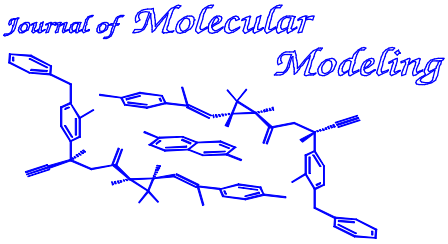
J. Mol. Model. **1995**, *1*, 123 – 142



Molecular Dynamics of the Ha-*ras* Protein: Nucleotide Atom-Centred Charges Within the AMBER Force Field.

Graham A. Worth[2], Colin Edge[1] and W. Graham Richards[*].

Oxford Centre for Molecular Sciences and Physical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, U.K.

1 SmithKline Beecham, Coldharbour Rd., The Pinnacles, Harlow, Essex, CM19 5AD, U.K.

2 Present address Theoretische Chemie, Universität Heidelberg, Im Neuenheimerfeld 253, 69120 Heidelberg, F.R.G. (gr@vax.ox.ac.uk)

Received: 31 May 1995 / Accepted: 09 June 1995

Abstract

Molecular dynamics simulations have become an essential tool for the study of biological systems. The Ha-*ras* protein, is a system suitable for such studies. Despite much recent progress, it is still not known exactly how the protein functions in the cell growth cycle. In this work atom-centred point charges for the guanosine nucleotide ligands are calculated and tested. To be compatible with the other AMBER force field parameters these are fitted to a molecular electrostatic potential derived from an *ab initio* wavefunction. The smallest basis set able to produce a stable wavefunction for the negatively charged GDP and GTP molecule ions was 3-21G* with diffuse functions added on the phosphate groups. To maintain force field integrity these charges were scaled to be equivalent to STO-3G derived values. This procedure is seen to produce a good magnesium-phosphate interaction potential when compared to 6-31++G* *ab initio* calculations. With the nucleotides fixed in the binding site conformation, it was found essential to include the electrostatics of the binding site in the calculation of the charges. It was also found to be inappropriate to divide the nucleotide into constituent parts for the calculations. From the calculated charges and experimental data, the nucleotide protonation states in the protein are deduced. It is unlikely that GDP is protonated, GTP probably binds one proton. The charges were tested in MD simulations of a protein modelled on the crystal structure of Tong *et al.* [1], during which the dynamics of the nucleotide and binding site residues were in good agreement with the crystal structure data. The model is seen to be sensitive, not only to the inclusion of explicit solvent, but to the number of waters ligating the magnesium ion and the conformation of the loop between residues 60 and 66; both pieces of information are lacking in the crystal structure data.

Keywords: Atom-centred charges, Molecular electrostatic potential, Ha-*ras* protein, guanosine nucleotides, molecular dynamics

Introduction

Classical molecular dynamics (MD) computer simulations have become increasingly important in elucidating the behaviour of biological systems such as proteins under conditions not easily examined by experimental techniques [2-6]. Using atomic parameters to reproduce the system potential energy surface (PES), Newton's equations of motion are solved numerically to produce a trajectory for the time evolution of the system. This trajectory can then reveal dynamic and structural information. A typical system for which useful information could be obtained from such simulations is the very important or Ha-*ras* protein.

This protein is a member of a family that has received a large amount of attention over recent years, primarily due to

* *To whom correspondence should be addressed*

its importance in the cell growth cycle and related oncogenesis. The proteins are membrane bound, responsible for transmitting growth factor signals from the hormone receptor to the cell interior. This process is mediated by a cycle that involves exchange of Guanosine-5'-diphosphate (GDP) and guanosine-5'-triphosphate (GTP) followed by dephosphorylation of the GTP [7, 8]. Recent work has started to shed light on the central position of these proteins in a whole range of signalling pathways [9]. However, while proteins both upfield and downfield of the *ras* protein in the signal path have been experimentally characterised [10], for an understanding of the mechanism and energetics of the cycle it is necessary to turn to theoretical methods. The cycle is a set of non-equilibrium processes, lending themselves to stucy using molecular dynamics.

A few studies of this nature have already been performed [11, 12], looking at the GTP hydrolysis mechanism. Many other parts of the cycle, such as the input needed to induce the nucleotide exchange that acts as a trigger, are still to be investigated. One major problem in this respect is the availability of structures from which such calculations can be started. While a full crystal structure (all heavy atoms) for the GTP complex due to Pai *et al*. [13] is now available in the Brookhaven databank (code 5p21), the structure for the GDP complex deposited there contains only alpha carbon and ligand heavy atoms (code 2p21) [1]. A higher resolution structure for this complex has however been published [14]. Various models of the GDP complex have also been published [15, 16]. As will be shown later, the dynamics of the protein are significantly affected by the starting structure.

Once a suitable starting structure has been generated, the problem is to choose a reasonable set of force field parameters for the simulation of the system. A typical, well-tested set of parameters for bio-molecular simulations, is provided by the AMBER force fields [17, 18]. These describe the potential energy function as a sum over terms for bond and bond angle distortions; torsional rotations, and interatomic interactions in the form of Lennard-Jones and coulombic terms. Both GDP and GTP have formal charges and so the interaction between them and the protein is likely to be dominated by the electrostatic energies. A good description of the charge distribution is thus essential for computer simulations.

The AMBER standard force field contains parameters for the RNA nucleotides. However the charges were optimised on fragments which may bear little relationship to the nucleotide in the protein binding site. This is especially true in the case of the phosphate groups, which in RNA are di-substituted monophosphates rather than mono-substituted di- or tri-phosphates of GDP and GTP. The charges on equivalent phosphate oxygens have also been averaged in the force field, which may remove an important anisotropic field within the protein binding site.

The development of a force field is a delicate balance between simplicity and accuracy. In particular the treatment of non-bonded interactions has proved to be the most sensitive part of molecular mechanics calculations. In most empirical force fields, of which AMBER is a typical example, these terms are represented as a set of point charges on the atomic centres combined with Lennard-Jones attraction and repulsion parameters. The parameters are optimised to reproduce experimental data. This corrects for differences between the electric field produced by the atom-centred charges and the more complex real molecular field, and at the same time corrects for missing energy terms such as those due to atom-atom polarisation.

In the AMBER force field, the Lennard-Jones non-bonded parameters were optimised with fixed charges and built up into a library connected with atom types to cover most chemical situations. To simulate a novel molecule, it is simply a matter of selecting the desired atom types and suitable charges, usually generated explicitly for the new molecule. Due to the optimisation process, it is necessary to use the original method for charge generation to preserve force field integrity and prevent the problems arising from having terms from different sources. For AMBER parameters, this means fitting the charges to a molecular electrostatic potential (MEP) derived from a STO-3G [19] basis set wavefunction. Using a higher basis set may produce a better description of the molecules electrostatic properties, but the atom-centred charges tend to be larger. For example in a set of quinone molecules 6-31G* derived charges [20] are on average 1.15 times as large as the corresponding STO-3G charges [21]. The Lennard-Jones terms should be re-optimised to compensate for this increased polarity. Approximate linear relationships have been found between basis set and atom-centred charges [21, 22]. It is therefore possible, in theory, to calculate charges with one method and scale to approximate those produced using a different basis set.

This problem can be seen on a practical level by a comparison using 6-31G* MEP generated charges [23] against the standard STO-3G MEP charges in AMBER simulations of DNA. Also in this case the charges from the higher basis set are more polar (see table 4 for comparisons of the two charge sets for GDP). Identical simulations on the DNA strand $d(ATATATATAT)_2$, starting in an idealised B-DNA conformation in a sphere of solvent containing placed counter ions, produce very different trajectories depending on the charges used [24]. In a trajectory using STO-3G charges, the DNA stays in a conformation predominantly described by B-DNA structural parameters, while in one using the 6-31G* charges the conformer changes to one that is better described by the experimentally less stable A-DNA structure. Another example of problems encountered with the use of 6-31G* charges in AMBER is the over estimation of the intra-molecular contribution to the free energy in a calculation of the equilibrium constant for histamine tautomerism [25].

Recently, a set of AMBER parameters has been published for simulations of the nucleotides GDP, GTP, GPPNP and GPPCP [26]. Standard Lennard-Jones parameters were taken and the charges were calculated by fitting to semi-empirical MNDO MEPs. This method, as well as being possibly incompatible with the AMBER force field, is known to be

unable to reliably treat hypervalent compounds such as sulphates and phosphates [27], which need the inclusion of d-orbitals. The situation is likely to be further complicated by the high negative charge; again something not well treated by minimal basis sets. A further problem is that these charges were calculated for the bare ions in vacuo in the protein binding site geometry, which as will be shown later may lead to spurious results. The parameters were also untested in the protein. Only a simulation in solution was made. This use of charges calculated in one geometry and used for simulations in another is known to cause problems [28, 29]. In a simulation of the ras: GTP complex, Foley *et al.* [12] also used standard AMBER parameters, but with 3-21G* [20] MEP fitted charges, subsequently scaled to mimic 6-31G* charges, for the monoprotonated GTP ion, but the conformation used is not reported and the charges were not published.

The aim of this paper is to calculate and test charges for GDP and GTP for use in molecular dynamics simulations of the Ha-*ras* protein complexes. Of greater interest to us is the GDP complex, and this will be the focus of the study; the GTP charges are merely presented as a by-product. A model of the complex was therefore built based on the available crystal structure. A protocol for the charge derivation of these large charged molecules is developed and tested on a small system. Charges for GDP in the binding site conformation are then calculated, firstly fragmenting the nucleotide, then treating the full molecule both with and without the surrounding protein. Finally some short simulations are run to check the performance of the charges against the experimental data of the crystal structure.

