research papers

Wavefunctions derived from experiment. III. Topological analysis of crystal fragments

Ian Bytheway,[a] Daniel J. Grimwood[a] and Dylan Jayatilaka[a,b]*

[a]Chemistry, School of Biomedical and Chemical Sciences, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia, and [b]Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia. Correspondence e-mail: dylan@crystal.uwa.edu.au

A constrained wavefunction model has been used to extract a Hartree–Fock wavefunction for $C_2H_2O_4 \cdot 4H_2O$ from both low-angle ($\sin\theta/\lambda < 0.71$ Å^{-1}, 571 reflections) and full ($\sin\theta/\lambda < 1.00$ Å^{-1}, 968 reflections) experimental X-ray diffraction data for crystalline α-oxalic acid dihydrate (α-$C_2H_2O_4 \cdot 2H_2O$) using polarized double-ζ and triple-ζ Gaussian basis sets. Properties obtained from the zero-flux partitioning of the total charge-density distribution derived from these wavefunctions, as well as from multipole refinement of the experimental data, are calculated and compared. This work represents the first calculation of integrated atomic properties derived from the fitting of Gaussian density functions to experimental X-ray diffraction data. In particular, atomic kinetic energies derived from experimental data are presented for the first time. The results obtained from the constrained (experimental) charge density show qualitatively similar properties to those obtained from conventional *ab initio* gas-phase calculations, though the quantitative differences are often substantial. The accuracy of integrated properties calculated using this procedure was established from the analysis of a wavefunction derived from simulated random-error diffraction data; that is, data obtained by adding normally distributed errors to the experimental structure factors. Analysis of this random-error wavefunction indicated that most topological properties are accurate to within approximately 5%, although the error is much larger for those properties that have a steep gradient in the region being evaluated [*e.g.* the value of $\nabla^2\rho(\mathbf{r}_b)$ at bond critical points] or are very small (*e.g.* the atomic dipole moment). Calculations of the constrained wavefunction using both the larger basis set and the complete set of experimental data yield results that agree quantitatively with the smaller calculations.

1. Introduction

The constrained wavefunction model proposed in recent papers (Jayatilaka, 1998; Jayatilaka & Grimwood, 2001; Grimwood & Jayatilaka, 2001) provides a new approach for the treatment of experimental structure-factor magnitudes obtained from elastic X-ray scattering experiments. In this model, conventional *ab initio* quantum-chemical calculations are used to provide a wavefunction, which is in turn used to generate theoretical structure-factor magnitudes. These theoretical and experimental structure factors are compared and successive constraints are applied to the wavefunction calculation until it provides structure-factor magnitudes that are in reasonable agreement with the experimental values. This iterative process is, therefore, a means of deriving an experimental wavefunction.

The wavefunction itself can be obtained from any of the variational techniques used in *ab initio* quantum chemistry.

The constrained wavefunction model has the advantage of being systematically improvable inasmuch as the complexity of the quantum-chemical calculation can, at least in principle, be readily increased. Initially, a wavefunction is obtained from a conventional *ab initio* calculation; such a wavefunction for our purposes can be thought of as 'unconstrained'. The method of Lagrange multipliers is used in subsequent wavefunction calculations to constrain the wavefunction, with the goal of improving the agreement between theoretical and experimental structure-factor magnitudes. It is important to note that this is not a least-squares fitting procedure and it is therefore possible, and indeed often the case, that there are many more parameters in the wavefunction than in the experimental data.

In this paper, our interest is in the chemical effects of the crystal lattice on the $C_2H_2O_4 \cdot 4H_2O$ cluster (Fig. 1). The theory of atoms in molecules (AIM) (Bader, 1990) has been chosen for this purpose in order to probe the changes in

properties that can be obtained from the total charge-density distribution $\rho(\mathbf{r})$. As well as topological analyses of $\rho(\mathbf{r})$ and the Laplacian of the charge density $\nabla^2\rho(\mathbf{r})$, the charge densities obtained from the constrained wavefunction model can be readily partitioned into the atomic basins defined by the zero-flux boundaries in $\rho(\mathbf{r})$. Atomic properties that require integration of atomic basins can then be readily calculated, and comparison of 'constrained' and 'unconstrained' wavefunctions provides a ready probe for the effects of crystal packing.

The calculation of total charge densities using the constrained wavefunction model contrasts the established procedure of deriving experimental charge densities from the fitting of atom-centred multipolar density functions to X-ray structure-factor magnitudes (Coppens, 1992, 1997; Coppens *et al.*, 1999). AIM analyses of the total charge-density distributions obtained from multipole fitting have typically been restricted to studies of only the topology of $\rho(\mathbf{r})$ and $\nabla^2\rho(\mathbf{r})$ [for example, amino acids (Flaig *et al.*, 1998), organic molecules (Destro *et al.*, 1991; Howard *et al.*, 1992, 1995; Koritsánszky *et al.*, 1996, 1999, 2000; Souhassou & Blessing, 1999; Macchi *et al.*, 2000), ammonia (Boese *et al.*, 1997), transition-metal molecules (Smith *et al.*, 1997; Iversen *et al.*, 1997; Macchi *et al.*, 1998*a*,*b*; Bianchi *et al.*, 2000; Bytheway *et al.*, 2001), as well as bulk periodic systems such as SiO_2 (Rosso *et al.*, 1999); the review by Koritsánszky & Coppens (2001) is recommended for a more detailed discussion] and not the calculation of integrated atomic properties. This is not surprising given the complexity associated with the definition of atomic regions defined by zero-flux surfaces in the charge density and their subsequent integration (Popelier, 1998*b*).

Partitioning of experimental multipolar charge densities is possible and has been accomplished in two separate studies by Flensburg & Madsen (2000), and Volkov, Gatti, Abramov & Coppens (2000). Using these methods, integrated atomic properties have been calculated for several organic molecules, including methylammonium hydrogen succinate monohydrate and methylammonium hydrogen maleate (Flensburg & Madsen, 2000), and *p*-nitroaniline (Volkov, Abramov, Coppens & Gatti, 2000)], and more recently the experimental charge density of cuprite, Cu_2O, has been obtained (Lippmann & Schneider, 2000).
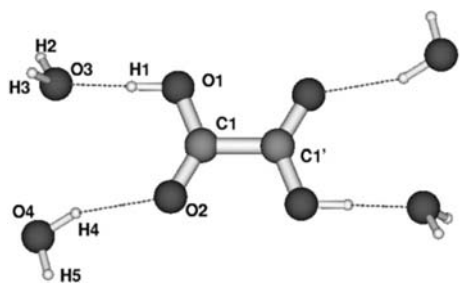

Figure 1
The $C_2H_2O_4 \cdot 4H_2O$ cluster used in these calculations, showing the numbering scheme used throughout (Schaftenaar & Noordik, 2000). A prime is used to denote atoms related by inversion symmetry.

It should be made clear at this point that the calculation of properties related to the density matrix from these methods that depend upon the multipole-fitted charge density is difficult. This is a disadvantage not shared by the constrained wavefunction model, where properties are calculated in exactly the same way as in a purely *ab initio* calculation. The constrained wavefunction model, therefore, provides both an alternative to the conventional route to experimental charge-density distributions *via* the least-squares refinement of multipolar functions, and a direct means of obtaining experimental values for properties that require virial partitioning of the total charge-density distribution.

2. Theoretical methodologies

2.1. Wavefunction calculations

The calculation of a wavefunction from experimental data using Gaussian density functions involves the iterative comparison of structure-factor magnitudes derived from the theoretical density [*via* the independent-molecule method (Chandler *et al.*, 1994) and corrected for the effects of thermal motion and extinction] with those observed experimentally. This procedure has been described previously (Jayatilaka, 1998; Jayatilaka & Grimwood, 2001; Grimwood & Jayatilaka, 2001) and will be referred to as the constrained wavefunction method in recognition of previous workers (Mukherj & Karplus, 1963; Henderson & Zimmermann, 1976) who also used this term to describe energy-variational calculations. Furthermore, the calculations described here utilize the Hartree–Fock method and are, therefore, denoted constrained Hartree–Fock calculations, abbreviated CHF.

Three constrained Hartree–Fock wavefunctions have been analysed in this work. The first, which is the main focus, is the same as that obtained previously (Grimwood & Jayatilaka, 2001). This calculation used the Dunning polarized double-ζ basis set (Dunning, 1970), denoted DZP, and the low-angle ($\sin\theta/\lambda < 0.71$ Å^{-1}, 571 reflections) experimental data of Zobel (1996). Constrained Hartree–Fock wavefunctions were also calculated using the Dunning correlation consistent polarized valence triple-ζ basis set (Dunning, 1989), denoted cc-pVTZ, in conjunction with the complete ($\sin\theta/\lambda < 1.00$ Å^{-1}, 968 reflections) experimental data set of Zobel (1996). Additionally, two procedures were used to partition the density: the Gaussian method, which was used exclusively in *all* of the DZP calculations as well as in unconstrained and constrained cc-pVTZ wavefunction calculations; and the Mulliken method, which was applied to the calculation of a triple-ζ wavefunction. Where necessary, 'G' and 'M' will be appended to the description of the cc-pVTZ calculations to denote which partitioning method, either Gaussian or Mulliken, respectively, was used. These partitioning procedures are described in more detail in the first paper in this series (Jayatilaka & Grimwood, 2001). Agreement statistics for the comparison of the experimental and theoretical structure factors obtained from the various wavefunction

Table 1
Agreement statistics and electronic energies (Hartree) based on the comparison of observed structure factors with those obtained from the different calculations.

