β Factor XII-a backbone based on the homology with parts of the 1rtf and 1lmw 3-D structures. The side-chains were positioned according to their most suitable similarities with the corresponding amino acids in the target 1rtf and 1lmw proteins. To add the missing parts, the loop searching procedure proposed in the *Homology* program was used.

The crude 3-D model of β Factor XII-a thus obtained was next refined using several rounds of energy minimisation procedures and the *Discover 93.0/3.0.0* [®] BIOSYM program. In the first step, the backbone of the regions of β Factor XII-a having a good similarity with those of 1rtf and 1lmw proteins were frozen. Next all constraints were removed and the structure fully energy-refined until convergence with the conjugate gradient algorithm. No cut-off was used in these calculations and the dielectric constant was chosen as distance dependent. The N- and C-terminal groups were considered as charged as well as the Arg, Lys, Asp and Glu residues.

The final configuration of the system, relaxed in the way described above, was then placed in a periodic MD box of side length 80 Å. Water molecules were added in order to fill the box to a realistic density (1001.75 kg m⁻³). The full system thus contained the β Factor XII-a (3574 atoms), 12 Na⁺ ions neutralizing the charge on the system, and 15682 water molecules, *i.e.* a total of 50632 atoms altogether. This system was subjected to MD simulations using the parallel program *ddgmc* with the potential based on the standard CVFF one. After equilibration at 300 K for 20 ps, the system was allowed to relax under NpT conditions at an applied pressure of 1 bar. After a period of 180 ps, the pressure had relaxed to the desired value and the volume was again fixed; the new density being 987.24 kg m⁻³. The simulation was then continued under NVT conditions using loose-coupling to a thermal bath and a coupling constant of 10 ps for another 800 ps for a total simulation duration of 1000 ps. With the following Ewald summation parameters ($\alpha=0.2$ Å⁻¹, $R_c=11.5$ Å, $K_{\max}=10$) and a tolerance of 10⁻⁵ used in the SHAKE routine maintaining all bond lengths rigid, the program used 1.68 s of CPU per 1 fs time step on 27 processors of the SGi Origin 2000. After the 1 ns MD simulation, the system was cooled to 1 K using MD before a final energy minimization was performed.

Preliminary analysis of the results of the MD simulation at 300 K show that the biopolymer chain relaxes to a stable conformation within the 1 ns. The final minimized configuration will be compared to models produced by two other labs in Porto and Rio de Janeiro. The findings of this study will then be published. The model of β Factor XII-a will then be used in studies of naturally occurring and synthetic inhibitors of trypsin and β Factor XII-a itself in order to better understand the inhibition of these serine proteases and, hence, ultimately produce more potent ones.

3.4.4 Glyceraldehyde-3 Phosphate Dehydrogenase (GAPDH)

The structure of GAPDH available from X-ray crystallographic studies does not contain the positions of the hydrogen atoms. It has thus been necessary to add these taking into account the following considerations. The active site of the enzyme contains the cysteine and histidine residues, Cys₁₄₉ and His₁₇₆, respectively. These are able to be considered as neutral or charged. For cysteine, the neutral form contains the S-H thiol group. When in the charged state the proton is lost and so the thiol becomes a thiolate group, S⁻. For histidine, there are two tautomeric forms of the imidazole ring (Im^δ or Im^ε), named according to which of the nitrogens (N^δ or N^ε) carries the hydrogen. Thus the histidine itself has two forms, histidine(δ) and histidine(ε). In the charged state, histidine(+), both nitrogens are associated with a hydrogen and the imidazole ring carries a charge of +1, Im⁺. There are then six possible pairings of His₁₇₆ and Cys₁₄₉.

At the optimum pH for the activity of the enzyme, it is thought that the pair exist as S⁻/Im⁺ in equilibrium with the neutral form SH/Im (either SH/Im^δ et SH/Im^ε). Under acidic conditions, the SH/Im⁺ pair should predominate whereas under alkaline conditions the forms S⁻/Im^δ and S⁻/Im^ε are more likely to occur.

The first stage of this project is therefore to optimize the structure of the GAPDH tetramer. The starting point for this work is the X-ray structure of GAPDH taken from the bacteria *stearothermophilus*. However, in the X-ray structure, the cysteine group is oxidized and carries a bulky group which distorts the system with respect to that required. If the thiol is simply substituted for the bulky group, an energy minimization runs the risk of finding a local minimum which does not correspond to the optimal structure. For this reason, the following procedure was carried out.

