Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 24 January 2001
Accepted 23 April 2001

Environmental influence on electron scattering from a molecule

Hiroshi Yamashita and Akinori Kidera*

Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan. Correspondence e-mail: kidera@qchem.kuchem.kyoto-u.ac.jp

The environmental influence on the electron scattering amplitudes of a molecule was evaluated by *ab initio* molecular-orbital calculations. The model system is formic acid in various states, *i.e.* the monomer, hydrogen-bonded dimer and ionized formate form. The model electrostatic potentials were calculated either *in vacuo* or with the polarizable continuum model as a simple model of an aqueous environment. It was found that charge compensation due to the environment affects the scattering amplitudes significantly. The resultant molecular electrostatic potential was fitted by six nucleus-centered Gaussians of site/environment-dependent atomic electrostatic potentials with small residual errors. Therefore, the site/environment-dependent atomic electrostatic potentials will give a good model for electron crystallography.

1. Introduction

The knowledge of electron scattering amplitudes is essential for the determination of molecular structures by electron crystallography. Electrons are scattered by the potential field of a nucleus, which is partially screened by the orbital electrons. Thus, electron crystallography produces a map of the electrostatic potential, while X-ray crystallography produces a map of the electron density. Conventional treatment of electron diffraction, following the conventional procedure in X-ray crystallography, assumes each atom to behave independently of its neighbors, so that the potential can be represented by a superposition of the spherically averaged potentials corresponding to isolated atoms. With this assumption, the general coherent scattering integral for a molecule,

$$F(\mathbf{s}) = \left(2\pi m_0 e/h^2\right) \int \varphi(\mathbf{r}) \exp(4\pi i\, \mathbf{s}\cdot\mathbf{r})\ \mathrm{d}\mathbf{r}, \qquad (1)$$

reduces to the simplified form

$$F(\mathbf{s}) \cong F_{\mathrm{isol}}(\mathbf{s}) = \sum_{\alpha} f_{\mathrm{isol},\alpha}(s) \exp(4\pi i\, \mathbf{s}\cdot\bar{\mathbf{r}}_{\alpha}), \qquad (2)$$

with

$$f_{\mathrm{isol},\alpha}(s) \equiv \left(8\pi^2 m_0 e/h^2\right) \int_0^{\infty} r^2 \varphi_{\mathrm{isol},\alpha}(r)[\sin(4\pi s r)/(4\pi s r)]\ \mathrm{d}r, \qquad (3)$$

where $s = |\mathbf{s}| = (\sin\theta)/\lambda$ [Å^{-1}], θ is half the angle of scattering and λ is the electron wavelength. In (2), the molecular potential function, $\varphi(\mathbf{r})$, is approximated by

$$\varphi(\mathbf{r}) \cong \varphi_{\mathrm{isol}}(\mathbf{r}) = \sum_{\alpha} \varphi_{\mathrm{isol},\alpha}(|\mathbf{r} - \bar{\mathbf{r}}_{\alpha}|), \qquad (4)$$

where $\bar{\mathbf{r}}_{\alpha}$ is the average position vector of the αth atom in the molecule, and the atomic potential function, $\varphi_{\mathrm{isol},\alpha}(|\mathbf{r} - \bar{\mathbf{r}}_{\alpha}|)$, is spherically symmetric. Here, we call $\varphi_{\mathrm{isol},\alpha}(|\mathbf{r} - \bar{\mathbf{r}}_{\alpha}|)$ and $f_{\mathrm{isol},\alpha}(s)$ the potential and scattering factor of the isolated-atom (IA) model, respectively. The values of $f_{\mathrm{isol},\alpha}(s)$ for all neutral atoms and most chemically significant ions are compiled in *International Tables for Crystallography*, Vol. C (Cowley, 1992). These values, of course, neglect the influence of redistribution of valence electrons in bound atoms and the environment surrounding the molecule.

The influence of redistribution of valence electrons on a potential map has already been pointed out by Mitsuoka *et al.* (1999) and Chang *et al.* (1999). Mitsuoka *et al.* (1999) found anomalies in a residual potential map for the backbone atoms of a protein, bacteriorhodopsin, when they included contributions from low-angle scattering. They concluded that the anomalies were caused by the polarization along bonds in the backbone. Chang *et al.* (1999) showed in several kinds of small molecules that there are significant differences between the scattering amplitudes calculated from *ab initio* molecular orbital (MO) calculations and those derived from the IA model, especially at low scattering angles. Grigorieff *et al.* (1996) attempted to account for bonding effects in protein electron crystallography by adjusting the scattering factors to fit the experimental data during the refinement process. This approach, however, did not succeed in achieving a marked improvement in the *R* factor.

In X-ray crystallography of organic molecules, more sophisticated models for atomic scattering factors to describe the deformed electron density have been proposed (Coppens, 1996, 1997). For example, Weiss & Freeman (1959) introduced non-spherical atomic scattering factors based on atomic orbital functions. Hansen & Coppens (1978) described the electron-density distribution as a nucleus-centered multipole expansion. It is expected in the Mott formula that only a small

change in the scattering factor for X-rays, $f^{X}(s)$, will bring a considerable change in $f(s)$ at low scattering angles, *i.e.*

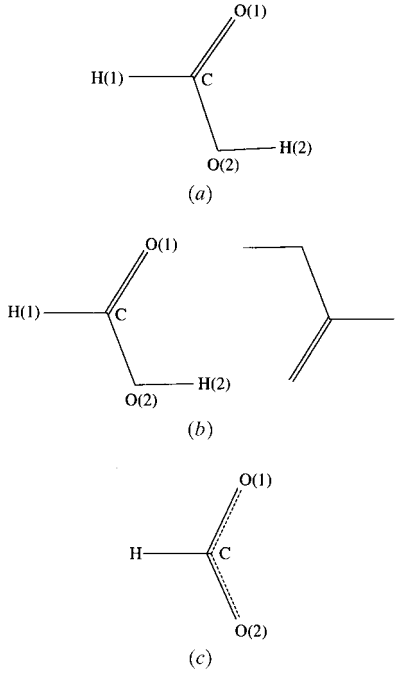

Figure 1
Molecular formula of (*a*) formic acid, (*b*) formic acid dimer and (*c*) the formate anion, showing atomic numbering.

$$f(s) = (me^2/2h^2)[Z - f^{X}(s)]/s^2. \quad (5)$$

Therefore, the development of a realistic model for the scattering factor is a crucial problem in electron crystallography.