	R (%)	R_w (%)	χ^2	GOF	Energy
HF/DZP	2.5	3.1	15.2	3.9	−680.606833
BLYP/DZP	2.1	2.7	11.5	3.4	−683.970232
CHF/DZP	1.4	1.5	3.2	1.8	−680.175480
CHF(re)/DZP	1.9	1.6	4.3	2.1	−680.170000
HF/cc-pVTZ	2.5	2.9	9.2	3.0	−680.711203
CHF(G)/cc-pVTZ	1.6	1.5	2.3	1.5	−680.425773
CHF(M)/cc-pVTZ	1.5	1.4	2.3	1.5	−680.432252
XD(low)	1.6	1.7	5.3	2.3	
XD(low, re)	2.1	1.9	6.2	2.5	
XD(full)	1.8	1.7	3.6	1.9	

Table 2
Monopole populations (P_v), κ and κ' parameters obtained from the multipole refinements of the full and low-angle data for $C_2H_2O_4 \cdot 2H_2O$.

Only values for H4 are shown, representing both of the hydrogen atoms of the water molecules as H4 and H5 were made equivalent during multipole refinement. Atom labels conform to those used in Fig. 1.

	C1	O1	O2	O3	H1	H4
XD(low)						
P_v	4.29	5.90	6.15	5.91	0.94	0.91
κ	0.99	1.02	1.00	1.01	1.2	1.2
κ'	0.99	0.81	0.80	0.88	1.2	1.2
XD(low, re)						
P_v	4.36	5.90	6.09	5.89	0.94	0.91
κ	0.99	1.02	1.01	1.01	1.2	1.2
κ'	0.98	0.75	0.80	0.94	1.2	1.2
XD(full)						
P_v	4.38	5.88	6.08	5.98	0.91	0.88
κ	0.96	1.01	1.00	1.00	1.2	1.2
κ'	0.94	0.80	0.77	0.88	1.2	1.2

calculations, along with the electronic energies, are given in Table 1.

The Hartree–Fock calculations described here, of both constrained and unconstrained wavefunctions, were performed using the *TONTO* software package (Jayatilaka & Grimwood, 2000). The *GAUSSIAN98* software (Frisch *et al.*, 1998) was used to perform the density functional calculations. Becke's exchange functional (Becke, 1988) in conjunction with the correlation functional of Lee, Yang & Parr (1988) were used to calculate the BLYP wavefunction.

2.2. Multipole refinement of experimental data

Least-squares refinements of

$$\sum w\,[F^o(\mathbf{k}) - k|F^c(\mathbf{k})|]^2, \tag{1}$$

where $w = \sigma^{-2}(F^o)$, were performed for both the full and low-angle data sets described above using the Hansen–Coppens multipole formalism (Hansen & Coppens, 1978; Coppens, 1997) as implemented in the *XD* software. Reflections with $F^o(\mathbf{k}) > 2\sigma(\mathbf{k})$ were used in the refinement and electroneutrality of the whole oxalic acid dihydrate unit was constrained throughout the refinement process, *i.e.* the $C_2H_2O_4$ and H_2O molecules were not separately constrained to be electroneutral.

The multipole refinement was carried out in three steps. In the first, multipole functions with $l \leq 4$ for carbon and oxygen atoms ($M1$, $D1\pm$, $Q0$, $Q2\pm$, $O1\pm$, $O3\pm$, $H0$, $H2\pm$, $H4\pm$) and $l \leq 2$ for hydrogen atoms ($M1$, $D0$, $Q0$) were refined. A κ refinement was then performed with the multipole functions fixed at the values obtained from the first step. Only the κ and κ' parameters for the carbon and oxygen atoms were refined, and the values for the hydrogen were fixed throughout at 1.2 (Boese *et al.*, 1997; Koritsánszky *et al.*, 2000). These parameters control the radial extent of the spherical (κ) and aspherical density (κ') and were refined independently with the constraint that the same κ' be used for each of the different sets of multipole functions. In the final refinement all parameters were refined, with the exception of the κ values for hydrogen atoms. Agreement statistics for these refinements, denoted XD(low) and XD(full), are given in Table 1; κ, κ' and monopole population parameters are given in Table 2. The

complete set of multipole population parameters from these refinements is available from the authors.

The agreement statistics in Table 1 are similar to those obtained previously (Zobel *et al.*, 1992), as expected. The χ^2 statistic indicates the better overall agreement for the full data set, while the χ^2 statistics for the multipole refinements are slightly higher than the values obtained from the CHF wavefunctions. The κ and κ' values in Table 2 are also similar to those obtained from a previous multipole refinement (Zobel, 1996) and show similar trends to those found for carbon and oxygen atoms in other systems (Abramov *et al.*, 1999; Koritsánszky *et al.*, 2000).

These multipole refinements were performed in order to provide a means of comparing the properties derived from the constrained Hartree–Fock wavefunction *via* the total charge density with those obtained from the conventional multipole-fitted charge density. The *TOPXD* software (Volkov, Abramov, Coppens & Gatti, 2000) was used to integrate the atomic basins derived from the multipole-fitted total charge densities and the results of these analyses are given in Tables 3–8.

2.3. Partitioning of the total charge densities

The theory of atoms in molecules (Bader, 1990, 1998) was used to analyse the constrained and unconstrained wavefunctions. All wavefunction analyses were performed with the *MORPHY98* software (Popelier & Bone, 1998). Differences between the integrated properties obtained from the various wavefunctions are simply

$$\frac{A_\Omega(\text{constrained}) - A_\Omega(\text{other})}{A_\Omega(\text{constrained})} \times 100\%, \tag{2}$$

where A_Ω is the integrated atomic property of interest, and 'constrained' refers to the value obtained from the constrained HF (hereafter denoted CHF) wavefunction, while 'other' refers to values from the HF, BLYP or random-error (re) wavefunctions.

2.4. Accuracy *versus* precision

The error associated with integration of the atomic basins is given by

$$L(\Omega) = \frac{-\hbar^2}{4m} \int_\Omega \nabla^2 \rho(\mathbf{r}) \, d\tau, \qquad (3)$$

which should be zero for an ideal integration (Bader, 1990). The deviation of $L(\Omega)$ from zero, therefore, provides a measure of the precision of the integrated results; in this work the values of $L(\Omega)$ ranged from approximately 3×10^{-6} to 5×10^{-4} for carbon and oxygen atoms, and from approximately 8×10^{-6} to 2×10^{-4} for hydrogen atoms.

These are not errors associated with the accuracy of the $F^o(\mathbf{k})$ and $\sigma(\mathbf{k})$ values; rather they relate to the precision of the numerical results. The accuracy of these properties is considered in a later section where a random-error algorithm is used to generate new sets of structure-factor magnitudes from $F^o(\mathbf{k})$ and $\sigma(\mathbf{k})$.

3. Topological analysis of DZP(low-angle data) charge densities

3.1. Covalent bonding interactions

Plots of the total charge densities for both CHF/DZP and HF/DZP wavefunctions are given in Fig. 2. Both plots show the same network of bond critical points and bond paths. Data for the bond and ring critical points are given in Table 3.

There is an increase in the amount of charge at the O1—H1 bond critical point of about 10% and a corresponding decreases in the charge densities at the O3—H2 (13%) and C1—C1′ (6%) bond critical points. Changes in $\rho(\mathbf{r}_b)$ at the C—O bond critical points are small, of the order of only 2%. These results are not unexpected, as a plot of the difference between the CHF/DZP and HF/DZP densities also showed the most significant differences in the charge density in the O—H bonding regions (Grimwood & Jayatilaka, 2001).

Bonding parameters obtained from the BLYP/DZP wavefunction are not significantly different from the HF/DZP values. This is reflected in the values of the χ^2 agreement statistic: 15.2 for the HF/DZP wavefunction and 11.5 for the BLYP wavefunction. While the BLYP/DZP wavefunction provides structure-factor magnitudes that are in better agreement with experiment, this agreement is still much worse than obtained after the constrained wavefunction calculation.

Comparison of bond critical point data obtained for the CHF/DZP and XD(low) charge densities shows the two to be quite similar. Values of $\rho(\mathbf{r}_b)$ differ by approximately 5%, although much larger differences are found for $\nabla^2 \rho(\mathbf{r}_b)$. These large differences in $\nabla^2 \rho(\mathbf{r}_b)$ can be explained by differences in the slope of the positive curvatures at the bond critical point, which in turn will depend on how the charge density is constructed. Differences are expected, however, and in previous studies *ab initio* charge densities have been compared with multipole-fitted charge densities by fitting the *ab initio* calculated structure factors with multipolar density functions (Howard *et al.*, 1992, 1995; Volkov, Abramov,

Coppens & Gatti, 2000; Volkov, Gatti, Abramov & Coppens, 2000). In this way, the experimental and theoretical charge densities are both expressed in terms of multipole functions and topological properties can be compared on an even footing. A distinct advantage of the procedure used in this paper is the direct comparison of properties obtained from the experimental data with *ab initio* values without the need for this projection of *ab initio* results into the multipole expansion.

