

Experimental Observation of Bonding Electrons in Proteins*

(Received for publication, April 15, 1999, and in revised form, May 25, 1999)

Victor S. Lamzin^{‡§}, Richard J. Morris[‡],
Zbigniew Dauter[¶], Keith S. Wilson^{‡||},
and Martha M. Teeter^{**}

From the [‡]European Molecular Biology Laboratory, 22603 Hamburg, Germany, the [¶]National Cancer Institute, Upton, New York 11973, the ^{||}University of York, Heslington, York, YO1 5DD, United Kingdom, and the ^{**}Department of Chemistry, Merkert Chemistry, Boston College, Chestnut Hill, Massachusetts 02167

We demonstrate with two examples the success and potential of recent developments in x-ray protein crystallography at ultra high resolution. Our preliminary structural analyses using diffraction data collected for the two proteins crambin and savinase show meaningful deviations from the conventional independent spherical atom approximation. A noise-reduction averaging technique enables bonding details of electron distributions in proteins to be revealed experimentally for the first time. We move one step closer to imaging directly the fine details of the electronic structure on which the biological function of a protein is based.

The principles underlying all molecular recognition processes and enzymatic reactions are based on chemistry involving the outer shell electrons of molecules, functional groups, or atoms (1). Detailed knowledge of electronic structure is a prerequisite for a deeper chemical understanding of biological processes.

Precise electron distributions may be obtained from quantum mechanical calculations. Schrödinger's equation provides the theoretical basis for the determination of wave functions from which the electron density may be computed. However even with rather extreme approximations, the numerical complexity of this approach excludes the possibility of modeling entire macromolecules and their interactions.

A qualitative image of electron distributions may also be obtained experimentally by the diffraction of x-rays from a crystal. The extent of the data (the number of unique reflections) in reciprocal (diffraction) space is quantified by the nominal resolution. There is an inverse relationship between direct space (the crystal) and diffraction space (the measurements): long range features of the molecule correspond to low resolution diffraction data, short range to high resolution. The minimum separation between the individual features in the electron density map, *i.e.* the level of detail, is approximately equal

to the nominal resolution of the data multiplied by 0.7 (2). Additional smearing of the electron density results from atomic vibration and static disorder and may increase this factor to about 1.1 (3). Roughly speaking, to resolve two points separated by 1.0-Å distance requires x-ray data extending to 1.0-Å resolution.

Observed distributions of core electrons are very well approximated by the spherical atomic density models employed routinely in crystallography. However, valence electrons show deviations from spherical symmetry, and their detection requires very high resolution diffraction data. Deformation density arising from such deviations can commonly be observed for small molecules.

In proteins, deformation density is generally not visible because of a number of factors. Macromolecules are flexible polymers and considerably disordered. Vibration of the atoms spreads the distribution of the electrons. The effect of this is two-fold. First, the resolution of the x-ray crystal diffraction is typically lower than that required to experimentally observe valence electrons (4). Second, modeling atomic vibration by isotropic or anisotropic atomic displacement parameters at least partially takes up the effects of the bonding electron features (4, 5). Particularly at high resolution, the standard crystallographic modeling, *i.e.* the approximation of the electron density distribution by a set of spherically symmetric atom-centered densities, cannot correctly account for important features such as bonding or lone pair electrons.

The fundamental observation of chemistry that properties of groups of atoms are to some extent conserved has led to an elegant approach to modeling electron density in macromolecules. Namely, the electron density multipole parameters obtained from small molecules have been proposed to be transferable to macromolecules (6). This may be viewed as a crystallographic adaptation of Bader's theory of atoms in molecules (7). Although the transferability has been demonstrated to give good results for polypeptides, its applicability to proteins has not previously been shown.

We show here that good quality x-ray diffraction data to atomic resolution (8) do permit detection of bonding electron information in proteins. We employ a simple technique of noise reduction for an averaged peptide bond to reveal this bonding information.

EXPERIMENTAL PROCEDURES

Relevant experimental details are presented in Table I. X-ray data were collected using synchrotron radiation from the DORIS storage ring (EMBL Hamburg, DESY) and a MAR Research imaging plate and processed using the HKL suite (9). The models were refined using SHELXL (10). The procedures used were those routinely applied in protein crystallography. Full description of the structures, together with data collection and refinement will be published elsewhere.

The extraction of bonding information relies on the fundamentally repetitive nature of macromolecular polymers. Proteins consist of linear chains of amino acids with a backbone built from a set of repeating peptide units. The electron density for the peptides repeats along the chain, and this fact was used to increase the signal-to-noise ratio. Provided the electron density distribution for all peptide units only differs in the level of statistical noise present, averaging of N such units is expected to increase the signal-to-noise ratio by a factor of \sqrt{N} . Although density for peptide units is not exactly the same over the protein chain and, in addition, some systematic errors always remain, averaging the peptide plane density proved to be powerful for the extraction of non-spherical valence information.