In the case of a protein crystal which is filled with aqueous solvent up to about half the total volume, ambient solvent is another important factor determining a potential map. When a solute molecule is transferred into a high dielectric medium, one would expect that the charge distribution of the solute becomes more localized, and that the diffuse tail of the charge distribution may be perturbed. The scattering amplitude of a free ion *in vacuo* approaches $\pm\infty$ as θ approaches zero, owing to the unscreened long-range electrostatic potential. In aqueous solution, on the other hand, the long-range electrostatic potential is screened by the solvent water, and it is expected that the ionization effect at low scattering angles should be mostly suppressed.

In order to study the influence of redistribution of the valence electrons and of environment on electron scattering amplitudes, we carried out *ab initio* MO calculations of a model compound. The model compound studied was formic acid in various states, *viz* monomer, hydrogen-bonded dimer and the ionized formate form (Fig. 1), either *in vacuo* or in aqueous solution. The hydrogen-bonded dimer is used to elucidate the influence of hydrogen bonding. For the MO calculations in aqueous solution, we employed the polarizable continuum model (PCM) (Miertus *et al.*, 1981) as a simple approximation.

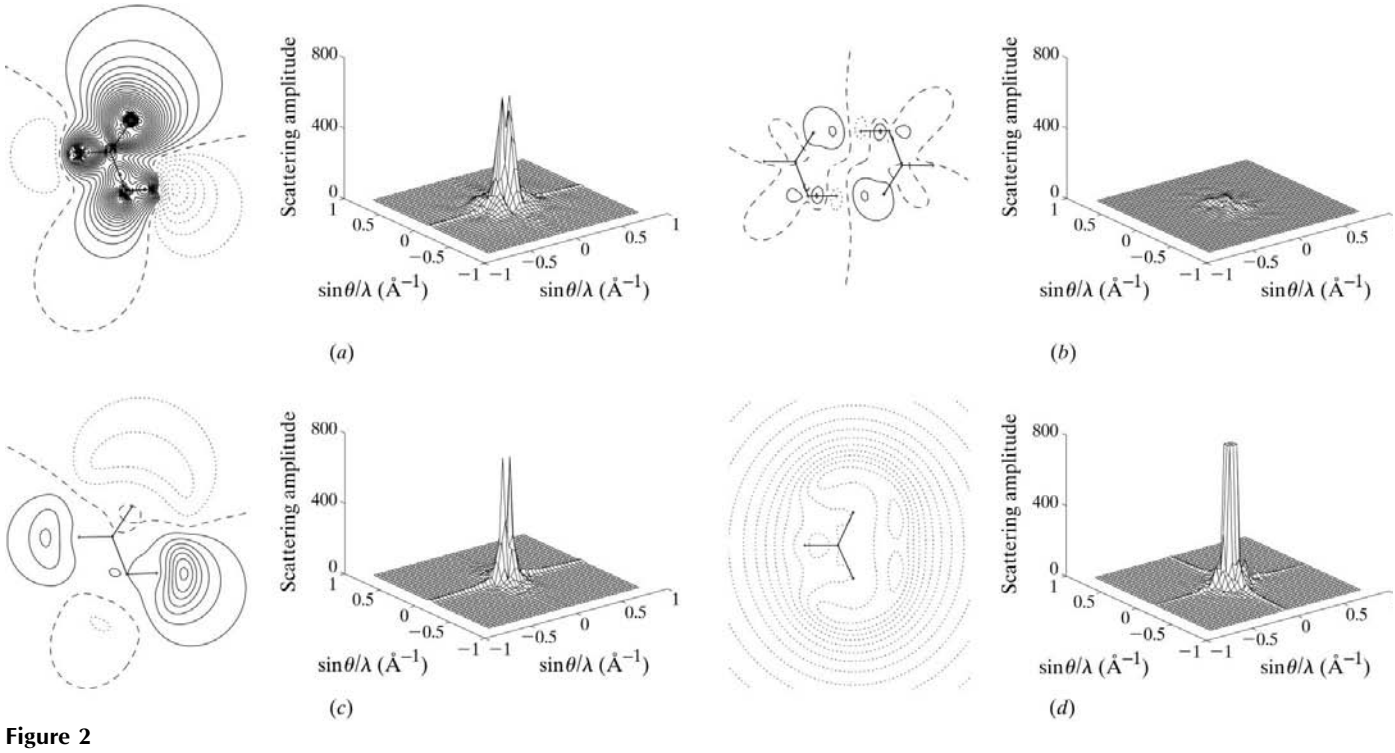

Figure 2
Difference potential maps in the molecular plane, shown on the left, and the central sections of squares of their Fourier transformation, shown on the right. (*a*) Formic acid – the IA model of formic acid. (*b*) Formic acid dimer – two monomers. (*c*) Formic acid in an aqueous environment ($\varepsilon = 78.39\,\varepsilon_0$) – formic acid *in vacuo*. (*d*) Formate anion in an aqueous environment ($\varepsilon = 78.39\,\varepsilon_0$) – formate anion *in vacuo*. Contour interval is 0.4 V; zero contour (broken), positive contours (dotted) and negative contours (solid).

