# Distinct response to dioxin in an arylhydrocarbon receptor (AHR)-humanized mouse

Takashi Moriguchi*, Hozumi Motohashi*†, Tomonori Hosoya*‡, Osamu Nakajima*, Satoru Takahashi*, Seiichiroh Ohsako†§, Yasunobu Aoki†§, Noriko Nishimura†§, Chiharu Tohyama†§, Yoshiaki Fujii-Kuriyama*†‡, and Masayuki Yamamoto*†‡¶

*Institute of Basic Medical Sciences, Center for Tsukuba Advanced Research Alliance, and ‡Exploratory Research for Advanced Technology Environmental Response Project, University of Tsukuba, Tsukuba 305-8575, Japan; §Environmental Health Sciences Division, National Institute for Environmental Studies, Onogawa, Tsukuba 305-8506 Japan; and †Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Kawaguchi 332-0012, Japan

Edited by Ronald M. Evans, The Salk Institute for Biological Studies, San Diego, CA, and approved March 19, 2003 (received for review December 26, 2002)

**There are large inter- and intraspecies differences in susceptibility to dioxin-induced toxicities. A critical question in risk assessment of dioxin and related compounds is whether humans are sensitive or resistant to their toxicities. The diverse responses of mammals to dioxin are strongly influenced by functional polymorphisms of the arylhydrocarbon receptor (AHR). To characterize responses mediated by the human AHR (hAHR), we generated a mouse possessing hAHR instead of mouse AHR. Responses of these mice to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) and 3-methylcholanthrene were compared with the responses of naturally sensitive (C57BL/6J) and resistant (DBA/2) mice. Mice homozygous for *hAHR* exhibited weaker induction of AHR target genes such as *cyp1a1* and *cyp1a2* than did C57BL/6J ($Ahr^{b\text{-}1/b\text{-}1}$) mice. DBA/2 ($Ahr^{d/d}$) mice were less responsive to induction of *cyp* genes than C57BL/6J mice. hAHR and DBA/2 AHR exhibit similar ligand-binding affinities and homozygous h*AHR* and $Ahr^{d/d}$ mice displayed comparable induction of AHR target genes by 3-methylcholanthrene. However, when TCDD was administered, a greatly diminished response was observed in homozygous h*AHR* mice compared with $Ahr^{d/d}$ mice, indicating that hAHR expressed in mice is functionally less responsive to TCDD than DBA/2 AHR. After maternal exposure to TCDD, homozygous h*AHR* fetuses developed embryonic hydronephrosis, but not cleft palate, whereas fetuses possessing $Ahr^{b\text{-}1}$ or $Ahr^{d}$ developed both anomalies. These results suggest that hAHR may define the specificity of the responses to various AHR ligands. Thus, the hAHR knock-in mouse is a humanized model mouse that may better predict the biological effects of bioaccumulative environmental toxicants like TCDD in humans.**

human | C57BL6/J | DBA/2 | CYP1A1

Polycyclic aromatic hydrocarbons (PAH) and halogenated aromatic hydrocarbons (HAH), including 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), benzo[*a*]pyrene, and polychlorinated biphenyls, are ubiquitous environmental toxicants whose chemical stability and lipophilicity make them highly persistent in the environment and in living organisms. These groups of chemicals cause various toxicological and biological responses, typified by teratogenesis, thymic atrophy, severe epithelial disorders, wasting syndrome, tumor promotion, and induction of xenobiotic-metabolizing enzymes in experimental animals (1, 2). The toxicities of these compounds are mediated by a conserved signaling pathway (1–4) through binding to and activation of the arylhydrocarbon receptor (AHR). AHR activation in turn mediates a transcriptional response for genes regulated by this transcription factor (5–8). Despite strong conservation of this pathway, there are wide inter- and intraspecies differences in the toxicological responses to AHR ligands (9–11). The molecular basis for these species and strain differences appears to relate to polymorphisms in AHR. Factors influencing susceptibility to the toxicity of TCDD have been studied in several animal models. There is a 10-fold difference in susceptibility between the dioxin-sensitive C57BL/6 and the resistant DBA/2 strains of mice that can be explained by polymorphic variations in the ligand-binding domain and in the C-terminal region of the AHR molecule of each strain (9, 12–14). Response to TCDD in the Long–Evans (sensitive) and Han/Wistar rats (resistant) differs by >1,000-fold due to a critical point mutation in the transactivation domain in the AHR of the Han/Wistar rat (15–17).