The GAPDH was placed in an MD box of side length 100 Å. The box was then filled with water. The system thus contained the tetramer (3 glyceraldehyde molecules of 70 atoms each plus the 20462 atoms in the protein chain), 68 Na⁺ ions, 52 Cl⁻ ions and 26146 water molecules, *i.e.* a total of 99230 atoms altogether at a density of 1029.928 kg m⁻³. Using the *ddgmq* program, this system was first energy minimized to remove high energy overlaps resulting from the introduction of the water molecules. A molecular dynamics simulation was then performed in which the temperature was raised gradually to 300 K at constant volume. This was achieved by initially randomizing the velocities using a Gaussian distribution characteristic of 300 K and setting the temperature loose-coupling constant to 1 ps. After 20 ps of MD, the system had relaxed at the desired temperature. It was then quenched by setting the required temperature to 1 K using MD. After 10 ps, the system was energy minimized. A second period of MD at 300 K followed in the same way. This time, the total period of dynamics was extended to 100 ps before the system was quenched and energy minimized in the same way as before. With the following Ewald summation parameters {α=0.2 Å⁻¹, R_c=11.5 Å, K_{max}=10} and a tolerance of 10⁻⁵ used in the SHAKE routine maintaining all bond lengths rigid, the program used the amounts of CPU per 1 fs time step shown in the following table on the 64 processor SGi Origin 2000 in Nancy.

No. of Procs. (N _p)	CPU time per step / s (τ _p)	%Efficiency 200*τ ₂ /N _p *τ _p	Speed-up 2*τ ₂ /τ _p
1	<i>Too much memory required for the Origin 2000 in Nancy</i>		
2	51.46	100	2
4	20.93	122.9	4.9
8	9.03	142.5	11.4
12	5.95	144.1	17.3
18	4.37	130.9	23.6
27	3.30	115.6	31.1
36	2.40	119.4	43.0
48	1.81	118.3	56.8
64	1.47	109.6	70.2

At the time of performing the tests the memory was insufficient to compile the code to run on just 1 processor so the efficiencies and speed-ups shown are all relative to the two-processor result. For this large system *ddgmq* delivers good performance up to the maximum number of processors available.

The MD calculations on the GAPDH will be continued to see if the structure relaxes further. However, from the structure already generated, the hybrid QM/MM calculations will be initiated. These will study the variation of the reactivity as the SH group is changed to an OH group. These results will be compared to *ab initio* studies of the model systems CH₃OH/methyl-Imidazole and CH₃SH/methyl-Imidazole. The results will also be compared to experimental data where possible. Ultimately, the reaction path should be amenable to study with the QM/MM method including all the various stages of the reaction, transition states, energy barriers *etc.*

3.5 Drug design

3.5.1. Introduction

Integrins are a widely expressed family of cell surface adhesion receptors consisting of α and β subunits ¹³⁸. Integrin α_{IIb}β₃ was the first integrin to be identified and has subsequently been the most studied. It is known that it facilitates binding between fibrinogen and platelets during the platelet aggregation process leading to thrombus formation ¹³⁹. Disintegrins are a family of low molecular weight proteins originally isolated from the venom of snakes belonging to the *viperidae* family ¹⁴⁰, but also later from the venom of certain leeches and ticks ¹⁴¹. The primary amino acid sequences of the disintegrins are highly homologous and all these proteins are cystein-rich, forming disulphide bridges which confer on them a structurally constrained core ^{140,142}. Disintegrins inhibit the adhesive functions of a variety of integrins on cell surfaces.

Almost all disintegrins contain the sequence RGD, which is known to be a recognition site for numerous integrins ^{138,141,143}. An exception to this is barbourin, a disintegrin which contains the KGD

sequence instead of RGD, and which selectively inhibits $\alpha_{IIb}\beta_3$ ¹⁴⁴. All the disintegrins inhibit the $\alpha_{IIb}\beta_3$ receptor present on platelet and megakaryocyte surfaces via their RGD (or KGD) sequence. It is interesting then to study their amino acid sequences and their three dimensional structures in order to elucidate the active conformation of the RGD binding sequence. Five structures have been obtained using two-dimensional NMR spectroscopy ; those of kistrin¹⁴⁵, decorsin¹⁴⁶, flavoridin^{147,148}, dendroaspin (also referred to as mambin)^{149,150} and echistatin¹⁵¹⁻¹⁵³. Except for decorsin, for which the RGD-containing region is well defined, this binding site is situated for the four other disintegrins at the apex of a relatively flexible loop protruding from the rigid core of the protein. Therefore, although the core structure could be relatively well resolved using NMR spectroscopy data, the three-dimensional structure of this most interesting region remains ill-defined for the majority of these disintegrins. Molecular modelling can then aid in the determination of the solution structures of the RGD loop. It was thus decided to carry out a simulation of a disintegrin in aqueous solution in order to elucidate the spatial conformations of the RGD sequence and the structures of the neighbouring amino acids forming the hairpin loop. To validate the molecular dynamics calculations, the results will be compared to NMR spectroscopy data.

