PERGAMON

Solid State Communications 116 (2000) 395–400

solid state communications

www.elsevier.com/locate/ssc

Electron density in the peptide bonds of crambin

M.V. Fernández-Serra[a,*], J. Junquera[a], C. Jelsch[b], C. Lecomte[b], E. Artacho[c]

[a]*Departamento de Física de la Materia Condensada, C-III, Universidad Autónoma de Madrid, 28049 Madrid, Spain*
[b]*Laboratoire Cristallographie et Modelisation des Materiaux Mineraux et Biologiques, CNRS, Universite H. Poincare, BP 239 54506 Vandoeuvre-les-Nancy Cedex, France*
[c]*Departamento de Física de la Materia Condensada, C-III, and Instituto Nicolás Cabrera, Universidad Autónoma de Madrid, 28049 Madrid, Spain*

Received 7 February 2000; accepted 31 July 2000 by F. Ynduràin

Abstract

The electron density in the peptide bonds of crambin, a plant-seed hydrophobic globular protein with 46 residues and 642 atoms, is studied both theoretically and experimentally. The results of density functional calculations of crambin in vacuo for the deformation electron density in its peptide bonds are compared with the electronic distribution obtained from ultra-high-resolution X-ray crystallography. The comparison is centered on the average peptide-bond map, where the experimental results are clearest. Theory is then used to ascertain on differences among peptide bonds in different chemical environments. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: A. Organic crystals; D. Electronic states

PACS: 71.20.Rv; 87.15.−v

If the knowledge of the electronic charge distribution is important in general in condensed matter systems [1], it is in the realm of biomolecules where the shape of the electron density is of greatest significance. Molecular recognition, drug design, enzymatic reactivity, docking, are all situations where the "electronic shape" of the molecule is crucial [2,3].

X-ray diffraction (XRD) from the experimental side, and first-principles calculations from the theoretical side, are the tools generally used for the characterization of the electron density. Recent developments in both fields (ultra-high-resolution XRD, and linear-scaling first-principles methodologies) have allowed to jump from small-sized molecules such as small peptides [4], to this first analysis of a whole protein. The feasibility of charge density studies on larger systems has been evaluated in Ref. [5]. Crambin is an hydrophobic globular protein found in the seeds of a plant, *crambe abyssinica*, and it is homologous to membrane-active plant toxins. It is composed of a chain of 46 amino acids with 642 atoms [6]. The crystal structure at atomic resolution has been extensively described [6–8], and the experimental charge density of the protein has already been presented elsewhere [9]. This paper is thus dedicated to the comparison with the predictions of first-principles calculations and the new information offered by them.

The present study is centered on the deformation density, where the sum of electron densities of isolated atoms is substracted from the total electron density of the system. The calculations were performed within the numerical atomic-orbital (NAO) method for linear-scaling density functional (DFT) [10–12], using the Siesta code implementation [11]. There was a previous linear-scaling Hartree–Fock calculation that focused on stereochemical properties [13]. The generalized gradient approximation (GGA) [14] of Kohn–Sham [15] theory was used for exchange and correlation. Core electrons were replaced by scalar-relativistic norm-conserving pseudopotentials [16] in their fully non-local formulation [17]. Valence electrons were described using a basis set of finite-range numerical atomic orbitals. They are based on the scheme of pseudo-atomic orbitals of Sankey and Niklewski [18,19], but generalized to arbitrarily complete bases with multiple-ζ, polarization, diffuse, and off-site orbitals [12]. A uniform mesh with a plane-wave cutoff of 100 Ry is used to represent the Hartree and exchange-correlation potentials and the local part of the pseudopotential. The method has been applied to a variety of systems, some of them reviewed in Ref. [12], including

* Corresponding author.

PII: S0038-1098(00)00334-3

biomolecules [20], where it has been thoroughly tested. The crambin coordinates used are derived from the crystallographic refinement [9]. In the disordered parts of the protein [8], only the main conformation was kept.