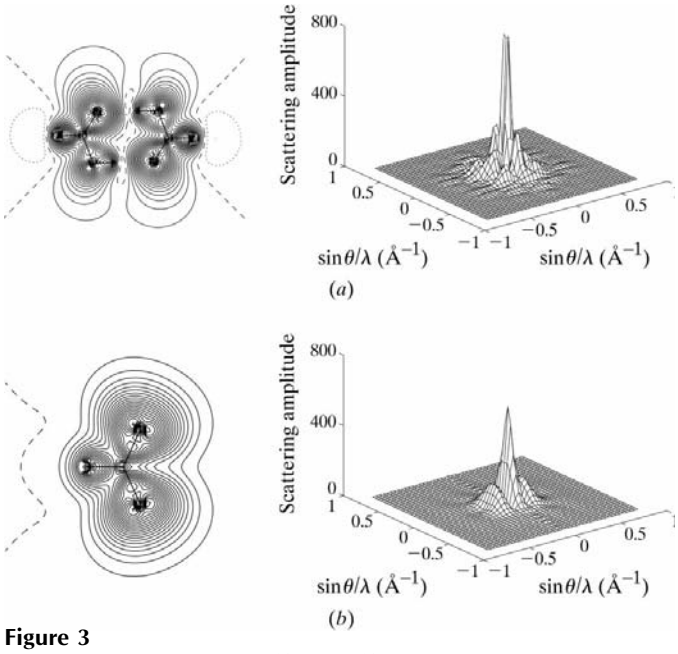

Figure 3
Difference potential maps, $\varphi(\mathbf{r}) - \varphi_{\rm isol}(\mathbf{r})$, in the molecular plane, shown on the left, and central sections of squares of their Fourier transformation, shown on the right. (*a*) Formic acid dimer; (*b*) formate anion. Contour interval is 0.4 V; zero contour (broken), positive contour (dotted) and negative contours (solid).

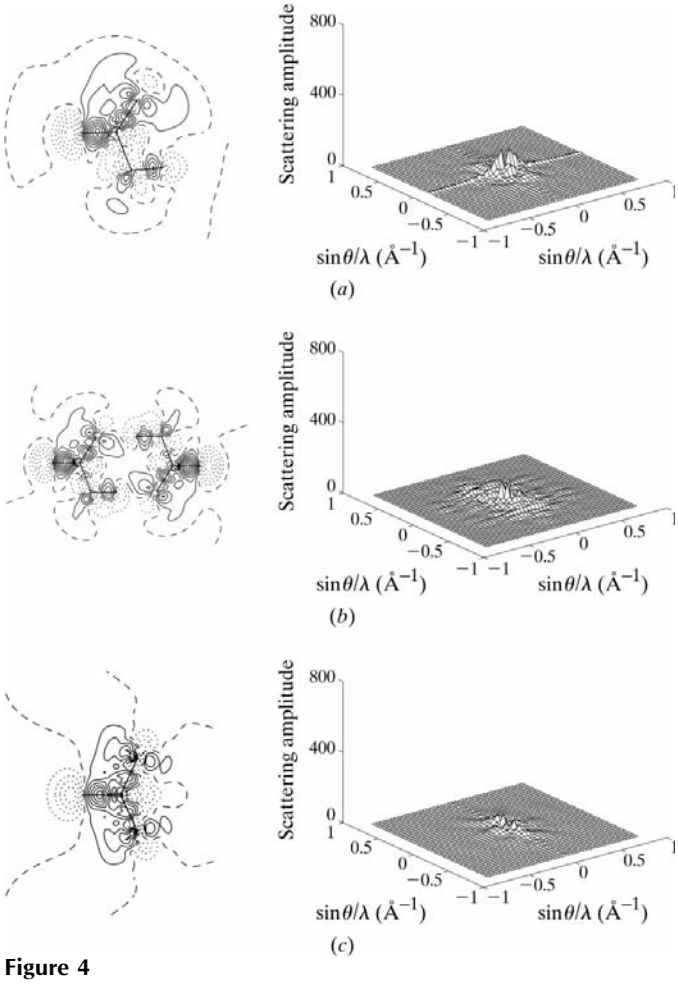

Figure 4
Same as Fig. 3, but using the SED model instead of the IA model.

Construction of a more realistic model of the electron scattering factors for electron crystallography is another objective of this paper. Chang *et al.* (1999) proposed molecular scattering factors, based on the MO calculations, without decomposing the molecular potential into atomic contributions. In the least-squares refinement using such a rigid model, however, fitting the rigid molecular fragments to a flexible molecule having slightly different bond angles and torsion angles is difficult. To avoid this difficulty, we consider the molecular potential as an aggregate of modified spherical atomic potentials depending on the molecular site and the environment. The site/environment dependence in scattering factors would be seen, for example, in the differences between the carboxylic O atoms in formic acid and the corresponding atoms in the hydrogen-bonded dimer. This model is expected to be flexible enough to accommodate any molecule. Moreover, the spherical potential does not require a major revision of the algorithms of the crystallographic refinement (Brünger, 1992; Sheldrick *et al.*, 1993). In spite of the simple form of the