3.5.2 Echistatin

Echistatin is a small peptide containing 49 amino acids (713 atoms) which is found in the venom of the saw-scaled viper, *Echis carinatus*. It is a member of the disintegrin family of natural peptides which inhibit blood coagulation by blocking the platelet-fibrinogen binding. Many members of this family have been studied since antiplatelet therapy became a useful means of preventing acute thrombo-embolic artery occlusions in cardiovascular diseases. Echistatin is known to be a particularly potent inhibitor of platelet aggregation, interacting with platelets through its binding sequence Arg24-Gly25-Asp26. In addition, the high affinity interaction of echistatin with platelets may depend upon its secondary structure too¹⁵⁴⁻¹⁵⁶. It is, therefore, interesting to study the structure of echistatin in solution using molecular dynamics.

Each of the eight 1H – NMR refined structures of echistatin in solution proposed by Atkinson *et al.*¹⁵¹⁻¹⁵³, available on-line from the Brookhaven National Laboratory Protein Data Bank WWW server (<http://www.pdb.bnl.gov>) with the reference 2ech, was first energy minimized in vacuum using the CVFF force-field to describe the interactions; these computations being carried out using the *Discover 93.0/3.0.0*® program from Biosym/MSI of San Diego. The lowest energy conformation (model 1 in the PDB file) was placed at the centre of a 60^3 \AA^3 cubic (periodic) cell filled with water molecules. Water molecules overlapping with the echistatin molecule were first removed before the entire system was energy minimized. A water molecule far from the peptide was then replaced by a chloride ion to neutralize the charge of the cell. The resulting system, containing 6903 water molecules, an echistatin molecule and one chloride ion, (21423 atoms in total at a density of 997.97 kg m^{-3}), served as a starting point for molecular dynamics simulations. With a time step of 1 fs, this system was followed for 2200 ps in total. These latter simulations used the parallel general purpose MD code *ddgmg*^{99,100} with the potential used being based on CVFF. The values used throughout for the Ewald summation were ($\alpha=0.29 \text{ \AA}^{-1}$, $R_c=8.5 \text{ \AA}$, $K_{\max}=14$). Although these values may seem sub-optimal for a scalar program, the reduction of the real space truncation to 8.5 \AA allows the system to be decomposed into $3\times 3\times 3=27$ domains each of $\sim 20 \text{ \AA}$ side length and each containing the minimum number of link-cells allowed by the code, *i.e.* 2 per dimension each with a length greater than R_c .

plus the shell-width (1.5 Å) used in the creation of the neighbour table. On 27 processors of the Origin 2000, the entire 2.2 ns simulation required ~25 days of CPU time, *i.e.* approximately 1 s of CPU time per 1 fs time step. The performance of *ddgmq* on this problem as a function of the number of processors is summarized in the following table.

No. of Procs. (N _p)	CPU time per step / s (τ _p)	%Efficiency 100*τ ₁ /N _p *τ _p	Speed-up τ ₁ /τ _p
1	24.20	100	1
2	12.19	99.2	2.0
4	4.98	121.4	4.9
8	2.51	120.3	9.6
12	1.80	111.9	13.4
18	1.32	101.5	18.3
27	1.02	88.0	23.8

In contrast to the much larger GAPDH system, the maximum number of domains, and hence processors, is limited in this system to 27. It is clear that the trade-off between real-space and reciprocal space calculations leads to some loss of efficiency the main point is that the CPU time can be reduced to a point which makes the calculation feasible.

During the first 20 ps of the simulation, the system was brought to thermal equilibrium at constant volume using loose-coupling ($τ_T=1$ ps) to a thermal bath ¹⁸ at the required temperature of 298 K. The system was then allowed to equilibrate at the required pressure, 1 bar, again using a loose-coupling technique ¹⁷. After a further 80 ps, the system density had relaxed in response to the required pressure and the volume was then fixed for the rest of the MD simulation; the corresponding density being 979.5 kg m⁻³. The temperature was controlled for a further 20 ps before the system was allowed to evolve under NVE conditions until an elapsed duration of 1200 ps. At this point, it was noted that a slow drift in energy was occurring despite the removal of all high frequency bond vibrations by SHAKE and a tolerance of 10⁻⁵. The drift amounted to a rise in the temperature of ~1 K over 1 ns, so it was not considered particularly alarming. Nevertheless for the remainder of the simulation, the loose-coupling algorithm was again used to correct this drift. It was found that a coupling constant of 10 ps achieved the desired effect.

The data from this simulation are currently being analysed and preliminary comparisons of experimental and simulation generated nuclear Overhauser effect (NOE) maps are quite encouraging. A paper presenting the entirety of the analyses is in an advanced state of preparation ¹⁵⁷. A stereo image of the final configuration of the echistatin molecules is shown in Fig. 3.5.1.