3.2. Hydrogen-bonding interactions

The characterization of hydrogen bonding in terms of the topology of $\rho(\mathbf{r})$ is well defined (Bader & Essén, 1984; Cheeseman *et al.*, 1988; Popelier & Bader, 1992; Popelier, 1998a) and has been observed directly in experimental charge densities (Destro *et al.*, 1991; Macchi *et al.*, 2000). The curvature of $\rho(\mathbf{r})$ along the bond at the bond critical point (denoted λ_3) corresponding to the hydrogen-bonding interaction has also been used to classify various types of hydrogen bond
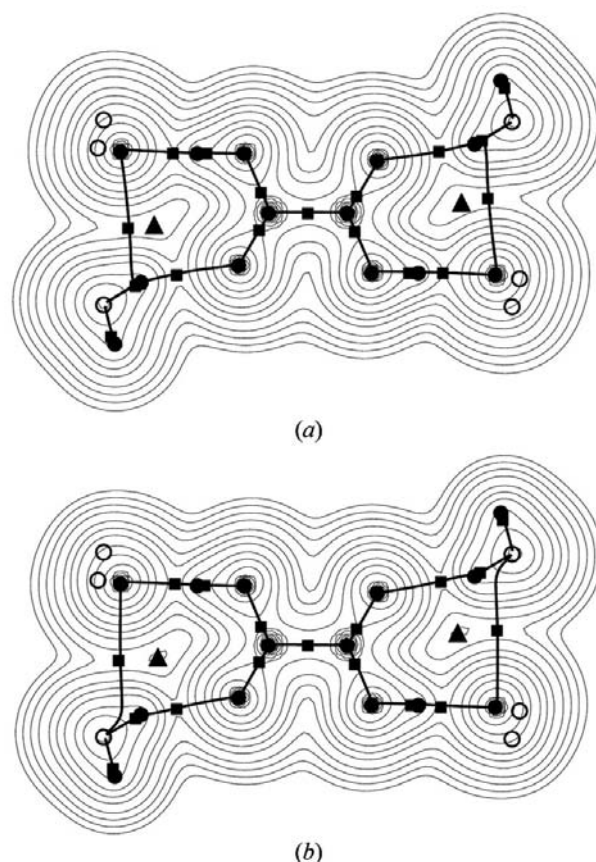

(a)

(b)

Figure 2
Plots of $\rho(\mathbf{r})$ for the (a) CHF/DZP and (b) HF/DZP wavefunctions of $C_2H_2O_4 \cdot 4H_2O$. The molecule is displayed in the same orientation as that shown in Fig. 1 and the plotting region is a 12 a.u. square centred on the inversion centre of the $C_2H_2O_4 \cdot 4H_2O$ molecule. Contours are drawn at intervals of $(2, 4, 8) \times 10^n$ a.u. ($n = -3$ to 3) with an additional contour at 0.001 a.u. Nuclei located in the plotting plane are denoted by solid circles, while those lying above and below the plotting plane are shown projected onto the plotting plane as open circles. Bond paths are overlaid onto the plots as thick lines, while bond critical points are denoted by solid squares and ring critical points as solid triangles.

Table 3
Bond critical point data for $C_2H_2O_4 \cdot 4H_2O$ obtained from the calculations of CHF, HF, BLYP and multipole-fitted total charge densities.

$\rho(\mathbf{r}_b)$ and $\nabla^2\rho(\mathbf{r}_b)$ are the values of the charge density and its Laplacian at the bond critical point, and $d(A)$ denotes the distance from the nucleus A to the bond critical point. All values are in atomic units. The numbering scheme in Fig. 1 is used to identify the nuclei.

	DZP basis set; low-angle data set						cc-pVTZ basis set; full data set			
	CHF	CHF(re)	HF	BLYP	XD	XD(re)	CHF(G)	CHF(M)	HF	XD
$\rho(\mathbf{r}_b)(C1-C1')$	0.26	0.26	0.27	0.25	0.25	0.26	0.26	0.26	0.27	0.26
$-\nabla^2\rho(\mathbf{r}_b)$	0.75	0.80	0.89	0.67	0.56	0.60	0.74	0.74	0.88	0.62
$d(C1)$	1.457	1.457	1.457	1.457	1.457	1.457	1.457	1.457	1.457	1.457
$\rho(\mathbf{r}_b)(C1-O1)$	0.35	0.35	0.34	0.34	0.38	0.39	0.35	0.35	0.35	0.43
$-\nabla^2\rho(\mathbf{r}_b)$	0.39	0.48	0.20	0.21	1.55	1.56	0.42	0.39	0.34	1.37
$d(C1)$	0.809	0.814	0.794	0.820	0.937	0.950	0.802	0.801	0.792	0.837
$\rho(\mathbf{r}_b)(C1-O2)$	0.41	0.41	0.40	0.39	0.44	0.44	0.42	0.40	0.42	0.43
$-\nabla^2\rho(\mathbf{r}_b)$	0.28	0.23	0.13	0.05	1.46	1.61	0.41	0.39	0.34	1.03
$d(C1)$	0.776	0.774	0.768	0.783	0.845	0.861	0.780	0.780	0.769	0.794
$\rho(\mathbf{r}_b)(O1-H1)$	0.41	0.41	0.37	0.37	0.42	0.41	0.43	0.43	0.38	0.40
$-\nabla^2\rho(\mathbf{r}_b)$	2.20	2.16	2.67	2.19	2.03	1.80	3.36	3.34	3.25	1.81
$d(O1)$	1.401	1.401	1.476	1.452	1.345	1.339	1.426	1.422	1.511	1.347
$\rho(\mathbf{r}_b)(O3-H2)$	0.44	0.44	0.50	0.49	0.46	0.46	0.49	0.49	0.52	0.45
$-\nabla^2\rho(\mathbf{r}_b)$	3.10	3.23	4.07	3.83	2.80	2.88	3.79	3.78	4.71	2.68
$d(O3)$	1.309	1.314	1.312	1.310	1.263	1.266	1.307	1.306	1.338	1.269
$\rho(\mathbf{r}_b)(O3-H3)$	0.40	0.40	0.42	0.41	0.40	0.40	0.43	0.43	0.43	0.39
$-\nabla^2\rho(\mathbf{r}_b)$	2.40	2.35	2.71	2.47	1.84	1.90	3.27	3.27	3.48	1.75
$d(O3)$	1.392	1.387	1.396	1.389	1.315	1.317	1.399	1.400	1.427	1.322
$\rho(\mathbf{r}_b)(O4-H4)$	0.40	0.41	0.42	0.41	0.40	0.40	0.43	0.43	0.43	0.39
$-\nabla^2\rho(\mathbf{r}_b)$	2.52	2.45	2.81	2.51	1.84	1.90	3.33	3.33	3.53	1.75
$d(O4)$	1.401	1.396	1.404	1.395	1.315	1.317	1.410	1.410	1.434	1.322
$\rho(\mathbf{r}_b)(O4-H5)$	0.44	0.44	0.50	0.49	0.46	0.46	0.49	0.49	0.52	0.45
$-\nabla^2\rho(\mathbf{r}_b)$	2.92	3.08	3.84	3.62	2.80	2.88	3.71	3.71	4.58	2.68
$d(O4)$	1.302	1.309	1.303	1.303	1.263	1.266	1.300	1.300	1.331	1.269
$\rho(\mathbf{r}_b)(O3\cdots H1)$	0.08	0.08	0.06	0.07	0.08	0.08	0.08	0.08	0.06	0.08
$\nabla^2\rho(\mathbf{r}_b)$	0.12	0.10	0.22	0.19	0.15	0.16	0.03	0.04	0.17	0.19
$d(O3)$	1.987	1.984	2.040	2.01	1.917	1.907	1.986	1.989	2.050	1.919
$\rho(\mathbf{r}_b)(O2\cdots H4)$	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.03
$\nabla^2\rho(\mathbf{r}_b)$	0.09	0.1	0.09	0.08	0.08	0.07	0.09	0.09	0.09	0.09
$d(O2)$	2.360	2.348	2.400	2.361	2.298	2.292	2.369	2.374	2.390	2.301
$\rho(\mathbf{r}_b)(O3\cdots O4)$	0.005	0.005	0.005	0.006	0.004	0.004	0.005	0.005	0.005	0.004
$\nabla^2\rho(\mathbf{r}_b)$	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.03
$d(O3)$	2.942	2.934	2.951	2.961	2.905	2.906	2.927	2.926	2.938	2.892
Ring critical point										
ρ	0.005	0.004	0.004	0.004	0.004	0.003	0.005	0.005	0.004	0.004

(Espinosa *et al.*, 1999). Characterization of the hydrogen-bonding interactions present in the α-oxalic acid dihydrate crystal, therefore, provides a further test of the applicability of this constrained wavefunction algorithm.

