Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An automatic method to generate force-field parameters for hetero-compounds

Kristina Nilsson,[a] David Lecerof,[b] Emma Sigfridsson[a] and Ulf Ryde[a]*

[a]Department of Theoretical Chemistry, Chemical Centre, University of Lund, PO Box 124, S-221 00 Lund, Sweden, and [b]Department of Molecular Biophysics, Chemical Centre, University of Lund, PO Box 124, S-221 00 Lund, Sweden

Correspondence e-mail: ulf.ryde@teokem.lu.se

Received 5 March 2002
Accepted 20 November 2002

A program, *Hess2FF*, has been developed that automatically constructs parameter and topology files to be used in crystallographic refinement for any molecule, based on a Hessian (force-constant) matrix estimated by any method. The program is tested by redefining hetero-compounds in five different proteins: the inhibitor *N*-methylmesoporphyrin bound to ferrochelatase, the haem group and its axial ligands in cyctochrome c_{553}, the active-site metal ion in iron superoxide dismutase, the catalytic zinc ion in alcohol dehydrogenase with a bound trifluoroethanol molecule and the 5′-deoxyadenosyl group in methylmalonyl coenzyme A mutase. It is shown that the resulting structures are improved in several aspects. In particular, the free $R_{\rm free}$ factor always decreases and it is shown that a 1.70 Å structure of cyctochrome c_{553} becomes more similar to a high-resolution (0.97 Å) structure of the same protein after re-refinement with *Hess2FF*. Thus, the force field used in crystallographic refinement significantly affects the final structure and therefore should be published together with the structure to ensure reproducibility. Various methods of obtaining the Hessian matrix employed by *Hess2FF* are discussed and some recommendations are given. *Hess2FF* allows the user to divide the atoms of the molecule into atom types that share the same force-field parameters. However, it seems to be favourable to assign a separate type to each atom, which can be performed automatically.

1. Introduction

X-ray crystallography is the major source of structural information, especially for large biomolecules such as proteins and DNA. An initial model built into an electron-density map usually contains many errors (Kleywegt & Jones, 1995*a*). To produce an accurate model, one must carry out several cycles of crystallographic refinement and rebuilding (Kleywegt & Jones, 1997). Refinement programs change the model (coordinates, occupancies, *B*-factors *etc.*) to improve the fit of observed and calculated structure-factor amplitudes. Many different refinement programs exist (Kleywegt & Jones, 1995*a*, 1997), but most contemporary programs use reciprocal-space methods. Because of the limited resolution typically obtained for biomolecules, the experimental data are usually supplemented by some sort of chemical information, typically in the form of a molecular-mechanics force field consisting of ideal bond lengths and angles, force constants of bonds, angles and torsions and the periodicity of dihedral angles.

For the normal amino acids in a protein, accurate force fields exist which are based on statistical analysis of small-molecule data (Engh & Huber, 1991). However, for unusual molecules, such as metal centres, substrates, inhibitors *etc.*, *i.e.*

hetero-compounds, experimental data is often partly lacking or is much less accurate. In particular, force-field parameters for hetero-compounds are typically not available, so the crystallographer has to construct them. This is a complicated and error-prone procedure which may make parts of the crystal structure less well determined (Kleywegt & Jones, 1998). It is also a serious bottleneck in high-throughput crystallography. Moreover, the force-field parameters are rarely tabulated or discussed when the structure is published, so the results are not reproducible.

Kleywegt and Jones have recognized this problem and set up the hetero-compound information centre at Uppsala (HIC-Up), which is a database of hetero-compounds available in old PDB files (Kleywegt & Jones, 1998). Moreover, the site provides the program *XPLO2D*, which automatically produces topology and parameter files for a compound, given a set of coordinates (Kleywegt & Jones, 1995*b*). Used wisely, this database and program can be powerful instruments for the design of a force field for hetero-compounds. On the other hand, there is also a risk that indiscriminate use of them will propagate errors in a hetero-compound from one structure to newer structures, possibly with new mistakes introduced by erroneous use of the *XPLO2D* output.

The *XPLO2D* program uses coordinates as the only necessary input and calculates from them reasonable ideal bond lengths, angles and torsions together with an appropriate topology file. However, the corresponding force constants are not available from the coordinates and are therefore given a set of standard values: 4184 kJ mol^{-1} Å^{-2} for bond lengths, 2092 kJ mol^{-1} rad^{-2} for angles and 3140 kJ mol^{-1} for proper and improper dihedral angles of constrained (*e.g.* aromatic) systems. Other dihedrals are assigned a force constant of 0, *i.e.* they are ignored.

In this paper, we design and test an alternative method to automatically generate topology and parameter files for crystallographic refinement of hetero-compounds. We base our force field on a calculation of the Hessian matrix (*i.e.* the second derivative of the energy with respect to the coordinates), which is easily obtained with any modern theoretical chemistry software. From this matrix, force constants can be extracted using the method suggested by Seminario (1996). We show that such a method may improve the crystal structure and that the interpretation of the final structure will be affected by the choice of the force field. Moreover, we test the level of theory that is necessary to obtain a good force field and how to best select atom types.

2. Methods

2.1. An automatic method to generate topology and parameter files

We have written a program, *Hess2FF*, that generates topology and parameter files from a Hessian matrix of a molecule calculated by any theoretical method. The input to the program consists of the coordinates, the Hessian matrix and, optionally, atom names and atom types. In the present version, the program automatically extracts data from *GAUSSIAN*, *SPARTAN* and *TURBOMOLE* output files (Frisch *et al.*, 1998; Spartan, 1997; Ahlrichs *et al.*, 2000). The output of the program is a topology and a parameter file. At present, only one output format is supported, *viz.* that of *Crystallography and NMR System* (*CNS*; Brünger *et al.*, 1998). In addition, the program writes a PDB coordinate file and three *CNS* input files: a file to generate the molecular-topology file, a file to run an energy calculation with the force field on the original structure (including the output of all large deviations from ideality) and a file for the calculation of an optimized structure using the new force field. These should be used to test the performance of the force field.

The program is based on the algorithm suggested by Seminario (1996). It considers the 3×3 submatrices of the Hessian matrix involving pairs of atoms. If there is a bond between two atoms, the corresponding submatrix has three positive eigenvalues, indicating that a restoring force will counteract any displacement from the equilibrium position of these two atoms. Using standard methods of geometry and algebra, force constants for any internal coordinate (bond, angle, dihedral angle, improper torsion *etc.*) can be calculated from the eigenvalues and eigenvectors of these submatrices. Such a procedure has the advantage of being automatic and fully invariant with respect to the choice of internal coordinates.

If a file with atom types and atom names is given, the program calculates the average value of all bonds, angles, dihedrals and force constants of each type. The statistics of the variation are written to a separate log file. This file should be carefully checked to see that the choice of atom types does not give rise to a large variation (and therefore inaccurate values) of the parameters. Ignoring some atoms (*e.g.* H atoms) in the force field is also allowed. If no atom-type file is provided, the program automatically assigns a separate type to each atom.

Bonds and angles are assigned a harmonic potential, as is employed in the *CNS* energy function,

$$V(x) = k_x(x - x_0)^2, \tag{1}$$

where x is the actual bond length or angle, x_0 is the ideal bond or angle and k_x is the corresponding force constant. For dihedral angles, the program may either use a similar harmonic potential or a trigonometric potential of the form

$$V(\varphi) = k_\varphi[1 + \cos(n\varphi + \delta)], \tag{2}$$

where φ is the actual dihedral angle, n is the periodicity of the torsion, k_φ is the force constant and δ is a phase shift. The force constant is determined by identifying the second-order term in the Taylor expansion of (2) with the harmonic force constant, *i.e.* $k_\varphi = 2k_x/n^2$. Moreover, the force constant is scaled down by the product of the number of atoms bound to the second and third atom in the dihedral angle (atoms excluded from the force field are not counted), because this is the number of interactions that change when the central bond is rotated. The harmonic force constant is given as a comment in the parameter file, together with the number of bonded atoms, so that

the force constant can easily be recalculated if the user decides to use another periodicity.

Similarly, the periodicity of the torsion is determined from the number of atoms bound to the second and third atom in the dihedral (this time also counting excluded atoms). If both atoms have two, three or four neighbours, a periodicity of one, two or three is used, respectively. If the two atoms have a different number of neighbours, the highest number of neighbours is used to determine the period, except when they have three and four neighbours, in which case a period of six is used. This simple algorithm works for most organic molecules, but fails for more complicated systems, *e.g.* metal complexes. The periodicity then has to be determined by hand. On the other hand, we will see below that dihedrals involving metal ions are normally not important to the structure and can therefore safely be ignored. This is the default behaviour for *Hess2FF* for torsions involving atoms with more than four neighbours.

Once the periodicity is decided, the phase shift δ is determined from the actual values of the torsion angle encountered in the structure. The program only considers phase shifts of 0 and 180°. Optionally, dihedral torsion terms involving central atoms that are sp^3 hybridized (which normally have small force constants) can be ignored. Another option is to describe the geometry of planar groups by improper torsions (described by a harmonic or cosine function). In the present version, improper dihedrals are defined and calculated only for atoms with exactly three neighbours.

van der Waals parameters of the atoms are taken directly from the *CNS* parameter file (protein_rep.param) on an atomic basis. C atoms are divided into three types depending on the number of H atoms attached. Following the approach of Engh and Huber, all force constants are scaled up by a factor of 3, making them comparable to force constants derived by statistical analysis of crystallographic data (Engh & Huber, 1991).

The Seminario algorithm is ambiguous in that the 3×3 submatrix of each pair of atoms can be selected in two different ways [the (i, j) submatrix is in general not identical to the (j, i) submatrix]. This problem is not discussed in the original paper (Seminario, 1996), but typically leads to an uncertainty in the force constants of less than 5%. We have solved the problem by taking the average of force constants obtained by the various possible selections. A measure of the uncertainty resulting from this ambiguity is printed in the log file. The program *Hess2FF* is available from the authors on request.

2.2. Hessian calculations

Hess2FF can in principle use a Hessian matrix and a molecular geometry obtained by any method, whether theoretical or experimental. In order to find an appropriate compromise between speed and accuracy, we tested five different levels of theory to optimize the structure and calculate the Hessian matrix.

Firstly, we used the three-parameter hybrid density-functional method B3LYP (Barone *et al.*, 1996) as implemented in the *GAUSSIAN*98 software (Frisch *et al.*, 1998). This method was combined with the medium-sized basis set 6-31G* (Hehre *et al.*, 1986). For metal ions, the double-ζ basis set of Schäfer *et al.* (1992), enhanced with *p*-, *d*- and *f*-type functions (exponents 0.162, 0.132 and 0.39 for Zn and 0.134915, 0.041843, 0.1244 and 1.339 for Fe with two *p* functions), was used. B3LYP is widely recognized as one of the best density-functional methods in general terms, providing good geometries, energies and frequencies at a modest cost (Bauschlicher, 1995; Siegbahn & Blomberg, 2000). It is applicable to molecules of up to about 100 atoms. As customary, the frequencies were scaled by a factor of 0.963 (*i.e.* the force constants were scaled by the square of this factor; Rauhut & Pulay, 1995). The time required for the B3LYP frequency calculations for the studied systems ranged from 20 h (5′-deoxyadenosyl) to 18 d (*N*-methylmesoporphyrin; MMP).