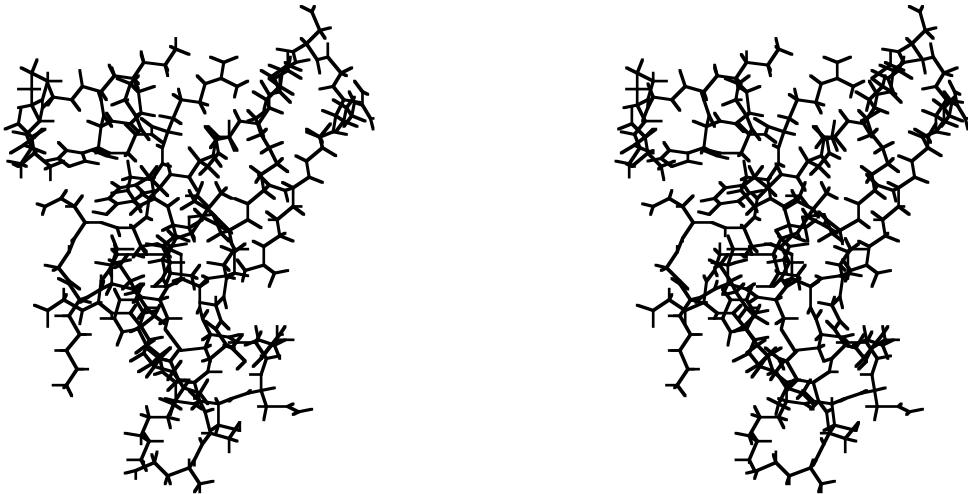


Figure 3.5.1 A stereo image of the structure of echistatin in aqueous solution following a 2.2 ns MD simulation at 298 K.

3.6 Conclusions

It is clear that this new parallel code in harness with the latest parallel machines delivers the kind of compute power that allows us to make a significant step forward in the size and complexity of problems that can be studied. The possibilities in the diverse areas of science are numerous. A few of these in the area of the biosciences and polymer materials have been alluded to in this summary. Many more can be envisaged. The prospects then for large scale atomistic simulations is indeed very promising in its own right. When combined with quantum mechanical methods and continuum approaches it offers the exciting possibility of studying reacting systems in reasonably realistic environments.