Self-consistent reaction field (SCRF) calculations on a simple dipep-

* This work was supported in part by the European Union through the Framework IV Biotechnology contracts BIO2-CT92-0524 and BIO4-CT96-1809, by the Boston College for 80% sabbatical support, and by the National Science Foundation for grant DMCB-9219857 (to M. M. T.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Tel.: 49 40 89902 121; Fax: 49 40 89902 149; E-mail: Victor@EMBL-Hamburg.DE.

TABLE I
Experimental details

Crystal structure	crambin	savinase
Temperature (K)	100	110
Resolution (Å)	0.67	0.90
Crystallographic <i>R</i> factor (%)	7.9	11.1
Mean main chain atomic displacement parameters (Å ²)	3.0	6.3
Number of peptide units averaged	40	229

tide model were carried out at Hartree-Fock level with a medium sized split valence basis set (HF/6-31+G) with the GAMESS software package (11).

RESULTS

The first example concerns the 46-residue plant protein crambin from *Crambe abyssinica* (12, 13). Its structure at 130 K was previously reported at 0.83-Å resolution with data from a rotating anode x-ray source (14). We collected synchrotron diffraction data to 0.67 Å at 100 K, and the model has been refined preliminarily against the complete set of unique reflections, Table I. The densities for individual amino acid residues show no significant evidence of deviation from the spherical atom approximation. However, exploiting the repetitive structure by superimposing the electron density of 40 peptide units (excluding the five peptides modeled with two conformations) does reveal detailed features.

The averaged density, Fig. 1, shows two peaks on both sides of the atoms within the peptide plane. These correspond to the highly populated regions of (φ , ψ) conformational space for L-amino acids (15), one of each pair representing a CB atom, the other the N or C atoms of the adjacent residues. In addition, a build-up of electron density between the atoms along the bonds is apparent.

Features in the averaged difference map, Fig. 2, are even more prominent. They provide clear experimental observation of the bonding electrons in the peptide moiety, namely the σ electrons in the middle of the CA—C, C—N, and N—CA bonds. Bonding density outside the peptide unit is visible next to the two CA atoms. Weaker density is visible for the C—O bond and even a trace of the N—H bonding density. With imagination, one can see traces of the lone pairs of the carbonyl oxygen. We approximated the peaks in the middle of the bonds between non-hydrogen atoms by a three-dimensional Gaussian function, and we estimated their content to be around 0.02 to 0.03 electrons. These values are lower than the comparable values for small molecules, typically about 0.4 to 0.5 electrons (5, 16) because of the lower resolution.

For comparison, we have calculated the deformation density of a peptide plane with purely theoretical methods, Fig. 3, as outlined above. Although this level of density detail is well beyond the scope of any x-ray diffraction experiment, we feel that the striking qualitative resemblance justifies our approach. The agreement exceeded our expectations and demonstrates the level of detail that can be extracted from diffraction data. It has now become feasible to image features as small as a fraction of an electron, thus recovering otherwise lost information.

We extended our analysis to a 27-kDa protein, savinase, a subtilisin-like proteinase (17), where the data only extended to a resolution of 0.90 Å, Table I. Despite the lower resolution, the larger molecular weight of savinase enhances the effectiveness of the density averaging. The results from savinase corroborate those obtained for crambin, with clearly visible bonding density, Fig. 4. Because of the more limited resolution and higher atomic displacement parameters, the bonding density is more pronounced for the longer bonds, CA—C and N—CA, and over-

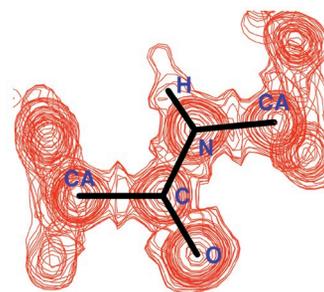


FIG. 1. Averaged ($3F_o - 2F_c, \alpha_c$) electron density Fourier synthesis for crambin. The lowest contour level is 0.4 electrons Å⁻³.

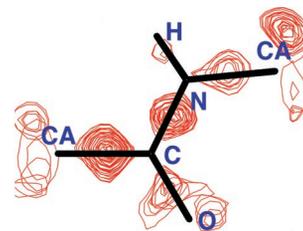


FIG. 2. Averaged ($F_o - 2F_c, \alpha_c$) difference density for crambin. The lowest contour level is 0.025 electrons Å⁻³.

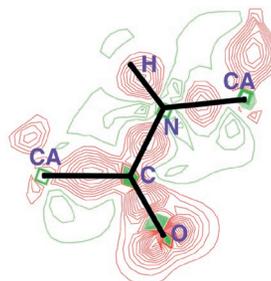


FIG. 3. Theoretical deformation density (the total molecular electron density minus the isolated atomic densities) from HF/6-31+G calculations. Positive density (magenta) shows where the electrons have moved; negative density (green) indicates where electrons have come from. The contouring is in steps of 0.06 electrons Å⁻³. The plot was produced with MOLDEEN.