X-ray diffraction data from a single crambin crystal were measured to ultra-high-resolution (0.54 Å) at low temperature $T = 100$ K [9]. The diffraction intensities were collected using synchrotron radiation and a 300 mm MarResearch imaging plate detector at the EMBL BW7A beamline at the DORIS storage ring, DESY, Hamburg. The data were reduced to a 97.6% complete set of 112,293 symmetry-independent reflection intensities. The diffraction data had an overall measurement redundancy of 4.3 and the average internal-reliability precision index was $R_{\mathrm{merge}}(\|F\|^2) = 5.5\%$. After the structure refinement, the final crystallographic reliability indexes were $R_{\mathrm{w}}(F) = 9.0\%$ and $R_{\mathrm{wfree}}(F) = 9.4\%$.

The analysis of the experimental data was performed using a simplified Hansen and Coppens model [21]:

$$\rho_{\mathrm{atom}} = \rho_{\mathrm{core}} + P_{\mathrm{val}}\rho_{\mathrm{val}} + \sum_{l=0}^{l_{\max}} R_l \sum_{m=0}^{l} P_{lm} Y_l^m \qquad (1)$$

The first two terms on the right describe spherically symmetric core-plus-valence density, and the third term describes nonspherical redistribution of valence-shell density due to chemical bonding. The R_l are Slater-type radial functions, and the Y_l^m are real spherical harmonic angular functions. This model describes the atomic electron density in terms of charges (P_{val} is the valence population) and multipoles P_{lm}.

The MOLLY [21] software was extensively modified for protein applications notably by introducing stereochemical and dynamical restraints. As previously shown [5,9,22], the spherical-atom approximation is not enough to describe these ultra-high-resolution data. The charge density parameters for the main-chain peptide moiety were refined as follows. The initial multipoles and charges were transferred from our database of average parameters derived from several crystal structures of amino acids and small peptides [23]. The average electron density parameters of the polypeptide main chain were refined against the crambin diffraction data. The repetitive nature of the polypeptide main-chain allowed to constrain all the $-C_a-(C{=}O)-NH-$ motifs to be equivalent and consequently to improve the deformation density signal. Although these motifs display a large variety of conformations, an average electron density can be obtained, as the multipoles definition is based on a local axis system which is unaffected by rotational degrees of freedom in the structure. As the oxygen lone pairs turned out to be underestimated in the refined density, the oxygen multipoles were set at their database value. During this refinement process, the progressive flattening of the peptide-averaged residual electron density features [9] was convincing physical evidence of real improvement in the modeling.

The peptide bond is made by the condensation of the carboxylic group of an amino-acid and the amine group of the next residue. The peptide moiety H–N–C=O is coplanar due to the π-like electron delocalization. The two adjacent C_a atoms are also nearly in this plane, the root-mean square deviation being 0.09 Å. The C_a–N–H–(C=O)–C_a motif of the polypeptide main chain displays a considerable rigidity, the relative disposition of these atoms being very similar for any peptide bond in a protein. Two-dimensional deformation-density maps are drawn for the plane containing the atoms involved in a peptide bond. These maps allow the best comparison with experiment, since they can be averaged over all the peptide bonds in the molecule, thereby reducing the experimental noise. In a second step, deviations from the average are analyzed.

The peptide group between the two amino-acids alanine and glycine in a dipeptide is used for testing the approximations used in the first-principles scheme directly on the magnitude of interest, the deformation-density map.

Exchange-correlation functional and pseudopotentials. Deformation-density maps for this peptide bond have been obtained for two different exchange-correlation (XC) potentials: the local-density approximation (LDA) [24] and the generalized-gradient approximation (GGA) of Perdew et al. [14]. The maps are essentially identical (figure not shown here[1]), suggesting that the deviations from the XRD maps (see below) are not to be attributed to DFT. In the following parts of this study, the GGA functional is always used. Similarly, different pseudopotentials (always reasonably transferable) have been compared giving the same sense of robustness: the different maps were indistinguishable.[1] This leaves the basis set as the main approximation that may affect the quality of the maps. It must be emphasized, however, that this conclusion is particular for the present system (good chemical bonds between light elements) and for the object of the study (the deformation density). The magnetization of iron, for instance, requires more care in the definition of the functional and the pseudopotential than for the basis [25].