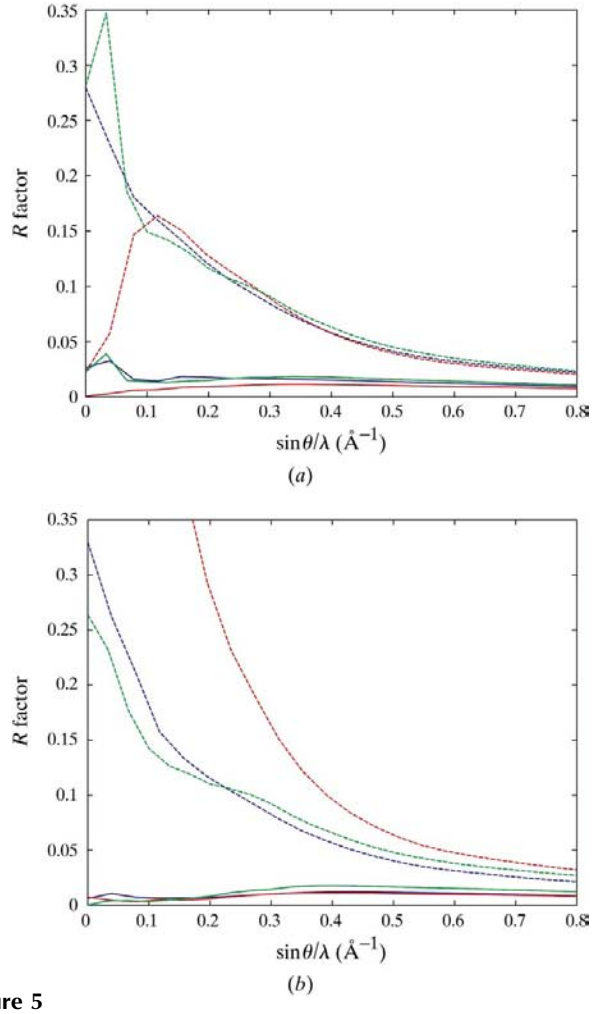

Figure 5
R factors for the SED models (solid lines) and those for the IA model (broken lines) of formic acid (blue lines), formic acid dimer (green lines) and the formate anion (red lines); (*a*) *in vacuo*; (*b*) in an aqueous environment ($\varepsilon = 78.39\,\varepsilon_0$).

model, it is precise enough to reproduce the results of *ab initio* MO calculations, as shown below.

2. *Ab initio* MO calculations

All geometrical parameters for formic acid, formic acid dimer and the formate anion were fully optimized by the density functional theory with the hybrid functional B3LYP (Becke, 1993; Lee *et al.*, 1988), and the basis set of 6-311++G(2*d*,2*p*), using the *Gaussian98* package (Frisch *et al.*, 1998). Molecular electrostatic potentials were evaluated at the same level of the theory on a cubic grid with a side of 12.8 Å and a grid spacing of 0.1 Å for formic acid and the formate anion, and on a $15.0 \times 12.8 \times 12.8$ Å rectangular grid for the formic acid dimer. The electrostatic potential is expressed as

$$\varphi(\mathbf{r}) = (1/4\pi\varepsilon_0) \int \mathrm{d}\mathbf{r}' \rho(\mathbf{r}')(|\mathbf{r} - \mathbf{r}'|)^{-1}, \quad (6)$$

where $\rho(\mathbf{r})$ is the total charge distribution including contribution from both the electrons and the nuclei.

The electrostatic potential of each molecule in a polar medium with dielectric constant $\varepsilon = 78.39\,\varepsilon_0$, ε_0 being the dielectric constant *in vacuo*, was calculated employing the polarizable continuum model (PCM) (Miertus *et al.*, 1981), using *Gaussian98*. In the PCM, the potential $\varphi(\mathbf{r})$ consists of contributions from the solute charge distribution $\rho(\mathbf{r})$ and those from a set of virtual charges q_k at $\mathbf{r}_k$ on the surface of a cavity embedded in an infinite polarizable dielectric continuum, that is,

$$\varphi(r) = (1/4\pi\varepsilon_0) \int \mathrm{d}r' \rho(r')(|r - r'|)^{-1} + (1/4\pi\varepsilon_0) \sum_k q_k (|r - r_k|)^{-1}. \quad (7)$$

Here the cavity is built up by spheres with appropriate radius around each solute atom.

3. Models for the electrostatic potential and evaluation of the scattering factor

The model potential, $\varphi_{\mathrm{site}}(\mathbf{r})$, describes the molecular electrostatic potential as an expansion of modified Gaussian functions, $\varphi_{\mathrm{site},\alpha}(r_\alpha)$, centered at the nucleus of each atom,