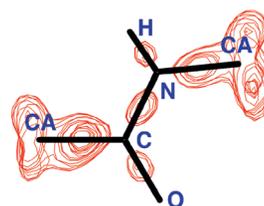


FIG. 4. Averaged ($F_o - 2F_c, \alpha_c$) difference density for savinase. The lowest contour level is 0.025 electrons Å⁻³.

laps somewhat with the bonding density outside the peptide unit. The peaks in the middle of the bonds C—O, C—N, and N—H are smaller but still clearly visible. The traces of lone pair electrons have vanished.

DISCUSSION

For the first time, bonding electrons have been observed experimentally for macromolecules. These results are qualitatively in good agreement with theoretical studies. Very high resolution x-ray data from crystals with low disorder and thermal motion are essential. Fortunately, this is now feasible not only for a small protein like crambin but for much larger proteins. The advent of high intensity synchrotron radiation sources, efficient two-dimensional detectors and cryogenic freezing techniques permits crystallographic x-ray data collection of unprecedented quality. By the middle of 1995, there

were about 30 x-ray crystal data sets recorded for proteins at atomic (1.2 Å or higher) resolution (18). By the end of 1998, there were about 130 such data sets collected on EMBL Hamburg beam lines alone. The proteins vary in size from 5–6 kDa (crambin or rubredoxin (19)) up to about 280 kDa for a heterohexameric methyl coenzyme M reductase.¹

Comparative studies of theoretical models for the electron density can now be performed for macromolecules, and modified non-spherical scattering factors may be derived. The repeating peptide motif allows an important means of enhancing the weak signal by averaging within the backbone. The methodological advance of averaging was thus essential for the unmasking of the bonding electron signal. This method may be extended to at least some of the side chains. One can also envision averaging the densities for a set of proteins. There is clearly a need to develop more elaborate methods to extract subtle information present in the x-ray diffraction data. Ideally one would like to visualize directly the valence electron density for an atomic group of interest, *e.g.* in an enzyme active site.

Exploitation of the techniques presented here will have an important biological impact because all macromolecular recognition and enzymatic processes are consequences of the valence electron distribution of interacting atoms or functional groups. Elucidation of detailed electronic structure is clearly beneficial for a deeper understanding of the chemical reactions underlying biological processes. The fine but subtle details of the stereochemical and electronic environment of key atoms are critical to function, catalysis, or ligand interactions. They are also crucial in knowledge-based drug design, enhancement of sub-

strate stereoselectivity, and development of biological molecules with desired and directed properties. This work moves us one important step closer to these goals.

Acknowledgment—We thank NOVO Nordisk (Denmark) for the provision of savinase samples.

REFERENCES

1. Mesecar, A. D., Stoddard, B. I., and Koshland, D. E., Jr. (1997) *Science* **277**, 202–206
2. James, R. W. (1948) *Acta Crystallogr.* **1**, 132–134
3. Swanson, S. M. (1988) *Acta Crystallogr. Sec. A* **44**, 437–442
4. Jelsch, C., Pichon-Pesme, V., Lecomte, C., and Aubry, A. (1998) *Acta Crystallogr. Sec. D* **54**, 1306–1318
5. Pichon-Pesme, V., Lecomte, C., Wiest, R., and Benard, M. (1992) *J. Am. Chem. Soc.* **114**, 2713–2715
6. Pichon-Pesme, V., Lecomte, C., and Lachekar, H. (1995) *J. Phys. Chem.* **99**, 6242–6250
7. Bader, R. F. W. (1990) *Atoms in Molecules. A Quantum Theory*, Oxford University Press, Oxford
8. Sheldrick, G. M. (1990) *Acta Crystallogr. Sec. A* **46**, 467–473
9. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
10. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol.* **277**, 319–343
11. Schmidt, M. W., Baldrige, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S. J., Windus, T. L., Dupuis, M., and Montgomery, J. A. (1993) *J. Comput. Chem.* **14**, 1347–1363
12. Teeter, M. M., and Hendrickson, W. A. (1979) *J. Mol. Biol.* **127**, 219–224
13. Teeter, M. M., Mazer, J. A., and L'Italien, J. J. (1981) *Biochemistry* **20**, 5437–5443
14. Teeter M. M., Roe, S. M., and Heo, N. H. (1993) *J. Mol. Biol.* **230**, 292–311
15. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95–99
16. Souhassou, M., Lecomte, C., Ghermani, N.-E., Rohmer, M.-M., Roland W., Benard, M., and Blessing, R. H. (1992) *J. Am. Chem. Soc.* **114**, 2371–2382
17. Betzel, C., Klupsch, S., Papendorf, G., Hasturp, S., Branner, S., and Wilson, K. S. (1992) *J. Mol. Biol.* **223**, 427–445
18. Dauter, Z., Lamzin, V. S., and Wilson, K. S. (1995) *Curr. Opin. Struct. Biol.* **5**, 784–790
19. Dauter, Z., Sieker, L. C., and Wilson, K. S. (1992) *Acta Crystallogr. Sec. B* **48**, 42–59

¹ W. Grabarse, M. Goubeaud, R. K. Thauer, V. S. Lamzin, and U. Ermler, manuscript in preparation.