Methods

Unless otherwise stated, the programs AMBER3.1 [30] and GAUSSIAN88 [31] and were used for the empirical molecular mechanics and *ab initio* quantum mechanics calculations respectively. The program AMBER was chosen from the various molecular mechanics programs available because of the availability of the source code. In all calculations, the protein was represented by a united-atom model, with explicit polar hydrogen atoms; all other molecules used an all-atom description. Water molecules were described by the TIP3P model [32]. The fitting of atom-centred charges to MEPs was made using computer code written by Reynolds (RATTLER) [33]. This uses a method [34, 35] based on the CHELP procedure [36] but puts random points in a region of space rather than over surfaces.

Calculation of atom-centred charges for use with the AMBER force field.

It is desirable when developing new charges for use with the AMBER force field, that they should be fitted to an MEP derived from a wavefunction calculated using a STO-3G basis set. However this basis set is too small to handle the high charge density on the triply negative GDP ion, a species unlikely to be stable in vacuo, leaving a number of unbound electrons in SCF orbitals with positive eigenvalues. It has been found that the addition of diffuse functions, which allow the wavefunction to expand in space, are required for such situations [36]. Unfortunately, the smallest basis set available in GAUSSIAN88 with such functions is 6-31+G*, far too large to treat the 40 atom GDP molecule.

For a compromise of speed and stability, the standard 3-21G* basis set was used with sp-diffuse functions taken from the 6-31+G* set. This basis set, 3-21+G* has been tested [37]. To save computer resources, for these calculations the diffuse functions were added only on the phosphate groups (i.e. just P and O atoms). This nonstandard set will in future be designated 3-21G*+. It might be thought that the omission of diffuse functions on the other heavy atoms is reasonable as the excess electron density will be almost entirely on the phosphate groups. A similar basis set, with the addition of d-functions on the oxygen atoms, was used in a recent paper on atomic charges for PO_3^- and related XO_3^n species [38]. In that study, it was noted that even with 6-31+G* (at the Hartree-Fock level) there were unbound electrons on XO_3^{2-} and XO_3^{3-} ions. This was however not a problem for any of the molecules studied here (GDP^{3-}, GTP^{4-} and $MePO4^{2-}$), the extra heavy atoms allowing a greater spread of charge.

The charges fitted to an MEP calculated with this basis set were then scaled to mimic what the STO-3G basis set would have produced. Assuming the 3-21G*+ charges scale approximately like 3-21G values, this involves dividing all charges [21] by a factor of 1.3. In conformity with the use of diffuse functions on only the phosphate groups, the overall charge lost by the scaling was then divided evenly over the phosphate atoms.

Testing the protocol for phosphate atom-centred charges

As a small test system for these charges, a methyl phosphate ion with a magnesium ion approaching along the ether P—O bond (i.e. the oxygen bonded to the methyl group) was set up. This is shown in figure 1. The model has the advantage of a symmetry of interaction between the magnesium and terminal oxygen atoms as the intermolecular distance changes. This is helpful due to the identical nature of chemically equivalent atoms in molecular mechanics.

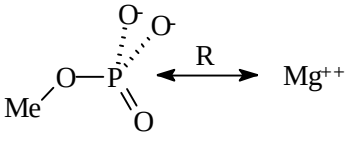


Fig. 1: *The magnesium - phosphate system used to calculate interaction potentials with various parameter sets.*

Initially a high quality ab initio interaction energy curve was calculated. The methylphosphate structure, initially set up as a random standard geometry, was optimised, at the 3-21G* level. An interaction energy curve with the magnesium ion at various values of R was then calculated using a 6-31++G* basis set, corrected for basis set superposition error (BSSE) with the Boys-Bernardi Counterpoise Correction method [39]. This correction was found to be significant for the phosphate ion as the magnesium approaches; the electron rich negative ion borrowing functions from the function rich positive ion.

This interaction energy was then recalculated using the AMBER program with various sets of parameters. The ab initio optimised methylphosphate geometry was kept to enable comparisons to be made. It was found that if the methylphosphate was energy minimised with AMBER it moved into a separate minimum which produced a different curve for the interaction energy. Interestingly this AMBER minimum corresponded to the geometry found in the GMP crystal structure [40]. Placing the magnesium at distances from the methylphosphate identical to the ab initio calculation, the ANALYSIS module was used to evaluate the interaction energy.

Building a model of the Ha-ras: GDP protein complex.

The Ha-ras: GDP complex alpha carbon coordinates produced by Tong et al. [1], the then only available Ha-ras crystal structure, were taken from the Brookhaven databank. This structure will be in future referred to as “the crystal structure”. The first stage was to turn this into a complete model. Initial modelling was done using the QUANTA / CHARMm software package [41]. A 171 residue polyalanine chain was taken. At appropriate positions alanines were changed to glycines. Using the crystal coordinates as a target, the chain was energy minimised using a linear constraint potential with a force constant of 50 kcal mol-1 to force the alpha carbons of the chain onto the experimentally measured positions. During this time the non-bonded pair list was never updated from the initial values calculated on the linear chain using a non-bond cutoff of 8 Å. This ensures atoms only interact with neighbours in the chain so allowing non-neighbouring parts of the chain to pass through one another. To ensure a reasonable backbone, the ω-torsion angles were kept planar with a stiff harmonic potential.

The model chain now had a fold similar to the crystal structure. In the next step, the alpha carbon atoms were given the exact crystal coordinates. Keeping these atoms fixed in space, the whole polyalanine / glycine protein was energy minimised, this time updating the non-bonded pair list every 100 steps, but still with the ω-torsion angles constrained to be planar. After this the alanine methyl groups were replaced with the appropriate sidechains to form Ha-ras. Charged residues are all on the protein surface, or in the exposed binding site, so protonation states were left as they are in solution. All four histidines are surface residues and so the N(τ)H tautomer, which is more abundant in solution, was chosen. This structure was then fully optimised, still keeping the alpha carbons fixed.

The GDP and magnesium ions from the crystal structure were then added, with appropriate protons, so that the model matched the primary sequence and measured x-ray coordinates. As the sidechains had been randomly added, it was now necessary to relax the model into an energy minimum where it would keep the crystal shape without constraints. A better conformation also had to be found for residues 60 to 66. These are not defined in the crystal structure and had been arbitrarily filled in by the above process.

Due to the known flexibility of this loop, it is unlikely that a database structure search would produce meaningful results. It is also likely that the surroundings play an important role in its conformation as it lies sandwiched between two more rigid loops. Conformational space was therefore sampled using high temperature molecular dynamics for all the experimentally uncharacterised regions of the model, keeping fixed the known crystal positions (i.e. the alpha carbons of residues 1- 61 and 66- 171, GDP heavy atoms and magnesium). SHAKE [42] was implemented to constrain the bond lengths to their equilibrium geometries and a time step of 0.002 ps was used. The GDP was treated with an all atom description, while the amino acids had non-polar hydrogens included in the heavy atoms.

The structure was initially quickly heated over 0.5 ps to 500 K. 20 ps of constant temperature dynamics were then run, during which the energy dropped by around 3600 kJmol-1 and remained stable for the last 5 ps. As an indication that the dynamics had searched a reasonable area of phase space during this time, the phi and psi angles of residues 60 to 66 were monitored and all varied by 80- 180° during the simulation. The lowest energy conformation found in the simulation was taken and energy minimised. Finally the constraints on the “crystal atoms” were removed and a full optimisation performed.

After this optimisation, the binding site of the model had significantly changed. The magnesium ion had moved to lie between the alpha and beta phosphates of the GDP ion, and the residues Asp33 and Asp57 had moved to bind to the magnesium, closing up the binding site and distorting the loop between residues 26 and 36, λ2. For the binding site to match the crystal structure, it was necessary to add four water molecules around the magnesium ion, as noted in the description of the high resolution crystal structure [14]. These waters were placed to provide octahedral coordination around the magnesium ion and to maximise hydrogen bonding; one lay between the magnesium ion and Asp33 and one between the magnesium ion and Asp57. The structure was then optimised, first allowing just the waters to move, and then a full optimisation. In Ha-ras: GDP formed after hydrolysis of GTP [43], Asp57 is a direct ligand to the magnesium. This indicates that the number of waters ligated to the magnesium ion may vary, and the binding site can become more or less tight depending on this.

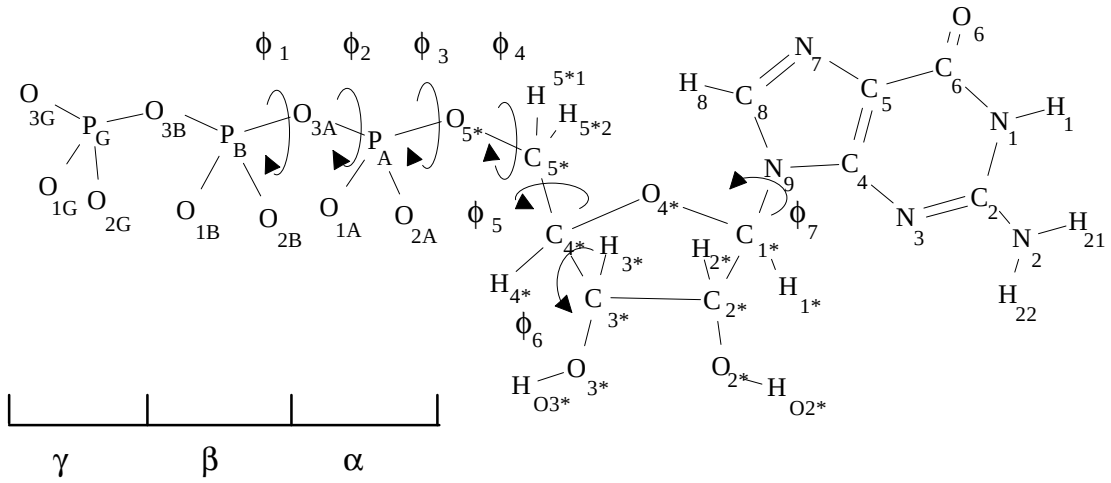


Fig. 2: *Naming protocol for the atoms in GTP and GDP. The torsion angles marked define the conformer of GDP. Values for these angles, in degrees, for the "extended" and Ha-*ras *protein binding site geometries used in the charge calculations are tabulated below.*