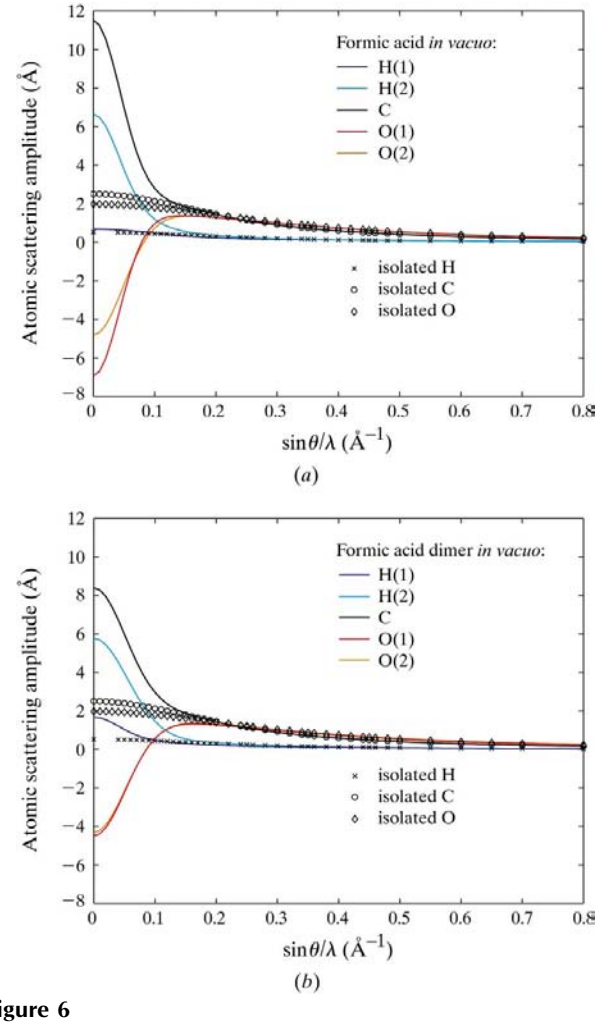

Figure 6
SED scattering factors for the atoms in (*a*) formic acid and (*b*) formic acid dimer *in vacuo*, together with IA scattering factors taken from *International Tables for Crystallography*, Vol. C.

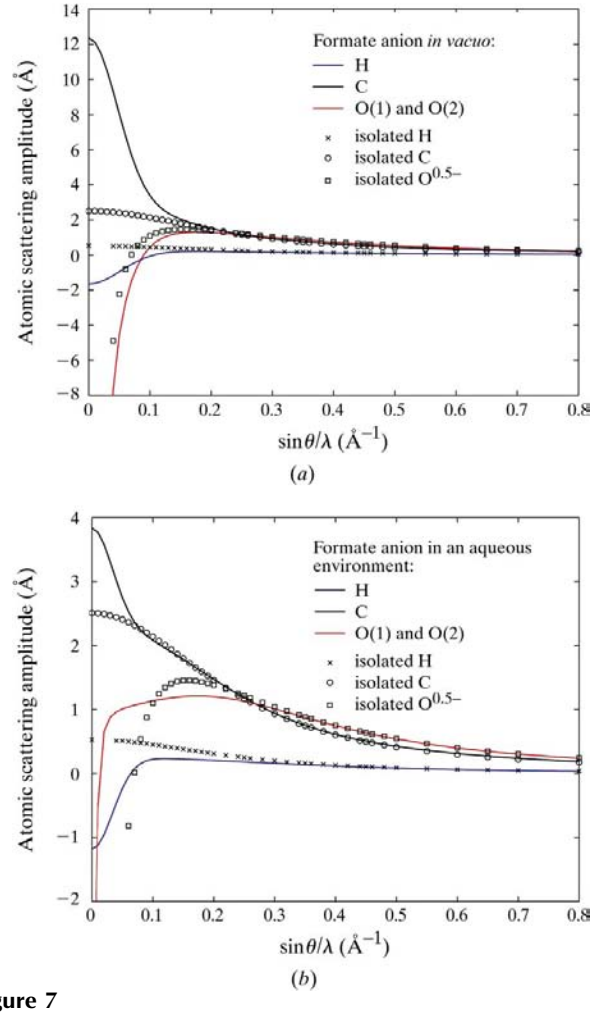

Figure 7
SED scattering factors for the atoms in formate anion (*a*) *in vacuo* and (*b*) in an aqueous environment ($\varepsilon = 78.39\,\varepsilon_0$), together with IA scattering factors taken from *International Tables for Crystallography*, Vol. C.

$$\varphi_{\rm site}(\mathbf{r}) = \sum_{\alpha} \varphi_{{\rm site},\alpha}(r_\alpha) = \sum_{\alpha} \sum_{i}^{n} a_{\alpha i} \exp(-b_{\alpha i} {r_\alpha}^2), \qquad (8)$$

where $r_\alpha = |\mathbf{r} - \mathbf{r}_\alpha|$, and $n = 6$ is used in this paper. The coefficients $a_{\alpha i}$ and $b_{\alpha i}$ were determined so as to reflect the molecular site and the environment by minimizing the target function,

$$\chi^2 = \sum_{i=1}^{N} [\varphi(\mathbf{r}_i) - \varphi_{\rm site}(\mathbf{r}_i)]^2, \qquad (9)$$

where N is the number of grid points and $\varphi(\mathbf{r}_i)$ is the molecular electrostatic potential on the ith grid point obtained by *ab initio* MO calculation either *in vacuo* or with the PCM. Therefore, we call (8) the potential of the site/environment-dependent (SED) model. A grid spacing of 0.1 Å used here was confirmed to be fine enough to reproduce the electrostatic potential, except near the centers of the nuclei. As discussed by Chang *et al.* (1999), since the potential near the nuclei has little relevance to the chemical bonding effects, the potential values were truncated at 0.1 Å from the nucleus centers, and replaced by the average value of the potential at a radius of 0.1 Å around each nucleus. As a reference, we calculated the IA model potentials, $\varphi_{\rm isol}(\mathbf{r})$, as a sum of the spherically averaged potentials of isolated atoms with the same nuclear positions as those in $\varphi_{\rm site}(\mathbf{r})$. In the formate anion, we assumed an ionic charge of -0.5 on both the O(1) and O(2) atoms (for atom names, see Fig. 1). Therefore, in the IA model, the potential of $O^{0.5-}$, the average potential of O and O^{1-}, was used for the atoms.

