Electronic structure of bacterial surface protein layers

Volodymyr V. Maslyuk* and Ingrid Mertig
Martin-Luther-Universität Halle-Wittenberg, Fachbereich Physik, D-06099 Halle, Germany

Thomas Bredow
Institut für Physikalische und Theoretische Chemie, Universität Bonn, D-53115 Bonn, Germany

Michael Mertig
Max-Bergmann-Zentrum für Biomaterialien, Technische Universität Dresden, D-01062 Dresden, Germany

Denis V. Vyalikh and Serguei L. Molodtsov
Institut für Festkörperphysik, Technische Universität Dresden, D-01062 Dresden, Germany
(Received 12 March 2007; revised manuscript received 11 September 2007; published 23 January 2008)

We report an approach for the calculation of the electronic density of states of the dried two-dimensional crystalline surface protein layer (*S* layer) of the bacterium *Bacillus sphaericus* NCTC 9602. The proposed model is based on the consideration of individual amino acids in the corresponding conformation of the peptide chain which additively contribute to the electronic structure of the entire protein complex. The derived results agree well with the experimental data obtained by means of photoemission (PE), resonant PE, and near-edge x-ray absorption spectroscopy.

DOI: 10.1103/PhysRevB.77.045419 PACS number(s): 61.82.Ms, 71.20.−b, 31.15.A−, 21.10.Ft

I. INTRODUCTION

The electronic properties of biomolecules are receiving increasing attention, motivated by both biological and technological concerns. The electronic states play an important role in determining the interatomic forces and, thus, both the conformations and particular functions of proteins and DNA. More recently, there has been great interest in the nature of the electronic structure of biomolecules because of their potential applications to nanotechnology and molecular electronics. However, with a few exceptions, this information is lacking. For example, whether DNA is a good conductor or not remains unsettled and the crucial experimental information about its electronic structure is still missing.[1] Standard solid-state spectroscopic techniques such as photoemission (PE)[2,3] and near-edge x-ray absorption fine structure (NEXAFS)[3,4] are difficult to apply since biological samples are extremely sensitive to photon irradiation and often experience structural changes when dried. Moreover, the interpretation of the spectroscopic data is not straightforward without electronic structure calculations. Theoretical studies, however, are computationally demanding because of the large size of the unit cells of biological systems. Concerning proteins, to the best of our knowledge, there are only a few calculations performed within density functional theory (DFT) for rather small peptides.[5]

Here, we present an extended building-block approach that allows us to calculate the electronic structure of large proteins in the dried state. Within the model, the protein is considered as an assembly of primary constituents which contribute additively to the total density of states (DOS). The constituents are the individual amino acids in the corresponding conformation of the peptide chain. The model is applied to calculate the DOS of the two-dimensional regular surface protein layer (*S* layer) of *Bacillus sphaericus* NCTC 9602 which was recently studied by PE and NEXAFS spectroscopy[3] and has the specific property that the two-dimensional crystal structure is preserved under dry conditions.[6] The crystallographic unit cell of the *S*-layer lattice is composed of four identical monomers of known primary structure.[7] Each monomer consists of 1050 amino acids which corresponds to a total of 62 864 atoms per unit cell clearly excluding a first-principles electronic structure calculation. The results of the proposed building-block model for the approximated DOS are in good agreement with the reported PE and NEXAFS data,[3] and those of resonant PE (RPE) spectroscopy at the C $1s \rightarrow \pi^*$ excitation threshold which provides specific bond-type information about the carbon-derived occupied DOS.

II. METHODS

RPE is performed in combination with NEXAFS experiments at the Berliner Elektronenspeicherring für Synchrotronstrahlung (BESSY) using radiation from the Russian-German beamline. Valence-band RPE spectra are acquired with a VG-CLAM4 electron-energy analyzer. The overall-system energy resolution accounting for the thermal broadening is set to 150 meV full width at half maximum (FWHM). The C 1*s* absorption spectrum is recorded in a total electron yield mode and normalized to the incident photon flux. The NEXAFS resolution is determined solely by the performance of the beamline and is set to 80 meV FWHM. The procedure of the *ex situ S*-layer deposition onto Si wafers is described elsewhere.[3]

Proteins have markedly large conformational degrees of freedom because of both their large size and chemical complexity. Figure 1 shows schematically the structure of the peptide chain of a protein. The chain consists of a backbone (BB) with α-carbons to which the amino acid residues are bound as side chains (SCs) in a sequence-specific manner.