The effects of TCDD on humans are less well understood, although high incidences of chloracne, teratogenicity, and abortion have been associated with high blood concentrations of dioxin and related compounds in residents of regions where industrial accidents or extensive use of dioxin-containing defoliants have resulted in human exposures (3). Increased levels of dioxin in the body have been reported recently to be associated with abnormal sex ratio of newborns nearly 25 years after the accident in Seveso, Italy (18). Because the AHR primarily mediates the pleiotropic manifestations of dioxin exposure, characterization of the structural and functional properties of the human AHR (hAHR) is critical for understanding the types and magnitudes of human responses to various PAH/HAHs.

To date, *in vitro* characterization of the hAHR has provided ambiguous insights into human sensitivity to dioxin. The dissociation constant ($K_d$) of hAHR for TCDD was comparable to that of TCDD-resistant DBA/2 AHR (9, 19), suggesting that humans might be resistant to TCDD. By contrast, high homology of the human receptor to the AHR of the guinea pig, which is the most sensitive animal to TCDD, suggests a high responsiveness of humans to the toxin (20). Ligand specificity of hAHR was also examined and compared with those of zebrafish and rainbow trout AHRs using polychlorinated dibenzo-*p*-dioxin, dibenzofuran, and biphenyl congeners as test ligands. These studies revealed that mono-ortho polychlorinated biphenyls activated hAHR but were not very effective in activating either zebrafish or rainbow trout AHRs (21).

Assessment of human responses *in vivo* to unintended exposures to various PAH/HAHs has been hampered by limited exposure assessments and toxicological follow-up. Observational studies after intentional exposures have not been and should not be conducted. To gain stronger insight into the hazards to human health posed by compounds interacting with the hAHR *in vivo*, we generated a mouse model that harbors the hAHR cDNA instead of the mouse *AhR* gene. This mouse may reveal a humanized susceptibility to chemical toxicities. In response to challenges with 3-methylcholanthrene (3-MC) and TCDD, two prototypical AHR ligands, the hAHR knock-in

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: AHR, aryl hydrocarbon receptor; hAHR, human AHR; h*AHR*, human AHR knock-in allele; $Ahr^{d}$, DBA/2 *Ahr* allele; $Ahr^{b\text{-}1}$, C57BL/6 *Ahr* allele; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; 3-MC, 3-methylcholanthrene; PAH, polycyclic aromatic hydrocarbons; HAH, halogenated aromatic hydrocarbons; ES, embryonic stem; GD, gestation day.

¶To whom correspondence should be addressed. E-mail: masi@tara.tsukuba.ac.jp.



www.pnas.org/cgi/doi/10.1073/pnas.1037886100

mouse displayed a distinct response profile compared with control animals harboring either the C57BL/6 *Ahr* allele ($Ahr^{b\text{-}1}$) (TCDD-sensitive C57BL/6J AHR) or the DBA/2 *Ahr* allele ($Ahr^{d}$) (TCDD-resistant DBA/2 AHR) in the same C57BL/6J genetic background. Although gene expression responses mediated by hAHR from 3-MC were comparable to that by DBA/2 AHR, the homozygous human AHR knock-in allele (h*AHR*) mouse was the weakest responder to TCDD among the three strains examined. These results suggest that hAHR molecules expressed in mice retain a functional human specificity that can be distinguished from the murine AHR and provide important insights into the toxicological susceptibility of humans to AHR ligands released into the environment.