References

- (1) Lees, A. W.; Edwards, S. F. *J. Phys.* **1972**, C5, 1921-1929.
(2) Brown, D.; Clarke, J. H. R. *Chem. Phys. Letts.* **1983**, 98, 579-583.
(3) Clarke, J. H. R.; Brown, D. *Molecular Physics* **1986**, 58, 815-825.
(4) Clarke, J. H. R.; Brown, D. *J. Chem. Phys.* **1987**, 86, 1542-1547.
(5) Brown, D.; Clarke, J. H. R. *J. Chem. Phys.* **1987**, 86, 6446-6453.
(6) Travis, K. P.; Brown, D.; Clarke, J. H. R. *J. Chem. Phys.* **1993**, 98, 1524-1530.
(7) Travis, K. P.; Brown, D.; Clarke, J. H. R. *J. Chem. Phys.* **1995**, 102, 2174-2180.
(8) Travis, K. P., Ph. D. thesis, *Computer simulation of liquid lubricants*, UMIST, **1993**.
(9) Liem, S. Y.; Brown, D.; Clarke, J. H. R. *Phys. Rev. A* **1992**, 45, 3706-3713.
(10) Heyes, D. M. *J. Chem. Soc., Faraday Trans. 2* **1986**, 82, 1365-1383.
(11) Brown, D.; Clarke, J. H. R. *J. Chem. Phys.* **1990**, 92, 3062-3073.
(12) Chandler, D. *J. Chem. Phys.* **1978**, 68, 2959.
(13) Edberg, R.; Evans, D. J.; Morriss, G. P. *J. Chem. Phys.* **1987**, 87, 5700.
(14) Brown, D.; Clarke, J. H. R. *J. Chem. Phys.* **1990**, 93, 4117-4122.
(15) Brown, D.; Clarke, J. H. R. *J. Chem. Phys.* **1990**, 94, 4684.
(16) Brown, D.; Clarke, J. H. R. *J. Chem. Phys.* **1986**, 84, 2858-2865.
(17) Brown, D.; Clarke, J. H. R. *Comput. Phys. Commun.* **1991**, 62, 360-369.
(18) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, 81, 3684-3690.
(19) Parrinello, M.; Rahman, A. *Phys. Rev. Letts.* **1980**, 45, 1196.
(20) Brown, D.; Clarke, J. H. R. *Macromolecules* **1991**, 24, 2075-2082.
(21) McKechnie, J. I.; Haward, R. N.; Brown, D.; Clarke, J. H. R. *Macromolecules* **1993**, 26, 198-202.
(22) Clarke, J. H. R.; Brown, D. *Molecular Simulation* **1989**, 3, 27-47.
(23) McKechnie, J. I.; Brown, D.; Clarke, J. H. R. *Macromolecules* **1992**, 25, 1562-1567.
(24) Brown, D.; Clarke, J. H. R.; Okuda, M.; Yamazaki, T. *J. Chem. Phys.* **1994**, 100, 6011-6018.
(25) Flory, P. J. *The Statistical Mechanics of Chain Molecules*; Hanser Publishers: New York, 1988.
(26) Flory, P. J. *Polymer Journal* **1985**, 17, 1.
(27) Brown, D.; Clarke, J. H. R.; Okuda, M.; Yamazaki, T. *J. Chem. Phys.* **1994**, 100, 1684-1692.
(28) Brown, D.; Clarke, J. H. R.; Okuda, M.; Yamazaki, T. *Comput. Phys. Commun.* **1993**, 74, 67-80.
(29) Brown, D.; Clarke, J. H. R.; Okuda, M.; Yamazaki, T. *Comput. Phys. Commun.* **1994**, 83, 1-13.
(30) Brown, D.; Clarke, J. H. R.; Okuda, M.; Yamazaki, T. *J. Chem. Phys.* **1996**, 104, 2078-2082.
(31) Neyertz, S.; Brown, D. *J. Chem. Phys.* **1995**, 102, 9725-9735.
(32) Neyertz, S.; Brown, D.; Clarke, J. H. R. *J. Chem. Phys.* **1996**, 105, 2076-2088.
(33) Gray, F. M. *Solid Polymer Electrolytes, Fundamentals and Technological Applications*; VCH: New York, 1991.
(34) Vincent, C. A. *Prog. Solid State Chem.* **1987**, 17, 145.
(35) Ratner, M. A.; Shriver, D. F. *Chem. Rev.* **1988**, 88, 109.
(36) Bruce, P. G.; Vincent, C. A. *J. Chem. Soc. Faraday Trans.* **1993**, 89, 3187.
(37) Takahashi, Y.; Tadokoro, H. *Macromolecules* **1973**, 6, 672-675.
(38) Wendsjö, Å. Ph. D. thesis, *The structure and dynamics of PEO-salt electrolytes*, Uppsala, **1992**.
(39) Berthier, C.; Gorecki, W.; Minier, M.; Armand, M. B.; Chabagno, J. M.; Rigaud, P. *Solid State Ionics* **1983**, 11, 91.
(40) Huq, R.; Farrington, G. C.; Koksbang, R.; Tonder, P. E. *Solid State Ionics* **1992**, 57, 277-283.
(41) Neyertz, S.; Brown, D.; Thomas, J. O. *J. Chem. Phys.* **1994**, 101, 10064-10073.
(42) Brown, D.; Neyertz, S. *Mol. Phys.* **1995**, 84, 577-595.
(43) Neyertz, S.; Brown, D.; Thomas, J. O. *Electrochimica Acta* **1995**, 40, 2063-2069.
(44) Neyertz, S.; Brown, D.; Thomas, J. O. *Comp. Poly. Sci.* **1995**, 5, 107-120.
(45) Neyertz, S.; Brown, D. *J. Chem. Phys.* **1996**, 104, 3797-3809.
(46) Neyertz, S.; Brown, D. *Electrochimica Acta* **1997**.
(47) Chatani, Y.; Okamura, S. *Polymer* **1987**, 28, 1815-1820.
(48) van Krevelen, D. W. *Properties of polymers : their correlation with chemical structure; their numerical estimation and prediction from additive group contributions*, 3rd completely revised ed.; Elsevier: Amsterdam, 1990.
(49) Roe, R. J. *J. Phys. Chem.* **1968**, 72, 2013.