Our assumption is that a modified atomic potential $\varphi_{{\rm site},\alpha}(r_\alpha)$ is spherically symmetrical. It follows immediately that the SED scattering factor is

$$f_{{\rm site},\alpha}(s) = (8\pi^2 m_0 e/h^2) \times \int_0^\infty r_\alpha^2 \varphi_{{\rm site},\alpha}(r_\alpha)[\sin(4\pi s r_\alpha)/(4\pi s r_\alpha)]\, {\rm d}r_\alpha. \qquad (10)$$

On the other hand, the scattering factor for an ionized atom has to be calculated in a different manner. To avoid the numerical Fourier transformation of the long-tailed unscreened long-range electrostatic potential, the unscreened part is treated separately from the potential $\varphi(\mathbf{r})$ as

$$\varphi_{{\rm site},\alpha}(r_\alpha) = \varphi^0_{{\rm site},\alpha}(r_\alpha) + (e\Delta Z/4\pi\varepsilon r_\alpha), \qquad (11)$$

where $\varphi^0_{{\rm site},\alpha}(r_\alpha)$ is the screened atomic potential and ΔZ indicates the ionic charge. Then, the scattering factor for the ionized atom becomes

$$\begin{aligned} f_{{\rm site},\alpha}(s) &= (8\pi^2 m_0 e/h^2) \int_0^\infty r_\alpha^2 \varphi^0_{{\rm site},\alpha}(r_\alpha)[\sin(4\pi s r_\alpha)/(4\pi s r_\alpha)]\, {\rm d}r_\alpha \\ &\quad + (m_0 e^2 \Delta Z/8\pi h^2 \varepsilon s^2) \\ &= f^{\,0}_{{\rm site},\alpha}(s) + (m_0 e^2 \Delta Z/8\pi h^2 \varepsilon s^2), \end{aligned} \qquad (12)$$

where $f^{\,0}_{{\rm site},\alpha}(s)$ arises from the screened atomic field. In this representation, the divergent part of the electron scattering amplitude of an ion, the second term of (12), comes from the unscreened long-range electrostatic potential, the second term of (11), as indicated by Doyle & Turner (1968).

4. Site and environment dependence of electron scattering amplitude

Various difference potential maps shown in Fig. 2 demonstrate the site dependence (redistribution of valence electrons in Fig. 2*a*) and the environment dependence (hydrogen bonding in Fig. 2*b*, and solvation in Figs. 2*c* and *d*) of electron scattering.

Fig. 2(*a*) shows the difference between the molecular potential of formic acid and the potential of an aggregate of the isolated atoms, $\varphi(\mathbf{r}) - \varphi_{\rm isol}(\mathbf{r})$, and the square of its Fourier transformation, $[F(\mathbf{s}) - F_{\rm isol}(\mathbf{s})]^2$, clearly demonstrating the influence of redistribution of valence electrons, or the site dependence. In the difference potential map (the left-hand side of Fig. 2*a*), the negative regions caused by charge accumulation can be observed in the lone-pair sites of the carboxylic O(1) and O(2) atoms. Charge depletion of the H(1) and H(2) atoms appears in the positive regions near those atoms. The bond regions are all marked by asymmetrical negative peaks originating from the bonding electrons. Such potential differences yield large changes in reciprocal space, particularly at low scattering angles, as shown on the right-hand side of Fig. 2(*a*). Fig. 3 shows similar results in the cases of the formic acid dimer and the formate anion.

The influence of hydrogen bonding is shown in the difference between the potential of the formic acid dimer and the potential calculated from two monomers (the left-hand side of Fig. 2*b*). In this comparison, the monomer potential was calculated by constraining the coordinates of formic acid at

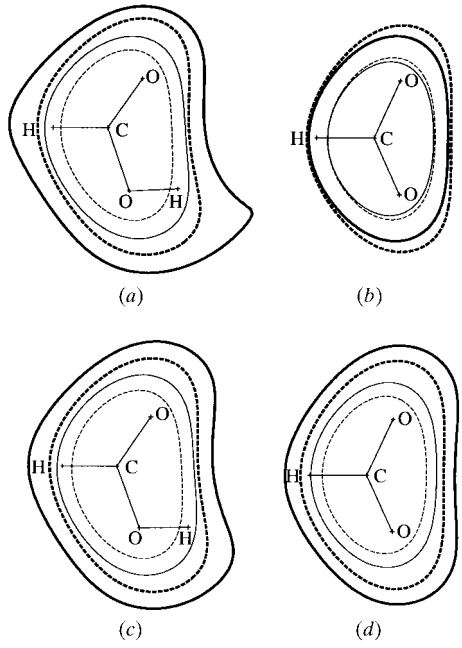

Figure 8
Potential maps on the molecular planes of (*a*) and (*c*) formic acid, and (*b*) and (*d*) the formate anion calculated from the SED atomic scattering factors with $0.0 \le s \le 0.2$ (solid lines), and those with $0.1 \le s \le 0.2$ (broken lines) omitting the low-angle scattering; (*a*) and (*b*) *in vacuo*, and (*c*) and (*d*) in an aqueous environment ($\varepsilon = 78.39\,\varepsilon_0$). Contours at 10.0 and 1.6 V are denoted by thin and thick lines, respectively.

the optimum position of the dimer. The redistribution of valence electrons upon hydrogen bonding is dominated by polarization interaction (Morokuma, 1977), which yields the induced polarization, $O^{-\delta}\cdots H^{+\delta}O^{-\delta}$. Accordingly, the difference potential map (the left-hand side of Fig. 2*b*) is positive around the H(2) atom, and negative around the O(1) and O(2) atoms. However, the polarization causes only small changes in reciprocal space as shown on the right-hand side of Fig. 2(*b*). The magnitude is one order smaller than the changes upon covalent bond formation (the right-hand side of Fig. 2*a*). This implies that changes in non-bonded interactions might not modify potential or the electron scattering intensity significantly.