Torsion no.	Extended	Protein
ϕ_1	168.2	298.4
ϕ_2	166.5	260.0
ϕ_3	185.8	327.9
ϕ_4	166.2	219.1
ϕ_5	182.5	65.4
ϕ_6	35.0	29.2
ϕ_7	287.1	297.0

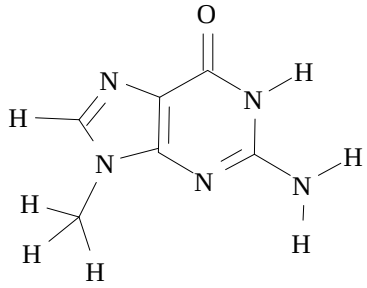
One final problem was found in the model. On full minimisation, even with the four water ligands present, the magnesium ion and the GDP β-phosphate moved to allow the magnesium to interact either with the O3B and O2B oxygen atoms or with the O3B and O1A. It is interesting to note that when the magnesium moved between the alpha and beta phosphates, the beta-phosphate rotated from its normal gauche position with respect to the alpha phosphate to an eclipsed conformation, allowing better interaction with the magnesium. When this happened the protein structure was virtually unchanged, with the exception of Gly12, which moved about 3 Å from its crystal position, and its two neighbouring residues; evidence of the connection between this residue, important in the mutagenesis of the protein, and the position of the beta-phosphate. On closer study, it was noticed that the O2B oxygen atom had no nearby positively charged ligand. During the initial modelling the loop with Gly12 and Gly13 had ended up with their backbone protons pointing away from the binding site. As Gly13 seems to be the only possible candidate for the role of ligand, this part of the chain was moved by hand to provide a better environment for the phosphate group. After a short energy minimisation of this altered segment, a full optimisation now produced a model with an RMSD of 1.67 Å compared to the crystal structure with no major deviations. A Ramachandran plot of the backbone phi and psi torsion angles showed that most of these were in the preferred regions.

Calculating atom-centred charges for GDP and GTP

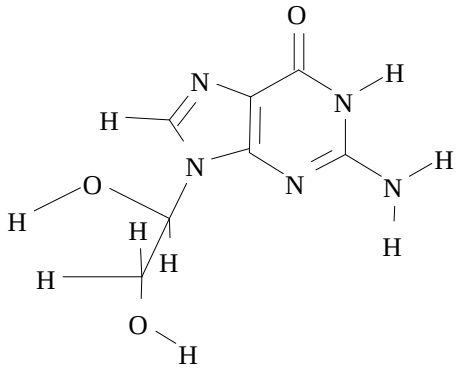
The naming convention employed for GDP and GTP is shown in figure 2. Using the scheme described above, charges for GDP and GTP were initially calculated in the same way as for the original RNA force field, splitting the nucleotide into parts for the base, sugar and phosphates. The nucleotide geometry taken as the basic structures were those from the above model based on the crystal structure. As well as the three parts used in the original study, a second set of fragments was taken one further step down the molecule as shown in figure 3.

To see the effect of treating the whole molecule and the importance of the conformation, charges were also calculated on the full nucleotide taken from the model crystal structure and a fully extended conformer of GDP. This has the phosphate chain in a minimum close to the all-trans conformation, taken to be the preferred geometry of GDP outside the protein. The flexible torsion angles for both geometries are shown in figure 2. The large size of the molecule prevented full ab initio optimisation and so the "all trans" conformer was energy minimised using the CHARMm force field. The final stage was to include the effects of the protein environment. This was done by taking all atoms in the protein model within 5 Å of the nucleotide and adding them as a set of point charges, taken from the AMBER force field, to the Hamiltonian.

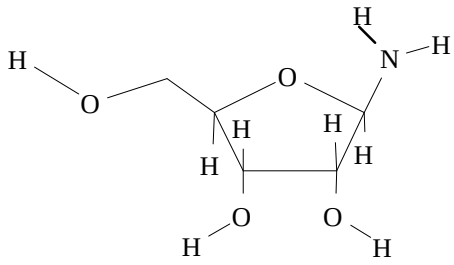
Charges were also calculated for GTP in the protein binding site. The coordinates for GPPNP complex crystal struc-



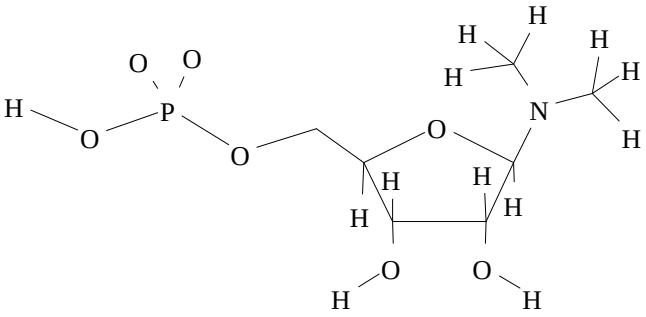
(a) Guanine 1



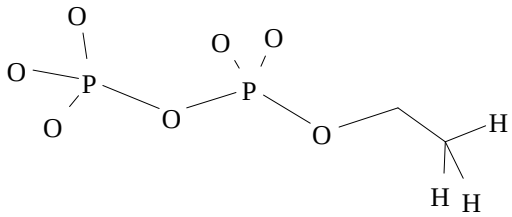
(d) Guanine 2



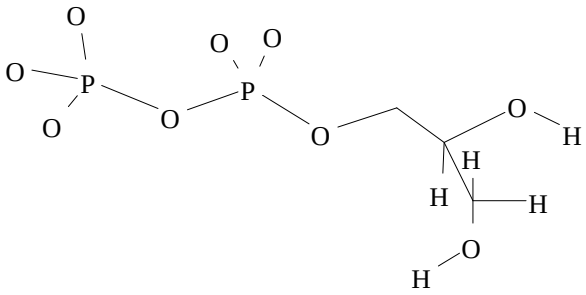
(b) Ribose 1



(e) Ribose 2



(c) Diphosphate 1



(f) Diphosphate 2

Figure 3. *Fragments used to test the effect of division schemes on calculating partial charges for GDP in the Ha-*ras *binding site conformation.*

ture [13], kindly provided by Pai, were taken; hydrogens were added using QUANTA and the nucleotide changed to GTP. It is known that there is little difference between the GPPNP and GTP bound protein structures [43], so the protein heavy atoms were fixed while the nucleotide and hydrogen atoms were minimised using CHARMm. Charges were then calculated on this GTP structure, taking the same 5 Å binding site point charge model as for GDP above.

Molecular dynamics simulations run using the new charges

The Ha-ras: GDP model was truncated to residues 1 to 166. This is known to have no effect on the biochemistry [44] and has the advantage that the protein is more spherical and so easier to solvate. The C-terminus so formed was capped with an N-methyl group, while the N-terminus, which would be in solution, was left as a charged -NH3+ group. This structure was then energy minimised using the AMBER force field. As a comparison, optimizations were made using both the 3-21G*+ derived charges and the standard AMBER STO-3G derived charges for the GDP ion.

The system was now ready to run short test simulations to check the new charges. Heating to 300 K and subsequent equilibration was performed over 10 ps, coupled to a temperature bath with the Berendsen velocity scaling algorithm [45] with a relaxation constant of 0.1 ps Both temperature and potential energy seemed stable after this. A further 50 ps of constant temperature dynamics were then run. During this time a problem was encountered in constraining the short, stiff "bonds" to the sulphur lone pairs with SHAKE. The mass of the lone pairs was increased from one to six and the sulphur decreased to twenty. This should have little effect on the sulphur motion and the vibrational frequency was lowered enough to enable SHAKE to keep the constraints. It is noticed that this problem has later been rectified in the force field used in AMBER4.0, with a different bond force constant. During the simulation, coordinates were saved every 0.2 ps This simulation is called the "crystal" simulation.

To test the sensitivity of the model to various factors, four further simulations were run. The first was to test the loop between residues 60 and 66, the regions modelled without any crystal data. This was done by taking the loop from before MD was used to search conformational space and grafting it onto the refined model structure (simulation "unrefined loop"). The second and third simulations were to investigate the importance of the waters bound to the magnesium ion for the protein dynamics. One simulation was therefore run with three water molecule ligands (simulation "3 water ligands"), corresponding to the Ha-*ras*: GDP structure captured after hydrolysis from Ha-*ras*: GTP, and one without any water molecule ligands (simulation "no water ligands"). Energy minimisation and 60 ps of dynamics in vacuo were then performed on all three new models.

Finally, the effect of adding solvation to the model was checked by the inclusion of explicit water molecules to the model. It was not possible to solvate the protein fully as a sphere of waters large enough to give a shell of solvent at least 8 Å from all solute atoms would contain at least 3000 water molecules. Instead, a 24 Å sphere was added centred on residue 9. This residue is approximately half way between the binding site and the exposed loop of residues 59- 65. This provided a shell approximately 5 Å thick around the protein on the side known to face into the cytosol, and the solvent shell decreased to around 3 Å for residues that are on the side of the protein that faces the membrane wall. This needed 1105 water molecules. Due to the small size of the solvation shell, no cap potential was used to hold the surface waters in place, but over the short period of the simulation performed, solvent evaporation was not a problem.

Fixing all the known crystallographic atoms (alpha carbons and ligand heavy atoms) and the four waters ligating the magnesium ion, energy minimisation of 100 steps of steepest descent, followed by a 15 ps simulation using the same conditions as before were run to relax the solvent and sidechains.