Secondly, we used the semiempirical AM1 method, also with *GAUSSIAN*98 (Dewar *et al.*, 1985). In semiempirical methods, all integrals that are needed to solve the Schrödinger equation are replaced by empirical parameters. This makes such methods about 1000 times faster than B3LYP, but also less accurate (Stewart, 1990). With AM1, the geometry and Hessian can be calculated for most small- and medium-sized molecules within a few hours. The frequency calculation for MMP, the largest molecule investigated here, took 35 min. For the metal-containing systems we instead used the semiempirical PM3 method, which also contains parameters for transition-metal ions (Spartan, 1997). In general terms, PM3 often gives slightly better results than AM1, but it is harder optimized to the calibration set of molecules and may therefore be dangerous to use for molecules that differ from those included in this set (Stewart, 1990).

Thirdly, we tested three different molecular-mechanics methods. In these, electrons are ignored and atoms are treated as classical balls connected with springs. The interactions between the atoms, both intermolecular and intramolecular, are described by an empirical force field of different degrees of sophistication. With molecular-mechanics methods, molecules with over 10 000 atoms can be studied and the force-constant matrix can be estimated for any hetero-compound within 1 h. All Hessian matrices in this paper could be calculated in less than 1 min. We tested three different force fields. The Dreiding (Mayo *et al.*, 1990) and UFF (Rappé *et al.*, 1992) force fields, available in *GAUSSIAN*98, are universal force fields with parameters for all elements. They can therefore be used for any molecule, providing mostly reasonable geometries but quite poor energies (Gundertofte *et al.*, 1996). In addition, we used the Merck molecular force field MMFF94 (Halgren, 1996) as implemented in *SPARTAN*5.0 (Spartan, 1997). This force field contains many more parameters and has been shown to be one of the most accurate force fields for conformational energies (Gundertofte *et al.*, 1996).

At first, it may seem strange to use molecular-mechanics calculations to obtain another molecular-mechanics force field (to be used in the refinement). Of course, direct use of the

original parameters in *CNS* would give a similar (and probably better) result. However, these force fields are not implemented in *CNS* and contain several terms not available in *CNS*, so the parameters could not be directly transferred to *CNS*. Moreover, *Hess2FF* can use a Hessian matrix obtained by any method (not only molecular mechanics) and it provides the necessary translation and sets up the needed files automatically.

For all methods, the geometry was optimized with the same method before the Hessian matrix was calculated by a frequency calculation. Default convergence criteria were employed for each program used.

2.3. *CNS* calculations

For each hetero-compound, theoretical method and set of atom types, we used *Hess2FF* to construct *CNS* parameter and topology files for the hetero-compound. For the other part of the protein, the standard *CNS* force field was used (protein_rep.param, water_rep.param, ion.param and dna-rna_rep.param). These data were then used to optimize the structure of the compound in the protein using the standard *CNS* script minimize.inp. The number of minimization steps was set to 200 and we used the default maximum-likelihood refinement target using amplitudes (*MLF*; Pannu & Read, 1996; Adams *et al.*, 1997). For the other entries, we used the data from the PDB file or the default values. After this minimization, the *B* factors of the hetero-compound were optimized with the *CNS* script bindividual.inp, using the same parameters but only 20 steps of optimization (default). We also tried to optimize the coordinates and *B* factors of the whole protein with *CNS*, but this did not improve the *R* factors. The final coordinates of the hetero-compounds, as well as the *CNS* topology and parameter files for the best calculation, are provided as supplementary material.[1]

2.4. Proteins

The performance of *Hess2FF* was tested by constructing force fields for hetero-compounds in five different crystal structures. Coordinates, occupancies, *B* factors and structure factors were downloaded from the Protein Data Bank. From these files we also obtained the space group, unit-cell parameters, resolution limits, *R* factors and the test set used for the evaluation of the free $R_{\rm free}$ factor.

The most thorough tests were performed on the inhibitor *N*-methylmesoporphyrin (MMP) in its complex with the enzyme ferrochelatase at 1.90 Å resolution. The calculations were based on the PDB file 1c1h (Lecerof *et al.*, 2000). The original topology and parameter files for MMP and the $Mg(H_2O)_6$ complex also present in the structure were obtained directly from the authors. This is not necessary, but it makes comparison with the published results easier. The full protein was used in all calculations, but for simplicity we excluded the alternative configuration of residues 33 and 120–122 (we retained the *A* conformation). This omission did not change the $R_{\rm free}$ factor, but the standard *R* factor increased from 0.181 to 0.183.

The force field obtained directly from *Hess2FF* performed well but not perfectly: the tilt of the *N*-methylated pyrrole ring was larger and the methyl group was out of the ring plane in the quantum-chemical structure, but not in that obtained by the *Hess2FF* force field. The reason for this discrepancy is that the tilt of the ring is determined by competition between the torsion force constants and the van der Waals parameters of all atoms involved in this ring. The methyl group does not fit into the central cavity of the porphyrin. Therefore, the methylated ring has to tilt out of the porphyrin plane, although this is unfavourable for the aromaticity of the system. Such competition is quite hard to describe in a molecular-mechanics simulation and this is the reason why the result was not fully satisfactory. The structure could be improved by increasing the van der Waals radius of the methyl C atom; an increase of about 1 Å gave the best result.

As a consequence of this competition between aromaticity and van der Waals repulsion, several torsions involving the methylated pyrrole ring give rise to large energies in the optimum structure. Therefore, we have not changed any torsion of the porphyrin ring and we have also assumed that all torsion in the porphyrin ring should ideally be planar, although they are not so in the optimized structure.

Next, we performed test calculations on the structure of the haem group and its axial methionine and histidine ligands in *Bacillus pasteurii* cyctochrome c_{553}. The Hessian was calculated for Fe^{III}(porphyrin)(imidazole)$[S(CH_3)_2]^+$ in the low-spin (doublet) state. Thus, the side chains of the haem group were not included in the calculations and the protein residues were truncated. This was necessary to make the quantum-chemical calculation possible (the B3LYP calculation still took 8 d). The side chains were described by the standard *CNS* force field (CH3E, CH2E, CH1E, C and OC atom types). Likewise, standard *CNS* parameters were used for the protein ligands, whereas the parameters involving iron and the porphyrin ring were obtained by *Hess2FF*. For the refinement, we used the 1.70 Å structure 1b7v (Benini *et al.*, 2000). The results of the re-refinement were examined by investigating how well the structure fitted the electronic density of the atomic resolution (0.97 Å) data of the same protein obtained by the same authors (PDB code 1c75; Benini *et al.*, 2000). To this end, the coordinates had to be rescaled to the slightly different unit-cell parameters of the latter structure by a transformation forth and back to fractional coordinates.

The third test case was the structure of alcohol dehydrogenase at 2.0 Å resolution (PDB code 1axe; Bahnson *et al.*, 1997). We constructed a force field for the catalytic zinc ion with the ligands encountered in the protein (trifluoroethanol, one histidine and two cysteine residues). The trifluoroethanol molecule was considered to be deprotonated, following experimental consensus (Pettersson, 1987). Thus, the force-constant matrix was calculated for the Zn^{II}(imidazole)-$(SCH_3)_2(OCH_2CF_3)^-$ complex. Standard parameters were used for the protein ligands, whereas parameters involving the

[1] Supplementary data have been deposited in the IUCr electronic archive (Reference: gr2257). Services for accessing these data are described at the back of the journal.

zinc ion and the trifluoroethanol molecule were extracted from the Hessian matrix. The force field for the coenzyme NAD^+ was taken from earlier simulations of alcohol dehydrogenase (Ryde, 1995). These Amber parameters were translated to *CNS* and the force constants were multiplied by a factor of 3, following the approach of Engh & Huber (1991). The structural zinc ion (with four cysteine ligands) was treated with only non-bonded (standard van der Waals) terms. The coordinates and *B* factors of NAD^+ and the structural zinc ion were not optimized in this investigation. Both catalytic sites in this dimeric protein were optimized simultaneously.

Next, we constructed a force field for the active iron site of *Sulfolobus solfataricus* iron superoxide dismutase. This site contains an iron ion bound to one aspartate and three histidine residues. The trigonal bipyramidal coordination geometry is completed by a solvent molecule. Following a thorough investigation (with quantum refinement; Ryde *et al.*, 2002) of the possible protonation states of this site, it was modelled as a hydroxide ion which is hydrogen bonded to the non-bonded and protonated O^{δ} atom of the Asp ligand (Nilsson & Ryde, manuscript in preparation). Thus, the frequency calculation was performed on $Fe^{III}(imidazole)_3(CH_3COOH)(OH)^{2+}$ in the high-spin sextet state. As in the other calculations on metal ions, the protein ligands were treated with standard *CNS* parameters, whereas parameters involving the iron ion were obtained using *Hess2FF*. For practical reasons, special atom types had to be defined for the coordinating atoms of two of the histidine ligands (two of them are equatorial ligands, whereas the third is an axial ligand, so the optimum angles are different for the three histidine ligands). However, all parameters for these new atom types were only copied from those of the standard histidine force field. Both active sites in the dimeric molecule were optimized simultaneously. The calculations were based on the PDB file 1sss (2.3 Å resolution; Ursby *et al.*, 1999).

Finally, we constructed a force field for 5′-deoxyadenosyl bound to the enzyme methylmalonyl coenzyme A mutase. The full 5′-deoxyadenosyl molecule was included in the frequency calculations. For the protein, we used the 2.2 Å structure 4req (Mancia *et al.*, 1999). The force field for the cobalamin coenzyme was obtained from an Amber force field (Marques *et al.*, 2001), converted in the same way as for NAD^+ above. For the succinate coenzyme A product molecule, standard *CNS* parameters were used. These molecules were *not* optimized. Both 5′-deoxyadenosyl molecules in the crystal were optimized simultaneously.

3. Results

3.1. MMP in ferrochelatase

As a first test of the performance of *Hess2FF*, we generated a force field for the hetero-compound *N*-methylmesoporphyrin (MMP) in its complex with the enzyme ferrochelatase (Lecerof *et al.*, 2000). We tested five different methods of obtaining the Hessian matrix (B3LYP, AM1, MMFF, UHF and Dreiding). The Dreiding force field yielded a strange optimized structure with the methylated pyrrole ring and the opposite ring almost perpendicular to the other two rings. Therefore, we did not attempt to extract a force field from this model. However, the other methods performed appreciably better. The UFF method gave a reasonable optimized structure, but with a quite small tilt of the methylated pyrrole ring, whereas both AM1 and MMFF gave optimized structures quite similar to that obtained with B3LYP.