## Materials and Methods

**Construction of the hAHR Knock-in Vector.** The hAHR knock-in vector was constructed by using 129SV/J mouse *Ahr* genomic clones and hAHR cDNA as described (22). A 2-kb *Bam*HI/*Hph*I fragment containing the 129SV/J *Ahr* promoter was ligated to the hAHR cDNA (9, 22). The *neo* gene cassette was fused to the 3′ end of the hAHR cDNA in a reverse orientation, followed by a 6.5-kb *Hind*III/*Eco*RI fragment of the 129SV/J *Ahr* gene. This construct was ligated to the thymidine kinase cassette on the 5′ end.

**Generation of hAHR Knock-in Mice.** The knock-in vector was electroporated into E14 embryonic stem (ES) cells (23). A pair of primers (sense, GTATGCATTACCATGCTCCCATTCTGCTGG; antisense, ACATCTTGTGGGAAAGGCAGCAGGCTAGCC) was used for PCR screening. After confirmation by Southern blot analysis, positive clones were injected into blastocysts. Heterozygous hAHR knock-in mice were backcrossed into a C57BL/6J background up to the seventh generation and interbred to yield heterozygous and homozygous h*AHR* and wild-type $Ahr^{b\text{-}1/b\text{-}1}$ mice. The genotype of each pup was determined by PCR, with a common sense primer; 5′-ATGAGCAGCGGCGCCAACAT-3′, an antisense primer for endogenous *Ahr* allele; 5′-GCTAGACGGCACTAGGTAGG-3′, and an antisense primer for targeted allele; 5′-CAGGTAACTGACGCTGAGCC-3′. PCR amplification was carried out for 30 cycles under the following conditions; 94°C for 30 sec, 62°C for 30 sec, and 72°C for 30 sec.

**Chemicals and Animals.** TCDD (99.5% pure) and 3-MC were purchased from Cambridge Isotope Laboratories (Andover, MA) and Wako Pure Chemical (Osaka), respectively. D2N-*Ahrd* mice and inbred C57BL6/J mice were procured from The Jackson Laboratory. *Ahr*-null mutant mice used in this study were generated by Y.F.-K (22).

**RNA Blotting Analyses.** We isolated total RNA by using ISOGEN (Nippon Gene, Tokyo) and purified polyA RNA by using an Oligotex-MAG mRNA purification kit (Takara Biotechnology, Tokyo). For detection of Ahr mRNA, 5 μg of polyA RNA per lane was applied, and a portion of mouse $Ahr^{b\text{-}1}$ cDNA (*Bpu*1102I-*Kpn*I; 734-bp) encoding the PAS domain was used for a probe. This nucleotide sequence is conserved with 83% homology to the corresponding hAHR cDNA (12, 24). To examine the inducibility of *CYP1A1* and *CYP1A2*, 6-week-old littermates ($Ahr^{b\text{-}1/b\text{-}1}$ and homozygous h*AHR*) and D2N-*Ahrd* ($Ahr^{d/d}$) mice were given a single i.p. injection of 80 mg/kg 3-MC or 100 μg/kg TCDD. Mice were killed by cervical dislocation 24 h after injection. Ten micrograms of total RNA per lane was hybridized with the appropriate mouse cDNA probes (25).