 ©2008 The American Physical Society

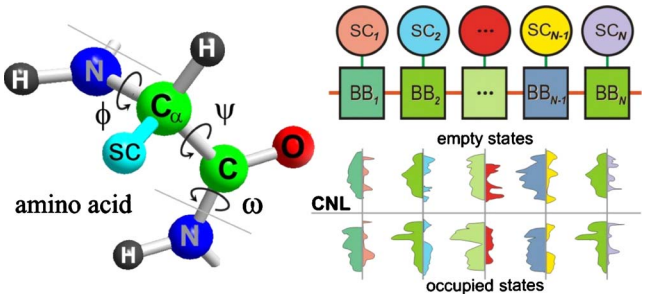

FIG. 1. (Color online) Schematic viewgraphs of the structure of the peptide chain (left) and of the extended building-block model (right).

The three dihedral angles ϕ, ψ, and ω define the conformation of the backbone and are responsible for its helicity. The electronic structure of the *S* layer is calculated in two steps. First, the DOS of the building blocks, which are the amino acids consisting of backbone units and side chains, is generated. To this end, we consider the four known conformations of the peptide-chain backbone (β-sheet, π-, 3_{10}-, and α-helices, usually referred to as secondary structure of proteins) and the 20 amino acid residues present in the *S* layer. In the second step, the individual contributions of the building blocks are superimposed energetically aligned at the charge-neutrality level (CNL)[8] to create the total DOS of the protein (Fig. 1). In accordance with the experimental UHV conditions, all calculations are carried out for dry polypeptides.

The electronic structure of the building blocks is calculated, starting from a polyglycine chain in all four secondary conformations generated by a supercell. The length of the corresponding supercells is determined by the helicity of the conformations. In the following step of calculation, one glycine (GLY) side chain in the polyglycine is replaced by any of the 19 other amino acid residues. Because of translational invariance of the supercell, all side chain positions are equivalent with respect to replacement. Interaction among the side chains is neglected. A multicode treatment is applied to calculate the electronic properties of the polypeptide chains. The semiempirical MSINDO method[9] is used to relax the fractional coordinates of the chain atoms, while the relaxation of the lattice constant and the electronic structure is calculated within the framework of DFT using a linear combination of atomic orbitals (CRYSTAL code[10]). For the C, N, O, H, and S atoms, a 6-311G** basis set[11] is used. Electron exchange and correlation are described by means of the B3LYP hybrid functional.[12] The optimized lattice constants, bond lengths, and bond angles of all relaxed polyglycine chain conformations obtained in our B3LYP/MSINDO calculations agree well with other theoretical results based on a plane-wave method with the Perdew-Becke-Ernzerhof (PBE) exchange-correlation potential.[13,14] Our lattice constant for poly-GLY in α-helical conformation is 16.42 Å in comparison with 16.16 Å in the PBE approximation.[13,14] The β-sheet lattice constant is 7.34 Å, which is very close to 7.21 Å.[13,14] The peptide bond length (see Fig. 1, C-N bond), in our case, is 1.39 Å for all conformations, which is in good agreement with 1.34 Å.[13,14] The C_α–N bond length varies between 1.45 and 1.47 Å depending on the considered side chain and agrees well with 1.45 Å.[13,14] The differences are related to the use of B3LYP with respect to PBE potentials.

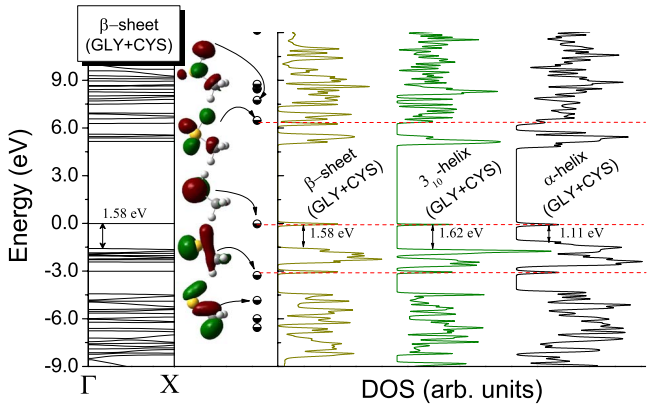

FIG. 2. (Color online) Band structure of the β-sheet ($7\cdot$GLY+CYS) conformation (left) and total DOS (right) of a cysteine-doped polyglycine chain in three different conformations (see text) compared with the molecular levels and orbitals of the free methyl mercaptan (CH_3SH) molecule shown in the middle panel.