Simulation name	Environment	No. Mg2+ water ligands	Region 66 - 74 [a]
"Crystal"	in vacuo	4	refined
"Unrefinedloop"	in vacuo	4	unrefined
"No water ligands"	in vacuo	0	refined
"3 water ligands"	in vacuo	3	refined
"Crystal in water"	aqueous solution	4	refined

Table 1. *Summary of MD simulations run. All were based on a model of the* Ha-*ras: GDP complex. 10ps of heating and equilibration to 300K were followed by 50ps of NVT dynamics.*

[a] Refinded indicates that loop initial geometry was found by high temperature MD, unrefined indicated it was the model geometry before the high temperature MD was used to search conformation space.

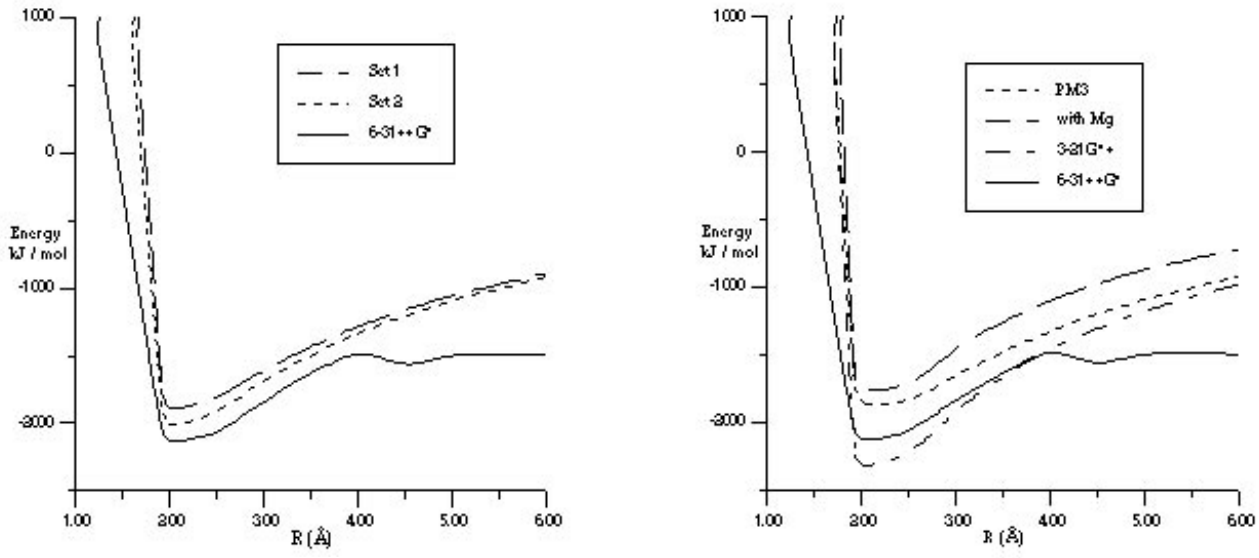


Figure 4. *Interaction potential energy between a magnesium ion and methylphosphate ion along the trajectory in figure 1. For both graphs, "6-31++G*" denotes* ab inito *energies calculated at this basis set.*
(a) "Set 1" and "Set 2" denote energies were calculated using AMBER with different Lennard -Jones parameters on the magnesium (see text for details).
(b) "3-21G+" denotes methylphosphate charges were fitted to an* ab initio *derived MEP using this basis set. "With Mg" means that the magnesium ion was explicitely included in the MEP calculation. "PM3" marks the energy calculated with charges derived from a semi-empirical MEP.*

Table 2. *Atom-centred charges used for calculating the interaction energy between a magnesium ion and a methylphosphate ion.*

Atom [a]	AMBER [b]	3-21G*+ [c]	With Mg [d]	PM3 [e]
Mg	2000	2000	1.800	2.000
O3	-0.943	-0.551	-0.609	-1.255
O2	-0.943	-0.551	-0.609	-1.255
O1	-0.943	-0.551	-0.609	-1.255
P	1.337	0.136	-0.323	2.300
O*	-0.748	-0.350	-0.073	-0.675
C	-0.015	0.206	-0.176	0.170
H3	0.085	0.056	0.199	-0.010
H2	0.085	0.056	0.199	-0.010
H1	0.085	0.056	0.199	-0.010

[a] Atom naming is with chemically equivalent atoms numbered, and O represents the ether oxygen.*
[b] Charges from the standard force field [18].
[c] Charges fitted to an ab initio *MEP calculated with the 3-21G*+ basis set.*
[d] Charges also fitted to a 3-21G+ basis set MEP, but magnesium ion was explicitely included in the calculation.*
[e] Charges fitted to a PM3 derived MEP.

Table 3. *Atom-centred charges fitted to a 3-21G*+ MEP calculated on GDP from the* Ha-ras*: GDP crystal structure and fragments of that structure. GDP atom names are those in figure 2 and the fragments are those shown in figure 3. An H in front of a charge shows that there was a terminating hydrogen in place of the GDP atom named.*

	guanine fragment 1	**guanine fragment 2**	**ribose fragment 1**	**ribose fragment 2**	**diphosphate fragment 1**	**diphosphate fragment 2**	**3-21G*+ Protein [a]**
O3B					-1.815	-0.129	-0.162
O2B					-1.543	-1.353	-1.659
O1B					-0.124	-0.715	-1.526
PB				H0.639	-0.643	-0.523	-0.771
O3A				-0.352	-0.354	-0.456	-0.556
O2A				-0.074	-0.042	-0.736	-0.956
O1A				-0.773	-0.110	-0.767	-0.944
PA			H0.420	-0.955	1.386	1.815	1.977
O5*			-0.681	0.053	-0.406	-0.556	-0.690
C5*			0.196	-0.320	0.554	-0.169	0.019
H5*1			0.072	0.294	-0.028	0.222	0.121
H5*2			0.013	0.193	0.045	0.116	0.097
C4*		H0.356	0.010	0.137	-0.437	0.283	0.137
H4*			0.104	0.225	H0.177	0.115	0.140
O4*	H0.131	-0.674	-0.523	-0.056	H0.155	-0.718	-0.504
C3*		H0.097	0.336	0.200	H0.183	0.172	0.272
O3*			-0.755	-0.733		-0.740	-0.676
HO3*			0.454	0.513		0.479	0.420
H3*			0.014	0.017		0.086	0.196
C2*	H0.112	-0.629	0.163	0.297		H0.069	-0.010
O2*		-0.678	-0.751	-0.742			-0.658
HO2*		0.157	0.426	0.447			0.455
H2*			0.118	0.038			0.222
C1*	-0.265	0.589	0.680	0.519		H0.485	0.431
H1*	0.165	0.045	0.020	0.086			0.099
N9	-0.104	-0.332	-1.167	-0.239			-0.716
C8	0.160	0.267	H0.403	-0.352			1.027
H8	0.155	0.123		0.080			0.072

Table 3. *Atom-centred charges fitted to a 3-21G*+ MEP calculated on GDP from the* Ha-ras *(continued).*

	guanine fragment 1	guanine fragment 2	ribose fragment 1	ribose fragment 2	diphosphate fragment 1	diphosphate fragment 2	3-21G*+ Protein [a]
N7	-0.532	-0.553		H0.143			-0.537
C5	-0.116	-0.045		H0.143			0.128
C6	0.866	0.831					0.850
O6	-0.650	-0.646					-0.323
N1	-0.899	-0.882					-0.943
H1	0.440	0.440					0.519
C2	1.055	1.047					1.141
N2	-1.077	-1.077					-0.918
H21	0.456	0.456		H0.102			0.499
H22	0.457	0.459		H0.168			0.511
N3	-0.776	-0.775		H0.124			-0.683
C4	0.423	0.430	H0.448	-0.314			0.896

*[a] GDP in the Ha-*ras*: GDP complex model geometry based on the crystal structure.*

At the end of this time a second short energy minimisation was performed. The same 60 ps simulation as run previously 10 ps of equilibration followed by 50 ps with data collection) was then made. This simulation is referred to as the "crystal in water" simulation. The simulation details are summarised in table 1.

Results

Comparisons of the methylphosphate-magnesium ion interaction energy

A comparison of the interaction energy curve between a magnesium ion and a methylphosphate ion calculated *ab initio* (6-31++G* basis set, BSSE corrected) and calculated using two sets of available AMBER parameters is shown in figure 4(a). The phosphate non-bonded parameters for both AMBER calculations were taken from the standard RNA all atom set. The parameters for the methyl group were from the serine-carbon, which is next to an electronegative oxygen, in the all-atom protein force field. The difference between the calculations was the choice of Lennard-Jones parameters used for the magnesium ion, which had a charge of 2+. One calculation used the parameters from the original force field (set 1) and the other values from Cieplak *et al.* [46] (set 2).

It can be seen that the curves follow the *ab initio* curve very well, but are higher in energy. The second set of magnesium parameters is seen to perform better and was taken in all subsequent calculations. The steepness of the r^{12} repulsion energy used by the AMBER potential is noticeable, as is the increasing divergence of the molecular mechanics and *ab initio* curves at large distances. Both these regions will not be sampled during simulations inside the protein and so this is not a significant problem.