**RT-PCR Analyses of hAHR and Murine Ahr mRNA Expression in Embryos.** Total RNA was isolated from palate and kidney of gestation day (GD)18.5 fetus by using ISOGEN. One microgram of the total RNA was reverse-transcribed into cDNA with Superscript-II reverse transcriptase (Life Technologies, Gaithersburg, MD) and random hexamers at 42°C for 50 min. The resulting cDNAs were subjected to 30 cycles of PCR by using the specific primers for the gene for the h*AHR* (5′ primer, 5′-GTAAGTCTCCCTTCATACC-3′; 3′ primer, 5′-AGGCACGAATTGGTTAGAG-3′), mouse *Ahr* (5′ primer, 5′-CTTTGCTGAACTCGGCTTGC-3′; 3′ primer, 5′-TTGCTGGGGGCACACCATCT-3′) and GAPDH (5′ primer, 5′-CCCCTTCATTGACCTCAACTACATGG-3′; 3′ primer, 5′-GCCTGCTTCACCACCTTCTTGATGTC-3′). The reaction was performed under the following conditions: 94°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec.

**Immunohistochemical Analysis of hAHR Expression.** Immunohistochemical analysis was performed as described (26). Lungs were fixed in 0.1 M phosphate buffer containing 4% paraformaldehyde for 24 h and embedded in paraffin. Sections were incubated with anti-AHR antibody in 1:200 dilution, which reacts with both human and mouse AHR (N-19; Santa Cruz Biotechnology). AHR immunoreactivity was visualized with the avidin-biotin-peroxidase system (Vector Laboratories).

**TCDD Treatment and Evaluation of Teratogenesis.** TCDD treatment was performed as described (22). On GD12.5, pregnant mice were given TCDD by i.p. administration at a dose of 40 μg/kg body weight (27). On GD18.5, the fetuses were taken out and fixed in 4% paraformaldehyde. The palatal structure was examined by cutting between the upper and lower jaws. The kidneys were sliced longitudinally and stained with hematoxylin/eosin. The presence and severity of hydronephrosis in each kidney was examined under a microscope as previously described (28) by using severity scores ranging from 0 to 3+ (0, normal kidney; 1+, slight decrease in length of papilla; 2+, marked decrease in length of papilla with some loss of renal parenchyma; 3+, complete absence of papilla, shell of kidney remaining with only a small amount of renal parenchyma). For statistical analysis, pairwise comparisons were made by Mann–Whitney *U* test, by using StatView for Macintosh version 5.0 (SAS Institute, Cary, NC).

## Results

**Replacement of the Mouse *Ahr* Gene with hAHR cDNA.** We hypothesize that the specific functional characteristics of the hAHR molecule form the principal basis for the pattern of human responses to xenobiotics that interact with the AHR. To characterize responses mediated by hAHR, we generated a mouse possessing hAHR instead of murine AHR. hAHR cDNA was introduced into the mouse *Ahr* locus by homologous recombination, thereby disrupting the mouse *Ahr* gene (Fig. 1*A*). The cDNA was recombined so that hAHR is expressed under the control of the endogenous mouse *Ahr* promoter. Sixteen independent G418-resistant ES clones were obtained of 240 by PCR screening, and seven clones were further confirmed as correctly targeted ES cells by genomic DNA blot analysis. *Eco*RI-digested genomic DNA from the three representative positive clones (nos. 14, 25, and 58) revealed 11.0- and 6.2-kb fragments derived from the intact and targeted alleles, respectively, when hybridized with the 5′-external probe (Fig. 1*B*).

These three clones harboring h*AHR* were used for the generation of chimeric offspring. The male chimeras were mated with C57BL/6J females to obtain heterozygotes of the h*AHR* allele. They were subsequently bred into a C57BL/6J genetic background through the seventh generation, and the backcrossed heterozygous animals were interbred to yield hAHR homozygous mutant mice. The transmission of the targeted allele to the offspring was confirmed by genomic DNA blot analysis, and the genotype was determined by PCR by using tail DNA as a template (Fig. 1 *C* and *D*). Of 124 offspring obtained