Fig. 2 shows the two distinct bond paths corresponding to hydrogen-bonding interactions. One connects O3 and H1 to form a hydrogen bond in which the oxalic acid molecule acts as the hydrogen donor; the other connects O2 and H4 to form a hydrogen bond in which the water molecule acts as the hydrogen donor. The $O3\cdots H1$ hydrogen bond is the shorter hydrogen bond and shows the most significant increase in the value of $\rho(\mathbf{r}_b)$, of about 25% in the CHF/DZP wavefunction compared with the HF/DZP value. This is offset somewhat by a reduction of $\rho(\mathbf{r}_b)$ at the $O3\cdots H1$ bond critical point by about 13%. The fitting procedure increases $\rho(\mathbf{r}_b)$ for the $O2\cdots H4$ bond by only 8%. This suggests that the longer hydrogen bond requires less reorganization of charge density and is adequately described at the HF/DZP level of theory without the guidance of experimental data.

The $O3\cdots H1$ and $O2\cdots H4$ hydrogen-bond critical points were also located in the XD(low) charge density. As with the bond critical points discussed above, the values of

$\rho(\mathbf{r}_b)(O3\cdots H1)$ and $\rho(\mathbf{r}_b)(O2\cdots H4)$ are similar to the *ab initio* values, with the largest differences occurring for the distance separating the bond critical point and the hydrogen nuclei, although this is only of the order of 5–7%.

Table 4 presents the characterization of these hydrogen bonds in terms of the λ_3 values, which are similar to those anticipated (Espinosa *et al.*, 1999). Interestingly, changes in λ_3 are small after the constrained wavefunction calculation (approximately 3%) even though the changes in $\rho(\mathbf{r}_b)$ were significant, indicating that this parameter is a robust descriptor of hydrogen bonding.

One interaction evident in Fig. 2 requires further comment. A bond critical point with a small value of $\rho(\mathbf{r}_b)$, of the order of 0.005 a.u. (atomic units), was located between the two water molecules and links O3 with O4. This bond critical point was located in *all* of the charge densities, as was the topologically necessary ring critical point. Although unexpected, and not noted in previous topological studies (Zobel *et al.*, 1992), the presence of these bond and ring critical points in all of the various total charge-density distributions suggests that it is indeed a real topological feature. This bond critical point is, therefore, indicative of a van der Waals bonding interaction

between these two oxygen nuclei of the sort often found in crystalline systems (Bader, 1998).

3.3. Integrated atomic properties

3.3.1. Net atomic and group charges. Net atomic, $q(\Omega)$, group and total charges are shown in Table 5. Overall, the charges obtained from each of the different wavefunctions are qualitatively similar: positively charged carbon and hydrogen atoms and negatively charged oxygen atoms. In the $C_2H_2O_4$ moiety, the CHF/DZP charges are as much as 30% and 19% lower for O1 and C1, respectively, and 30% higher for H1 when compared with the HF/DZP values. Comparison with the BLYP/DZP values shows that the charges for O1 and H1 are smaller by 13% and 19%, respectively, and 9% larger for O2. The charges on C1 are essentially the same and differ by only 0.5%. Differences in the charges for the two H_2O molecules show a similar trend.

These similarities in atomic charges suggest that the CHF/DZP wavefunction reproduces the effects of electron correlation inherent in the correlated BLYP/DZP calculation. The constrained wavefunction procedure results in modification to the HF/DZP wavefunction through the use of experimental

structure factors which 'inform' the *ab initio* HF/DZP calculation.

The integrated charges obtained from the CHF/DZP and XD(low) charge densities are qualitatively similar, although the CHF/DZP charges have larger absolute values, *i.e.*, $|q(\text{CHF/DZP})| > |q(\text{XD(low)})|$. Both predict the formation of anionic $C_2H_2O_4$ and cationic H_2O groups, while the unconstrained wavefunction does not show the same degree of charge separation. The charges obtained from the CHF/DZP and XD(low) charge densities show that even in the crystalline environment the $C_2H_2O_4 \cdot 2H_2O$ molecule has a tendency toward the formation of the $C_2O_4^{2-}$ and H_3O^+ ions found in solution.

3.3.2. Atomic kinetic energies. Integrated atomic kinetic energies, $K(\Omega)$, obtained from experimental charge densities have not previously been reported. Values of $K(\Omega)$ from the CHF/DZP calculations are compared with the corresponding unconstrained values, from HF/DZP and BLYP/DZP, in Table 6. Sums of individual atomic energies and values obtained directly from the wavefunction calculations are also shown. The agreement between the calculated total kinetic energy and that obtained from the sum of atomic fragments provides a further indication that the integration accuracy is acceptable.

The atomic kinetic energies from the CHF/DZP wavefunction show some differences with respect to those obtained from the HF/DZP wavefunction. The most notable difference is for the H1 atom which belongs to the oxalic acid molecule and is involved in a hydrogen-bonding interaction with a water molecule. The difference in energy of this atom is large: an increase of 40% on going from the unconstrained to the constrained wavefunction. The other hydrogen atom involved in hydrogen bonding, H4, is similarly more energetic in the constrained wavefunction, though by only 7%. This is consistent with the values of $\rho(\mathbf{r}_b)$ for these bonds, which also indicate that the shorter $O3 \cdots H1$ hydrogen bond is stronger than the longer $O2 \cdots H4$ bond.

The kinetic energies obtained from the BLYP/DZP wavefunction show closer agreement with the values obtained from the CHF/DZP wavefunction. The CHF/DZP value of $K(\text{H1})$ is

Table 4
Values of the positive curvature of the charge density at the bond critical points λ_3 (e Å^{-5}) for the hydrogen-bonding interactions in $C_2H_2O_4 \cdot 4H_2O$.

$d(\text{O})$ and $d(\text{H})$ are the distances from the bond critical point to the O and H nuclei, respectively (Å).

	CHF	CHF(re)	HF	BLYP	XD(low)	XD(low, re)
$O3 \cdots H1$						
λ_3	11.8	11.6	11.4	11.2	12.0	12.1
$d(\text{O})$	1.051	1.034	1.080	1.060	1.014	1.009
$d(\text{H})$	0.494	0.496	0.466	0.484	0.531	0.536
$O2 \cdots H4$						
λ_3	3.7	3.83	3.6	3.7	4.0	4.0
$d(\text{O})$	1.249	1.242	1.270	1.249	1.216	1.213
$d(\text{H})$	0.713	0.719	0.689	0.710	0.752	0.755

Table 5
Integrated net atomic charges, $q(\Omega)$ (a.u.), for $C_2H_2O_4 \cdot 4H_2O$ obtained from the various CHF, HF, BLYP and multipole-fitted total charge densities.

	DZP basis set; low-angle data set						cc-pVTZ basis set; full data set			
	CHF	CHF(re)	HF	BLYP	XD	XD(re)	CHF(G)	CHF(M)	HF	XD
C1	1.66	1.66	1.98	1.65	1.26	1.21	1.73	1.73	2.00	1.51
O1	−1.07	−1.08	−1.40	−1.21	−0.91	−0.89	−1.11	−1.11	−1.44	−1.01
O2	−1.26	−1.25	−1.34	−1.14	−1.06	−1.00	−1.25	−1.25	−1.35	−1.20
H1	0.56	0.56	0.73	0.67	0.47	0.44	0.58	0.58	0.76	0.45
$\sum q(C_2H_2O_4)$	−0.22	−0.20	−0.06	−0.06	−0.48	−0.48	−0.10	−0.10	−0.06	−0.50
O3	−1.07	−1.06	−1.28	−1.20	−0.77	−0.76	−1.11	−1.12	−1.32	−0.77
H2	0.57	0.58	0.66	0.64	0.49	0.50	0.59	0.59	0.69	0.53
H3	0.59	0.58	0.65	0.62	0.54	0.54	0.60	0.61	0.67	0.47
$\sum q(H_2O)$	0.09	0.10	0.03	0.06	0.26	0.28	0.08	0.08	0.04	0.23
O4	−1.15	−1.14	−1.32	−1.26	−0.77	−0.76	−1.18	−1.19	−1.36	−0.77
H4	0.62	0.61	0.67	0.62	0.49	0.50	0.62	0.62	0.69	0.47
H5	0.54	0.55	0.64	0.61	0.54	0.54	0.55	0.56	0.66	0.53
$\sum q(H_2O)$	0.01	0.02	−0.01	−0.03	0.26	0.28	−0.01	−0.01	0.01	0.23
$\sum q(\text{all})$	−0.02	−0.04	−0.02	−0.00	0.04	0.08	0.04	0.04	0.04	−0.04

Table 6
Integrated atomic kinetic energies, $K(\Omega)$ (a.u.), for $C_2H_2O_4 \cdot 4H_2O$ obtained from the various CHF, HF and BLYP wavefunctions.