Figure 4(b) compares the same *ab initio* interaction energy curve with AMBER calculations using three different charge sets with the set 2 AMBER Lennard-Jones terms. One set of charges was derived by fitting to a 3-21G*+ derived MEP, averaging over the terminal oxygens and scaling by a factor of 1.3 to mimic STO-3G charges. The total ionic charge lost by this scaling was divided over the five atoms of the phosphate group. The second set was to study the effect of including the magnesium in the charge fitting procedure. Charges were again fitted to a 3-21G*+ MEP, but this time from the wavefunction of the supermolecule with the magnesium at the *ab initio* minimum energy distance shown in figure 4(a). The same averaging and scaling as above was

Table 4. *Atom centre charges for the GDP and GTP molecular ions. Atom names are those in figure 2. The sums at the bottom of the table are the totals on the nucleotide component groups.*

	AMBER STO-3G [a]	**AMBER 6-31G* [b]**	**3-21G*+ Extended [c]**	**3-21G*+ Protein [c]**	**3-21G*+ Protein inc. site [d]**	**3-21G*+ Protein inc. site [d]**
O3C						-0.978
O2C						-1.028
O1C						-1.278
PC						1.401
O3B	-0.981	-1.054	-1.049	-0.162	-1.194	-0.727
O2B	-0.981	-1.054	-1.049	-1.659	-1.011	-0.742
O1B	-0.981	-1.054	-1.049	-1.526	-0.927	-0.715
PB	1.222	2.049	1.697	-0.771	1.440	1.465
O3A	-0.623	-1.055	-0.771	-0.556	-0.748	-0.833
O2A	-0.850	-0.997	-0.865	-0.956	-0.744	-0.807
O1A	-0.850	-0.997	-0.865	-0.944	-0.777	-1.028
PA	1.429	2.049	1.422	1.977	1.211	1.477
O5*	-0.509	-0.763	-0.563	-0.690	-0.398	-0.178
C5*	0.180	0.197	0.140	0.019	-0.209	-0.049
H5*1	0.008	0.012	0.049	0.121	0.185	0.121
H5*2	0.008	0.012	-0.015	0.097	0.142	0.005
C4*	0.100	0.300	-0.068	0.137	0.075	-0.106
H4*	0.061	0.028	0.143	0.140	0.101	0.010
O4*	-0.343	-0.564	-0.409	-0.504	-0.549	-0.188
C3*	0.303	0.310	0.302	0.272	0.149	0.585
O3*	-0.509	-0.822	-0.606	-0.676	-0.538	-0.629
HO3*	0.306	0.438	0.351	0.420	0.341	0.287
H3*	0.007	0.018	0.030	0.196	0.119	-0.111
C2*	0.101	-0.003	0.156	-0.010	0.066	0.019
O2*	-0.546	-0.704	-0.576	-0.658	-0.522	-0.424
HO2*	0.324	0.466	0.345	0.455	0.353	0.287

Table 4. *Atom centre charges for the GDP and GTP molecular ions (continued)*

	AMBER STO-3G [a]	AMBER 6-31G* [b]	3-21G* + Extended [c]	3-21G*+ Protein [c]	3-21G*+ Protein inc. site [d]	3-21G*+ Protein inc. site [d]
H2*	0.008	0.026	0.033	0.222	0.094	0.140
C1*	0.117	0.240	0.284	0.431	0.616	-0.063
H1*	0.054	0.033	0.068	0.099	-0.017	0.000
N9	-0.042	-0.010	-0.018	-0.716	-0.175	-0.010
C8	0.266	0.134	0.199	1.027	0.135	0.155
H8	0.046	0.153	0.194	0.072	0.114	0.127
N7	-0.543	-0.547	-0.474	-0.537	-0.441	-0.403
C5	-0.060	0.000	0.061	0.128	0.028	0.020
C6	0.690	0.741	0.591	0.850	0.646	0.713
O6	-0.458	-0.602	-0.484	-0.323	-0.537	-0.477
N1	-0.729	-0.831	-0.701	-0.943	-0.771	-0.735
H1	0.336	0.422	0.336	0.519	0.426	0.417
C2	0.871	1.062	0.868	1.141	0.888	0.850
N2	-0.778	-1.120	-0.806	-0.918	-0.863	-0.754
H21	0.325	0.471	0.326	0.499	0.414	0..238
H22	0.339	0.471	0.326	0.511	0.329	0.366
N3	-0.709	-0.766	-0.641	-0.683	-0.628	-0.739
C4	0.391	0.312	0.167	0.896	0.176	0.318
Phosphate[e]	-2.615	-2.113	-2.259	-4.597	-2.750	-3.792
Ribose [f]	-0.330	-0.776	-0.335	0.074	0.009	-0.293
Guanine [g]	-0.055	-0.110	-0.136	1.523	-0.259	0.162

[a] Charges from the AMBER force field [18].
[b] From Hausheer et al. *[23]*
[c] Charges calculated from a 3-21G+ MEP on the "extended" and protein binding site geometries defined in figure 2.*
[d] Charges calculated from a 3-21G+ MEP on the nucleotides from the protein models including the binding site atoms within 5Å from the nucleotide as point charges.*
[e] Atoms O3C to PA inclusive.
[f] Atoms O5 to H1* inclusive.*
[g] Atoms N9 to C4 inclusive.

performed. As a contrast, a final set of charges fitted to an MEP from a PM3 wavefunction, calculated with the program MOPAC5.0 [47], and averaged over the terminal oxygens without scaling is also shown.

The 3-21G*+ charges (calculated without the magnesium ion present) produce a curve with a lower minimum energy than the *ab initio* curve. Considering the lack of electron correlation in the quantum mechanical calculation, this is a better potential than obtained using the original AMBER charges indicating the greater ability of the 3-21G*+ basis set to represent the electron density on a phosphate ion. The

inclusion of the magnesium ion into the charge calculation might be thought to represent perfectly the electronic distribution at the minimum energy geometry. However the curve in figure 2(b) shows that this is not the case and the energy is much too high. Hence while the modelled electron density might be closer to the *ab initio* calculation, the non-bond parameters do not produce enough Lennard-Jones energy to compensate for the transfer of charge density onto the magnesium ion. The interaction using unscaled PM3 charges also compares badly to the *ab initio* calculation, with a curve much too high in energy. All the charges used above are shown in table 2.

Atom-centred charges for the GDP and GTP molecule ions

The results for the fragments formed from the GDP nucleotide taken from the model crystal structure are shown in Table 3. The charges calculated on the full structure are given as a comparison. The values vary considerably in the boundary regions depending on the dividing scheme. Hence calculating charges on constituent fragments is not applicable for the molecule in this conformation and it is necessary to calculate charges on the whole.

Table 4 shows the results of the charge calculations on the complete GDP and GTP molecules. The charges calculated for GDP in vacuo are given for the extended conformation described in the methods section and the binding site conformation. Charges for both GDP and GTP calculated in the binding site, with the environment included as point charges are also given. For comparison the standard AMBER charges, both from STO-3G and 6-31G* MEPs are also given. These have been modified by smoothing the extra ionic charge over the phosphate groups. The new charges from the 3-21G*+ MEP have been scaled as before, and in the extended conformer charges for chemically equivalent atoms have been averaged over, with the exception of the H5*1 and H5*2 atoms which are significantly different. When compared to the standard AMBER STO-3G charges, the 3-21G*+ extended charges are much closer than the 6-31G* charges, having root mean square deviations (RMSD) of 0.11 and 0.22 respectively over all the charges. The binding site electrostatics thus allow the electron density of the GDP molecule to be similar to that found in other situations, supporting the concept of chemical groups.

It is obvious that the conformer is important, very different charges being produced. In fact simulations using the charges calculated on GDP in the binding site conformation, but without including the binding site charges produced large deviations from the crystal structure. Due to the large charges on the phosphate groups, the magnesium ion always moved to lie between the two phosphates, with a rotation of the beta-phosphate. This was found not to occur for the charges calculated including the binding site. This problem is also seen in solution phase simulations of Cannon [26] with MNDO charges calculated on the binding site conformer. In that study, GTP exhibited a preference for the C2' exo sugar pucker; something not observed experimentally, attributed to forces between C5' and N9, mediated by H5'1 and H8. The charges in this region are seen here to be most sensitive to the conformation chosen for the calculation.

Protonation state of GDP and GTP in the Ha-ras binding site

One final problem needs to be solved before the parameters can be used in simulations; the protonation state of the nucleotides. Ignoring the ribose hydroxyl groups and the base carbonyl groups, GDP has three protonation sites on the phosphate chain and one on the guanine amine group. Likewise GTP has four phosphate protonation sites and the amine. With such molecules a knowledge of the equilibrium constants for the various sites measured in the relevant environment is needed for an accurate evaluation of the relative population of the states. This information is rarely available for groups inside proteins. A recent molecular dynamics study of the Ha-*ras*: GTP complex used the monoprotonated species, but it was not stated where or by what criterion the proton was added [12].

In aqueous solution, the proton equilibria have been measured and standard values for the GDP dissociation constants are listed in table 5, along with those for the related pyrophosphoric acid, $H_4P_2O_7$. Comparing the nucleotide values with the phosphoric acid data, it seems plausible that the values above 9 are not due to the phosphates but the base or ribose. From pK_1 and pK_2 for GDP, one dissociation constant is for the phosphates and one for the protonated amine group on the base. This may seem a very low basicity for an amine, but it should be remembered that it is conjugated to the guanine ring system. As a comparison, the anilinium ion-aniline equilibrium has a pK_a of 4.6. It is known empirically that on the same centre, pK_a's differ by around 5 units. Hence the first two phosphate protons come from different phosphates.