III. RESULTS AND DISCUSSIONS

DFT calculations based on the obtained structural data show that the electronic structure of the considered peptide chains depends strongly on their particular backbone conformation. To illustrate this fact, both the band structure and the DOS of a polyglycine chain with one GLY replaced by cysteine (CYS) in the α-helical ($10\cdot$GLY+CYS), the 3_{10}-helical ($6\cdot$GLY+CYS), and the β-sheet ($7\cdot$GLY+CYS) conformation are calculated (Fig. 2). The numbers indicate the length of the individual unit cells. The relative positions of the cysteine-derived electron levels are similar to those in a free methyl mercaptan molecule (dashed lines) but take charge transfer between the backbone and the side chain into account. Note that a sulfur-related energy level of the CYS radical appears as expected[15] in the energy gap of the electronic structure of the backbone. The highest occupied molecular orbital (HOMO) of the considered peptide chain almost exclusively consists of backbone states forming bonds between neighboring C atoms. Figure 2 shows that both the shape and the energy position of the projected backbone DOS depend on the secondary structure of the considered peptide chain. For the α-, 3_{10}-helical, and β-sheet conformations, the energy differences between the sulfur state localized in the energy gap and the top of the backbone valence band are 1.11, 1.62, and 1.58 eV, respectively. These differences can be understood by the fact that, depending on the secondary structure, the backbone atoms have different coordinations.

The full electronic structure of the protein complex is constructed as a weighted sum of the projected DOS (PDOS) of the building blocks aligned at the CNL. The position of the CNL for all building blocks is derived in the way pro-

posed in Ref. 8. The DOS of each polyglycine chain with one side chain exchanged to one particular amino acid residue ($10 \cdot \mathrm{GLY} + R$) in the α-helical conformation is convoluted with a Gaussian function with a FWHM of 3 eV. Variation of FWHM (between 2 and 8.5 eV) has only a small effect on the position of the CNL (smaller or equal 0.13 eV), which is determined by integration of the convoluted DOS up to the energy that corresponds to the number of electrons in the supercell. The same procedure is applied to find the CNLs of the other secondary conformations. Thereafter, the PDOSs of the building blocks—the side chains and the corresponding backbone units—are extracted and superimposed with respect to the CNL by means of the following equations to derive the total protein DOS:

$$n(E) = n_{BB}(E) + n_{SC}(E),$$

where $n_{BB}(E)$ and $n_{SC}(E)$ are the PDOSs of the backbone and the side chain, respectively.

$$n_{SC}(E) = \sum_{\alpha}^{side\ chains} B_\alpha n_\alpha(E),$$

where B_α is the number of side chains of the type (α) with the PDOS $n_\alpha(E)$. The influence of the backbone conformation to the PDOS of the side chains is very weak and was neglected in this calculation.

$$n_{BB}(E) = \sum_{i}^{sequence} [p_H^i n_H^i(E) + p_L^i n_L^i(E) + p_E^i n_E^i(E)],$$

where p_k^i is the probability ($p_H + p_L + p_E = 1$) that the amino acid i is a part of one of the three conformations k of the backbone [p_H, p_L, and p_E—the probabilities of helical (α-helix), loop (3_{10}- and π-helices), and sheet (β-helix) configurations, respectively]; $n_k^i(E)$ is the PDOS of the backbone unit i in the conformation k. The values for p_k^i were taken from the PredictProtein server.[16]

The calculated total DOS of the studied *S* layer and its projections onto different parts of the polypeptide chain are shown in Fig. 3. Very good agreement of the theoretical one-electron levels with the valence-band PE and NEXAFS spectra[3] is obtained. The total and PDOS of the protein are broadened by a Gaussian function with FWHM $\sigma = 1.5$ eV. For comparison, the fine structure of the total DOS is also shown in Fig. 3(a) for FWHM $\sigma = 0.1$ eV. The projections of the DOS on both the side chains and the backbone show that the pronounced features of the occupied states are mainly caused by the backbone, namely, by its parts originating from the sheet and loop structures which are depicted at the bottom of Fig. 3(a). There are certainly considerable contributions to all spectral features from the side chains. The line shape of the residue-determined valence-band signal is similar to that of the main chain, pointing in both cases to similar types of the contributing chemical bonds. The PE response originating from the amino acid residues is, however, much smoother as compared with the backbone signal. Due to nonequivalent chemical environments of different residues, individual contributions of even the same bond type are energetically shifted relative to each other which complicates the