Basis size. In spite of the vast literature available on the subject, our inclusion of a basis-size study here is justified by the novelty of the numerical atomic orbitals employed. Fig. 1 shows the results for the six basis sets tested. The deformation density, at the level pursued in this work, is adequately converged within a double-ζ singly polarized scheme (DZP) [12], similar to what was found for other biomolecules [20]. The oxygen lone pairs are overestimated when using a poor basis set, like single-ζ, as already reported [26,27]. The maps show the importance of the polarization functions for the accurate description of the electron density. This fact was already pointed out for Gaussian-type basis sets [4,28]. To quantify the convergence, the distance between two densities (or deformation densities), defined as $N_{\mathrm{e}} D_{ij}^2 = \Omega \int_\Omega [\rho_i(\mathbf{r}) - \rho_j(\mathbf{r})]^2 \, \mathrm{d}^3\mathbf{r}$, is computed for the basis sets considered, with respect to the largest one, a triple-ζ with double polarization. Ω is the

[1] The maps are available from the authors upon request.

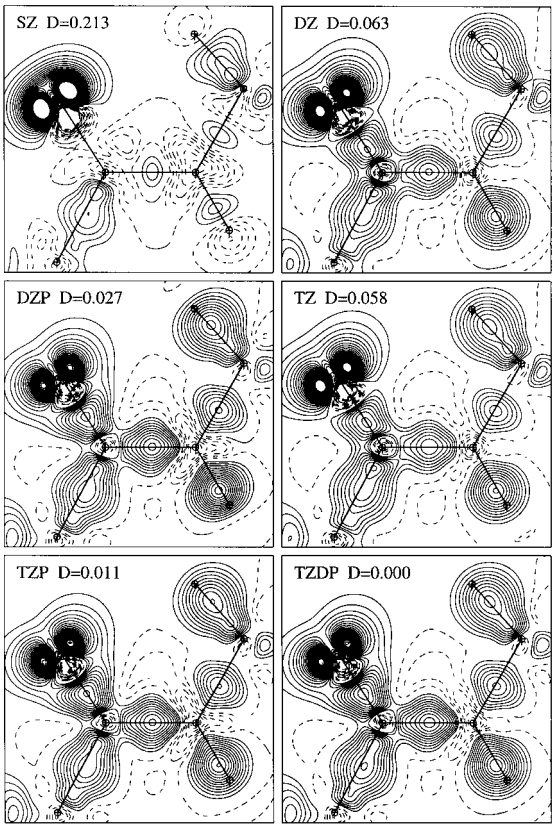

Fig. 1. Deformation-density maps for the peptide bond of the Ala–Gly dipeptide for different basis sets. SZ, DZ, and TZ stand for single-ζ, double-ζ, and triple-ζ, respectively. The split-norm [12] is 15%, except for hydrogen 50%. P and DP stand for single and double polarization, respectively. The values of D measure the distance between the corresponding density and the one of TZDP. The range of the basis is long in all cases. The contour level spacing is 0.05 $e^-/\text{Å}^3$, solid line standing for positive deformation, dashed for negative.

volume enclosed by the isosurface of constant density $\rho = 10^{-4}\ e^-/\text{Å}^3$, and N_e is the total number of electrons of the molecule. The D_{ij} distance is thus an adimensional number, between 0 and 1. The results are included in Fig. 1.

Basis range. The finite range of the basis functions is essential for achieving linear-scaling computations. Furthermore, the efficiency in the calculations depends quite dramatically on the range. Fig. 2 shows the results for two ranges, a short and a long one. It is clear from the figure that the short range is not converged for the deformation density. However, the electron density of the *molecule* (the one expensive to calculate) is well converged even for the short-range basis, the problem being the convergence of the sum of atomic densities. This is realized when the deformation density is built from the molecular electron density with short-range basis, and the atomic densities with short-range basis, and the atomic densities with long-range basis (see (b) and (c) in Fig. 2). The calculation of the protein can therefore be performed with a short-range basis, the atomic calculations being then done with a range as long as desired.