We also tested three different ways of selecting atom types for MMP (similar atom-type sets will also be used for the other proteins below).

(i) A separate type for each atom. This set of 43 atoms is called C_1.

(ii) A set of atom types exploiting the C_s symmetry of the *N*-methylporphine ring without side groups (a plane of reflection through CCA, NA and NC; the names of the atoms are defined in Fig. 1). Thus, most atom types of the porphine ring involved two atoms. For the side chains, standard *CNS* atom types were used (CH3E, CH2E, C and OC for CH_3, CH_2, carbonyl C and carboxyl O, respectively). This set is called C_s and contains 18 atom types.

(iii) 11 atom types obtained by an approach similar to that used by *CNS* for normal amino acids, *i.e.* we tried to find a minimum number of atom types that gives a small variation in the ideal bond lengths or angles and the corresponding force constants. The C atoms in the porphyrin ring were divided into three types: CPM for the methine linkages and CPA or CPB for the C atoms in the pyrrole rings α or β with respect to the N atoms. The N atoms themselves where divided into three types: NPC for the methylated N atom, NPH for (implicitly) protonated N atoms and NPE for N atoms with a lone-pair orbital. Finally, the methyl C atom was assigned an atom type of its own, so that its van der Waals radius could be varied separately. For the side chains, we used the same atom types as

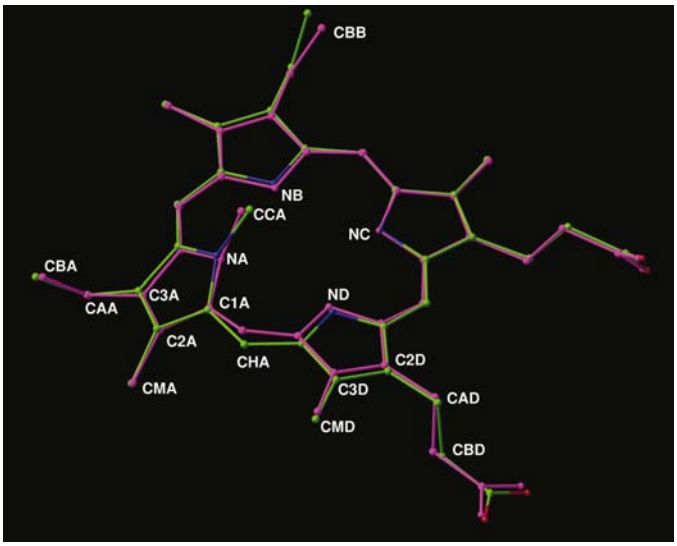

Figure 1
A comparison of the geometry of MMP in the original crystal structure of MMP bound to ferrochelatase (magenta) and in the re-refined structure using the *Hess2FF*/C_1 force field based on a B3LYP calculation.

Table 1
R factors for MMP in ferrochelatase using force fields obtained by *Hess2FF* from a Hessian matrix calculated with various methods and with different sets of atom types.

Method	Atom type	R_{free}	R
Crystal		0.23171	0.1828
B3LYP	CI	0.23128	0.1838
B3LYP	C_s	0.23126	0.1838
B3LYP	C_1	0.23115	0.1838
B3LYP/crystal	C_1	0.23123	0.1838
AM1	C_1	0.23189	0.1843
UFF	C_1	0.23133	0.1833
MMFF94	CI	0.23160	0.1839
MMFF94	C_s	0.23147	0.1838
MMFF94	C_1	0.23126	0.1839
B3LYP/MMFF94†	C_1	0.23190	0.1845
B3LYP/MMFF94‡	C_1	0.23176	0.1847
XPLO2D	*XPLO2D*	0.23138	0.1831
XPLO2D	CI	0.23135	0.1831
XPLO2D	C_1	0.23176	0.1834
PRODRG	*PRODRG*	0.23255	0.1832
ComQum-X		0.23118	0.1839

† Angles and dihedral involving the first atom in the side chains were taken from the B3LYP Hessian. ‡ Angles and dihedral involving the first atom in the side chains were taken from the MMFF Hessian.

in the C_s set. This choice of atom types will be denoted CI (chemical intuition) in the following.

As customary, H atoms were ignored in all refinements (but are present in the calculation of the Hessian matrix). As is described in §2, we had to increase the van der Waals radius of the CCA atom to obtain a MMP structure similar to that obtained with the best quantum-chemical calculations.

The *R* factors obtained after re-refinement of MMP in ferrochelatase using *Hess2FF* force fields from the various methods and atom types are gathered in Table 1. They show a rather small variation: 0.231–0.233 for R_{free} and 0.183–0.185 for *R*. The results are discussed more below together with the result for the other proteins.

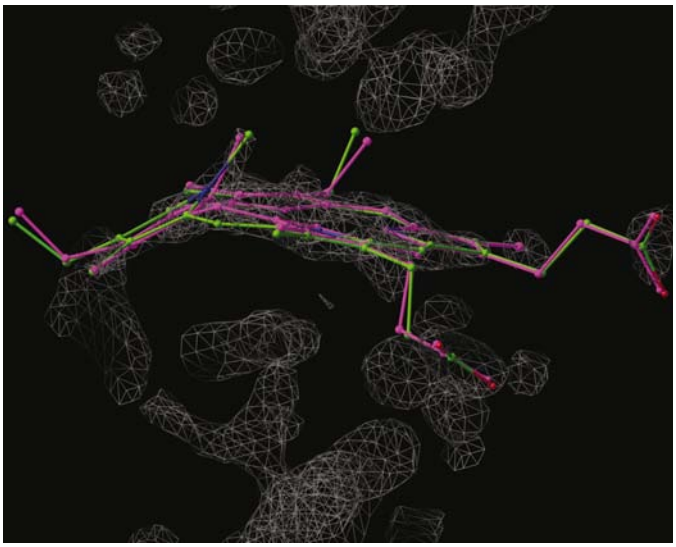

Figure 2
The B3LYP/C_1 structure and the original crystal structure (magenta) of MMP in ferrochelatase compared with the electron density ($2F_o - F_c$ omit map at the 2.0σ level).

In spite of the small difference in the *R* factors, Fig. 1 shows that the reoptimization leads to appreciable changes in the structure of MMP, especially in the distortion of the porphyrin ring and in the orientation of the ethyl groups.

Ferrochelatase is the terminal enzyme in haem synthesis, *i.e.* the enzyme that incorporates the iron ion into the porphyrin ring. MMP is a potent inhibitor of this enzyme and, as its name suggests, is a tetrapyrrole with a methyl group bound to one of the central pyrrolic N atoms. Steric repulsion between this methyl group and the other three pyrrole N atoms forces this pyrrole ring (the *A* ring) to bend out of the porphyrin plane. A similar distortion is believed to be enforced by the enzyme onto the porphyrin substrate, exposing the lone-pair electrons of the N atoms to the other substrate of the enzyme, an Fe^{2+} ion, thereby accelerating its insertion into the porphyrin ring (yielding a haem group; Lavallee, 1988). In fact, antibodies raised towards MMP catalyse the same reaction as ferrochelatase, although an order of magnitude slower than for the enzyme (Cochran & Schultz, 1990).

The re-refined structure fits excellently into the electron density, as can be seen in Fig. 2. The density is well defined for the porphyrin ring, whereas the side groups, especially the ethyl groups and the β atoms of the propionate side chains, are not so easy to position. Consequently, the largest differences between this *Hess2FF* structure and the original crystal structure are seen for the CBB, CBA and CBD atoms, which have moved 0.50, 0.47 and 0.35 Å, respectively. The average movement of all atoms in the two structures is 0.16 Å.

Another conspicuous difference is seen around the CHA atom, where the original structure shows a sharp kink. In the *Hess2FF* structure, there is a more gradual transition between the methylated *A* ring and the rest of the porphyrin. Similarly, the CCA and NA atoms have also moved significantly (0.23–0.25 Å), giving a slightly better fit to the electron density, as can be seen in Fig. 2. These differences are also present in the *ComQum-X* structure (Ryde *et al.*, 2002) and are therefore not an artifact of the new force field. There are some further differences between the two structures, but these are smaller and probably of less significance.

The differences are caused by small inconsistencies in the original force field. In particular, the improper torsions around atoms C2A, C3A, C2D and C3D are not correctly defined (the authors seem to have overlooked the asymmetry of the side chains and exchanged CMA and CAA, as well as CMD and CAD). This gives rise to large energy terms even in the final crystal geometry, *e.g.* 188 kJ mol^{-1} for the improper dihedral C2A—C1A—C3A—CAA. Thus, we see that small errors in the force field directly propagate to the final geometry, illustrating the importance of the force field.

Interestingly, in the optimized vacuum structures of MMP, the porphyrin ring is completely planar except for the tilt of the *A* ring. However, this is not the case in the crystal, where the structure is strongly ruffled (Fig. 1; it looks like the figure '8' seen from the edge) (Jentzen *et al.*, 1997). This ruffling is still seen in our reoptimized structure. However, it is notable

Table 2
Bond lengths to the Fe ion and *R* factors for haem in cytochrome c_{553}, using force fields obtained by *Hess2FF* from a Hessian matrix calculated with various methods and with different sets of atom types.

Method	Atom type	Fe—N_{Por} (Å)	Fe—N_{His} (Å)	Fe—S_{Met} (Å)	ΔR_{low}	ΔR_{high}
1c75		1.97–2.00	1.99	2.33	0.0054	
1b7v		2.02–2.08	2.31	2.21	0.0000	0.0000
B3LYP	Vacuum	2.00–2.01	2.00	2.35		
B3LYP	CI	2.00–2.02	2.06	2.30	−0.0078	−0.0152
B3LYP	CI†	2.01–2.02	2.05	2.30	−0.0062	−0.0148
B3LYP	C_1	2.01–2.02	2.05	2.31	−0.0073	−0.0147
PM3	Vacuum	1.92–1.93	1.89	2.40		
PM3	CI	1.96	2.04	2.36	−0.0075	−0.0155
UFF	Vacuum	1.95–1.96	1.95	2.36		
UFF	CI	1.96	1.99	2.34	−0.0068	−0.0149
Dreiding	Vacuum	1.96	1.96	2.35		
Dreiding	CI	1.98–1.99	1.99	2.34	−0.0043	−0.0140
MMFF	Vacuum	2.05–2.06	2.12	4.30		
MMFF	CI	2.06–2.07	2.10	2.30	−0.0069	−0.0141

† In this calculation, dihedrals involving sp^3 atoms were ignored and improper dihedrals were included.

that the energy involved in this ruffling is minimal. A *ComQum-X* calculation shows that the ruffled structure is destabilized by only 4–6 kJ mol^{-1} compared with the optimum vacuum structure (Ryde *et al.*, 2002). This shows that distortions of the porphyrin ring are low-energy modes.