	DZP basis set; low-angle data set				cc-pVTZ basis set; full data set		
	CHF	CHF(re)	HF	BLYP	CHF(G)	CHF(M)	HF
C1	36.974	36.961	36.468	36.778	37.037	37.025	36.471
O1	75.795	75.793	75.820	75.790	75.895	75.809	75.828
O2	75.980	75.996	75.567	75.559	75.886	75.914	75.569
H1	0.464	0.466	0.280	0.328	0.448	0.451	0.257
O3	75.694	75.627	75.727	75.765	75.648	75.666	75.775
H2	0.364	0.355	0.353	0.364	0.393	0.394	0.337
H3	0.362	0.374	0.338	0.358	0.372	0.373	0.321
O4	75.627	75.559	75.665	75.720	75.631	75.634	75.732
H4	0.354	0.365	0.329	0.355	0.365	0.367	0.257
H5	0.384	0.373	0.374	0.380	0.413	0.413	0.355
$\sum K(\Omega)$	683.996	683.738	681.842	682.794	683.996	684.092	681.919
K(wfn)†	683.99	683.74	681.833	682.793	684.009	684.097	681.918

† The value obtained from the *ab initio* wavefunction calculation.

larger by 29% while K(H4) is smaller by only 0.4% compared with the BLYP values. This suggests that, while the effects of electron correlation are included in the BLYP calculation, there appear to be contributions to the experimental atomic energy that are not accounted for by the BLYP method. Application of the constraint algorithm in conjunction with other quantum-chemical models should help clarify these differences in the atomic kinetic energies.

3.3.3. Magnitudes of atomic dipole moments. Atomic dipole moments are a measure of the displacement of the centroid of an atom's charge density from its nucleus (Bader, 1990). Magnitudes of the atomic dipole moments $|\mathbf{M}(\Omega)|$ are given in Table 7 and show marked differences between the CHF/DZP, HF/DZP, BLYP/DZP and XD(low) values.

Similar values for $|\mathbf{M}(\text{C1})|$ and $|\mathbf{M}(\text{O2})|$ were obtained from each of the three wavefunctions; however, for $|\mathbf{M}(\text{O1})|$, the CHF/DZP value is about 11% smaller than the HF/DZP value and 19% larger than the BLYP/DZP value. $|\mathbf{M}(\text{H1})|$ is also markedly different in the two HF calculations, the CHF/DZP value being 24% larger. The BLYP/DZP value is smaller than the CHF/DZP value, by 5%, which is consistent with the trend that BLYP/DZP values of $|\mathbf{M}(\Omega)|$ tend to be closer to the CHF/DZP values than the HF/DZP values. The XD(low) values of $|\mathbf{M}(\Omega)|$ are generally smaller than those obtained from the *ab initio* calculations. Such differences are probably related to the essentially unrestricted κ refinement model used here and its effect on the definition of the interatomic surfaces. Further investigation may require the use of a restricted κ-refinement model using suitably defined κ values (Abramov *et al.*, 1999).

3.3.4. Atomic volumes. Values of atomic volumes, $v(\Omega)$, are given in Table 8. The HF/DZP wavefunctions show changes in the atomic volumes of the carbon and hydrogen atoms compared with the CHF/DZP and BLYP/DZP volumes, which are similar. The volumes obtained from the XD(low) charge density are generally smaller than the CHF/DZP values, which may be a consequence of how the different calculations take into account the crystalline environment. In fact, the increased atomic volumes of the C1 and hydrogen atoms may be overly

pronounced since the use of $C_2H_2O_4 \cdot 4H_2O$ as the model cluster does not provide a wavefunction that takes into account neighbouring molecules, and it is these that in turn provide the conditions for the proper zero-flux boundaries for the atoms in the crystalline system. Of the properties calculated from the CHF/DZP wavefunction, therefore, $v(\Omega)$ is probably the least reliable. A solution to this problem is to include the neighbouring molecules in the constrained wavefunction calculation. While some attempt has been made to calculate constrained wavefunctions from oxalic acid dihydrate clusters (Grimwood & Jayatilaka, 2001), a major disadvantage of this approach is a significant increase in the computational resources required to obtain and analyse such a wavefunction.

3.3.5. The Laplacian of the charge density. It is possible, and desirable given the importance of electron pairs to chemistry, to analyse a wavefunction to find regions where the probability of finding electrons is greatest. One such analysis is provided by the 'electron localization function' (Becke & Edgecombe, 1990; Savin *et al.*, 1991), used in the previous study of α-oxalic acid dihydrate (Grimwood & Jayatilaka, 2001). Here the focus is upon the Laplacian of charge density, $L = -\nabla^2 \rho(\mathbf{r})$, which shows regions where charge density is locally concentrated ($L > 0$) or depleted ($L < 0$) (Bader, 1990).

Plots of L for the CHF/DZP and HF/DZP wavefunctions are given in Fig. 3. The plots are similar and show regions of charge concentration along the C—C, C—O and O—H bonds as well as in regions corresponding to the 'lone-pair' charge density of the oxygen atoms. Charge depletion is also apparent in both plots, in regions close to nuclei or between the $C_2H_2O_4$ and H_2O molecules. It is in these latter regions that the bond critical points corresponding to the different hydrogen bonds are located, as expected [see, for example, Table 2 of Espinosa *et al.* (1999)]. Although not shown, a plot of $\nabla^2 \rho(\mathbf{r})$ obtained from the BLYP wavefunction shows features analogous to those in Fig. 3.

The atomic graph of C1 and O2 has the expected trigonal arrangement of three bonding charge-concentration maxima, while O3 and O4 have the quasi-tetrahedral arrangement expected for two bonding and two non-bonding maxima (Bader, 1990; Gillespie & Popelier, 2001). Differences between the HF and CHF values of L at the charge-concentration maxima are of the order of 15–20%.[1]

The atomic graph for the O1 atom is of some interest, however, as it consists of only three charge-concentration maxima: one along each of the O1—C1 and O1—H1 bonds, and a single maximum attributable to non-bonding electron density. The presence of only one non-bonding charge-

---

[1] A detailed list of atomic graph information for the HF, CHF, CHF(re) and BLYP wavefunctions is available from the IUCr electronic archives (Reference: AU0274). Services for accessing these data are described at the back of the journal.

concentration maximum, rather than the two that might be expected from the Lewis electron-dot model, is similar to the situation in bent OX_2 molecules. Only in those molecules where X is more electronegative than oxygen (*e.g.* $X = NH_2$, OH or F) does the atomic graph for oxygen consist of two bonding and two non-bonding charge-concentration maxima. OX_2 molecules with less-electronegative X ligands (*e.g.* $X = BH_2$, SiH_3 or PH_2) have atomic graphs for the central oxygen atom consisting of two bonding and only one non-bonding charge-concentration maximum (Gillespie & Johnson, 1997; Gillespie & Popelier, 2001). The similarity of atomic graphs obtained for O1 from each of the wavefunctions is an indication that the valence density of this oxygen atom is not polarized sufficiently to form a quasi-tetrahedral arrangement of valence-shell charge-concentration maxima.

4. Estimates of the accuracy of topological properties

The relationship between these integrated atomic properties and variability in the experimental structure factors was investigated by constructing a quasi-experimental data set. This is a random-error data set that was synthesized by the application of a small perturbation to the observed structure-factor magnitudes:

$$F^{re}(\mathbf{k}) = F^{o}(\mathbf{k}) + A\,\sigma(\mathbf{k}), \qquad (4)$$

where $F^{re}(\mathbf{k})$ are the random-error structure factors, $F^{o}(\mathbf{k})$ are the low-angle experimental structure factors, and $\sigma(\mathbf{k})$ are the estimated errors in $F^{o}(\mathbf{k})$ obtained from the original experiment. A is a normally distributed random variable with a variance of unity and a mean of zero (Press *et al.*, 1992). Furthermore, it is also known (Applebaum, 1996) that the application of this algorithm to experimental data requires doubling of the experimental $\sigma^2(\mathbf{k})$ values and, consequently, the termination of the CHF procedure should be at $\chi^2 = 2$ when random-error data are used. As expected from this, and from the fact that the CHF/DZP calculations did not terminate at $\chi^2 = 1$, the final χ^2 value for the CHF(re)/DZP calculation is also greater than the optimum value of 2. Agreement statistics for the CHF(re)/DZP calculations and multipole refinements are included in Table 1 and the results of the analyses of the total charge densities derived from these random-error reflections are given in Tables 3–8.

Comparison of bond critical point data obtained from the CHF/DZP and CHF(re)/DZP wavefunctions, presented in Table 3, shows small differences in the values of $\rho(\mathbf{r}_b)$, of the order of only 1%. Differences in the distances from the nuclei to the bond critical points are similarly small: less than 1% for all bonds within the $C_2H_2O_4$ molecule as well as for the hydrogen bonds and the O3$\cdots$O4 bond. Within the two water molecules, there is a difference of approximately 2% in the distance of the hydrogen nuclei to the O—H bond critical point, while the difference is still less than 1% for the distance from the O nuclei to the O—H bond critical point. Values of $\rho(\mathbf{r}_b)$ at bond critical points obtained from the XD(low) and XD(low, re) charge densities are all similar, differing by 0.01 a.u. at most.