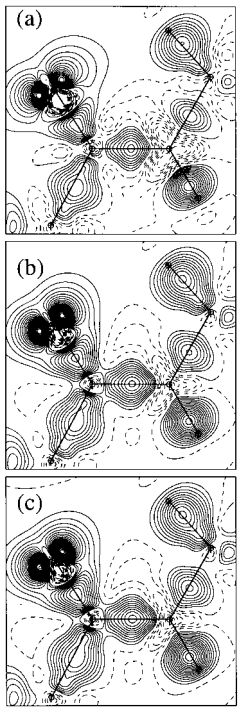

Fig. 2. Alanine–glycine dipeptide: effect of basis finite range. (a) Short range for both the system and the atomic densities. (b) Long range for both. (c) Short range for the system density and long for the atomic densities. The orbital radii for the short-range basis are (in a.u.): 4.2 for H, 4.1 for C, 3.6 for N, 3.2 for O, and 4.2 for S. The radii of the long-range basis are of 9 a.u. for all species. Contour spacing as in Fig. 1, the basis is DZP in all cases.

Average map for crambin. Fig. 3 shows the comparison of (a) experiment and (b) theory, both for the electron deformation density map averaged over all the peptide bonds in the crambin molecule. The DFT calculation has been done with the DZP basis with short range for the molecule and long range for the atoms. In the comparison, the regions of the two C_a–H_a bonds (upper right and lower-left in Fig. 3) should not be considered, as the differences are essentially of geometric origin: the C_a–H_a bond has various orientations with respect to the plane of the rigid peptide bond.

The general agreement is quite satisfactory, notably in the atomic bond regions (Fig. 3d). The two maps display a correlation coefficient [9] of 86%, if the C_a–H_a regions are omitted. Also the root mean square value of the deformation density is quite similar: 0.196 and 0.226 in the experimental and theoretical maps, respectively. The experimental map seems to be about 10% underestimated as already pointed out [9]. The clearest differences are:

(i) The deformation density around the oxygen atom is

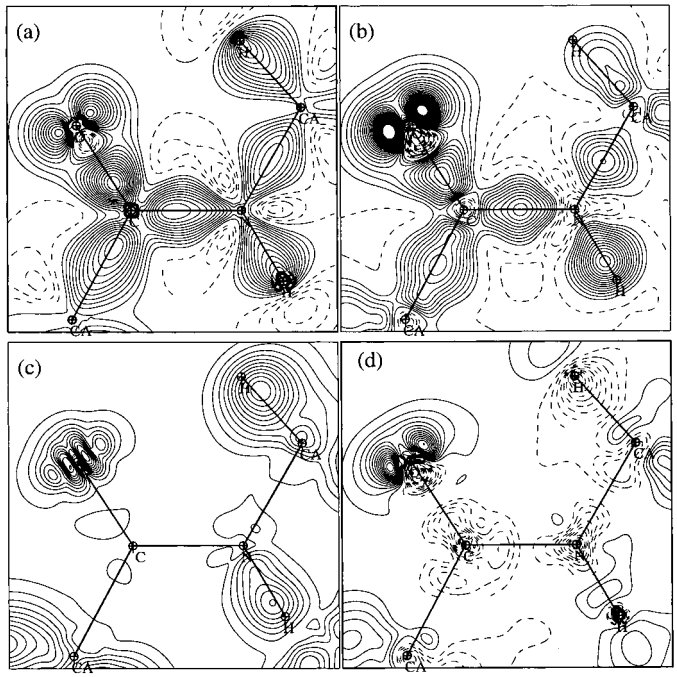

Fig. 3. Deformation density in the peptide bonds of crambin. Average map, (a) XRD-derived and (b) theoretical, with contour spacing of 0.05 $e^-/Å^3$. (c) Theoretical dispersion map $D(\vec{r}) = \sqrt{\sum_i [\rho_i(\vec{r}) - \tilde{\rho}(\vec{r})]^2}$ with contour spacing of 0.02 $e^-/Å^3$. (d) Difference between theoretical and XRD-derived.

larger for DFT than for XRD, in both the region of negative deformation on the C=O bond and the region of positive deformation corresponding to the lone pairs. It should be stressed that the oxygen lone pairs, constitute a fine feature in the electron density at around 0.3 Å from the nucleus only. Accurate high-resolution diffraction data and a low thermal motion are required for a proper refinement of the charge density parameters describing them [5,9]. This conclusion agrees with other XRD observations of underestimation of the deformation around lone pairs in peptides as compared to quantum mechanical results [4].