- (50) Chadwick, A. V.; Hanmer, P.; Coppola, L.; Ranieri, G.; Terenzi, M. *Solid State Ionics* **1994**, *72*, 147-151.
- (51) Fauteux, D.; Lupien, M. D.; Robitaille, C. D. *J. Electrochem. Soc.:Electrochemical Science and Technology* **1987**, *134*, 2761-2767.
- (52) Müller-Plathe, F. *Acta Polymer* **1994**, *45*, 259-293.
- (53) Stouch, T. R.; Alper, H. E.; Bassolino-Klimas, D. *International Journal Of Supercomputer Applications and High Performance Computing* **1994**, *8*, 6-23.
- (54) Egberts, E.; Marrink, S. J.; Berendsen, H. J. C. *European Biophysics Journal* **1994**, *22*, 423-436.
- (55) Venable, R. M.; Zhang, Y. H.; Hardy, B. J.; Pastor, R. W. *Science* **1993**, *262*, 223-226.
- (56) Sheng, Q.; Schulten, K.; Pidgeon, C. *Journal Of Physical Chemistry* **1995**, *99*, 11018-11027.
- (57) Heller, H.; Schaefer, M.; Schulten, K. *Journal Of Physical Chemistry* **1993**, *97*, 8343-8360.
- (58) Sok, R. M.; Berendsen, H. J. C.; van Gunsteren, W. F. *J. Chem. Phys.* **1992**, *96*, 4699-4704.
- (59) Pohorille, A.; Wilson, M. A. *Origins Of Life and Evolution Of the Biosphere* **1995**, *25*, 21-46.
- (60) Bassolino-Klimas, D.; Alper, H. E.; Stouch, T. R. *Journal Of the American Chemical Society* **1995**, *117*, 4118-4129.
- (61) Bassolino-Klimas, D.; Alper, H. E.; Stouch, T. R. *Biochemistry* **1993**, *32*, 12624-12637.
- (62) Forester, T. R.; Smith, W.; Clarke, J. H. R. *Journal Of Physical Chemistry* **1995**, *99*, 14418-14423.
- (63) Chiu, S. W.; Novotny, J. A.; Jakobsson, E. *Biophysical Journal* **1993**, *64*, 98-108.
- (64) Elber, R.; Chen, D. P.; Rojewska, D.; Eisenberg, R. *Biophysical Journal* **1995**, *68*, 906-924.
- (65) Woolf, T. B.; Roux, B. *Proceedings Of the National Academy Of Sciences Of the United States Of America* **1994**, *91*, 11631-11635.
- (66) Engels, M.; Bashford, D.; Ghadiri, M. R. *Journal Of the American Chemical Society* **1995**, *117*, 9151-9158.
- (67) Wong, Y. T.; Clark, T. W.; Shen, J.; McCammon, J. A. *Molecular Simulation* **1993**, *10*, 277.
- (68) Fincham, D. *Molecular Simulation* **1987**, *1*, 1-45.
- (69) Rapaport, D. C. *Phys. Rev. A* **1987**, *36*, 3288-3299.
- (70) Bruge, F.; Fornili, S. L. *Comput. Phys. Commun.* **1990**, *60*, 31-38.
- (71) Liem, S. Y.; Brown, D.; Clarke, J. H. R. *Comput. Phys. Commun.* **1991**, *67*, 261-267.
- (72) Petersen, H. G.; Perram, J. W. *Molecular Physics* **1989**, *67*, 849-860.
- (73) Raine, A. R. C.; Fincham, D.; Smith, W. *Comput. Phys. Commun.* **1989**, *55*, 13.
- (74) Pinches, M. R. S.; Tildesley, D. J.; Smith, W. *Molecular Simulation* **1991**, *6*, 51-87.
- (75) Smith, W. *Comput. Phys. Commun.* **1991**, *62*, 229-248.
- (76) Smith, W. *Theoretica Chimica Acta* **1993**, *84*, 385-398.
- (77) Plimpton, S.; Hendrickson, B. *International Journal Of Modern Physics C Physics and Computers* **1994**, *5*, 295-298.
- (78) Plimpton, S.; Hendrickson, B. *J. Comp. Chem.* **1996**, *17*, 326-337.
- (79) Bruge, F. *Comput. Phys. Commun.* **1995**, *90*, 59-65.
- (80) Tamayo, P.; Mesirov, J. P.; Boghosian, B. M. In *Proceedings of Supercomputing '91*; IEEE Computer Society Press: Albuquerque, New Mexico, 1991; pp 462-470.
- (81) Rapaport, D. C. *Comput. Phys. Commun.* **1993**, *76*, 301-317.
- (82) Kalia, R. K.; Nakano, A.; Greenwell, D. L.; Vashishta, P.; Deleeuw, S. W. *Supercomputer* **1993**, *10*, 11-25.
- (83) Hedman, F.; Laaksonen, A. *International Journal Of Modern Physics C Physics and Computers* **1993**, *4*, 41-48.
- (84) Beazley, D. M.; Lomdahl, P. S. *Parallel Computing* **1994**, *20*, 173-195.
- (85) Plimpton, S. *Journal Of Computational Physics* **1995**, *117*, 1-19.
- (86) Esselink, K.; Hilbers, P. A. J. *Journal Of Computational Physics* **1993**, *106*, 108-114.
- (87) Lin, S. L.; Pettitt, B. M.; Phillips, G. N. *Faseb Journal* **1992**, *6*, A 464-a 464.
- (88) Sato, H.; Tanaka, Y.; Yao, T. *Fujitsu Scientific & Technical Journal* **1992**, *28*, 98-106.
- (89) Janak, J. F.; Pattnaik, P. C. *Journal Of Computational Chemistry* **1992**, *13*, 1098-1102.
- (90) Debolt, S. E.; Kollman, P. A. *Journal Of Computational Chemistry* **1993**, *14*, 312-329.
- (91) Mattson, T. G.; Ravishanker, G. *Acs Symposium Series* **1995**, *592*, 133-150.
- (92) Hwang, Y. S.; Das, R.; Saltz, J. H. *Ieee Computational Science & Engineering* **1995**, *2*, 18-29.
- (93) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43-56.
- (94) Chynoweth, S.; Klomp, U.; Scales, L. *Comput. Phys. Commun.* **1991**, *62*, 297-306.
- (95) Clark, T. W.; van Hanxleden, R.; McCammon, J. A.; Scott, L. R. In *Proc. Scalable High Performance Computing Conference-94*; IEEE Computer Society Press, Los Alamitos, CA: Knoxville, TN, 1994; p 95.