Table 5. *The acid dissociation constants for GDP, GTP and the related polyphosphoric acids.*

GDP [a]	**$H_4P_2O_7$**	**GTP [a]**	**$H_5P_3O_{10}$**
			$pK_{a\,1}$ = <1
	$pK_{a\,1}$ = 2.5		$pK_{a\,2}$ = 2.2
$pK_{a\,3}$ = 2.9	$pK_{a\,2}$ = 2.7	$pK_{a\,4}$ = 3.3	$pK_{a\,3}$ = 2.6
$pK_{a\,4}$ = 6.3	$pK_{a\,3}$ = 6.0	$pK_{a\,5}$ = 6.3	$pK_{a\,4}$ = 5.6
$pK_{a\,5}$ = 9.6	$pK_{a\,4}$ = 8.3	$pK_{a\,6}$ = 9.3	$pK_{a\,5}$ = 7.9

[a] taken from Dissociation Constants for Organic Bases in Aqueous Solution, *Butterworth, 1965.*

The fraction of the molecule present as the jth protonated species is given by

$$\alpha_j = \frac{[H_jA]}{[A_{TOT}]} = \beta_j [H^+]^j$$

where $_j$ is the formation constant of the species

$$jH^+ + A > H_jA$$
$$\beta_j = K_1 K_2 .... K_j \qquad (2)$$

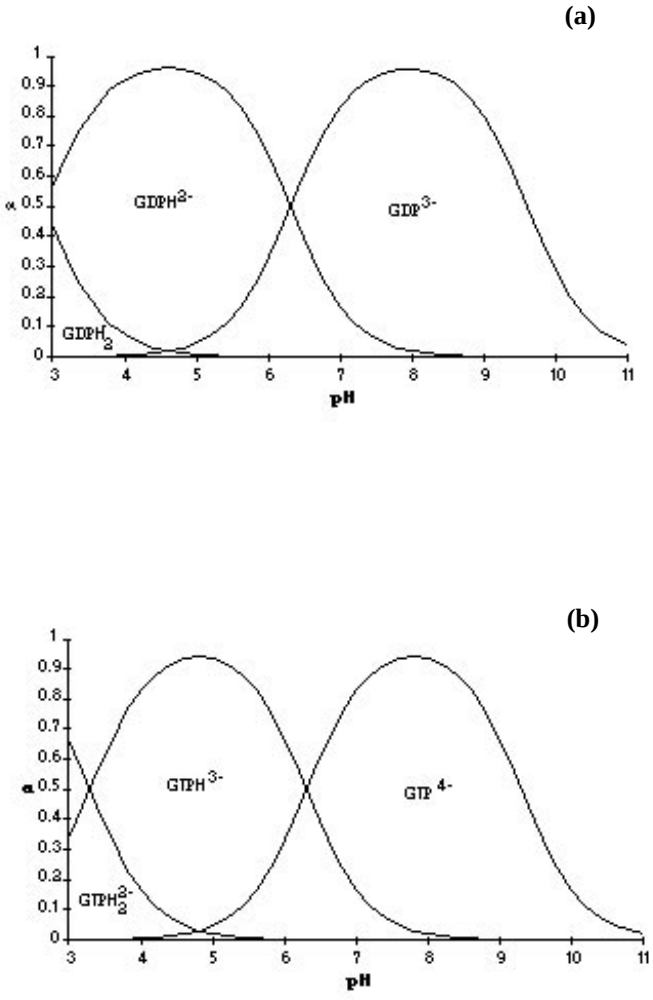


Figure 5. *The fraction, α, of different protonated forms of (a) GDP and (b) GTP present in solution around pH7. The numbers of hydrogens shown are the number of protons on the phosphate groups.*

which can be related to the stepwise association constants, the inverse of the dissociation constants given in table 5

$$H^+ + H_{j-1}A > H_jA \qquad K_j = \frac{1}{K_{a,n-j}} \qquad (3)$$

The ratios for the GDP and GTP species with different number of phosphate protons present around pH 7 are plotted in figure 5. Hence GDP and GTP both have on average between 0 and 1 protons on the phosphate groups.

In the protein binding site, the environment is of course different from that in solution. A good guess can however be made by comparing the charges calculated on the "solution phase" extended structure and in the binding site. For GDP, there is no reason for protonation. All oxygens have virtually identical charges to the extended "solution phase" structure, hence the acidity (electron density) is not enhanced by the binding site. Also from inspection of the crystal structure, there is no obvious advantage to be gained from protonation, all the oxygens are close to the magnesium ion or a hydrogen bond donor. On examining the GTP charges, two oxygens have comparatively high electron density, O1A and O1C. In the crystal structure, the former is 1.7 Å from the backbone amide proton of Ala18. The latter however is not particularly close to any hydrogen bond donors (the peptide proton of Thr35 is 3.2 Å away) and so it is quite possible that this oxygen is protonated. Further calculations with a proton in this position are needed to clarify this.

The performance of the new charges

Comparing the protein crystal structure to the corresponding atoms in the energy minimised model structures, there is an RMSD of 0.62 Å when using the 3-21+G* charges, and an RMSD of 0.65 Å with the standard charges. Figure 6 shows the difference in the absolute deviations of the two minimised structures from the crystal structure, defined as deviation (standard charges) - deviation (3-21+G* charges). In general there is a tendency for positive differences, meaning the standard charge calculation has larger deviations from the crystal structure than when using the 3-21+G* charges. Major differences are noticeable for the GDP phosphate oxygens and residues 14, 15, 32, 33 and 59 which are in the binding site near the phosphate groups. These results indi-

Figure 6. *(next page) Difference in deviation of alpha carbon atoms from their crystal structure co-ordinates after energy minimisation of the Ha-ras:*
GDP model using two different charge sets. Difference, in Å, is defined as deviation (STO-3G) – deviation(3-21G+). Atom numbers 1 to 166 are the alpha carbon atoms for these residue numbers, atom numbers 167 to 194 are the GDP heavy atoms in the order given in table 3, atom number 195 is the magnesium ion.*

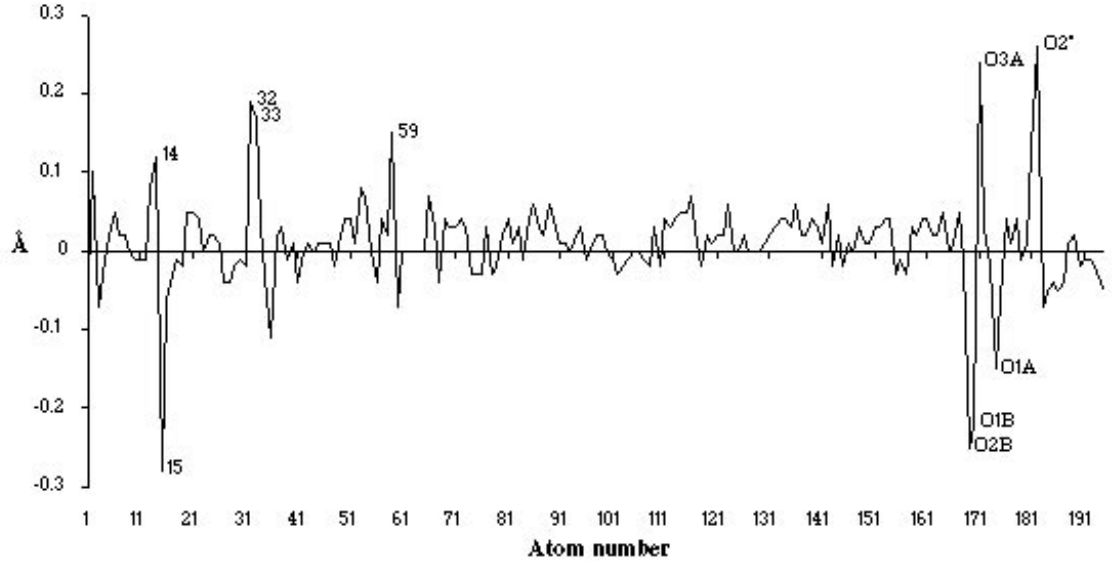


Table 6. *The regions of secondary structure (alpha helices and beta sheets only) of the model* Ha-ras: *GDP protein complex compared to the experimental structure.*

structural unit	Crystal structure [a]	model after EM [b, c]	model after MD [b, d]
β1	2 – 9	2 – 7	3 – 10
α1	16 – 25	16 – 22	17 – 24
β2	37 – 46		38 – 42
β3	49 – 58	51 – 56	52 – 57
α2	66 – 68	70 – 73	67 – 68 3-turn
			71 – 73 5-turn
β4	77 – 83	77 – 80	77 – 81
α3	87 – 103	87 – 103	92 – 103
β5	111 – 116	111 – 115	111 – 116
α4	127 – 136	127 – 136	127 – 135
β6	141 – 143	140 – 145	141 – 144
α5	152 – 165	152 – 164	153 – 165

*[a] The length of structural elements for the Ha-*ras*: GDP complex are nowhere reported and so these values are taken from the structurally similar, excepting the region 66- 74, Ha-* ras*: GPPNP crystal structure [50].*

[b] Calculated using the Kabsch and Sander method [48], as implemented in the program QUANTA.

[c] Values for the model structure after energy minimisation with the AMBER force field and 3-21G+ GDP charges*

[d] Values for the model structure after 50ps of molecular dynamics (simulation "crystal").

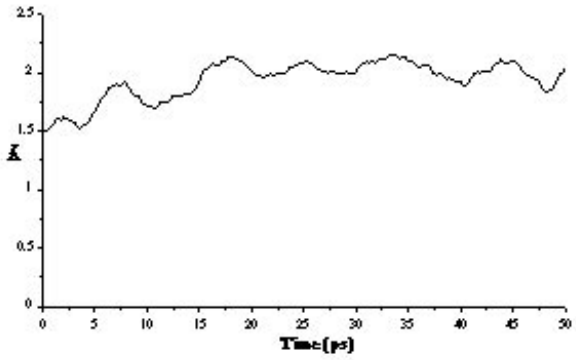


Figure 7. *RMSD of Ha-ras: GDP model alpha carbon atoms compared to the crystal structure during a 50ps in vacuo MD simulation.*

cate that the whole protein, but especially the binding site geometry is sensitive to changes on the GDP charges.

Figure 7 shows the change in RMSD of the model alpha carbon atoms compared to the known crystal structure atoms over the 50 ps simulation. During this time, the RMSD converges to around 2 Å Table 6 compares the protein structural units before and after the simulation, analysed by the Kabsch and Sanders method [48], with the elements in the crystal structure. It can be seen that during the simulation the elements change slightly and correspond better to the crystal structure afterwards, for example residues 38 to 42 are now a definite beta sheet. Of particular interest is the region between residues 66 and 74. In the GTP bound form this is a triple turn alpha helix, while the GDP crystal structure indicates that this is a single turn helix between 70- 73 (from the model of fitted to the crystal structure). After the simulation, this region is now characterised by two turns, but a significant rearrangement would be necessary before the full alpha helix seen in the GTP structure is possible.