(ii) The XRD map displays a more pronounced elongation of the deformation density associated to bonds. In the present XRD study, the model was simplified as the dilatation coefficients for the spherical and multipolar valence electron density were not refined but set to standard values. The XRD static map can be biased by the atomic model used. This points to the simplified multipolar model as the possible responsible of the discrepancies in the shape of the bonding deformation density in Fig. 3.

(iii) The DFT map shows negative deformation in the core regions, especially for N. This last point is easy to rationalize when considering that, on one hand, there is a high electron density there, from core electrons, so that the deformation is more sensitive to errors, and, on the other hand, the calculations replace the core electrons by pseudopotentials. The difference map (Fig. 3d) between theory and experiment shows indeed systematic negative peaks around the atomic nuclei.

Both the limitations of the XRD experiment and the approximations used in the theoretical computations may be responsible for the discrepancies in the average deformation densities shown in Fig. 3a and b. The tests performed above show that the differences due to deficiencies of the basis set used are smaller than these discrepancies; the same applies for the approximate density functional and for the pseudopotentials used. It is important to keep in mind, however, that the protein in the crystalline state is hydrated by a layer of solvent molecules, whereas the computation was carried out in vacuo; this simplification of the system may affect the electron density on the polar groups C=O and N–H.

On the other hand, the intensities of X-ray reflections corresponding to different points in Fourier space at a limited resolution might be affected by systematic errors. The XRD map (Fig. 3a) is not the direct result of experiment, as the intensities are translated into a deformation density in real space through the Hansen–Coppens model [21]. Furthermore, the intensities of X-ray reflections originate from a crystal structure with thermal motion, the deconvolution of the thermal displacement parameters and the deformation density being not straightforward.

Beyond the average. The absence of noise in the theoretical calculations allows the analysis of how the electron

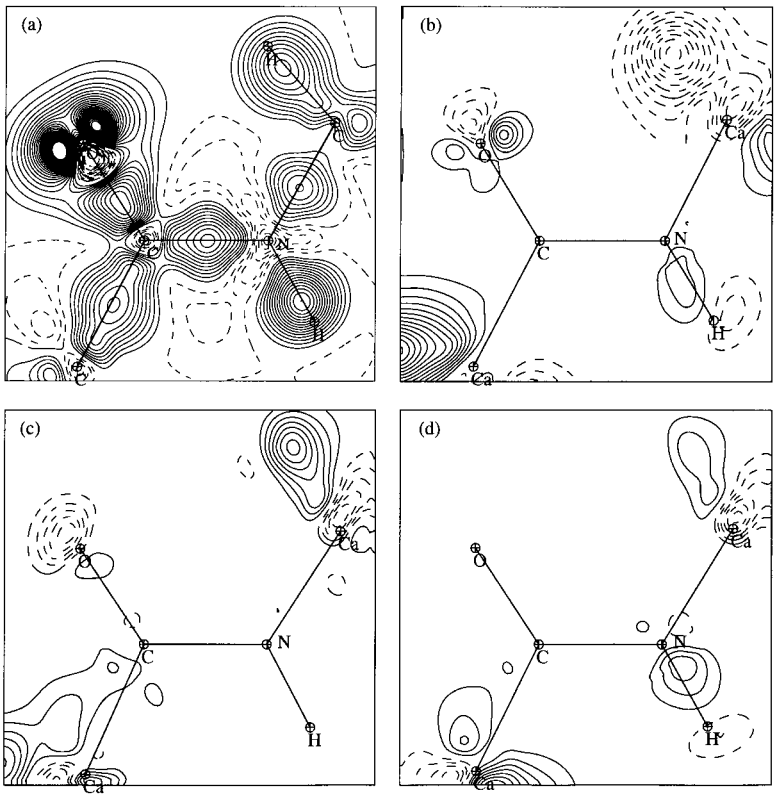

Fig. 4. Comparison of the DFT deformation density in several peptide bonds: between two alanine–threonine pairs and two consecutive isoleucine–isoleucine pairs in different parts of crambin. Discrepancies due to geometric distorsions were removed in the O=C–N–H region by superposing strictly all the atoms of the C_a–(O=C)–(N–H)–C_a moiety. (a) Reference deformation density in the plane Ala38.Thr39. The following deformation-density differences are shown: (b) Ala27.Thr28–Ala38.Thr39; (c) Ile33.Ile34–Ala38.Thr39; (d) Ile34.Ile35–Ala38.Thr39; The contour level spacing is 0.05 $e^-/Å^3$.

density in a peptide bond changes depending on the environment. Fig. 3c illustrates the amount of these changes by plotting the dispersion around the average. As mentioned above, the dispersion around C_a is mainly due to geometrical reasons, since the positions of the H_a atoms (and their deviation from the plane of the map) are highly dependent on the particular local conformation of the amino acids they belong to.