On the other hand, solvation appears to have a large influence on the potential, as shown in Figs. 2(*c*) and 2(*d*), which show differences between potentials calculated by the PCM and *in vacuo*. When going from the gas phase to a polar medium ($\varepsilon = 78.39\,\varepsilon_0$), formic acid develops greater local charge concentration, which makes potential tails spread over the solvent region. The potential tails, however, are compensated by oriented water dipoles represented by the virtual charges. This potential difference corresponds to a large change at low scattering angle in reciprocal space, as shown on the right-hand side of Fig. 2(*c*). The change in scattering amplitude is almost the same level as in covalent bond formation (the right-hand side of Fig. 2*a*). Such a solvation effect is much more prominent in the formate anion (the left-hand side of Fig. 2*d*). The PCM potential can be divided into the contributions from the two terms in (7). The difference between the first term, the solute potential, and the potential *in vacuo* is in the same level as the change upon hydrogen-bond formation in Fig. 2(*b*) (data not shown). Therefore, the large influence of solvation should be attributed to the second term in (7) or the charges of oriented water dipoles. This means that ignorance of surrounding solvents in a model produces serious errors in the interpretation of potential maps, particularly when the model contains ions.

5. Accuracy of the SED model

The site/environment dependence of the potential and electron scattering have been illustrated above. Here we analyze the reproducibility of the variation in potential by the SED model.

In order to judge the accuracy in reproducing the potentials of formic acid, formic acid dimer and the formate anion, the difference potential maps and the squared differences of their Fourier transformations for the IA model are shown in Figs. 2(*a*) and 3. In Fig. 4, the corresponding quantities for the SED model are shown. The accuracy of the fit was also evaluated by the *R* factor, $R = \sum ||F| - |F_{\rm model}|| / \sum |F|$, where $F_{\rm model}$ is the structure factor for either the IA or SED model. The values of the *R* factor are summarized in Fig. 5 for the cases *in vacuo* and in the PCM.

For the IA model, as explained above, the redistribution of the valence electrons significantly affects the potential map and the electron scattering. The *R* factors for formic acid and its dimer *in vacuo* exceed 0.25 at low scattering angles. On the other hand, in the case of the anion, the *R* factor decreases at low scattering angles (Fig. 5*a*). This is because the contribution from the unscreened potential arising from the excess charge prevails over the local charge contribution outside the molecule, and is well reproduced by the potential of the $O^{0.5-}$ atom in the IA model.

After optimization of parameters in the atomic scattering factors, the accuracy of the SED model considerably improves that from the IA model, as shown in Figs. 4 and 5. The *R* factors *in vacuo* are reduced from 0.120, 0.116 and 0.127 for formic acid, formic acid dimer and the formate anion, respectively, in the IA model, to 0.018, 0.015 and 0.009 in the SED model if the data up to $s = 0.1$ Å^{-1} are included. In the PCM model, the charge compensation by the solvent dipoles alleviates the anisotropy of the potential, and yields better fits: the *R* factors are 0.007, 0.003 and 0.004, respectively. These excellent fits were attained by the flexibility of the model potential, $\varphi_{\rm site}(\mathbf{r})$ in (8), whose parameters, $a_{\alpha i}$ in (8), can be either positive or negative. Even though each Gaussian function is isotropic, the superposition of such functions with different centers corresponding to each atomic position can produce an anisotropic molecular potential. However, the strong anisotropy in the potential around H(1) or H cannot be reproduced by the spherical Gaussian model in all three cases; positive and negative regions near the H atom shown in Fig. 4. This is a limitation of the spherical Gaussian model. A further improvement may be achieved by placing pseudo point charges at the center of the positive and the negative regions near the H atom as suggested by Hirshfeld & Rabinovich (1967).

6. SED atomic scattering factors

In Figs. 6 and 7, we show the behavior of the SED scattering factors, $f_{{\rm site},\alpha}(s)$, for the atoms in the molecules.

Considering the fact that the partial charges, *e.g.* the electrostatic potential (ESP) charges (Besler *et al.*, 1990), are derived by decomposing the molecular electrostatic potential into atomic contributions, we notice that the SED atomic scattering factors at $s = 0$, calculated by decomposing the molecular potential, should reflect the local potentials in the molecules. Actually, the excess values, $f_{{\rm site},\alpha}(0) - f_{{\rm isol},\alpha}(0)$, correlate well with the ESP charges, the linear correlation coefficient being 0.984. Here, the limiting value, $f_{{\rm site},\alpha}(0)$, was obtained from (10) as

$$f_{{\rm site},\alpha}(0) = (8\pi^2 m_0 e/h^2) \int_0^\infty r_\alpha^2 \varphi_{{\rm site},\alpha}(r_\alpha)\,{\rm d}r_\alpha. \qquad (13)$$

As seen in Fig. 6(*a*), the curves for the O(1) and O(2) atoms in formic acid become negative at low scattering angles, and thus the difference increases when compared with the corresponding scattering curves for the isolated atom. On the other hand, at larger *s* values the curves of $f_{{\rm site},\alpha}(s)$ approach the values of $f_{{\rm isol},\alpha}(s)$. This can be explained by the charge accumulation in the lone-pair regions, whose influence spreads

Table 1
R factors defined in (14).

s range	*In vacuo*	In the aqueous environment
$0.0 \leq s \leq 0.1$	0.594	0.242
$0.1 \leq s \leq 0.2$	0.129	0.137

over a wide area, but have little effect on the local potential around the nuclear position. However, the good agreement between the curve for the H(1) atom and the values of the isolated H atom may reflect the failure to fit the asymmetric potential (the left-hand side of Fig. 4*a*) by the symmetrical model, as pointed out above. The curve should have increased at low scattering angles as in H(2).

The curves for the O(1) and O(2) atoms in the formate anion (Fig. 7*a*) decrease rapidly at $s = 0.1$ and diverge as s approaches zero. This is attributed to the unscreened long-range electrostatic part of the scattering factor, the second term of (12), which dominates at sufficiently low scattering angles. At higher scattering angles, on the other hand, the screened electrostatic part of the scattering amplitude, the first term of (12), becomes dominant, and converges to the values of the isolated atom.