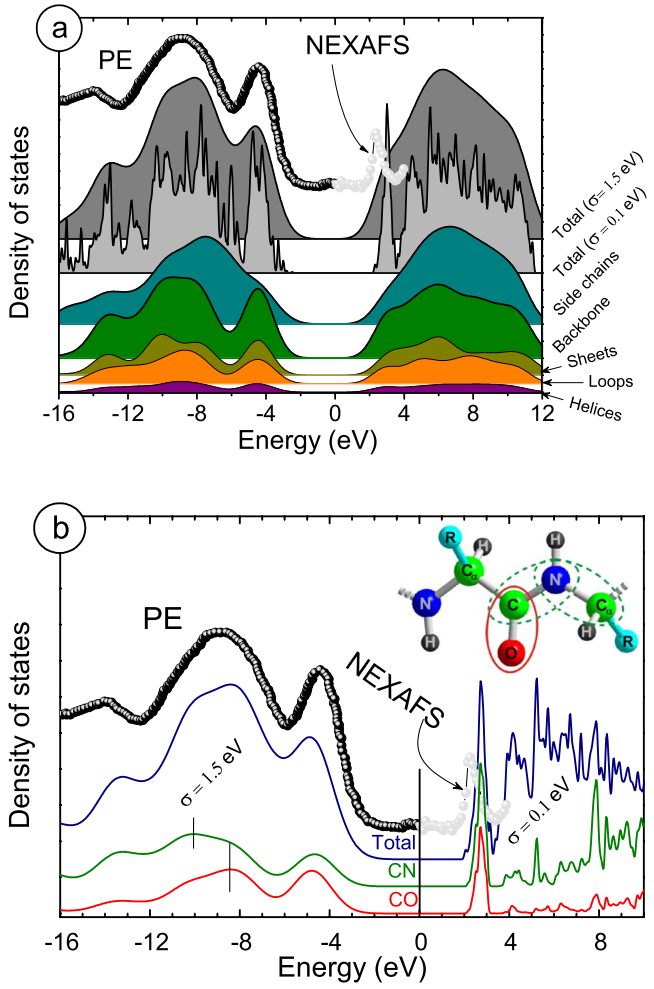

FIG. 3. (Color online) Experimentally determined electronic structure of the *S* layer of *B. sphaericus* NCTC 9602 in comparison with the calculated total DOS and (a) its projections on different parts of the protein structure or (b) contributions of the charge densities located around the C and O atoms ("C-O bonds," solid line in the inset) as well as around the C and N atoms ("C-N bonds," dashed lines) of the backbone.

analysis of the experimental data. Therefore, in the following, we will solely discuss the backbone PDOS which contains all characteristic features of the occupied valence-band structure of the *S*-layer protein. A detailed analysis of the PDOS related to the C-N and C-O contributions of the backbone bonds [Fig. 3(b)] allows to identify the nature of the valence-band PE features. The HOMO corresponds to signals stemming from both the C-N and C-O bonds, with slightly dominating contributions from the charge density concentrated around the C and O atoms. The peak at ~ -13 eV reflects mainly the C-N electron densities. The broad structure at ~ -8 eV can be related to the contributions from the C-N and C-O bonds with their two main peaks shifted relative to each other by approximately 2 eV.

The calculated data are also in good agreement with the RPE spectra, taken upon the C 1*s* electron excitation into different unoccupied molecular orbitals and indicated by numbers in the NEXAFS spectrum (Fig. 4). The RPE phe-

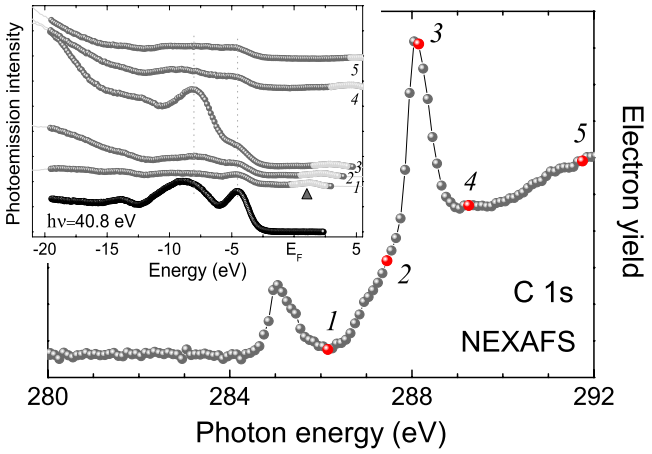

FIG. 4. (Color online) Experimental spectra of the *S* layer of *B sphaericus* NCTC 9602. Main panel: C 1*s* NEXAFS. Inset: resonant PE spectra taken with photon energies labeled in the NEXAFS plot (Ref. 3). The C 1*s* core-level contribution excited with the second-order light is depicted by light color and marked additionally with a filled triangle.