The highest dispersion occurs very close to the oxygen atom. The deformation density varies very rapidly in that region. Small changes in the position of the oxygen atom in the different maps can thus originate a strong dispersion with respect to the average map (average oxygen position). Another region of significant dispersion is the N–H bond. The O=C–N–H peptide group, although rigid, displays a certain variability in its stereochemistry, for example the dihedral angle C_a–C–N–C_a displays a standard deviation of 5.4° around −180°. From Fig. 3c, it is, therefore, not easy to ascertain whether the dispersion is of chemical or geometric origin.

It is important to identify the regions of the peptide bond which are most sensitive to the environment, together with the main character of their sensitivity, either structural or chemical. The main modifiers of the electron density on the polypeptidic main chain are presumably: (i) the stereochemistry (interatomic distances, angles); (ii) the nature of the particular residues at each side of a peptide bond, and (iii) the polar interactions: the C=O and N–H group can form hydrogen bonds with other protein atoms or the solvent molecules. To analyze the relative importance of the different effects, peptide bonds joining the same or different amino-acids in different locations of the protein are compared in Fig. 4. The distortions due to the variability of the atomic positions in the O=C–N grid plane were removed from the maps: the coordinates of the atoms were superposed to those in the reference peptide bond plane and the deformation density value was readjusted by interpolation.

The peptide bond Ala38–Thr39 was chosen for reference, since the polar groups C=O and N–H are devoid of polar interactions as they are located at the protein surface and exposed to the vacuum (in the DFT calculations). Again the upper-right and lower-left regions in the difference maps are not to be considered here, as the density difference of

geometric origin (orientation of C_a–H_a bond) cannot be removed by the method employed. The largest deviations are located near the N–H bond and around the oxygen atom. The N–H group of Ile33 and the Ile34 carbonyl are not involved in any hydrogen bond, which is to be related to the absence of electron density deviation in Fig. 4b and c.

The oxygen atoms of residue Ala27 and residue Ile33 form, respectively, two and one hydrogen bonds with N–H groups, these polar interactions lead to an alteration of the electron density of the oxygen. On the other hand, the N–H groups of Thr28 and Ile35 are each hydrogen bonded to a carbonyl group. This correlates with a strengthening of the N–H bond and a depletion of electrons in the H···O region: the negatively charged oxygen is responsible of the electron repulsion (Fig. 4b and d). The electronic shift generated on the N–H group is moreover along the same direction as the H···O interaction.

The conclusions of this study are as follows:

(i) A short-range DZP basis gives enough convergence for the calculation of the electron density for the maps studied here. The deformation density requires the calculation of the atomic reference densities with a long-range basis.
(ii) In spite of the convergence of the approximations involved in the DFT calculation, discrepancies persist in the comparison with XRD-derived map. Notably the lone pairs are attenuated in the XRD maps, which might be due to the thermal motion in the crystal structure and to the limits in resolution of the diffraction data. The different shape of the bonding density is probably related to the simplified multipolar model used in the derivation of the real-space deformation-density to map out the XRD reflection intensities.
(iii) The theoretical analysis of specific peptide bonds along the protein suggests that the carbonyl oxygen and N–H regions display a significant variability of their electron density. The presence of polar interactions like hydrogen bonds seems to be an environment effect of primary importance that perturbs the electron cloud of these moieties.

These results suggest that the computation methods described in this study may be employed to analyze molecular interactions in macromolecules. More systematic and comparative work could be done on specific peptide bonds to ascertain more precisely on the deformation electron density discrepancies along the polypeptide main chain. Also different planes may be analyzed, especially those involving the C_a atoms, which are more likely to be perturbed by the nature of the local amino acid due to their proximity to the side chain.