Finally, we note that the tilt of the *A* ring relative to the porphyrin plane (defined as the angle between the NB—NC—ND and the NA—C2A—C3A planes) is similar in the re-optimized and crystal structures: 36 and 37°, respcetively. This is slightly larger than in vacuum (30°) and in various crystal structures of *N*-substituted porphyrins (20–30°; Lavallee, 1988).

3.2. Haem in cyctochrome c_{553}

Our next test case was cyctochrome c_{553} from *Bacillus pasteurii*. This protein was selected because its structure has been solved by the same group at 0.97 Å resolution with *ab initio* phasing and independently at 1.70 Å resolution in a multiple anomalous dispersion experiment (Benini *et al.*, 2000). The crystals were obtained under similar conditions. We can therefore use *Hess2FF* to re-refine the low-resolution structure and compare the result with the high-resolution structure and electron-density map. Thereby, we obtain highly objective criteria for the improvement of the structure. In Table 2, we show how the *R* factor changes (compared with the original low-resolution structure) for the low-resolution (ΔR_{low}) and high-resolution data (ΔR_{high}).

This small protein contains a haem group, where the central Fe^{III} ion binds to two axial ligands from the protein: a histidine $N^{\varepsilon 2}$ atom and a methionine S^{δ} atom. This group can be studied by *Hess2FF*, even if it is somewhat large for the B3LYP method. However, the Hessian calculations can be restricted to the haem group without any side chains, because these are flexible and contain only normal atom types. Likewise, the protein ligands also contain only standard amino-acid atom types. Therefore, we can calculate the Hessian matrix for Fe(porphine)(imidazole)$[S(CH_3)_2]^+$, which took 8 d at the B3LYP level.

We will concentrate the discussion on the axial Fe—N_{His} and Fe—S_{Met} distances, which are quite flexible; they are tabulated in Table 2. In the high-resolution crystal structure, the two distances are 1.99 and 2.33 Å, respectively. However, in the 1.7 Å structure they are 2.31 and 2.21 Å, respcetively, *i.e.* 0.32 Å longer and 0.12 Å shorter than in the better structure. We will see if these errors can be decreased by the use of *Hess2FF* methods.

The result of *Hess2FF*/B3LYP calculations are included in Table 2. They give Fe—N_{His} and Fe—S_{Met} distances of 2.05 and 2.30 Å, respectively. Thus, *Hess2FF* can partly correct the errors in the axial iron distances, but not fully.

As for MMP, we also tested several different choices of atom types and theoretical methods. Two types of *Hess2FF*/B3LYP calculations are included in Table 2, *viz.* atom types equivalent to CI (two different atom types for the equatorial N atom had to be used, because they form N—Fe—N angles of either 90 or 180° to each other) and C_1 of MMP. Both calculations gave very similar results in terms of the Fe—ligand distances and *R* factors. For both sets, we also tested ignoring all sp^2 dihedrals and instead including improper torsions. In all cases, this led to slightly worse results in terms of the *R* factors (in Table 2, only the results for the CI calculation are included).

Calculations with semiempirical and molecular-mechanics methods are also included in Table 2. For the isolated model, the PM3 method gives quite a poor result with a non-planar porphyrin ring and errors in the axial iron distances of 0.07–0.10 Å. Both UFF and Dreiding force fields give excellent Fe—ligand distances (an error of less than 0.04 Å), but the latter method gives a non-planar haem ring. In the protein (with CI atom types and no improper dihedrals), all these

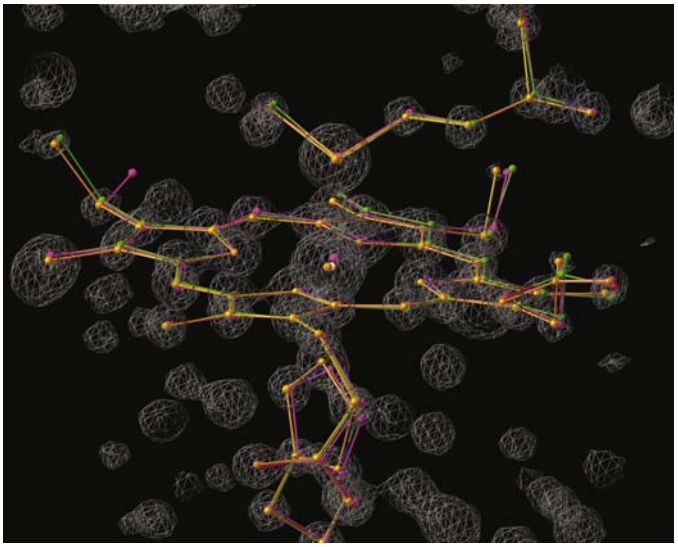

Figure 3
The B3LYP/CI structure of haem in cyctochrome c_{553}, compared with the original low-resolution crystal structure (magenta) and the high-resolution structure (orange). The figure also shows a $2F_o - F_c$ omit map (at the 2.5σ level) from the high-resolution data.

Table 3
Bond lengths to the Zn ion and *R* factors for the active site in alcohol dehydrogenase, using force fields obtained by *Hess2FF* from a Hessian matrix calculated with various methods.

Method	Zinc treatment	Zn—S (Å)	Zn—N (Å)	Zn—O (Å)	R_{free}	*R* factor
Crystal		2.08–2.29	2.13–2.20	2.05–2.07	0.23866	0.1909
B3LYP	Vacuum	2.33–2.33	2.27	1.93		
B3LYP	Non-bonded	2.32–2.34	2.15–2.19	2.02–2.09	0.23840	0.1911
B3LYP	Bonds	2.32–2.35	2.15–2.18	1.97–1.98	0.23844	0.1911
B3LYP	Bonds + angles	2.37–2.39	2.18–2.19	1.97–1.98	0.23823	0.1912
B3LYP	+ Dihedrals	2.34–2.39	2.18–2.19	1.98–2.00	0.23825	0.1911
PM3	Vacuum	2.39–2.36	2.11	2.03		
PM3	Bonds + angles	2.37–2.40	2.10–2.11	2.04–2.05	0.23845	0.1918
UFF	Vacuum	2.24–2.24	2.86	1.84		
UFF	Bonds + angles	2.31–2.33	2.15–2.16	1.90–1.92	0.23828	0.1915
Dreiding	Vacuum	2.36–2.37	3.38	1.99		
Dreiding	Bonds + angles	2.37–2.38	2.09–2.13	2.01	0.23831	0.1917

methods give quite good Fe—N_{His} (1.99–2.04 Å) and Fe—S_{Met} bond lengths (2.34–2.36 Å). MMFF, on the other hand, gives a flat haem group but over-long Fe—N bonds (2.05–2.12 Å) and a non-bonded methionine model (4.30 Å). In the protein, this leads to over-long Fe—N_{Por} and Fe—N_{His} bond lengths (2.07 and 2.10 Å). Interestingly, however, the Fe—S_{Met} bond length becomes quite good, 2.30 Å, which shows that the length of this flexible bond is mainly determined by the crystallographic data.

The main difference between the original low-resolution structure and that obtained with the *Hess2FF*/B3LYP force field is in the position of the iron ion. As can be seen in Fig. 3, the iron ion has moved more into the porphyrin plane in the latter structure. This is in accordance with the position of the iron ion in the high-resolution structure and also with the high-resolution electron-density map. The ligand atoms have also moved, but to a smaller extent. Once again, the *Hess2FF* structure is closer to the high-resolution than the low-resolution structure. Another conspicuous difference is in the position of the side-chain CBB atom (upper left corner), where again the *Hess2FF* structure is much closer to the high-resolution than the low-resolution structure.

3.3. Fluoroethanol and zinc in alcohol dehydrogenase

Next, we studied the catalytic zinc ion in horse liver alcohol dehydrogenase with the inhibitor trifluoroethanol (Bahnson *et al.*, 1997). We used *Hess2FF* to obtain a force field for this inhibitor and for the zinc ion, whereas the protein ligands (one histidine and two cysteine residues) were treated with standard *CNS* parameters.

For this system, we primarily studied how a metal force field is best obtained. A metal can either be described by only non-bonded (*i.e.* van der Waals) interactions or by including bonds, angles and dihedrals to its ligands. Using a *Hess2FF*/B3LYP force field, these variants were tested. The results are gathered in Table 3 and include the Zn–ligand distances and *R* factors.

From the results, it can be seen that the four types of calculations give quite similar results for the geometry around the zinc ion. In the crystal structure, the Zn—S, Zn—N and Zn—O bond lengths are 2.08–2.34, 2.13–2.20 and 2.05–2.07 Å, respectively; *i.e.* they show an appreciable variation. However, with the *Hess2FF* force field, irrespectively of whether any bond lengths, angles and dihedral terms for the zinc ion are included in the force field, these bond lengths are 2.32–2.39, 2.15–2.19 and 1.97–2.00 Å, respectively. Only the Zn—O bond shows a clear effect from the *Hess2FF* force field, decreasing by ∼0.1 Å (from 2.03 Å) when a Zn—O bond term is included.

It can also be seen that the re-refined Zn—S distances in the protein are similar to the quantum-chemical vacuum results, even if no force field for Zn is included. For the Zn—N bond, the difference is ∼0.1 Å. The reason for this is that this histidine is hydrogen bonded to an aspartate group, which introduces some imidazolate character to this ligand, giving a shorter bond (Gervasio *et al.*, 2001).

For this system, we have also studied semiempirical and molecular–mechanics methods. These results are included in Table 3. All methods give quite strange vacuum structures in one or another aspect. MMFF failed totally for this system (the model dissociated into many parts) and no force field was therefore abstracted for this method. The histidine ligand dissociates with the Dreiding force field and binds very weakly with the UFF method (2.86 Å). On the other hand, PM3 and Dreiding give quite long Zn—O bonds. However, all these methods give quite reasonable structures of the zinc site in the protein with similar R_{free} factors.

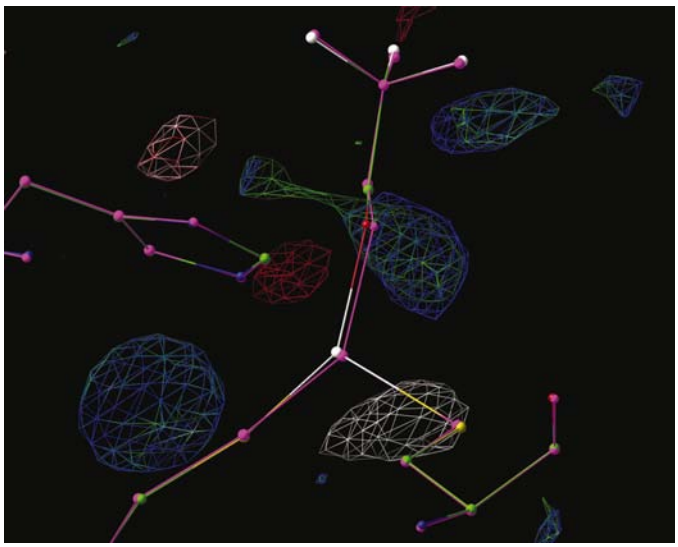

Figure 4
The *Hess2FF*/B3LYP structure (including zinc bonds and angles) and the original crystal structure (magenta) of the catalytic zinc site in alcohol dehydrogenase compared with $F_o - F_c$ difference maps at the $\pm 3.0\sigma$ level *Hess2FF* maps are blue (positive) and red; crystal structure maps are green and white.