The effects of charge compensation due to hydrogen bonding and solvation can be seen in Figs. 6(*b*) and 7(*b*), respectively, where a reduction in the magnitude of the atomic scattering factors are shown.

Comparison between the values of $f_{\mathrm{site},\alpha}(s)$ and $f_{\mathrm{isol},\alpha}(s)$ shows that the influence of bonding and the environment are significant at low scattering angles. The differences appear at an experimentally observable range of scattering angles (Kimura *et al.*, 1997), and indicate the importance of a more sophisticated model of the scattering functions like the SED atomic scattering factors proposed here.

7. Identification of ionization state of an atom

Kimura *et al.* (1997) and Mitsuoka *et al.* (1999) attempted to identify the ionization states of glutamic acids and aspartic acids in bacteriorhodopsin by comparing two kinds of potential maps, one calculated by all the data and the other obtained by excluding the low-angle data. They expected that when the negative contribution of an anion appearing in the low-angle data is excluded, the potential becomes more positive at the charge site. Here, we compared the two kinds of potential maps, one calculated by the structure factor with $0.0 \leq s \leq 0.2$ and the other with $0.1 \leq s \leq 0.2$.

In the case of neutral formic acid, when the low-angle data were removed in the calculation, the potential decreased both *in vacuo* and in the aqueous environment (Figs. 8*a* and 8*c*). On the other hand, for the ionized formate anion, the removal of the low-angle data increased the potential *in vacuo* (Fig. 8*b*), but decreased the potential in the aqueous environment (Fig. 8*d*). These observations suggest that when a charge is located in a low dielectric medium it is possible to identify the ionization state in the local potential map, as seen in the difference between Figs. 8(*a*) and 8(*b*). However, when the charge is surrounded by the aqueous environment, the charge compensation by oriented water molecules fully masks the charge effect on the local potential map.

According to the results given in Figs. 2–7, the contribution of ionization appears in the potential tail spread over the solvent region, rather than in the local potential near the nucleus. In other words, the ionization effect appears more clearly in the low-angle scattering. Therefore, the *R* factor for the low-angle data should be a better measure for identification of the ionization state than the local potential map. To examine this possibility, we calculated the *R* factor, defined by

$$R = \sum ||F(\mathbf{s})_{\mathrm{acid}}| - |F(\mathbf{s})_{\mathrm{anion}}|| / \sum |F(\mathbf{s})_{\mathrm{acid}}|, \qquad (14)$$

where $F(\mathbf{s})_{\mathrm{acid}}$ and $F(\mathbf{s})_{\mathrm{anion}}$ are the molecular structure factors for formic acid and the formate anion, respectively. The *R* values thus calculated are listed in Table 1. These values indicate that there is an obvious distinction between the molecular structure factors for formic acid and the formate anion, particularly in the low-angle region, even in the aqueous environment.

However, in the case of a protein having many anions and cations in the side chains, it is expected that the contribution to the *R* factor from a single atom becomes small and depends on the choice of the ionization states of the other atoms. Hence, there will be a difficulty in identifying the charged states of ionizable side chains of a protein from the *R* factor. An application of the SED scattering factors to a protein, bacteriorhodopsin, and an attempt to identify the ionization states are in progress.

We are grateful to Drs K. Mitsuoka and Y. Fujiyoshi for fruitful discussions, and Dr G. Basu for his critical comments on the manuscript. This work was supported by a grant from MESC to AK. The computations were performed at the Computer Center of the Institute for Molecular Science, Center for Promotion of Computational Science and Engineering of JAERI.

References

Becke, A. D. (1993). *J. Chem. Phys.* **98**, 5648–5652.
Besler, B. H., Merz, K. M. Jr & Kollman, P. A. (1990). *J. Comput. Chem.* **11**, 431–439.
Brünger, A. T. (1992). *X-PLOR*. Version 3.1. *A System for Crystallography and NMR*. Yale University Press.
Chang, S., Head-Gordon, T., Glaeser, R. M. & Downing, K. H. (1999). *Acta Cryst.* A**55**, 305–313.
Coppens, P. (1996). *International Tables for Crystallography*, Vol. B, edited by U. Shmueli, pp. 10–21. Dordrecht: Kluwer Academic Publishers.
Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. Oxford University Press.
Cowley, J. M. (1992). *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson, pp. 223–245. Dordrecht: Kluwer Academic Publishers.

Doyle, P. A. & Turner, P. S. (1968). *Acta Cryst.* A**24**, 390–397.
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., 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, PA, USA.
Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. (1996). *J. Mol. Biol.* **259**, 393–421.
Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.
Hirshfeld, F. L. & Rabinovich, D. (1967). *Acta Cryst.* **23**, 989–1000.
Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T. & Fujiyoshi, Y. (1997). *Nature (London)*, **389**, 206–211.
Lee, C., Yang, W. & Parr, R. G. (1988). *Phys. Rev. B*, **37**, 785–789.
Miertus, S., Scrocco, E. & Tomasi, J. (1981). *Chem. Phys.* **55**, 117–129.
Mitsuoka, K., Hirai, T., Murata, K., Miyazawa, A., Kidera, A., Kimura, Y. & Fujiyoshi, Y. (1999). *J. Mol. Biol.* **286**, 861–882.
Morokuma, K. (1977). *Acc. Chem. Res.* **10**, 294–300.
Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hope, H. & Sieker, L. C. (1993). *Acta Cryst.* D**49**, 18–23.
Weiss, R. J. & Freeman, A. J. (1959). *J. Phys. Chem. Solids* **10**, 147–161.