The deviation of the simulation average structure from the crystal structure is shown in figure 8(a). In general the deviations are less than 2 Å The larger deviations are for surface regions, probably due to the lack of crystal environment in the simulation; the large deviation at residue 60 could however indicate poor modelling of the 60- 66 loop. In particular it can be seen that the nucleotide and magnesium ion are very close to the crystal structure geometry. Table 7 shows the RMSD for various parts of the protein; the whole protein, ligands (GDP and magnesium ion), binding site residues (as defined in table 9) and the major structural elements (alpha helices and beta sheets as defined in table 8, the remaining protein defined as loops), from which it is clear that the major deviations are in the surface loops.

In addition, the variance (average square deviation) of the atom coordinates from the average during the simula-

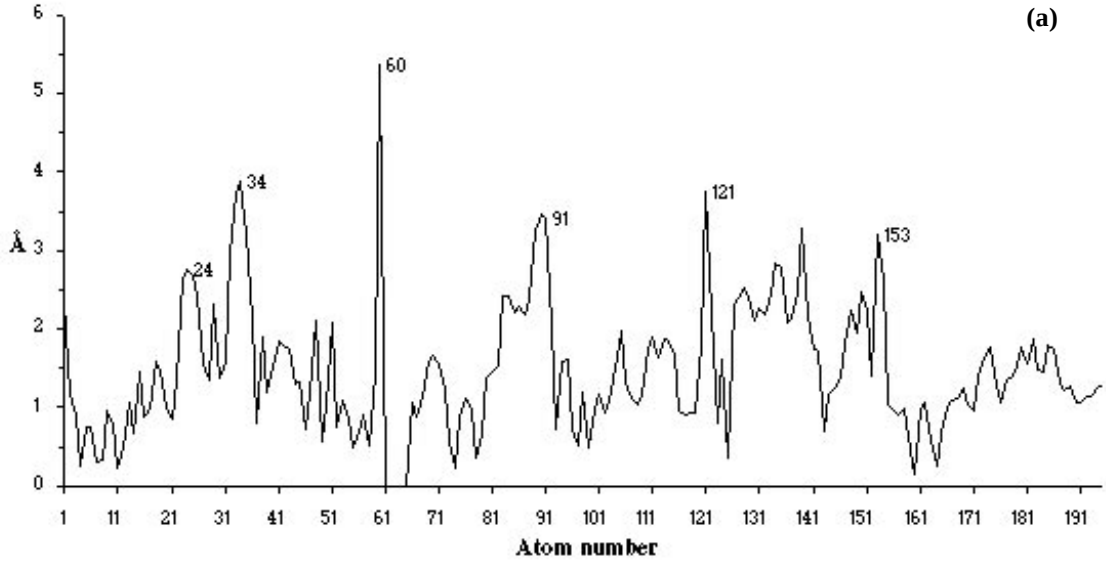


*Figure 8. Comparison of behaviour of Ha-*ras*: GDP model with the crystal structure. Atom numbers 1 to 166 are the alpha carbon atoms for these residue numbers, atom numbers 167 to 194 are the GDP heavy atoms in the order given in table 3, atom number 195 is the magnesium ion.*

(a) Positive deviation of the atoms in the 50ps simulation average structure from their crystal structure positions.
*(b) Difference in B-factors for the Ha-*ras *atoms, defined as experimental - calculated.(next page)*

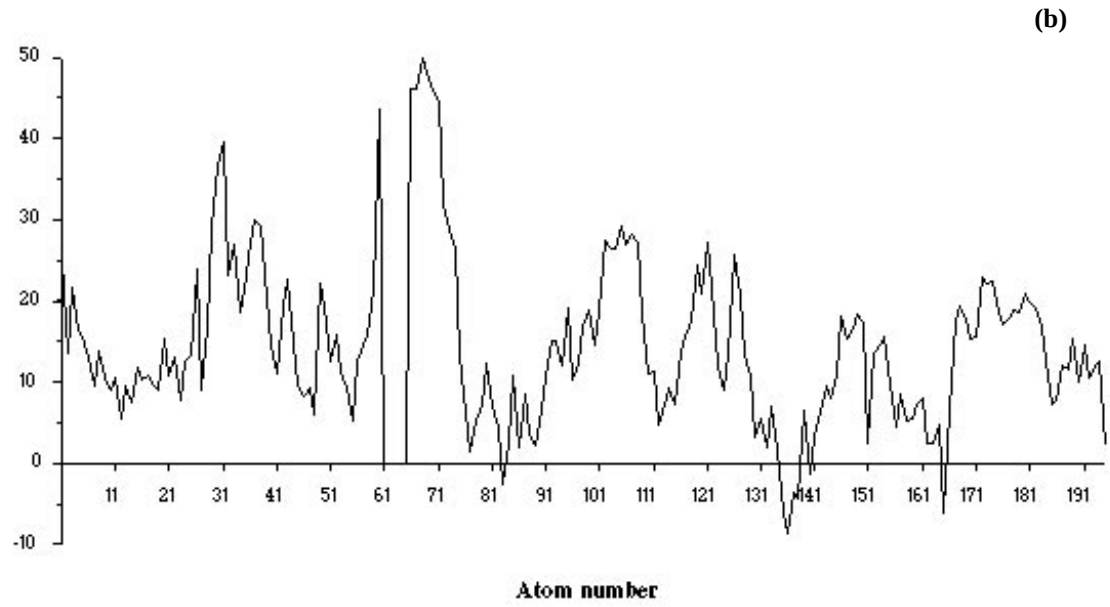


Table 7. *Comparison of simulations with crystal structure*

	"Crystal"	"Unrefined loop"	"No water ligands"	"3 water ligands"	"Crystal in water"
Modulus of the average absolute deviation from the crystal structure (Å)					
Total [a]	1.72	1.88	2.14	2.01	1.68
ligands [b]	1.39	1.14	1.78	2.19	1.29
binding site [c]	1.66	2.12	2.07	2.02	1.74
helix [d]	1.83	1.78	2.03	1.97	1.55
sheet [e]	1.29	1.50	1.39	1.64	1.48
loops [f]	1.99	2.41	2.74	2.19	2.05
Average difference in B-factors; Experimental - calculated (Å)					
Total [a]	14.69	15.17	16.44	14.36	14.28
ligands [b]	15.51	14.39	15.73	14.85	14.16
binding site [c]	14.49	12.83	13.08	15.81	13.46
helix [d]	11.81	12.50	14.60	12.21	12.65
sheet [e]	12.33	12.79	13.33	10.95	9.67
loop [f]	18.55	19.58	20.60	18.45	19.06

[a] All atoms present in the crystal structure.
[b] GDP heavy atoms and magnesium ion.
[c] Binding site residues, listed in table 8.
[d] Alpha helices, as listed in column2 table 6, ignoring the region 66- 74.
[e] Beta sheets, as listed in column2 table 6.
[f] All residues not defined as helices or sheets.

Residue number	Function
17	Binds Mg
35	Binds gamma phosphate and Mg in GTP protein
59 – 61	Bind gamma phosphate in GTP
10 – 17	Bind beta phosphate
18	Binds alpha phosphate
29, 117	Bind ribose
28, 116 – 119, 145 – 147	Bind guanine
12, 13 and 61	Mutation sites enabling oncogenesis
32 – 40	Bind GAP protein (regulates dephosphorylation)
58	Important for auto-phosphorylation
60 – 76	Bind Y13-259 antibody (prevents activation)
66 – 74	Loop with single turn helix in GDP protein andtriple turn helix in GTP protein

*Table 8. Summary of residues with known functional importance in the Ha-*ras *proteins bound to GDP and GTP.*

tion were calculated. These are related to the experimentally determined thermal, or B, factors by the equation [49]

$$B = (8^2/3)\langle(r)^2\rangle \quad (4)$$

where $\langle(r)^2\rangle$ is the variance of the displacement with respect to the average position. Differences in calculated and experimental B-factors are plotted in figure 8(b). Again deviations away from the average difference of 14.7 Å, which can be taken as the crystal lattice disorder factors ignored in equation (4), are exhibited by the surface residues, and in particular the region around loop 60- 66. The averages of the differences in B-factors, given for the different parts of the protein, are summarised in table 7.

The four further simulations showed that the dynamics of the protein are sensitive to the starting structure. The results for the absolute deviation of the average structure from the crystal structure and the difference in experimental and calculated B-factors are again shown in table 7 divided into the various protein parts. Using the unrefined loop 60- 66, the deviations relative to the crystal structure have all increased. In comparison to the above "crystal" simulation, residue 60 is as far away from its crystal position, with a deviation of 5.2 Å compared to 5.4 Å, but residue 66 and the following sequence has moved considerably, now the deviation is 3.7 Å, compared to a deviation of 1.1 Å in the simulation "crystal". The ligands have not appreciably changed geometry, but both the binding site and the surface loops have been affected. In particular the region around the GAP binding loop (residues 32-40) has moved further away from the crystal structure; in the simulation "crystal", the RMSD for residues 26 to 31 and 32 to 40 are 1.3 Å and 1.5 Å respectively, while the corresponding RMSDs for this simulation "unrefined loop" are 4.4 Å and 2.6 Å

Changing the number of waters attached to the magnesium ion, also produce larger deviations in all parts of the protein from the crystal structure. With three water ligands, the magnesium ion moves 2.9 Å from its binding site and Asp33 becomes a new ligand, causing a distortion of the binding site, which produces further changes throughout the protein. The important residue Gly12 is now 2.4 Å away from its crystal site, compared to the 0.5 Å deviation with four ligand waters. The RMSD for residues 26 to 31 and 32 to 40 have also increased to 2.5 Å and 1.7 Å respectively. With no water ligands, the magnesium ion moves to sit between the two phosphate groups. Asp33 and Asp57 have now both become magnesium ion ligands. This generally produces greater deviations from the crystal structure.