Table 4
Bond lengths to the Fe ion (Å) and *R* factors for the active site in superoxide dismutase, using force fields obtained by *Hess2FF* from a Hessian matrix calculated with various methods.

Method	Iron treatment	F—N_{His}	Fe—O_{Asp}	Fe—O_{solv}	R_{free}	*R* factor
Crystal		2.13–2.32	2.01–2.02	2.30–2.34	0.1896	0.1601
B3LYP	Vacuum	2.05–2.13	2.03	1.89		
B3LYP	Non-bonded	2.12–2.25	1.96–1.99	2.22–2.28	0.1886	0.1625
B3LYP	Bonds	2.08–2.17	1.99–2.00	1.92	0.1876	0.1622
B3LYP	Bonds + angles	2.10–2.21	2.02	1.94	0.1880	0.1621
B3LYP	+ Dihedrals	2.09–2.19	2.01–2.02	1.92	0.1875	0.1623
B3LYP	+ Dihedrals†	2.08–2.19	2.02–2.03	1.92	0.1877	0.1624
PM3	Vacuum	1.88–1.95	2.03	1.86		
PM3	+ Dihedrals	1.91–1.99	2.04–2.05	1.87	0.1882	0.1631
UFF	Vacuum	1.92–2.97	2.72	1.85		
UFF	+ Dihedrals	1.94–2.40	2.09	1.75	0.1879	0.1641
Dreiding	Vacuum	1.93–3.72	3.65	1.94		
Dreiding	+ Dihedrals	1.98–2.24	1.90–1.92	1.98–1.99	0.1891	0.1639
MMFF	Vacuum	2.27–2.55	2.58	2.51		
MMFF	+ Dihedrals	2.21–2.45	2.54–2.55	2.33–2.35	0.2010	0.1751

† In this calculation, dihedrals involving sp^3 atoms were ignored and improper dihedrals were included.

Fig. 4 shows a comparison of the crystal structure and the *Hess2FF*/B3LYP calculation (with zinc bonds and angles). It can be seen that the two structures are very similar; the only clear differences are in the positions of the zinc ion and the alcohol O atom. However, in the $F_o - F_c$ difference map this leads to appreciable changes, especially around the zinc ion, where a large negative density (white) has been removed, whilst another smaller density has appeared above the zinc ion.

3.4. The catalytic iron site in superoxide dismutase

Metal sites are common in proteins and often play an active role in catalysis. Unfortunately, no generally applicable and accurate force field exists for metal sites and they therefore pose serious problems in the determination of crystal structures. Consequently, metal sites constitute one of the most important areas of application of the *Hess2FF* method. Therefore, our next application is also a metal site, *viz.* the catalytic site in iron superoxide dismutase. It consists of an iron ion bound to one aspartate and three histidine residues. The trigonal bipyramidal coordination sphere of the ion is completed by a solvent molecule.

Several structures of iron superoxide dismutase exist, as well as structures of the closely related manganese superoxide dismutase with the same ligand sphere. They show an appreciable variation in the Fe—O_{solv} distance (1.95–2.35 Å; Ursby *et al.*, 1999), even if the mechanistic consensus is that the solvent molecule should be deprotonated to a hydroxide ion and therefore give a Fe—O_{solv} distance at the lower end of this range (Holm *et al.*, 1996). We have studied the protein from the hypothermophile *S. solfataricus* determined at 2.3 Å resolution (Ursby *et al.*, 1999), which has the longest Fe—O_{solv} bond in the compilation above (2.35 Å). A detailed investigation (Nilsson & Ryde, manuscript in preparation) with the recently developed *ComQum-X* program (Ryde *et al.*, 2002) shows that the solvent molecule is also undoubtedly deprotonated in this structure and that the Asp ligand is probably protonated and forms a hydrogen bond to the hydroxide ion through the O^{δ} atom not bound to the zinc ion. Therefore, we base the present *Hess2FF* calculations on the Hessian for the $Fe^{III}(imidazole)_3(CH_3COOH)(OH)^{2+}$ model.

The results (Fe—ligand distances and *R* factors) of the various calculations are gathered together in Table 4. As for the previous protein, we test whether the iron ion is best described by only non-bonded interactions or with bond, angle or dihedral terms obtained by *Hess2FF*. For iron superoxide dismutase, the results vary somewhat more than for alcohol dehydrogenase. The results with only non-bonded interactions are quite similar to those of the original crystal structure, even if all iron bonds are slightly (∼0.05 Å) shorter. However, this does not mean that the resulting distances are those preferred by the reflections alone. It must be remembered that also this treatment involves a molecular-mechanics force field for the iron ion, *viz.* the non-bonded van der Waals terms. Also, these terms impose constraints on the Fe–ligand distances: if the active site is optimized with only these parameters and no crystal data, the optimum Fe—N_{His}, Fe—O_{Asp} and Fe—O_{solv} distances are 2.31–2.54, 2.27 and 2.81 Å, respectively.

If the *Hess2FF*/B3LYP force field with bonded interactions between iron and its ligands is included in the calculations, all the Fe–ligand distances become shorter and more similar: 2.08–2.21, 1.99–2.03 and 1.92–1.94 Å for the Fe—N_{His}, Fe—O_{Asp} and Fe—O_{solv} bonds, respectively. There is only a small difference (∼0.02 Å) between the various calculations, which differ in whether the angle and dihedral terms are defined for iron or not. The largest change compared with the crystal structure is the almost 0.4 Å shortening of the Fe—O_{solv} bond length. All calculations have a ∼0.002 lower R_{free} factor than the crystal structure, with only a small variation between the various calculations. Thus, for iron superoxide dismutase, the bond terms are most important, whereas the other terms do not change the results significantly.

We have also tested the four other theoretical methods on this system. As for alcohol dehydrogenase, all methods have problems giving a proper structure of the active site in vacuum. With the molecular-mechanics methods at least the Asp and two of the His ligands tend to dissociate from the iron ion. PM3 gives a better structure, but with short Fe—N_{His} bonds (1.88–1.95 Å). This is also reflected in the re-refined structures, which give quite strange structures and larger *R* factors than the B3LYP calculations.

In Fig. 5, we compare the *Hess2FF*/B3LYP structure (including bond, angle and dihedral terms) with the original crystal structure. It can be seen that the main difference between the two structures is in the position of the iron-bound hydroxide ion, which is closer to the iron ion in the *Hess2FF*

structure. The difference map shows that this gives an improved description of the density: a prominent volume of positive density (green) disappears between the iron ion and the water molecule when the structure is re-refined with the *Hess2FF* force field (blue).

3.5. 5′-deoxyadenosyl in methylmalonyl coenzyme A mutase

Finally, we constructed a force field for 5′-deoxyadenosyl bound to the enzyme methylmalonyl coenzyme A mutase (Mancia *et al.*, 1999). This investigation was mostly performed to also illustrate the importance of the force field for non-bonded interactions. This protein employs 5′-deoxyadenosyl cobalamin as a coenzyme. During catalysis, the Co^{III}—C bond breaks and a 5′-deoxyadenosyl radical is formed, which rapidly abstracts a proton from the substrate (methylmalonyl-coenzyme A). After a reorganization of the substrate radical, the product radical (succinyl-coenzyme A) abstracts the H atom back and the Co^{III}—C bond is reformed.

The studied crystal structure is suggested to contain a mixture of the substrate and the product, together with a dissociated 5′-deoxyadenosyl group with about 50% occupancy (assumed not to be a radical; Mancia *et al.*, 1999). However, the structure of the latter group is quite strange, with a short interaction between the C5′ and the C8 atoms (2.11–2.35 Å in the two subunits; *cf.* Fig. 6). This is intermediate between the expected values for a covalent bond (∼1.5 Å) and a non-bonded interaction (∼3 Å). In vacuum, the methyl group is oriented in the opposite direction and the distance is 4.9 Å.

We have re-refined the structure of the 5′-deoxyadenosyl group in methylmalonyl coenzyme A mutase using various force fields. The results are gathered in Table 5. As might be expected, the refined distance between the C5′ and C8 atoms strongly depends on the van der Waals parameters used for these atoms, whereas all other parameters have little influence. Thus, calculations with a *Hess2FF* force field based on a B3LYP calculation, but with two different sets of atom types (CI and C_1) and with or without improper torsion, all give C5′—C8 distances of 3.79–3.83 Å.

These calculations involve van der Waals parameters of extended atoms with three implicit H atoms for C5′ (0.758 kJ mol^{-1} and 3.8576 Å) and with one implicit H atom for C8 (0.203 kJ mol^{-1} and 4.2140 Å; taken from the *CNS* libraries). If we instead use the parameters in the *CNS* library file dna-rna_rep.param, which actually contains all parameters needed for 5′-deoxyadenosyl and which assigns a special atom type for each atom (*i.e.* C_1), the C5′—C8 distance becomes shorter (3.32–3.43 Å). This calculation involved the same van der Waals parameters for both atoms, *viz.* 0.419 kJ mol^{-1} and

Table 5
Distance between the C5′ and C8 atoms in 5′-deoxyadenosyl and *R* factors for methylmalonyl coenzyme A mutase, refined with various force fields.

Method	Atom types	C5′—C8 (Å)	R_{free}	*R* factor
Crystal		2.11–2.35	0.26817	0.22740
B3LYP	Vacuum	4.87		
B3LYP	CI	3.79–3.82	0.26811	0.22738
B3LYP	CI†	3.79–3.80	0.26812	0.22737
B3LYP	C_1	3.81–3.83	0.26814	0.22740
B3LYP	C_1†	3.79–3.80	0.26815	0.22739
Amber	C_1	3.77–3.82	0.26812	0.22738
CNS	C_1	3.32–3.43	0.26816	0.22737

† In this calculation, dihedrals involving sp^3 atoms were ignored and improper dihedrals were included.