Values of the Laplacian of the charge density at the bond critical point, $\nabla^2\rho(\mathbf{r}_b)$, differ more significantly between CHF/DZP and CHF(re)/DZP and between XD(low) and XD(low, re), in some cases by as much as 24%. This is by far the largest difference and is a consequence of the nature of $\nabla^2\rho(\mathbf{r})$ at the bond critical point. The plots in Fig. 3 show that the change in $\nabla^2\rho(\mathbf{r})$ in the region of the bond critical points is
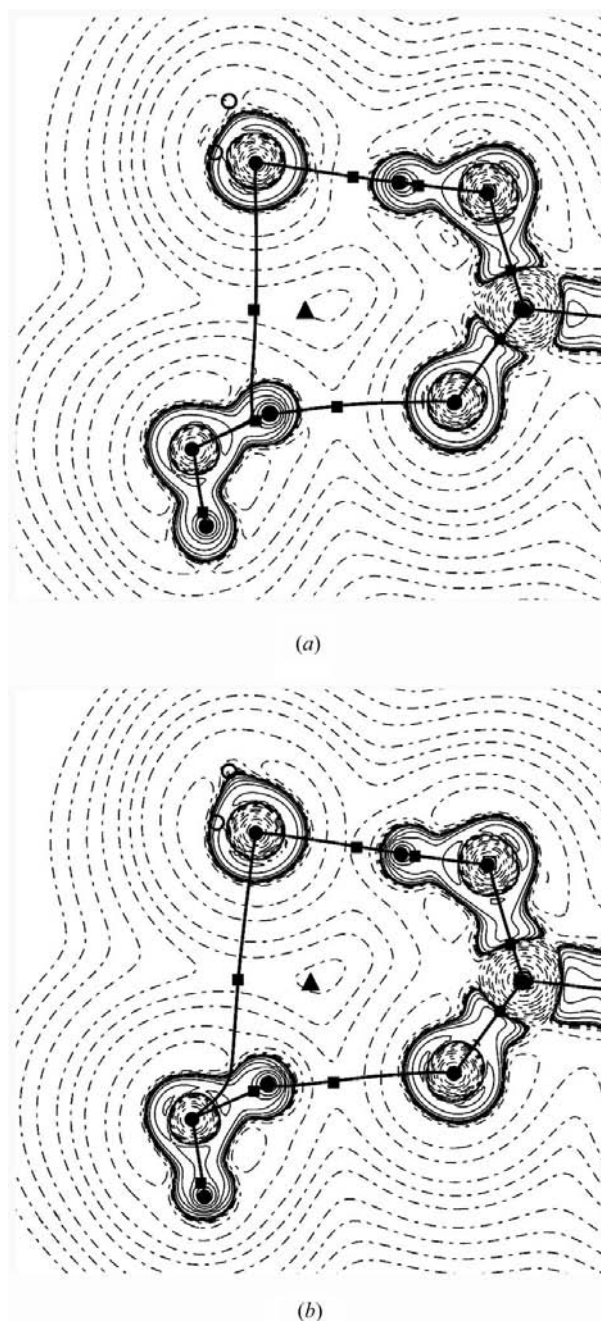

(a)

(b)

Figure 3
Plots of L for the (a) CHF/DZP and (b) HF/DZP wavefunctions of $C_2H_2O_4\cdot 4H_2O$. The plane of the plot is the same as that used in Fig. 2, showing only the unique portion of the molecule. Contours are drawn at intervals of $(\pm 2, \pm 4, \pm 8) \times 10^n$ a.u. ($n = -3$ to 3) with an additional contour at 0.001 a.u. Full lines denote regions of charge concentration ($L > 0$) and broken lines denote regions of charge depletion ($L < 0$).

Table 7
Atomic dipole moment magnitudes, $|\mathbf{M}(\Omega)|$ (a.u.), for $C_2H_2O_4 \cdot 4H_2O$ wavefunctions obtained from the CHF, HF, BLYP and multipole-fitted total charge densities.

	DZP basis set; low-angle data set						cc-pVTZ basis set; full data set			
	CHF	CHF(re)	HF	BLYP	XD	XD(re)	CHF(G)	CHF(M)	HF	XD
C1	0.82	0.82	0.80	0.75	0.73	0.74	0.81	0.81	0.80	0.82
O1	0.39	0.36	0.43	0.32	0.22	0.21	0.37	0.38	0.49	0.38
O2	0.53	0.56	0.66	0.49	0.37	0.36	0.56	0.56	0.70	0.56
H1	0.10	0.10	0.08	0.10	0.10	0.11	0.10	0.10	0.07	0.11
O3	0.17	0.20	0.30	0.25	0.03	0.02	0.26	0.25	0.40	0.03
H2	0.23	0.23	0.14	0.16	0.17	0.17	0.20	0.19	0.14	0.17
H3	0.18	0.18	0.14	0.16	0.16	0.16	0.17	0.17	0.14	0.18
O4	0.13	0.17	0.25	0.20	0.03	0.02	0.19	0.18	0.35	0.03
H4	0.15	0.15	0.12	0.15	0.16	0.16	0.14	0.14	0.15	0.18
H5	0.25	0.25	0.15	0.17	0.17	0.17	0.22	0.21	0.12	0.17

Table 8
Atomic volumes, $v(\Omega)$ (bohr^3), for $C_2H_2O_4 \cdot 4H_2O$ wavefunctions obtained from the CHF, HF, BLYP and multipole-fitted total charge densities.

	DZP basis set; low-angle data set						cc-pVTZ basis set; full data set			
	CHF	CHF(re)	HF	BLYP	XD	XD(re)	CHF(G)	CHF(M)	HF	XD
C1	42.2	42.3	31.0	38.0	38.6	42.2	37.7	37.6	29.5	37.9
O1	112.8	113.4	117.0	112.9	94.9	95.8	115.1	115.3	117.8	100.9
O2	129.4	127.7	128.4	123.2	105.0	103.6	129.3	129.6	129.0	108.9
H1	10.1	10.6	6.4	8.3	10.8	12.6	9.7	9.9	5.3	13.0
O3	124.9	124.7	126.3	123.7	102.4	102.0	128.5	126.9	129.3	105.8
H2	23.0	22.8	14.4	16.3	14.4	13.4	19.7	19.4	13.9	14.6
H3	21.5	21.2	16.5	18.8	16.3	16.4	20.1	20.0	15.8	16.9
O4	142.1	142.5	140.9	139.2	102.4	102.0	146.2	145.0	144.5	102.4
H4	13.6	14.6	11.2	13.4	13.9	13.4	13.2	13.2	15.2	16.9
H5	25.4	25.1	15.8	17.7	16.3	16.4	22.1	21.6	10.5	14.6

rapid; note the tightly spaced contours near the bond critical points. Large changes in $\nabla^2\rho(\mathbf{r})$ are, therefore, a consequence of steep gradients in the bonding region, despite only small changes in the location of the bond critical point. It appears that $\nabla^2\rho(\mathbf{r}_b)$ is not well defined and dependent on the structure-factor magnitudes used in the fitting procedure.

In all cases the distances from nuclei to the critical points corresponding to the hydrogen bonds deviate by only 0.5 to 2%. The λ_3 descriptor of these hydrogen bonds also shows only small differences between charge densities, of 1 to 4%.

The integrated atomic properties showed generally good agreement between the original and random-error charge densities. Net atomic charges obtained from the CHF/DZP wavefunctions agreed to within approximately 2%, while the largest difference in charges obtained from the XD densities was 6%. Atomic kinetic energies obtained from the CHF/DZP wavefunctions agreed to within 1% for the atoms in the $C_2H_2O_4$ moiety and to within 3% for the atoms in the H_2O molecules. Atomic volumes also show similarly small variation in the comparison of CHF/DZP values, with the majority (six) agreeing to within 1%, two varying by less than 2% and two by as much as 4%. Comparison of $v(\Omega)$ obtained from the XD densities shows generally good agreement (less than 2%)

although the volume of the C1 atom differs by almost 9%, which is probably because of the complex nature of the interatomic surfaces defining this atom which made it the most difficult to integrate.

Of the integrated atomic properties it is $|\mathbf{M}(\Omega)|$ that varies most between the two CHF/DZP wavefunctions. Within the $C_2H_2O_4$ portion of the molecule, the differences in the magnitude of the atomic dipole moment are of the order of 5% and as high as 8%. For the oxygen atoms within the H_2O molecules, the differences are 19 and 28%, although this represents differences of only 0.03 and 0.04 a.u., respectively. It should also be noted that the values of the individual components of these dipole moments are themselves quite small. For example, the values of the x component of the dipole moment for O3 are -0.041 and -0.077 for the CHF/DZP and CHF(re)/DZP wavefunctions, respectively. This represents a large relative difference, but only a small absolute difference in the value. Similarly, the values of $|\mathbf{M}(\Omega)|$ obtained from the XD densities differ by only 0.01 a.u., which also represents large percentage differences in some cases. In general, it is the components of the dipole moment with absolute values of less than about 0.1 a.u. (*i.e.* close to zero) that vary the most, and it would appear that, of the integrated atomic properties considered here, the dipole moment is the hardest to obtain accurately from experiment in this way.

Differences in the values of the (3, -3) critical points in L for the various atoms are also of interest in light of the variability found for $\nabla^2\rho(\mathbf{r}_b)$. As might have been anticipated, since the values of interest represent stationary points in L, the variation between wavefunctions is much smaller than was found for $\nabla^2\rho(\mathbf{r}_b)$ discussed above. Values of charge-concentration maxima in the valence shells of carbon and oxygen atoms differ by approximately 3 to 6%. The radii of these critical points show much smaller variation, of the order of 0.3%, indicating that the locations of these charge-concentration maxima are very well defined.