Acknowledgements

We acknowledge support of the Spanish Ramón Areces Foundation. E.A. is grateful for the hospitality of the Ecole Normale Supérieure de Lyon (France) where part of this work was done. The calculations where done on the SESC computer at the PSMN of the ENS Lyon, and were funded by CNRS.

References

[1] R.F.W. Bader, Atoms in Molecules: a Quantum Theory, Clarendon Press, Oxford, 1990.
[2] A.D. Mesecar, B.L. Stoddard, D.E. Koshland, Science 277 (1997) 202.
[3] D. Housset, V. Pichon-Pesme, C. Jelsch, F. Benabicha, A. Maierhofer, S. David, J.C. Fontecilla-Camps, C. Lecomte, Acta Crystallogr. D 56 (2000) 151.
[4] M. Souhassou, C. Lecomte, N. Ghermani, M.M. Rohmer, R. Wiest, M. Bénard, R.H. Blessing, J. Am. Chem. Soc. 114 (1992) 2371.
[5] C. Jelsch, V. Pichon-Pesme, C. Lecomte, A. Aubry, Acta Crystallogr. D 54 (1998) 1306.
[6] U. Rao, B. Stec, M.M. Teeter, Acta Crystallogr. D 51 (1995) 914.
[7] B. Stec, R. Zhou, M.M. Teeter, Acta Crystallogr. D 51 (1995) 663.
[8] A. Yamano, N.H. Heo, M.M. Teeter, J. Biol. Chem. 272 (1997) 9597.
[9] C. Jelsch, M.M. Teeter, V. Lamzin, V. Pichon-Pesme, R.H. Blessing, C. Lecomte, Proc. Natl Acad. Sci. (USA) 97 (2000) 3171.
[10] P. Ordejón, E. Artacho, J.M. Soler, Phys. Rev. B 53 (1996) R10441.
[11] D. Sánchez-Portal, P. Ordejón, E. Artacho, J.M. Soler, Int. J. Quant. Chem. 65 (1997) 453.
[12] E. Artacho, D. Sánchez-Portal, P. Ordejón, A. García, J.M. Soler, Phys. Status Solidi B 215 (1999) 809.
[13] C. Van Alsenoy, C.-H. Yu, A. Peeters, J.M.L. Martin, L. Schaefer, J. Phys. Chem. A 102 (1998) 2246.
[14] J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77 (1996) 3865.
[15] W. Kohn, L.J. Sham, Phys. Rev. 140 (1965) 1133.
[16] N. Troullier, J.L. Martins, Phys. Rev. B 43 (1991) 1993.
[17] L. Kleinman, D.M. Bylander, Phys. Rev. Lett. 48 (1982) 1425.
[18] O.F. Sankey, D.J. Niklewski, Phys. Rev. B 40 (1989) 3979.
[19] D. Sánchez-Portal, E. Artacho, J.M. Soler, J. Phys.: Condens. Matter 8 (1996) 3859.
[20] M. Machado, P. Ordejón, D. Sánchez-Portal, E. Artacho, J.M. Soler, submitted for publication (preprint in xxx.lanl.gov/physics/9908022).
[21] N.K. Hansen, P. Coppens, Acta Crystallogr. A 34 (1978) 909.
[22] V.S. Lamzin, R.J. Morris, Z. Dauter, K.S. Wilson, M.M. Teeter, J. Biol. Chem. 274 (1999) 20753.
[23] V. Pichon-Pesme, C. Lecomte, H. Lachekar, J. Phys. Chem. 99 (1995) 6242.
[24] J.P. Perdew, A. Zunger, Phys. Rev. B 23 (1981) 5075.
[25] J. Izquierdo, A. Vega, L.C. Balbás, D. Sánchez-Portal, J. Junquera, E. Artacho, J.M. Soler, P. Ordejón, Phys. Rev. B (2000) (in press).
[26] P.E. Cade, Trans. Am. Crystallogr. Assoc. 8 (1972) 1.
[27] V.D. Smith, I. Absar, Isr. J. Chem. (1977) 87.
[28] K.L. Kunze, M.B. Hall, J. Am. Chem. Soc. 108 (1986) 5122.