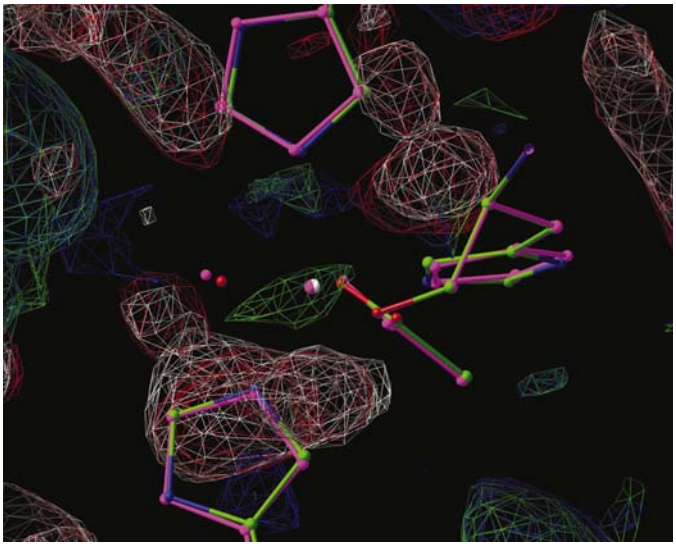

Figure 5
The *Hess2FF*/B3LYP structure (including iron bond, angles and dihedrals) and the original crystal structure (magenta) of the active iron site in superoxide dismutase compared with $F_o - F_c$ difference maps at the $\pm 2.0\sigma$ level. *Hess2FF* maps are blue (positive) and red; crystal structure maps are green and white.

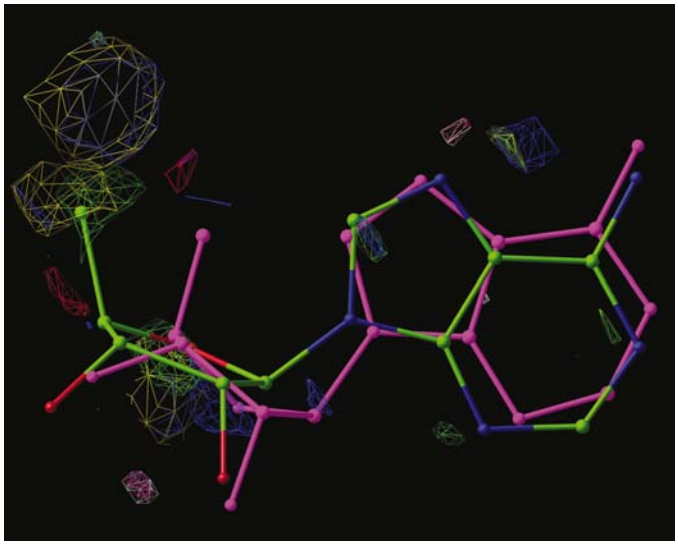

Figure 6
The *Hess2FF*/B3LYP/CI structure and the original crystal structure (magenta) of trifluoroethanol in methylmalonyl coenzyme A mutase compared with a $2F_o - F_c$ gradient map at the $\pm 2.5\sigma$ level. *Hess2FF* maps are blue (positive) and red; crystal structure maps are green and white. The figure also includes a map calculated from the structure optimized by the *CNS* force field (yellow and magenta).

3.296 Å. Thus, the shorter distance in the refined structure is caused by the smaller van der Waals radius. Finally, we also tried the van der Waals parameters from the Amber force field (Cornell *et al.*, 1995), 0.458 kJ mol^{-1} and 3.816 Å for both atoms. All other parameters were taken from the B3LYP/CI calculation. This gave a C5′—C8 distance of 3.77–3.82 Å, *i.e.* close to the *Hess2FF* value.

This illustrates that the refined structure strongly depends on the force field used in the refinement. It is hard to judge which force field gives the best result from only the results in Table 5, because they give very similar *R* factors, *viz.* 0.2681 for $R_{\rm free}$ and 0.2274 for the standard *R* factor. Therefore, we look instead at the electron-density map in Fig. 6. The best contrast is seen in the $2F_o - F_c$ gradient map, but it is still hard to decide which structure is best. The *Hess2FF* structure (blue and red) seems to give an improvement around the ribose ring, whereas the crystal structure (green and white) is best around the C5′ atom (upper left); the *CNS* structure (yellow and magenta) is poor around ribose, but best around the adenine ring. However, it is clear that the C5′—C8 distance in the crystal structure is unrealistic and for this reason, the re-refined structures must be considered to be improvements.

4. Discussion

We have tested *Hess2FF* on five proteins containing hetero-compounds using various method and atom types. The results are gathered in Tables 1–5 and are briefly described above. In this section, we will compare and discuss various aspects of these results.

4.1. Quality criteria

The first and primary question is whether *Hess2FF* improves the crystal structures significantly. From the results in Tables 1–5, we can see that for all proteins, the structures re-refined with a force field from *Hess2FF* give a smaller $R_{\rm free}$ factor than the original crystal structure. However, the improvement is not very large, ranging from 0.00006 for methylmalonyl coenzyme A mutase (with 22 224 atoms) to 0.008 for cyctochrome c_{553} (with 667 atoms). In spite of these small differences, Figs. 1–6 show that there are appreciable changes in the structure of the hetero-compound. The reason for this is that $R_{\rm free}$ is a global factor that is insensitive to local changes in one group of the whole protein.

Interestingly, the standard *R* factor does not show the same improvement. On the contrary, it normally increases by a slightly larger amount (by 0.0003–0.0022), methylmalonyl coenzyme A mutase being the only exception by showing a slight improvement (0.00002). This reflects that the original crystal structures are strongly optimized with respect to the standard *R* factor. Ideally (without overfitting), the two *R* factors should be equal. Therefore, *both* the decrease in $R_{\rm free}$ and the decrease in the difference between $R_{\rm free}$ and *R* flag an improvement in the structure. This is also the reason why we cannot use as the quality criterion the residue (real-space) *R* factor (Jones *et al.*, 1991) (which is almost 100 times more sensitive to the variations in the structure than the *R* factors): it is strongly correlated to the normal *R* factor but not to the $R_{\rm free}$ factor and therefore tends to increase for the structures refined with the *Hess2FF* force field. Moreover, the residue *R* factor (as implemented in the *CNS*, *MAPMAN* and *O* programs) is sensitive to details in the calculation (*e.g.* the type of density map, resolution range and parameters in the equation used when the factor is calculated; Jones & Kjeldgaard, 1995). We have tested several combinations, but with little success.

Another way to judge the quality of the re-refined structures is to study electron-density maps. Figs. 2–6 show density maps for the various proteins. Electron-density maps such as those in Figs. 2 and 3 show how well the models fit into the density and which regions of the structure are well defined by the experimental data. We have used omit maps, where the hetero-compound has been left out during the calculation of the map, to avoid bias of the density from the coordinates of these groups.

However, in many cases it is still hard to decide which model fits best into the map, especially when the differences in the structure are small. A more sensitive way to judge the differences between the structures is to use $F_o - F_c$ difference maps. Such maps show where a certain model has a surplus or deficit of electron density. If densities are calculated for the original and re-refined model, a comparison directly shows whether or not the structure has been improved. Thus, even if such maps become more complicated (with four different densities), they normally give a clearer picture of the improvement, as can be seen in Figs. 4 and 5.

A similar picture can also be obtained from gradient maps, which also show areas of surplus or deficit not directly in the electron density, but instead in the gradient of the refinement target function. Thus, they show in which direction the atoms should be moved to give an improved fit to the density. Such a map is shown in Fig. 6 for methylmalonyl coenzyme A mutase.

An important problem with both the *R* factors and the electron-density maps is that they are sensitive to errors in the reflections. If the goal was only to obtain the structure that fits the reflections as well as possible, the molecular-mechanics force field should be ignored during the refinement. However, experience shows that this gives unrealistic structures except for the most accurate structures. The force field is used to make the structure chemically reasonable (*i.e.* with reasonable bond lengths and angles): it supplements and corrects the raw data. Therefore, the best structure (*i.e.* the one that is closest to the correct model) does not necessary give the smallest *R* factor, the best fit to the density or the smallest deviations in the difference or gradient maps, owing to small errors in the raw data.

A way to solve this dilemma is to find a protein that has been solved at both low or medium resolution and at atomic resolution (where geometric restraints have a small influence on the structure), but otherwise under as similar conditions as possible. It could then be investigated how close the model resulting from refinement of the low-resolution data is to the model obtained from the high-resolution data and how well

the low-resolution model fits to the high-resolution density map. This is the reason why we have studied the cyctochrome c_{553} protein, the structure of which has been solved by the same authors under similar conditions at both 0.97 and 1.70 Å resolution (Benini *et al.*, 2000).

From the data in Table 2, we can directly see that the structures refined with the *Hess2FF*/B3LYP force field are appreciably closer to the high-resolution crystal structure than the original low-resolution crystal structure, at least for the bonds to the iron ion (the axial $Fe-N_{His}$ and $Fe-S_{Met}$ distances are 2.05 and 2.30 Å in the re-refined structure, 1.99 and 2.33 Å in the high-resolution structure, and 2.31 and 2.21 Å in the low-resolution structure, respectively).

Interestingly, the isolated quantum model optimized by B3LYP in vacuum gives even better distances, 2.00 and 2.33 Å, respectively, only 0.01 and 0.02 Å from the distances observed in the atomic resolution crystal structure. This shows that the B3LYP method accurately reproduces the geometry around the iron ion in cyctochrome c_{553} and that stronger restraints could have been used to actually *improve* the medium-resolution structure locally. This also illustrates the problem of refinement: even if the force field is correct, it cannot fully compensate for errors in a low- or medium-resolution structure unless the force constants are set to unphysically high values. We are currently investigating how large force constants are physically motivated.

Another way to study the quality of the force field is to compare the re-refined model of the low-resolution structure with the high-resolution structure. In Fig. 3, we directly see that the *Hess2FF*/CI structure is much closer to the high-resolution crystal structure than the low-resolution model. Of course, the *Hess2FF* model then also fits the high-resolution electron density better than the original low-resolution structure.

This can also be seen by calculating the *R* factors of the high-resolution data using models obtained from the low-resolution data. Such factors are listed in the ΔR_{high} column of Table 2. It is clear that the structures refined using a *Hess2FF* force field are better models for the high-resolution structure than the original low-resolution structure.

Interestingly, there is rather a strong correlation between these *R* factors and the *R* factors obtained directly during the refinement using the low-resolution data (ΔR_{low} column in Table 2). This shows that the R_{free} factor gives a reasonable criterion for the quality of the re-refined structures. This is in agreement with our observations with both *Hess2FF* and the *ComQum-X* program (Ryde *et al.*, 2002) that changes in the R_{free} factor, even if small, are consistent and make sense (*i.e.* the R_{free} factor decreases when the theoretical method is improved and it also improves as the geometry optimization in *ComQum-X* progresses). Therefore, we recommend the use of the R_{free} factor as a quality guide during the use of *Hess2FF* and *ComQum-X*. However, there are also some examples where a rather poor theoretical method gives a lower R_{free} factor than better ones. Therefore, other quality criteria must also be used in parallel, *e.g.* electron-density maps and the standard *R* factor.