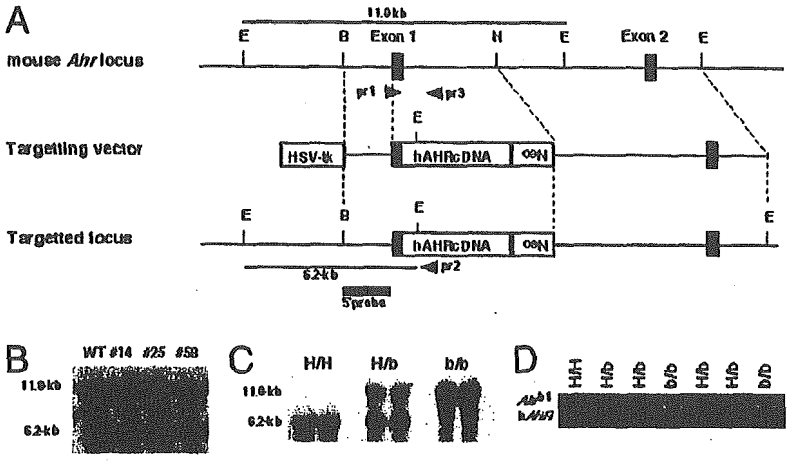



**Fig. 1.** Generation of the hAHR knock-in mouse. (*A*) Strategy for hAHR cDNA knock-in by homologous recombination. E, H, and B are restriction sites for *Eco*RI, *Hind*III, and *Bam*HI, respectively. Neo indicates the neomycin-resistance gene, and HSV-tk is the thymidine kinase gene under control of the herpes simplex virus promoter. The 5′-genomic probe used for DNA blot analysis is indicated by the hatched box. The positions of wild-type (pr 3) and mutant allele-specific (pr 2) primers and the common primer (pr 1) used in the genotyping PCR are indicated by arrowheads. The *Eco*RI restriction fragments detected with the 5′-genomic probe in the wild-type and targeted allele are denoted by horizontal bars. (*B*) DNA blot analyses of three recombinant ES clones. Genomic DNA was prepared from the ES clones (nos. 14, 25, and 58), and aliquots (10 μg) were digested by *Eco*RI. *Eco*RI digestion generated 11.0- and 6.2-kb bands for the wild-type and targeted alleles, respectively, by using the 5′-genomic probe. (*C*) Genotyping of the *Ahr* gene by DNA blot analysis. Genomic DNA was extracted from the tails of heterozygous and homozygous h*AHR* mice and wild-type $Ahr^{b\text{-}1/b\text{-}1}$ mice and digested by *Eco*RI for DNA blot analysis. (*D*) Genotyping of littermates from the intercrosses of heterozygotes. PCR fragments of wild-type amplified with pr1 and pr3 ($Ahr^{b\text{-}1}$; 280 bp) and mutant allele with pr1 and pr2 (h*AHR*; 240 bp) as depicted in *A*. H/H, H/b, and b/b indicate homozygous and heterozygous h*AHR* mice and wild type ($Ahr^{b\text{-}1/b\text{-}1}$), respectively.

from heterozygous matings, wild-type ($Ahr^{b\text{-}1/b\text{-}1}$), heterozygous, and homozygous h*AHR* mutant mice numbered 29, 71, and 24, respectively, conforming to the expected Mendelian inheritance ratio. Homozygous h*AHR* mice were viable, and no abnormalities were observed.

**Expression of hAHR in hAHR Knock-in Mice.** The expression of hAHR and mouse Ahr mRNAs was examined by RNA blot analysis by using polyA RNAs isolated from major AHR-expressing organs including liver, lung, kidney, intestine, and thymus (Fig. 2*A*). A cDNA fragment encoding the PAS domain of C57BL/6 AHR, which shows 83% homology with the corresponding human molecule, was used as a common probe for detecting both mouse Ahr and hAHR mRNAs. The larger band detected in heterozygous h*AHR* mice and wild-type $Ahr^{b\text{-}1/b\text{-}1}$ mice corresponds to the 5.4-kb transcript derived from the endogenous $Ahr^{b\text{-}1}$ gene, and the shorter 5.0-kb transcript observed in heterozygous and homozygous h*AHR* is derived from the hAHR knock-in allele. This result establishes that, whereas the homozygous h*AHR* mouse lacks mRNA for murine *Ahr*, it expresses mRNA for h*AHR*. Further, the level of expression of hAHR mRNA is comparable to that of endogenous murine Ahr mRNA in the other strains.