nomenon is based on the configuration mixing of two different states: a continuum final state representing the outgoing photoelectron and a hole in the valence band with a discrete intermediate state, populated by a dipole transition and decaying via an Auger process. Since the probability of transitions into the intermediate state varies drastically when tuning the photon energy across a core-level threshold, large variations of the photoionization cross section of valence-band states with orbitals overlapping the initial- and intermediate-state orbitals are obtained known as the Fano-Beutler resonance.[17]

In the present study, we use mainly the excitation into the unoccupied C-O-type orbital (feature 3 in the NEXAFS plot) in order to emphasize the C-O-derived features in the valence-band PE spectra. In the inset of Fig. 4, the RPE data are compared with the valence-band PE spectrum taken at a lower photon energy. The spectrum taken at the photon energy indicated with 1 reveals almost no signal since this photon energy is located in the minimum of the NEXAFS intensity between the lowest unoccupied molecular orbital and the C-O responses and therefore does not lead to resonance phenomena. Direct PE excitation of the 2*p* valence electrons of C, O, and N is characterized by very small cross sections at these photon energies.[18]

In resonance (NEXAFS feature 3), a strong enhancement of the low-energy side of the broad structure located at ~−8 eV is observed. As follows from our calculations, particularly this part of the −8 eV feature consists of the C-O contributions to the PE intensity. The HOMO signal is also emphasized. The on-resonance growth of the HOMO intensity is, however, not that pronounced as the enhancement of the broad structure, although the C-O/C-N ratio of the contributions to the HOMO is larger than that in the case of the −8 eV structure. This finding can be understood taking into account the lower spatial localization of the HOMO originating electronic states compared to the deeper states. The partly non-localized character of the upper electrons reduces the spatial overlap of the HOMO-derived final states and the unoccupied C-O intermediate states damping the RPE channel. The obtained behavior of the HOMO electrons along the backbone is in agreement with our band-structure calculations which reveal noticeable dispersion of the electronic states related to the HOMO (Fig. 2). The structure at −13 eV is only slightly emphasized upon on-resonance excitation with $h\nu$=288.1 eV, indicating presumably the non-C-O origin of this PE signal. The overall enhancement of the distinguished C-O contributions decreases on both sides of the C-O resonance (compare spectra taken at the NEXAFS features 2, 4, and 5). The entire RPE data are in accordance with the calculated partial C-O densities of the occupied electronic states presented in Fig. 3(b).

The partial C-N contributions comprising the −13 eV structure and the left-side part of the broad −8 eV structure are not enhanced even upon excitation applying the NEXAFS feature 2, which was attributed in our previous studies either to the C-N or C-H resonances.[3] Thus, the presented data favor any other than a C-N origin of this NEXAFS feature.

IV. SUMMARY

In the present paper, a theoretical approach to calculate the electronic structure of proteins is presented. The method is based on an extended building-block model that allows to calculate the electronic structure of large proteins in the dried state. The building blocks are individual amino acids in the corresponding conformation of the peptide chain. The electronic structure of the building blocks is calculated by state of the art *ab initio* electronic structure calculations. The electronic structure of the total protein is constructed by superposition of the individual contributions of the building blocks aligning them at a common charge-neutrality level. The capability of the method is demonstrated by calculating the density of states for the surface protein layer of *Bacillus sphaericus* NCTC 9602. The theoretical results are in very good agreement with the experimental data measured in PE, resonant PE, and NEXAFS experiments. The calculations allow insight into the complex electronic structure down to the atomic scale. We anticipate that the considered approach may generally be applied to describe the electronic structure of large biomolecules.

ACKNOWLEDGMENTS

V.V.M. and I.M. acknowledge financial support through the Deutsche Forschungsgemeinschaft (DFG), Priority Programme 1165: "Nanowires and Nanotubes." M.M. thanks the DFG (Grant No. ME 1256/13-1) and the BMBF (Grant No. 03X0004A). The experiments were supported by the bilateral program "Russian-German Laboratory at BESSY."