4.2. Comparison of Hessians obtained at various levels of theory

We have seen that *Hess2FF* allows us to improve the structure of a hetero-compound according to several quality criteria. Next, we want to decide an appropriate level of theory to use for the calculation of the Hessian matrix as a proper compromise between accuracy and speed.

The density-functional B3LYP method performs well for most systems (Bauschlicher, 1995; Siegbahn & Blomberg, 2000). However, it is probably too time-consuming for systems of this size; the frequency calculation for MMP took 9 d on an AMD 1.47 GHz processor and the geometry optimization of this 81-atom model took almost the same time. Therefore, we tested two other levels of theory: the semiempirical AM1 and PM3 methods (Dewar *et al.*, 1985; Stewart, 1989), which are fast but with an appreciably lower accuracy, and molecular-mechanics calculations, which are even faster and have a varying accuracy (Gundertofte *et al.*, 1996). For the molecular-mechanics calculations, we used three different force fields: UFF, Dreiding and MMFF (Rappé *et al.*, 1992; Mayo *et al.*, 1990; Halgren, 1996). The first two force fields are universal, *i.e.* parameters exist for all elements, but their accuracy is rather low. MMFF contains parameters for fewer molecules, but it is appreciably more accurate (Gundertofte *et al.*, 1996).

The results of the various methods are included in Tables 1–4. For all systems except one, B3LYP gives the best result in terms of the R_{free} factor. It is also notable that the other methods often give significantly higher *R* factors than the B3LYP method. However, as noted above, the difference is not very large. Therefore, it seems to be possible to obtain a reasonable force field with cheap methods for all systems of interest, but for very accurate results, *e.g.* to improve the structure, B3LYP should be used.

An alternative to the full B3LYP method could be to mix calculations at different levels of theory. For example, we could derive parameters for the porphyrin ring in MMP from a B3LYP calculation of *N*-methylporphine ring without side chains (which can be calculated within a few days) and combine them with the MMFF force field for the side chains. However, as can be seen from Table 1 (entry B3LYP/MMFF), this did not give very impressive results. In fact, the R_{free} factors are larger than for the B3LYP and MMFF calculations and even larger than in the crystal, 0.2318–0.2319, irrespective of whether angles and dihedrals involving the first atom in the side chains are taken from the B3LYP or the MMFF calculation.

Therefore, it seems to be best to use one method for the whole hetero-compound. The method of choice varies with the type of the molecule, as we have seen. The most important thing is that the optimized structure of the isolated molecule should be reasonable and close to the structure encountered in the protein.

4.3. Comparison of various sets of atom types

Next, we tested various ways of selecting atom types. Atom types are used to reduce the number of parameters in the force

field (to reduce the parameterization effort and the chance of making mistakes). Thus, force-field parameters are not defined for all the individual bonds, angles and dihedrals, but only once for each interaction involving the same atom types. Therefore, only atoms with the same chemical environment should share the same atom types, *i.e.* they should have the same preferences for the bond lengths, angles *etc.*

For MMP, haem and 5′-deoxyadenosyl, we have studied three ways of defining the atom types, defined above as CI (chemical intuition, *i.e.* a minimum set of atom types), C_s (employing approximate symmetry in the molecule) and C_1 (employing a separate atom type for each atom). For MMP the results in Table 1 are clear and consistent: with both the B3LYP and MMFF Hessians, the result neatly follows the number of atom types. Thus, CI gave the highest $R_{\rm free}$ factors, whereas C_1 gave the lowest factors. However, as usual the difference is not very large; the $R_{\rm free}$ factor varies between 0.2311 and 0.2313 with the various sets of atom types at the B3LYP level, and between 0.2313 and 0.2316 at the MMFF level. Moreover, for the other two systems, the CI and C_1 atom-type set gave similar results, with the $R_{\rm free}$ factor of the former being slightly lower. The same also applies when we compare the re-refined structure of cyctochrome c_{553} with the high-resolution data of the same protein.

However, for the general use of *Hess2FF*, there is no reason to collect similar atoms into groups of atom types. On the contrary, assigning a separate atom type to all atoms is both easier (it can be performed automatically) and should give the best result. In fact, *Hess2FF* always calculates parameters for all the individual interactions; parameters for the atom types are obtained by averaging. The only disadvantage of the C_1 atom types is that the topology and parameter files becomes larger (a small problem) and that it increases the chance of introducing errors. On the other hand, it removes the risk of making a poor choice of atom types.

As is discussed in §2, *Hess2FF* gives the option of ignoring dihedrals involving central atoms that are sp^3 hybridized. The reason for this is that the force constants of these terms are normally small and therefore of little significance. Moreover, in vacuum calculations the geometry of such flexible groups are often determined by weak internal CH$-X$ hydrogen bonds, which gives ideal dihedrals that are not the optimum for the isolated group (*e.g.* staggered conformations). Such dihedrals are always ignored in the *XPLO2D* program (Kleywegt & Jones, 1995*b*). Moreover, this program also describes the geometry of planar groups by an improper dihedral angle rather than by normal dihedral angles. For most systems, a full set of dihedral angles also gives a proper structure of planar systems. However, in some cases ignorance of H atoms may lead to non-planar structures.

We have tested whether such a treatment improves the *Hess2FF* force field for the investigated proteins. However, as can be seen in Tables 1–5, we see no improvement for any system. On the contrary, in all cases investigated the $R_{\rm free}$ factor increases.

Metal centres provide a related problem. When the metal has more than three ligands, the normal routines for determining the period of the torsion do not work and the periodicity becomes complicated and depends on the precise geometry of the site. Moreover, the force constants of these terms are normally small. For this reason, dihedral terms involving metal ions as one of the central atoms are normally ignored (Merz, 1991; De Kerpel & Ryde, 1999). Therefore, for the tetrahedral zinc ion in alcohol dehydrogenase, we tested the effect of including such dihedral angles. As can be seen in Table 3, they do not improve the structure; instead, their inclusion leads to a slightly increased $R_{\rm free}$ factor (by 0.00005). On the other hand, inclusion of bond angles around the zinc ion leads to an improved force field of the site ($R_{\rm free}$ factor decreased by 0.0002). Therefore, torsions with a central metal ion were ignored in all the other calculations.

On the other hand, the results for the iron site in superoxide dismutase (in Table 4) show that dihedral angles involving the metal ion as one of the distal atoms leads to an improvement of the force field ($R_{\rm free}$ decreases by 0.0005). Therefore, we recommend that metal centres are described by all bond and angle terms, but only with dihedrals involving the metal ion as peripheral atoms. Such a force field was used in the study of cyctochrome c_{553}.

4.4. Comparison with other methods of obtaining a force field

Finally, we compared the *Hess2FF* method with other methods of automatically obtaining a force field for refinement of hetero-compounds. Firstly, we tested the *XPLO2D* program from Uppsala Software factory (a part of HIC-Up; Kleywegt & Jones, 1995*b*). It automatically constructs *CNS* topology and parameter files given a set of coordinates. In contrast to *Hess2FF*, it does not use a force-constant matrix, but uses a small set of default values for the force constants (as described above). Moreover, it does not employ dihedral angles but rather improper dihedrals to constrain groups to be planar or tetrahedral.

We downloaded *XPLO2D* from HIC-Up and applied it to the B3LYP-optimized structure of MMP in ferrochelatase using three different approaches. Firstly, we used the default choice of atom types in *XPLO2D* and applied them without any further change. This gave an $R_{\rm free}$ factor of 0.2314, *i.e.* slightly worse than B3LYP/CI. Secondly, we enforced the program to choose atom types identical to our CI set. This did not change the $R_{\rm free}$ factor. Finally, we instead asked *XPLO2D* to choose a separate atom type for each atom (*i.e.* the C_1 atom type set). This led to an increase of the $R_{\rm free}$ factor to 0.2318, *i.e.* slightly worse than in the starting structure. Thus, at least for MMP in ferrochelatase, *Hess2FF* gives better result than *XPLO2D*, which was expected as *Hess2FF* is more involved than *XPLO2D*.

XPLO2D can also be used for the other systems, but it will not give any improvement of the metal-containing hetero-compounds. It treats the metals as any other atom and defines bonds and angles involving the metal with ideal values taken directly from the input coordinates. Thus, if we use the original crystal coordinates as the input to the program, we will obtain

the same bond lengths and angles after the refinement (or the average if there are several bonds or angles of the same type), especially as the force constants used by *XPLO2D* are those typical for a covalent bond and are therefore appreciably higher than those expected for bonds to a metal. Therefore, *XPLO2D* will not provide any improvement of the original crystal structure, unless it is based on an accurate structure of a small model complex or a theoretical calculation. *Hess2FF*, on the other hand, is based on theoretical calculation and therefore automatically provides an optimized structure, which may improve the crystal structure if the theoretical method was appropriate.

Secondly, we tested another automatic program available on the Internet, *viz. PRODRG* (van Aalten *et al.*, 1996). This program can provide *CNS* topology and parameter files for any molecule inserted in PDB format. Unfortunately, the original paper does not provide any information about the selection of force-field parameters. We submitted the crystal coordinates of MMP in ferrochelatase to the *PRODRG* server (http://davapc1.bioch.dundee.ac.uk/programs/prodrg/prodrg.html) and used the resulting *CNS* files directly for a re-refinement of the structure. Unfortunately, this gave much worse *R* factors than any other method applied, as can be seen in Table 1 ($R_{\rm free}$ = 0.233). *PRODRG* does not support any metal ions, so it could not be used for the three metal-containing systems.

Thirdly, we tested the possibility of combining experimental geometries with calculated force constants. Thus, we used the B3LYP/C_1 force constants, but replaced all ideal bond lengths and angles by the corresponding values from the crystal structure of a *N*-methylated porphyrin. Unfortunately, there is no crystal structure of MMP. Instead, we used the crystal structure of *N*-ethoxycarbonylmethyl-octaethylporphyrin (McLaughlin, 1974). It differs from MMP in that all the side chains are ethyl groups (therefore, the side-chain geometry was taken from the B3LYP calculation) and that the *A*-ring nitrogen does not bind a methyl group but an ethoxycarbonylmethyl group. Interestingly, this did not improve the force field significantly: the structure optimized with this force field had a slightly worse (by 0.00008) $R_{\rm free}$ factor than that based completely on the B3LYP calculation (B3LYP/crystal in Table 1). This shows that B3LYP gives an excellent structure.

Finally, we also used the recently developed *ComQum-X* program (Ryde *et al.*, 2002). This program couples the quantum-chemical geometry-optimization software *TURBOMOLE* (Ahlrichs *et al.*, 2000) with *CNS*, so that quantum-chemical calculations replace the force field in the crystallographic refinement. In the original article (Ryde *et al.*, 2002), we thoroughly tested different sets of parameters for the calculations. From Table 1, it can be seen that the best result obtained, which gives a $R_{\rm free}$ factor of 0.2312, is almost identical to the *Hess2FF*/B3LYP results. In principle, *ComQum-X* provides the result of a *Hess2FF* calculation with a perfect force-field parameterization. Thus, it is clear that *Hess2FF* gives excellent results and that it is a competitive alternative for the average crystallographer with access to simple quantum-chemical software (the *ComQum-X* calculations typically take twice as much time as the the frequency calculation for *Hess2FF*).