These results show that some integrated atomic properties are less variable than others, with the electron population and atomic kinetic energies changing the least. This finding is similar to previous work (Aicken & Popelier, 2000) which also showed that integrations of purely theoretical data using different integration grids resulted in stable values for $q(\Omega)$ and $K(\Omega)$. Additionally, the error estimates for these integrated atomic properties, obtained using random-error structure factors, are also similar to the 5% estimated by Flensburg & Madsen (2000) in their AIM analyses of multipolar charge-density distributions.

Table 9
Average absolute differences in topological properties between different charge densities.

The symbols are the same as those used in the previous tables. All values are in atomic units.

	CHF/DZP *versus* CHF(re)/DZP	CHF/DZP *versus* CHF(G)/cc-pVTZ	HF/DZP *versus* HF/cc-pVTZ	XD(low) *versus* XD(low, re)	XD(low) *versus* XD(full)
$\rho(\mathbf{r}_b)$	0.00	0.02	0.01	0.00	0.01
$\nabla^2\rho(\mathbf{r}_b)$	0.06	0.42	0.35	0.07	0.12
$d(A)$	0.005	0.007	0.017	0.006	0.018
$q(\Omega)$	0.01	0.02	0.03	0.02	0.06
$K(\Omega)$	0.021	0.040	0.028	–	–
$|\mathbf{M}(\Omega)|$	0.01	0.03	0.04	0.01	0.05
$v(\Omega)$	0.5	2.3	2.1	1.0	2.2

5. Topological analysis of cc-pVTZ(full data) charge densities

The systematic improvability of the CHF procedure within the conventional *ab initio* framework is perhaps the greatest strength of this approach and the major reason for pursuing it as a method for extracting information from diffraction data. Thus far, this work has focused upon the extraction of topological information from the charge density *via* the use of a double-ζ basis set and low-angle X-ray diffraction data. In order to improve the model used to study α-oxalic acid dihydrate, a constrained wavefunction was obtained using the larger cc-pVTZ basis set in conjunction with the full data set. Additionally, two different fragment partitioning methods (Jayatilaka & Grimwood, 2001) were used in the calculation of $F^c(\mathbf{k})$. A multipole refinement of the full data set was also performed for the sake of further comparison.

Bond critical point data obtained from the cc-pVTZ wavefunctions are given in Table 3. These data show that the different fragment partitioning models do not result in significantly different topological parameters. The CHF/cc-pVTZ bond critical point data are generally similar to the CHF/DZP results, $\rho(\mathbf{r}_b)$ change by less than 10% and the average difference in $\rho(\mathbf{r}_b)$ from the two charge densities is 0.02 a.u. (see Table 9). Distances of the bond critical points to nuclei differ by less than 1%; however, the values of $\nabla^2\rho(\mathbf{r}_b)$ changed significantly.

Comparison of the unconstrained HF wavefunctions shows similar trends in the agreement of bond critical point data: values of $\rho(\mathbf{r}_b)$ agree to within approximately 5%, while values of $\nabla^2\rho(\mathbf{r}_b)$ vary by as much as 40%. Such basis-set dependence of bond critical point data is consistent with other work, which also noted the dependence of some bond critical point properties on the level of theory used to calculate electron-density distributions (O'Brien & Popelier, 1999). Absolute differences in topological parameters are given in Table 9 and show that variation in bond critical point parameters between the CHF/DZP(low) and CHF/cc-pVTZ(full) data is similar to the differences encountered when only the basis set was changed and no constraints were applied; *cf.* data columns two and three in Table 9. While the effects of including higher-angle data in the constrained wavefunction procedure should not be ignored, change in the bond critical point data appears to result largely from the change in basis set.

The XD(low) and XD(full) charge densities are also compared in Table 9. The bond critical point data agree in much the same way as was found for the *ab initio* wavefunctions: close agreement in both $\rho(\mathbf{r}_b)$ and the distance between bond critical points and nuclei (see Table 9), and marked changes in $\nabla^2\rho(\mathbf{r}_b)$. As is the case with different quantum-chemical models, $\nabla^2\rho(\mathbf{r}_b)$ is sensitive to changes in the data used in the multipole refinement. This general agreement of bond critical point data between the two multipole models, $\nabla^2\rho(\mathbf{r}_b)$ excepted, is anticipated since the high-angle reflections should have only a small influence on the valence electron density, with which these topological descriptors of bonding are primarily concerned.

Integrated atomic charges from the cc-pVTZ wavefunctions are also given in Table 5. It appears that the choice of partitioning method is not critical for the determination of AIM properties. Wavefunctions derived from the Gaussian and Mulliken partitioning algorithms for the calculation of $F^c(\mathbf{k})$ are similar, and $q(\Omega)$ differ from the unconstrained values in much the same way as above for the DZP calculations.

Differences in integrated charges between the CHF/DZP and CHF/cc-pVTZ data are of the order of 5%, and differences between unconstrained wavefunctions are also approximately 5%, representing an average difference of about 0.02 a.u. (Table 9). Since these changes are approximately the same, whether or not we are dealing with constrained or unconstrained wavefunctions, it also suggests that changes in basis set have a greater influence on integrated charges than the inclusion of the high-angle reflections in the calculation of the CHF wavefunctions. This is not the case for $q(\Omega)$ obtained from the multipole-fitted densities, which differ by as much as 17% upon inclusion of the high-angle data in the refinement. It appears that, without further refinement of thermal parameters, multipole refinement of experimental data produces results that are much less stable with respect to the addition of more experimental data, and that the dominant error in the CHF calculations arise from the controllable assumptions, *e.g.* the choice of basis set.

Atomic kinetic energies from the integration of the cc-pVTZ wavefunctions are given in Table 6. The CHF/cc-pVTZ values agree to within 0.2% with the CHF/DZP values for the carbon and oxygen atoms, and to within 3–7% for the hydrogen atoms. In absolute terms, these percentage differences represent energies of the order of 250 kJ mol^{-1} for

carbon and oxygen atoms, and 75 kJ mol^{-1} for hydrogen atoms, with an average difference of about 105 kJ mol^{-1} (Table 9). Similar differences in $K(\Omega)$ from the unconstrained DZP and cc-pVTZ wavefunctions were also found for the heavy atoms, although hydrogen atoms differed by as much as 20%. The agreement between the CHF/DZP and CHF/cc-pVTZ wavefunctions is encouraging as it suggests convergence; however, these small percentage differences represent amounts that are large in chemical terms and so are probably not suitable for quantitative purposes involving energetics.

The magnitudes of atomic dipole moments calculated from the cc-pVTZ wavefunctions are shown in Table 7. Comparison of the constrained cc-pVTZ results with the corresponding DZP values shows reasonable agreement for the $C_2H_2O_4$ portion of the molecule and differences on atoms of the H_2O molecules in the range 5–35%. Differences between the other pairs of results [XD(low) and XD(full), HF/DZP and HF/cc-pVTZ, and CHF/DZP and CHF(re)/DZP] show more variability over the whole molecule, although values of $|\mathbf{M}(\Omega)|$ are generally small and absolute differences are of the order of 0.01–0.05 a.u.

Atomic volumes from the CHF/cc-pVTZ calculations, given in Table 8, show reasonable agreement with the CHF/DZP values. The exceptions are $v(\Omega)$ of the hydrogen atoms, which differ by as much as 20%. Similar trends are found in the comparison of the XD(low) and XD(full) results, as well as the comparison of integrated atomic volumes obtained from the unconstrained wavefunctions. As mentioned above, the atomic volumes are probably the least accurate of the integrated properties, as a boundary at 0.001 a.u. is required to envelope regions not bounded by zero-flux surfaces. Nevertheless, the agreement between the DZP and cc-pVTZ constrained wavefunctions, particularly for the heavy atoms, indicates that experimental values for atomic volumes can be extracted.

Taken *in toto*, the results in this section demonstrate that the results obtained from any constrained wavefunction calculations can be systematically improved by increasing both the basis-set size and by including more experimental data in the constrained wavefunction procedure. Such improvement is not readily available in a multipole refinement of experimental data, where basis functions cannot be easily modified (Chandler & Spackman, 1982; Volkov *et al.*, 2001; Volkov & Coppens, 2001) and one must always be careful about the ratio of experimental data to fitted parameters. These CHF/cc-pVTZ results also provide valuable information about the variability of AIM properties derived from CHF wavefunctions and, importantly, suggest that the wavefunction obtained from the DZP/low-angle calculations is already close to the single-determinant experimental wavefunction.

6. Conclusions

The most important point of the work described here is that experimental data have been used to calculate integrated atomic properties that were previously restricted to the domain of *ab initio* quantum theory. The constrained wave-function procedure, therefore, provides more than just an alternative to the conventional approach of fitting multipolar density functions to the same experimental data.

Topologies of calculated and constrained charge densities are all qualitatively similar, although there are quantitative differences. That is, the number and type of critical points in both $\rho(\mathbf{r})$ and L obtained from the different charge-density distributions are the same, although the values themselves vary, in some cases substantially.

An interesting and previously uncharacterized feature common to all of the charge densities was the location of a bond critical point between the oxygen atoms of two water molecules, as well as the topologically necessary ring critical point. While the deeper significance of this unanticipated feature in the total charge density is unclear, it does highlight the need to examine the topology of $\rho(\mathbf{r})$ in a systematic fashion, guided by chemical intuition but not bound by it.