The embryonic expressions of mouse Ahr and hAHR mRNAs were examined by RT-PCR at the stage of GD18.5. As observed in the RNA blot analysis of adult tissues, the hAHR mRNA was expressed in the embryonic palate and kidney of homozygous and heterozygous h*AHR* mice. The abundance was comparable with that of the mouse Ahr mRNA expressed in $Ahr^{b\text{-}1/b\text{-}1}$ and heterozygous h*AHR* mice (Fig. 2*B*). These results demonstrate that hAHR mRNA is transcribed under the control of the mouse *Ahr* promoter in both adult and embryonic hAHR knock-in mice.

To ascertain that hAHR protein is expressed from the knock-in allele, immunohistochemical analysis was performed on lung sections obtained from hAHR knock-in homozygous mouse and the *Ahr*-null mutant (22). Intense signals were detected in the alveolar epithelial cells of hAHR knock-in animals (Fig. 2*C*). The signal intensity of *Ahr*-null mutant lung (Fig. 2*D*) was as faint as the hAHR knock-in lung without the antibody (data not shown). Thus, hAHR protein is expressed from the knock-in allele.

**The hAHR Knock-in Mouse Displays a Distinct Induction Profile of AHR Target Genes to Different AHR Ligands.** The response of the hAHR knock-in mouse to two prototypical AHR ligands, 3-MC and TCDD, was examined. To characterize the distinct properties, if any, of the hAHR, two strains of control mice were used for the analysis. One strain is a wild-type mouse in the C57BL/6J genetic background, which possesses AHR with high affinity for TCDD. The other strain is a congenic mouse, D2N-*Ahrd*, possessing AHR with low affinity (from DBA/2 mouse) in the C57BL6J genetic background. Because the hAHR knock-in mouse was backcrossed into C57BL/6J, these two strains of mouse enabled us to compare the characteristics of hAHR to those of C57BL/6J and DBA/2 AHR in the same genetic background.

Robust expression of the *CYP1A1* and *CYP1A2* genes was observed in the liver of $Ahr^{b\text{-}1/b\text{-}1}$ mice after administration of 3-MC, whereas the magnitudes of induction in homozygous h*AHR* and $Ahr^{d/d}$ mice were much weaker and comparable to each other (Fig. 3*A*). The relative mean band intensities for *CYP1A1* were 1.0 and 0.9 and were 1.0 and 1.1 for *CYP1A2* in homozygous h*AHR* and $Ahr^{d/d}$ mice, respectively. After treatment with TCDD, the induction of the two genes was strongest in $Ahr^{b\text{-}1/b\text{-}1}$ mice, intermediate in $Ahr^{d/d}$ mice, and weakest in homozygous h*AHR* mice (Fig. 3*B*). The fold inductions in homozygous h*AHR*, $Ahr^{d/d}$, and $Ahr^{b\text{-}1/b\text{-}1}$ mice were 1.0, 4.9, and 14.6 for *CYP1A1*, and 1.0, 5.7, and 8.4 for *CYP1A2*, respectively.