4.5. Recommended use

The results in the previous sections can be summarized in the following recommendations for general use of *Hess2FF*.

(i) Build the molecule of interest with your favourite molecular-builder software. Note that H atoms must be present in the theoretical calculation.

(ii) Select an appropriate level of theory. The following rules of thumb may be useful.

(*a*) For an organic compound with only normal functional groups and no severe steric repulsion, molecular-mechanics calculations with the MMFF or MM2/MM3 force fields are probably appropriate.

(*b*) For organic compounds with unusual functional groups and other more complicated organic systems, a semi-empirical method such as AM1 and PM3 can be used. However, it should not be used for systems where hydrogen bonds are important.

(*c*) For high accuracy and for systems including transition metals, we recommend the density-functional B3LYP method with the 6-31G* basis set.

(iii) Optimize the geometry of the molecule. If there is high-resolution crystallographic data available for the molecule of interest, it can be used instead.

(iv) Check that the optimized structure looks reasonable; if not, use a higher level of theory.

(v) Calculate the Hessian matrix by a frequency calculation.

(vi) Construct a file with the desired atom names and possibly with atom types. Ensure that there is no collision with the standard *CNS* atom types and with the names used for the scattering data (*i.e.* that the two first letters do not coincide with the chemical abbreviation of another element). We recommend a separate type for each atom.

(vii) Run *Hess2FF* using default values of all parameters.

(viii) Run the generate, energy and minimization jobs with *CNS*.

(ix) Check the output file of the energy calculation for large energy terms (>4 kJ mol^{-1}). Such terms signal problems with the force field and should be removed by zeroing the force constant, changing the atom types or changing the period of a dihedral angle.

(x) Check that the structure minimized with the new force field is reasonable and does not differ from the original structure. If it does, the force field has to be improved either by using a different set of atom types or by changing some of the force-field parameters.

(xi) Perform the crystallographic refinement using the final topology and parameter files.

4.6. Concluding remarks

We have developed a method to automatically obtain force-field parameters of a hetero-compound for crystallographic refinements. The method is based on a quantum-chemical

calculation of the Hessian matrix (the force-constant matrix) of the molecule of interest. We show that the resulting force field improves the structure of the hetero-compound in the protein in terms of the $R_{\rm free}$ factor, as well as in the fit to various electron-density maps, and that it changes a low-resolution structure towards a high-resolution structure of the same protein.

We have compared the method with other automatic methods of calculating topology and parameter files for hetero-compounds and showed that the method is competitive, in particular for metal sites. We have also tested how the Hessian is best calculated in terms of accuracy and speed of calculation and what sort of atom types should be used. The results of this investigation are summarized in the previous section.

It should be noted that the force field obtained from *Hess2FF* involves some approximations which make it less useful in pure molecular simulations (molecular dynamics, Monte Carlo *etc.*). Firstly, electrostatics are ignored, or rather electrostatic interactions are implicitly included in the terms of the force field. This follows the crystallographic custom of ignoring electrostatics. It is probably a reasonable approximation in these calculations, where the general structure is determined by the crystallographic raw data. However, in classical simulations of polar systems (such as a protein in solution), ignorance of electrostatics would lead to unreliable results.

In a similar manner, van der Waals interactions are included in the Hessian and therefore implicitly also in the bond and angles terms in the force field. However, for these interactions there are also explicit terms in the *CNS* force field. Therefore, there will be some double counting of such interactions in the resulting force field. Ideally, an iterative approach should have been used which subtracts the effects of the *CNS* van der Waals interactions before the Hessian is used to extract the bonded force-field parameters (Norrby & Liljefors, 1998). This would lead to a much more complicated program and longer execution times. For the present use, this approximation seems reasonable and clearly leads to improved results compared with a force field constructed by hand or using standard values for the force constants.

The most important conclusion of this paper is that the force field used in crystallographic refinements for a hetero-compound has a strong influence on the final structure and for the mechanistic implications of the structure, at least for medium-resolution structures (our test models had a resolution of 1.7–2.3 Å). We have seen that small changes in the force field may change a metal–ligand bond length by 0.4 Å and a non-bonded interaction by 1.7 Å without any significant change in the *R* factors. Therefore, the topology and parameter files used for a hetero-compound should be published together with the crystal structure so that a reader can reproduce the structure or judge if the structure can be trusted. This is not normally done today. For the five structures used in this investigation, either nothing is said about the force fields for the hetero-compound or it is stated that the structure does not change significantly if the force constants of some of the parameters for the hetero-compound are removed (Mancia *et al.*, 1999; Bahnson *et al.*, 1997; Benini *et al.*, 2000; Lecerof *et al.*, 2000). Only for superoxide dismutase is a more detailed account given of how the iron ions were treated (Ursby *et al.*, 1999).

Moreover, our results show that it is worthwhile to spend some computer time in order to obtain an appropriate force field if there is a hetero-compound present in the structure. Clearly, an automatic method to obtain an accurate force field for hetero-compounds is of great use, especially in high-throughput crystallography.

This investigation has been supported by grants from the Crafoord Foundation and the Swedish Research Council (VR). It has also been supported by the computer resources of the Swedish Council for High Performance Computing (HPDR), Parallelldatorcentrum (PDC) at the Royal Institute of Technology, Stockholm and Lunarc at Lund University. We thank Professor G. J. Kleywegt for useful discussions.

References

Aalten, D. M. F. van, Bywater, R., Findlay, J. B. C., Hendlich, M., Hooft, R. W. W. & Vriend, G. (1996). *J. Comput. Aided Mol. Des.* **10**, 255–262.

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.

Ahlrichs, R. *et al.* (2000). *TURBOMOLE Version 5.3.* Universität Karlsruhe, Germany.

Bahnson, B. J., Colby, T. D., Chin, J. K., Goldstein, B. M. & Klinman, J. P. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 12797–12802.

Barone, V., Adamo, C. & Mele, F. (1996). *Chem. Phys. Lett.* **249**, 290–296.

Bauschlicher, C. W. (1995). *Chem. Phys. Lett.* **246**, 40–44.

Benini, S., González, A., Rypniewski, W. R., Wilson, K. S., Van Beeumen, J. J. & Ciurli, S. (2000). *Biochemistry*, **39**, 13115–13126.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Cochran, A. G. & Schultz, P. G. (1990). *Science*, **249**, 781–783.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). *J. Am. Chem. Soc.* **117**, 5179–5197.

De Kerpel, J. O. A. & Ryde, U. (1999). *Proteins Struct. Funct. Genet.* **36**, 157–174.

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. (1985). *J. Am. Chem. Soc.* **107**, 3902–3909.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Frisch, M. J. *et al.* (1998). *GAUSSIAN*98, Revision A.5. Gaussian, Inc., Pittsburgh, PA, USA.

Gervasio, F. L., Schettino, V., Mangani, S., Carloni, P. & Parrinello, M. (2001). *J. Inorg. Biochem.* **86**, 233.

Gundertofte, K., Liljefors, T., Norrby, P.-O. & Pettersson, I. (1996). *J. Comput. Chem.* **17**, 429–449.

Halgren, T. A. (1996). *J. Comput. Chem.* **17**, 490–641.

Hehre, W. J., Radom, L., Schleyer, P. von R. & Pople, J. A. (1986). *Ab Initio Molecular Orbital Theory.* New York: Wiley–Interscience.

Holm, R. H., Kennepohl, P. & Solomon, E. I. (1996). *Chem. Rev.* **96**, 2239–2314.

Jentzen, W., Song, X.-Z. & Shelnutt, J. A. (1997). *J. Phys. Chem. B*, **101**, 1684–1699.

Jones, T. A. & Kjeldgaard, M. (1995). *O – The Manual.* Uppsala University, Sweden.

Jones, T. A., Zou, J.-Y., Cowan, W. E. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Jones, T. A. (1995*a*). *Structure*, **3**, 535–540.
Kleywegt, G. J. & Jones, T. A. (1995*b*). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 45–50.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **227**, 208–230.
Kleywegt, G. J. & Jones, T. A. (1998). *Acta Cryst.* D**54**, 1119–1131.
Lavallee, D. K. (1988). *Mol. Struct. Energ.* **9**, 279–313.
Lecerof, D., Fodje, M., Hansson, A., Hansson, M. & Al-Karadaghi, S. (2000). *J. Mol. Biol.* **297**, 221–232.
McLaughlin, G. M. (1974). *J. Chem. Soc. Perkin Trans. II*, pp. 136–140.
Mancia, F., Smith, G. A. & Evans, P. R. (1999). *Biochemistry*, **38**, 7999–8005.
Marques, H. M., Ngoma, B. & Brown, K. L. (2001). *J. Mol. Struct.* **561**, 71–91.
Mayo, S. L., Olafson, B. D. & Goddard, W. A. III (1990). *J. Phys. Chem.* **94**, 8897–8909.
Merz, K. M. (1991). *J. Am. Chem. Soc.* **113**, 406–411.
Norrby, P.-O. & Liljefors, T. (1998). *J. Comput. Chem.* **19**, 1146–1166.
Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.
Pettersson, G. (1987). *CRC Crit. Rev. Biochem.* **21**, 349–389.
Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. III & Skiff, W. M. (1992). *J. Am. Chem. Soc.* **114**, 10024–10035.
Rauhut, G. & Pulay, P. (1995). *J. Phys. Chem.* **99**, 3093–3100.
Ryde, U. (1995). *Protein Sci.* **4**, 1124–1132.
Ryde, U., Olsen, L. & Nilsson, K. (2002). *J. Comput. Chem.* **23**, 1058–1070.
Schäfer, A., Horn, H. & Ahlrichs, R. (1992). *J. Chem. Phys.* **97**, 2571–2577.
Seminario, J. M. (1996). *Int. J. Quantum Chem. Quantum Chem. Symp.* **30**, 59–65.
Siegbahn, P. E. M. & Blomberg, M. R. A. (2000). *Chem. Rev.* **100**, 421–437.
Spartan (1997). *SPARTAN Version* 5.0. Wavefunction, Inc., 18401 Von Karman Avenue, Suite 370, Irvine, CA 92612, USA.
Stewart, J. J. P. (1989). *J. Comput. Chem.* **10**, 209–220.
Stewart, J. J. P. (1990). *J. Comput. Aided Mol. Des.* **4**, 1–105.
Ursby, T., Adinolfi, B. S., Al-Karadaghi, S., De Vendittis, E. & Bocchini, V. (1999). *J. Mol. Biol.* **286**, 189–205.