The accuracy of the integrated atomic properties derived from the CHF/DZP calculation was probed by using a set of random-error structure factors derived from the experimental data, which were also subjected to a constrained HF calculation. The resultant wavefunction showed changes to properties of approximately 5%, though in some cases, most noticeably those of $\nabla^2\rho(\mathbf{r}_b)$ and $|\mathbf{M}(\Omega)|$, the changes were much larger. Values of properties such as $\nabla^2\rho(\mathbf{r}_b)$, which vary steeply in the region being evaluated, or $|\mathbf{M}(\Omega)|$, which may be quite small, should, therefore, be used cautiously; otherwise it appears that the properties are accurately determined from the data and the assumptions of the model.

The accuracy of topological properties was also assessed by performing calculations with the larger cc-pVTZ basis set and experimental data extended to include high-angle reflections. Bond critical point data and integrated atomic properties generally agreed with the DZP/low-angle constrained wavefunction, with the exception of $\nabla^2\rho(\mathbf{r}_b)$ and $|\mathbf{M}(\Omega)|$ for the reasons already mentioned. The overall general agreement of properties calculated using the two levels of theory is an encouraging finding as it shows how the use of experimental data can be used to overcome any basis-set deficiencies and, importantly, suggests that essentially the same experimental wavefunction is obtained from the two calculations.

References

Abramov, Y. A., Volkov, A. & Coppens, P. (1999). *Chem. Phys. Lett.* **311**, 81–86.
Aicken, F. M. & Popelier, P. L. A. (2000). *Can. J. Chem.* **78**, 415–426.
Applebaum, D. (1996). *Probability and Information: an Integrated Approach.* Cambridge University Press.

Bader, R. F. W. (1990). *Atoms in Molecules: a Quantum Theory*. Oxford University Press.

Bader, R. F. W. (1998). *J. Phys. Chem. A*, **102**, 7314–7323.

Bader, R. F. W. & Essén, H. (1984). *J. Chem. Phys.* **80**, 1943–1960.

Becke, A. D. (1988). *Phys. Rev. A*, **38**, 3098–3100.

Becke, A. D. & Edgecombe, K. (1990). *J. Chem. Phys.* **92**, 5397–5403.

Bianchi, R., Gervasio, G. & Marabello, D. (2000). *Inorg. Chem.* **39**, 2360–2366.

Boese, R., Niederprüm, N., Bläser, D., Maulitz, A., Antipin, M. & Mallinson, P. R. (1997). *J. Phys. Chem. B*, **101**, 5794–5799.

Bytheway, I., Figgis, B. N. & Sobolev, A. N. (2001). *J. Chem. Soc. Dalton Trans.* pp. 3285–3294.

Chandler, G. S., Figgis, B. N., Reynolds, P. A. & Wolff, S. K. (1994). *Chem. Phys. Lett.* **225**, 421–426.

Chandler, G. S. & Spackman, M. A. (1982). *Acta Cryst.* A**38**, 225–239.

Cheeseman, J. R., Carroll, M. T. & Bader, R. F. W. (1988). *Chem. Phys. Lett.* **143**, 450–458.

Coppens, P. (1992). *Annu. Rev. Phys. Chem.* **43**, 663–692.

Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. Oxford University Press.

Coppens, P., Volkov, A., Abramov, Y. & Koritsánszky, T. (1999). *Acta Cryst.* A**55**, 965–967.

Destro, R., Bianchi, R., Gatti, C. & Merati, F. (1991). *Chem. Phys. Lett.* **186**, 47–52.

Dunning, T. H. (1970). *J. Chem. Phys.* **53**, 2823–2833.

Dunning, T. H. (1989). *J. Chem. Phys.* **90**, 1007–1023.

Espinosa, E., Souhassou, M., Lachekar, H. & Lecomte, C. (1999). *Acta Cryst.* B**55**, 563–572.

Flaig, R., Koritsánszky, T., Zobel, D. & Luger, P. (1998). *J. Am. Chem. Soc.* **120**, 2227–2238.

Flensburg, C. & Madsen, D. (2000). *Acta Cryst.* A**56**, 24–28.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Stratmann, R. E. Jr, Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S. & Pople, J. A. (1998). *GAUSSIAN98*, Revision A.7. Gaussian, Inc., Pittsburgh, USA.

Gillespie, R. J. & Johnson, S. A. (1997). *Inorg. Chem.* **36**, 3031–3039.

Gillespie, R. J. & Popelier, P. L. A. (2001). *Chemical Bonding and Molecular Geometry*. Oxford University Press.

Grimwood, D. J. & Jayatilaka, D. (2001). *Acta Cryst.* A**57**, 87–100.

Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.

Henderson, G. A. & Zimmermann, R. K. (1976). *J. Chem. Phys.* **65**, 619–622.

Howard, S. T., Hursthouse, M. B., Lehmann, C. W., Mallinson, P. R. & Frampton, C. S. (1992). *J. Chem. Phys.* **97**, 5616–5630.

Howard, S. T., Hursthouse, M. B., Lehmann, C. W. & Poyner, E. A. (1995). *Acta Cryst.* B**51**, 328–337.

Iversen, B. B., Larsen, F. K., Figgis, B. N. & Reynolds, P. (1997). *J. Chem. Soc. Dalton Trans.* pp. 2227–2240.

Jayatilaka, D. (1998). *Phys. Rev. Lett.* **80**, 798–801.

Jayatilaka, D. & Grimwood, D. J. (2000). *TONTO. A Research Tool for Quantum Chemistry*. The University of Western Australia, Nedlands, Western Australia, Australia.

Jayatilaka, D. & Grimwood, D. J. (2001). *Acta Cryst.* A**57**, 76–86.

Koritsánszky, T., Buschmann, J., Lentz, D., Luger, P., Perpetua, G. & Röttger, M. (1999). *Chem. Eur. J.* **5**, 3413–3420.

Koritsánszky, T., Buschmann, J. & Luger, P. (1996). *J. Phys. Chem.* **100**, 10547–10553.

Koritsánszky, T. & Coppens, P. (2001). *Chem. Rev.* **101**, 1583–1627.

Koritsánszky, T., Zobel, D. & Luger, P. (2000). *J. Phys. Chem. A*, **104**, 1549–1556.

Lee, C., Yang, W. & Parr, R. G. (1988). *Phys. Rev. B*, **37**, 785–789.

Lippmann, T. & Schneider, J. R. (2000). *Acta Cryst.* A**56**, 575–584.

Macchi, P., Iversen, B. B., Sironi, A., Chakoumakos, B. C. & Larsen, F. K. (2000). *Angew. Chem. Int. Ed. Engl.* **39**, 2719–2722.

Macchi, P., Proserpio, D. M. & Sironi, A. (1998a). *J. Am. Chem. Soc.* **120**, 13429–13435.

Macchi, P., Proserpio, D. M. & Sironi, A. (1998b). *J. Am. Chem. Soc.* **120**, 1447–1455.

Mukherj, A. & Karplus, M. (1963). *J. Chem. Phys.* **38**, 44–48.

O'Brien, S. E. & Popelier, P. L. A. (1999). *Can. J. Chem.* **77**, 28–36.

Popelier, P. L. A. (1998a). *J. Phys. Chem. A*, **102**, 1873–1878.

Popelier, P. L. A. (1998b). *Comput. Phys. Commun.* **108**, 180–190.

Popelier, P. L. A. & Bader, R. F. W. (1992). *Chem. Phys. Lett.* **189**, 542–548.

Popelier, P. L. A. & Bone, R. G. A. (1998). *MORPHY98*. UMIST, Manchester, England.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in Fortran 77*, 2nd ed., pp. 279–280. Cambridge University Press.

Rosso, K. M., Gibbs, G. V. & Boisen, M. B. (1999). *Phys. Chem. Miner.* **26**, 264–272.

Savin, A., Becke, A. D., Flad, J., Nesper, R., Preuss, H. & von Schnering, H. G. (1991). *Angew. Chem. Int. Ed. Engl.* **30**, 409–412.

Schaftenaar, G. & Noordik, J. H. (2000). *J. Comput. Aided Mol. Design*, **14**, 123–134.

Smith, G. T., Mallinson, P. R., Frampton, C. S., Farrugia, L. J., Peacock, R. D. & Howard, J. K. (1997). *J. Am. Chem. Soc.* **119**, 5028–5034.

Souhassou, M. & Blessing, R. H. (1999). *J. Appl. Cryst.* **32**, 210–217.

Volkov, A., Abramov, Y. A. & Coppens, P. (2001). *Acta Cryst.* A**57**, 272–282.

Volkov, A., Abramov, Y., Coppens, P. & Gatti, C. (2000). *Acta Cryst.* A**56**, 332–339.

Volkov, A. & Coppens, P. (2001). *Acta Cryst.* A**57**, 395–405.

Volkov, A., Gatti, C., Abramov, Y. & Coppens, P. (2000). *Acta Cryst.* A**56**, 252–258.

Zobel, D. (1996). Private communucation.

Zobel, D., Luger, P., Dreissig, W. & Koritsánszky, T. (1992). *Acta Cryst.* B**48**, 837–848.