- (96) Plimpton, S.; Pollock, R.; Stevens, M. In *SIAM Conference on Parallel Processing for Scientific Computing*, 1997.
- (97) Brown, D. In *Proceedings of Faraday Symposium 27, J. Chem. Soc. Farad. Trans.*, 1992; Vol. 88, p 1783.
- (98) Tuckerman, M. E.; Martyna, G. J.; Berne, B. J. *J. Chem. Phys.* **1990**, 93, 1287-1291.
- (99) Brown, D.; Minoux, H.; Maigret, B. *Comput. Phys. Commun.* **1997**, 103, 170-186.
- (100) Brown, D., The *gmx* User Manual Version 3, **1999**, available at <http://www.univ-savoie.fr/labos/lmops/brown/gmx.html>.
- (101) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.
- (102) Eisenhauer, G.; Schwan, K. *Journal Of Parallel and Distributed Computing* **1996**, 35, 76-90.
- (103) Hammonds, K. D.; Ryckaert, J.-P. *Comput. Phys. Commun.* **1991**, 62, 336-351.
- (104) Schreiber, H.; Steinhauser, O. *Chemical Physics* **1992**, 168, 75-89.
- (105) York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, 99, 8345-8348.
- (106) Lau, K. F.; Alper, H. E.; Thacher, T. S.; Stouch, T. R. *Journal Of Physical Chemistry* **1994**, 98, 8785-8792.
- (107) Delbuno, G. S.; Cohen, T. S.; Rossky, P. J. *Journal Of Molecular Liquids* **1994**, 60, 221-236.
- (108) Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. *Journal Of Physical Chemistry* **1996**, 100, 17011-17020.
- (109) Smith, W. *Comput. Phys. Commun.* **1992**, 67, 392-406.
- (110) Pollock, E. L.; Glosli, J. *Comput. Phys. Commun.* **1996**, 95, 93-110.
- (111) Hockney, R. W.; Eastwood, J. W. *Computer Simulation using Particles*; IOP: Bristol, 1987.
- (112) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, 98, 10089-10092.
- (113) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, 103, 8577-8593.
- (114) Beckers, J. V. L.; Lowe, C. P.; de Leeuw, S. W. *Molecular Simulation* **1998**, 20, 369-383.
- (115) Anfinsen, C. B. *Science* **1973**, 181, 223-230.
- (116) Zhang, X. J.; Baase, W. A.; Matthews, B. W. *Biochemistry* **1991**, 30, 2012-2017.
- (117) Geisow, M. J. *TIBTECH* **1992**, 10, 141-144.
- (118) Sauer, U. H.; Dao-pin, S.; Matthews, B. W. *Journal of Biological Chemistry* **1992**, 267, 2393-2399.
- (119) Remington, S. J.; Anderson, W. F.; Owen, J.; Ten Eyck, L. F.; Grainger, C. T.; Matthews, B. W. *J. Mol. Biol.* **1978**, 118, 81-98.
- (120) Eriksson, A. E.; Baase, W. A.; Matthews, B. W. *J. Mol. Biol.* **1993**, 229, 747-769.
- (121) Ueda, T.; Nakashima, A.; Hashimoto, Y.; Miki, T.; Yamada, H.; Imoto, T. *J. Mol. Biol.* **1994**, 235, 1312-1317.
- (122) McLeish, M. J.; Nielsen, K. J.; Wade, J. D.; Craik, D. J. *FEBS Lett.* **1993**, 315, 323-328.
- (123) Duo-Ping, S.; Soderlind, E.; Baase, W. A.; Wozniak, J. A.; Sauer, U. S.; Matthews, B. W. *Biochemistry* **1991**, 30, 7142-7153.
- (124) Hurley, J. H.; Baase, W. A.; Matthews, B. W. *J. Mol. Biol.* **1992**, 224, 1143-1159.
- (125) Seckler, R.; Jaenicke, R. *The FASEB J.* **1992**, 6, 2545-2552.
- (126) Maigret, B.; Brown, D. **1997**.
- (127) Wong, K. B.; Freund, S. M.; Fehrst, A. R. *J. Mol. Biol.* **1996**, 259, 805-818.
- (128) Nash, D. P.; Jonas, J. *Biochem. Biophys. Res. Commun.* **1997**, 238, 289-291.
- (129) Akasaka, K.; Tezuka, T.; Yamada, H. *J. Mol. Biol.* **1997**, 271, 671-678.
- (130) Zhang, J.; Peng, X.; Jonas, A.; Jonas, J. *Biochemistry* **1995**, 34, 8631-8641.
- (131) Takeda, N.; Kato, M.; Taniguchi, Y. *Biochemistry* **1995**, 34, 5980-5987.
- (132) Tanaka, N.; Nishizawa, H.; Kunugi, S. *Biochim. Biophys. Acta* **1997**, 1338, 13-20.
- (133) Stockman, B. J.; Euvrard, A.; Scahill, T. A. *J. Biomol. NMR* **1993**, 3, 285-296.
- (134) Briggs, M. S.; Roder, H. *Proc. Natl. Acad. Sci. USA* **1992**, 89, 2017-2021.
- (135) Alonso, D. O.; Daggett, V. *J. Mol. Biol.* **1995**, 247, 501-520.
- (136) Thery, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *Journal Of Computational Chemistry* **1994**, 15, 269-282.
- (137) Pearson, W. R.; Lipman, D. J. *Proc. Natl. Acad. Sci. USA* **1988**, 85, 2444.
- (138) Hynes, R. O. *Cell* **1992**, 69, 11-25.
- (139) Phillips, D. R.; Charo, I. F.; Scarborough, R. M. *Cell* **1991**, 65, 359-362.
- (140) Gould, R. J.; Polokoff, M. A.; Friedman, P. A.; Huang, T.-F.; Holt, J. C.; Cook, J. J.; Niewiarowski, S. *Proc. Soc. Exp. Biol. Med.* **1990**, 195, 168-171.
- (141) Cox, D.; Aoki, T.; Seki, J.; Motoyama, Y.; Yoshida, K. *Med. Res. Rev.* **1994**, 14, 195-228.