*volodymyr.maslyuk@physik.uni-halle.de

[1] R. G. Endres, D. L. Cox, and R. R. P. Singh, Rev. Mod. Phys. **76**, 195 (2004).

[2] H. S. Kato, M. Furukawa, M. Kawai, M. Taniguchi, T. Kawai, T. Hatsui, and N. Kosugi, Phys. Rev. Lett. **93**, 086403 (2004).

[3] D. V. Vyalikh, S. Danzenbächer, M. Mertig, A. Kirchner, W. Pompe, Y. S. Dedkov, and S. L. Molodtsov, Phys. Rev. Lett. **93**, 238103 (2004); D. V. Vyalikh, A. Kirchner, S. Danzenbächer, Y. S. Dedkov, A. Kade, M. Mertig, and S. L. Molodtsov, J. Phys. Chem. B **109**, 18620 (2005); D. V. Vyalikh, A. Kirchner, A. Kade, S. Danzenbächer, Yu. S. Dedkov, M. Mertig, and S. L. Molodtsov, J. Phys.: Condens. Matter **18**, S131 (2006).

[4] J. Boese, A. Osanna, C. Jacobsen, and J. Kirz, J. Electron Spectrosc. Relat. Phenom. **85**, 9 (1997); K. Kaznacheyev, A. Osanna, C. Jacobsen, O. Plashkevych, O. Vahtras, H. Ågren, V. Carravetta, and A. P. Hitchcock, J. Phys. Chem. B **106**, 3153 (2002); A. P. Hitchcock, C. Morin, Y. M. Heng, R. M. Cornelius, and J. L. Brash, J. Biomater. Sci., Polym. Ed. **13**, 919 (2002); M. L. Gordon, G. Cooper, C. Morin, T. Araki, C. C. Turci, K. Kaznatcheev, and A. P. Hitchcock, J. Phys. Chem. A **107**, 6144 (2003); C. Morin, A. P. Hitchcock, R. M. Cornelius, J. L. Brash, S. G. Urquhart, A. Scholl, and A. Doran, J. Electron Spectrosc. Relat. Phenom. **137-140**, 785 (2004); G. Cooper, M. Gordon, D. Tulumello, C. Turci, K. Kaznatcheev, and A. P. Hitchcock, *ibid.* **137-140**, 795 (2004).

[5] J. Ireta, M. Galvan, K. Cho, J. D. Joannopoulos, F. Aparicio, J. Ireta, A. Rojo, L. Escobar, A. Cedillo, and M. J. Galvan, J. Phys. Chem. B **107**, 1692 (2003).

[6] B. Aichmayer, M. Mertig, A. Kirchner, O. Paris, and P. Fratzl, Adv. Mater. (Weinheim, Ger.) **18**, 915 (2006).

[7] Nucleotide Sequence Database of the European Molecular Biology Laboratory (http://www.ebi.ac.uk).

[8] H. Vázquez, R. Oszwaldowski, P. Pou, J. Ortega, R. Pérez, F. Flores, and A. Kahn, Europhys. Lett. **65**, 802 (2004); H. Vázquez, F. Flores, R. Oszwaldowski, J. Ortega, R. Pérez, and A. Kahn, Appl. Surf. Sci. **234**, 108 (2004); H. Vázquez, W. Gao, F. Flores, and A. Kahn, Phys. Rev. B **71**, 041306(R)(2005).

[9] B. Ahlswede and K. Jug, J. Comput. Chem. **20**, 563 (1998); **20**, 572 (1998).

[10] V. R. Saunders, R. Dovesi, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, N. M. Harrison, K. Doll, B. Civalleri, I. J. Bush, Ph. D'Arco, and M. Llunell, *CRYSTAL2003 User's Manual* (University of Turin, Turin, 2003).

[11] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).

[12] A. D. Becke, Chem. Phys. **98**, 5648 (1993).

[13] J. Ireta, J. Neugebauer, M. Scheffler, A. Rojo, and M. Galván, J. Phys. Chem. B **107**, 1432 (2003).

[14] J. Ireta (private communication).

[15] K. Taken and K. Shiraishi, J. Phys. Soc. Jpn. **6**, 421 (1996).

[16] B. Rost, Methods Enzymol. **266**, 525 (1996); B. Rost and C. Sander, J. Mol. Biol. **232**, 584 (1993); Proteins **19**, 55 (1994).

[17] U. Fano, Phys. Rev. **124**, 1866 (1961).

[18] J. J. Yeh and I. Lindau, At. Data Nucl. Data Tables **32**, 1 (1985).