The inclusion of explicit solvent molecules in the simulation "crystal in water" reduces the overall deviation and in particular the deviations for the ligands and helices, both regions sensitive to electrostatic environment. In general, the deviations of the surface residues have also been reduced, especially residue 60 which now has a deviation of 3.3 Å The binding site has however opened up slightly by loop $_2$ (residues 26 to 36) moving out into the solvent. This leads to larger deviations for the residues Gly12 and Gly13, which are both now 3.4 Å away from their crystal positions. A change is also seen in the GAP binding region, with residues 26 to 31 and loop 32 to 40 now having RMSDs of 1.5 Å and 3.6 Å respectively.

Conclusion

A set of atom-centred charges for the nucleotide ligands of the Ha-*ras* protein compatible with the AMBER force field has been calculated. Simple tests on the GDP charges show that these produce a model which during simulations is compatible with the crystal structure. The method used and the simulations run have produced a number of interesting observations.

The use of a basis set larger than STO-3G is necessary to handle the delocalised phosphate groups, and in particular diffuse functions are needed to handle the large charge density. Scaling to mimic STO-3G derived charges can then be used to make these charges compatible with the AMBER force field. Further, it is necessary to include the binding site electrostatics to produce reasonable charges for the binding site conformation. Using this protocol, the atom-centred charges for GDP in the Ha-*ras* binding site are similar to those taken from standard AMBER values, leading to good compatibility with the other parameters. The protein model is however sensitive to the exact charges on the phosphate groups, and using the 3-21G*+ scaled charges produces a model structure closer to the crystal structure than the standard values.

During a simulation, the model behaves in agreement with the crystal structure, both structurally and from the B-factors. Major deviations are due to ignoring the crystal environment. In particular the ligand and binding site are found to reproduce the experimental structure in a satisfactory manner, showing the applicability of the new charges.

Further simulations have shown that the model is also sensitive to the exact starting structure and environment. For example, the binding site is sensitive to the number of waters binding to the magnesium ion and changing the number of ligands from four to three particularly affects residue 12, a known oncogenic site. The binding site is also sensitive to the presence of solvent. The GAP binding loop is influenced by external solvent and the geometry of loop 60- 66, which is undefined in the crystal structure. As these regions are of particular interest for the mechanism of the protein and hence to simulations, care must be taken in the choice of initial model and conditions to produce sensible results. Further work is still needed to study this important and interesting protein system, but the ground work presented here should provide the basis for genuinely revealing simulations.

Acknowledgements

GAW thanks SmithKline Beechams for funding with a CASE studentship and to Rebecca Wade (EMBL, Heidelberg) in whose group this work was finished. Thanks are also due to professor E.F. Pai of the Max-Planck-Institut für medizinische Forschung, Heidelberg, for providing the Ha-*ras*: GTP complex crystal structure coordinates on request.

References

1. Tong, L.; Milburn, M.V.; DeVos, A.M.; Kim, S.-H. *Science* **1989**, *245*, 244.
2. McCammon J.A.; Harvey S.C. *Dynamics of proteins and nucleic acids*; Cambridge University Press, 1987.
3. Brooks III, C.L.; Karplus, M.; Pettitt, B.M. *Proteins: A theoretical perspective of dynamics, structure and thermodynamics*; John Wiley and sons, 1988 and *Adv. Chem. Phys.* **1988**, *71*.
4. vanGunsteren, W.F.; Berendsen, H.J.C. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992.
5. vanGunsteren, W.F.; Mark, A.E. *Eur. J. Biochem.* **1992**, *204*, 947.
6. vanGunsteren, W.F. *Curr. Op. Struct. Biol.* **1993**, *3*, 277.
7. Bourne, H.R.; Sanders D.A.; McCormick, F. *Nature* **1990**, *348*, 125.
8. Lowry, D.R.; Willumsen, B.M. *Ann. Rev. Biochem.* **1993**, *62*, 851.
9. Egan, S.E.; Weinberg, R.A. *Nature* **1993**, *365*, 781.
10. Boguski, M.S.; McCormick, F. *Nature* **1993**, *366*, 643.
11. Langen, R.; Schweins, T.; Warshel, A. *Biochemistry* **1992**, *31*, 8691.
12. Foley, C.K.; Pedersen, L.G.; Charifson, P.S.; Darden, T.A.; Wittinghofer, A.; Pai, E.F.; Anderson, M.W. *Biochemistry* **1992**, *31*, 4951.
13. Pai, E.F.; Krengel, U.; Petsko, G.A.; Goody, R.S.; Kabsch, W.; Wittinghofer, A. *EMBO J.* **1990**, *9*, 2351.
14. Tong, L.; DeVos, A.M.; Milburn, M.V.; Kim, S.-H. *J. Mol. Biol.* **1991**, ***217***, 503.
15. Dykes, D.C.; Brandt - Rauf, P.; Luster, S.M.; Chung, D.; Friedman, F.K.; Pincus, M.R. *J. Biomol. Struct. and Dyn.* **1992**, *9*, 1025.
16. Stouten, P.F.W.; Sander, C.; Wittinghofer, A.; Valencia, A. *FEBS Letts*. **1993**, *320*, 1.
17. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta Jr., S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
18. Weiner, S.J.; Kollman, P.A.; Nguygen, D.T.; Case, D.A. *J. Comp. Chem.* **1986**, *7*, 230.
19. Hehre, W.J.; Ditchfield, R. Pople, J.A. *J. Chem. Phys.* **1969**, *51*, 2657.
20. Binkley, J.S.; Pople, J.A. *J. Chem. Phys.* **1977**, *66*, 879.
21. Lister, S.G.; Reynolds, C.A.; Richards, W.G. *Int. J. Quant. Chem.* **1992**,. *41*, 293.
22. Merz, K.M. *J. Comp. Chem.* **1992**, *13*, 749.
23. Hausheer, F.H.; Singh, U.C.; Palmer, T.C.; Saxe, J.D. *J. Am. Chem. Soc.* **1990**, *112*, 9468.
24. Haworth, I.; Elcock, A. *Private communication*, **1991**.
25. Worth, G.A.; Richards, W.G. *J. Am. Chem. Soc.* **1994**, *116*, 239.
26. Cannon, J.F. *J. Comp. Chem.* **1993**, *8*, 995.
27. Stewart, J.J.P. *J. Comp-Aided Mol. Design* **1990**, *4*, 1.
28. Essex, J.W.; Reynolds, C.A.; Richards, W.G. *J. Am. Chem. Soc.* **1992**, *114*, 3634.

29. Reynolds, C.A.; Essex, J.W.; Richards, W.G. *Chem. Phys. Lett.* **1992**, *199*, 257.
30. Singh, U.C.; Weiner, P.K.; Caldwell, J.; Kollman, P.A. AMBER version 3.1., Dept. of Pharmaceutical chemistry, University of California, San Francisco, CA 94143, U.S.A., 1988.
31. Frisch, M.J.; Head - Gordon, M.; Schlegel, H.B.; Raghavarchi, K.; Binkley, J.S.; Gonzalez, C.; Defrees, D.F.; Fox, D.J.; Whiteside, R.A.; Seeger, R.; Melius, C.F.; Baker, J.; Martin, R.; Kahn, L.R.; Stewart, J.J.P.; Fluder, E.M.; Topiol, S.; Pople, J.A. GAUSSIAN88, Gaussian inc., Pittsburgh, P.A., U.S.A., 1988.
32. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
33. RATTLER, Oxford Molecular limited, Magdalen centre, Oxford Science Park, Sandford - on thames, Oxford, OX4 4GA, U.K.
34. Ferenczy, G.; Reynolds, C.A.; Richards, W.G. *J. Comp. Chem.* **1990**, *11*, 159.
35. Reynolds, C.A.; Ferenczy, G.; Richards, W.G. *J. Mol. Struct. (Theochem)* **1992**, *256*, 249.
36. Chirlian, L.E.; Francl, M.M. *J. Comp. Chem.* **1987**, *8*, 894 and references cited therein.
37. Spitznagel, G.W.; Clark, T.; von R. Schleyer, P. *J. Comp. Chem.* **1992**, *8*, 1109.
38. Evleth, E.M.; Kassab, E.; Colonna, F.; Akacem, Y.; Ouamerli, O. *Chem. Phys. Lett.* **1992**, *199*, 513.
39. Boys, S.F.; Barnardi, F. *Mol. Phys.* **1970**, *19*, 553.
40. Emerson, J.; Sundralingham, M. *Acta Cryst. Sect. B.* **1980**, *36*, 1510.
41. QUANTA / CHARMm version 3.2., Polygen corporation, 200 Fifth Ave., Waltham, MA 02154, U.S.A.
42. Ryckaert, J.P.; Ciccoti, G.; Berendsen, H.J.C. *J. Comp. Phys.* **1977**, *23*, 327.
43. Schlichting, I.; Almo, S.C.; Rapp, G.; Wilson, K.; Petratos, K.; Lentfer, A.; Wittinghofer, A.; Kabsch, W.; Pai, E.F.; Petsko, A.; Goody, R.S. *Nature* **1990**, *345*, 309.
44. John, J.; Schlichting, I.; Schiltz, E.; Rösch P.; Wittinghofer, A. *J. Biol. Chem.* **1989**, *264*, 13086.
45. Berendsen, H.J.C.; Postma, J.P.M.; vanGunsteren, W.F.; Di Nola, A.; Haak, J.R. *J. Chem. Phys.* **1984**, *81*, 3684.
46. Cieplak, P.; Lybrand T.P.; Kollman, P.A. *J. Chem. Phys.* **1987**, *86*, 6393.
47. Stewart, J.J.P. MOPAC5.0, Frank J. Seiler Research Lab., U.S. Air Force Academy, Colorado 80840, U.S.A.
48. Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
49. Willis, B.T.M.; Pryor, A.W. *Thermal Vibrations in crystallography*; Cambridge University Press, 1975.
50. Pai, E.F.; Kabsch, W.; Krengel, U.; Holmes, K.C.; John, J.; Wittinghofer, A. *Nature* **1989**, *341*, 209.