- (142) Scarborough, R. M.; Rose, J. W.; Naughton, M. A.; Phillips, D. R.; Nannizzi, L.; Arfsten, A.; Campbell, A. M.; Charo, I. F. *J. Biol. Chem.* **1993**, *268*, 1058-1065.
- (143) Cook, N. S.; Kottirsch, G.; Zerwes, H.-G. *Drugs of the Future* **1994**, *19*, 135-159.
- (144) Scarborough, R. M.; Rose, J. W.; Hsu, M. A.; Phillips, D. R.; Fried, V. A.; Campbell, A. M.; Nannizzi, L.; Charo, I. F. *J. Biol. Chem.* **1991**, *266*, 9359-9362.
- (145) Adler, M.; Lazarus, R. A.; Dennis, M. S.; Wagner, G. *Science* **1991**, *253*, 445-448.
- (146) Krezel, A. M.; Wagner, G.; Seymour-Ulmer, J.; Lazarus, R. A. *Science* **1994**, *264*, 1944-1947.
- (147) Senn, H.; Klaus, W. *J. Mol. Biol.* **1993**, *232*, 907-925.
- (148) Klaus, W.; Broger, C.; Gerber, P.; Senn, H. *J. Mol. Biol.* **1993**, *232*, 897-906.
- (149) Sutcliffe, M. J.; Jaseja, M.; Hyde, E. I.; Lu, X.; William, J. A. *Nat. Struct. Biol.* **1994**, *1*, 802-807.
- (150) Jaseja, M.; X., L.; Williams, J. A.; Sutcliffe, M. J.; Kakkar, V. V.; Parslow, R. A.; Hyde, E. I. *Eur. J. Biochem.* **1994**, *226*, 861-868.
- (151) Saudek, V.; Atkinson, R. A.; Pelton, J. T. *Biochemistry* **1991**, *30*, 7369-7372.
- (152) Saudek, V.; Atkinson, R. A.; Lepage, P.; Pelton, J. T. *Eur. J. Biochem.* **1991**, *202*, 329-338.
- (153) Atkinson, R. A.; Saudek, V.; Pelton, J. T. *Int. J. Pept. Protein Res.* **1994**, *43*, 563-572.
- (154) Savage, B.; Marzec, U. M.; Chao, B. H.; Harker, L. A.; Maraganore, J. M.; Ruggeri, Z. M. *J. Biol. Chem.* **1990**, *265*, 11766-11772.
- (155) Garsky, V. M.; Lumma, P. K.; Freindinger, R. M.; Pitzenberger, S. M.; Randall, W. C.; Veber, D. V.; Gould, R. J.; Friedman, P. A. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 4022-4026.
- (156) Gan, Z. R.; Gould, R. J.; Jacobs, J. W.; Friedman, P. A.; Polokoff, M. A. *J. Biol. Chem.* **1988**, *263*, 19827-19832.
- (157) Minoux, H.; Brown, D.; Maigret, B. **1997**.