When the responses of $Ahr^{b\text{-}1/b\text{-}1}$ and $Ahr^{d/d}$ mice were compared, the *CYP1A1* expression levels were higher in $Ahr^{b\text{-}1/b\text{-}1}$ than in $Ahr^{d/d}$ mice, which is consistent with previous reports (9, 12, 13). It is noteworthy that the responsiveness of homozygous h*AHR* mice to 3-MC was almost comparable to that of $Ahr^{d/d}$ mice, whereas the responsiveness to TCDD was much weaker. The differential response between $Ahr^{d/d}$ and homozygous h*AHR* mice was unexpected, because a previous study indicated

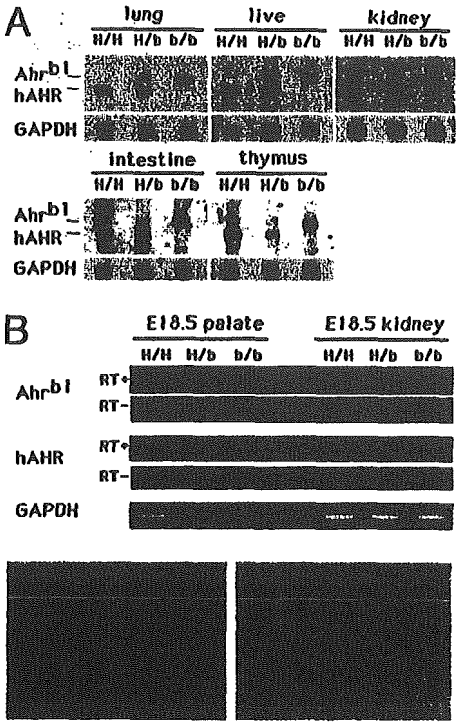



Fig. 2. *Expression of hAHR in multiple tissues of the hAHR knock-in mouse.* (*A*) RNA blot analysis of polyA RNA (5 μg/lane) extracted from five representative organs of homozygous and heterozygous h*AHR* mice and $Ahr^{b\text{-}1/b\text{-}1}$ mice. Human and mouse *Ahr* transcripts (hAHR and $Ahr^{b\text{-}1}$, respectively) are indicated (*Left*). The same membrane was rehybridized with $^{32}$P-labeled cDNA of mouse *GAPDH*. H/H, H/b, and b/b are described in the Fig. 1 legend. (*B*) RT-PCR analyses of hAHR and murine Ahr mRNA expression in kidney and palate of GD18.5 fetuses. The reverse transcription was conducted either in the presence (+) or absence (−) of reverse transcriptase. PCR products representing the transcripts derived either from h*AHR* or $Ahr^{b\text{-}1}$ are indicated on the left. (*C* and *D*) Immunohistochemical analysis of hAHR protein in the lung of a homozygous h*AHR* mouse. Immunoreactivity of AHR protein was observed in the alveolar epithelial cells of homozygous h*AHR* lung (*C*), whereas no immunoreactivity was observed in the lung of $Ahr^{-/-}$ mouse (*D*). Original magnifications, ×400 (*C* and *D*).

that hAHR and DBA/2 AHR exhibit similar dissociation constants for TCDD binding as measured *in vitro* (9, 19). This result suggests that ligand binding does not fully define the integrated function of hAHR.

**hAHR Knock-in Mouse Is Relatively Resistant to TCDD-Induced Teratogenicity.** The responses to TCDD mediated by hAHR are weaker than that by DBA/2 and C57BL/6 AHR when measured

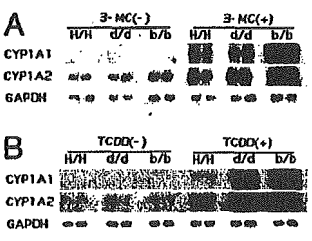



Fig. 3. *Inducible expression of AHR target genes. Northern blot analysis of* AHR-regulated *CYP1A1* and *CYP1A2* was performed. Six-week-old homozygous h*AHR*, $Ahr^{d/d}$, and $Ahr^{b\text{-}1/b\text{-}1}$ mice were treated with 80 mg/kg 3-MC (*A*) or 100 μg/kg TCDD (*B*). Total hepatic RNA was isolated 24 h after treatment and subjected to Northern analysis (10 μg/lane). Equal loading was confirmed by the abundance of *GAPDH* transcripts.

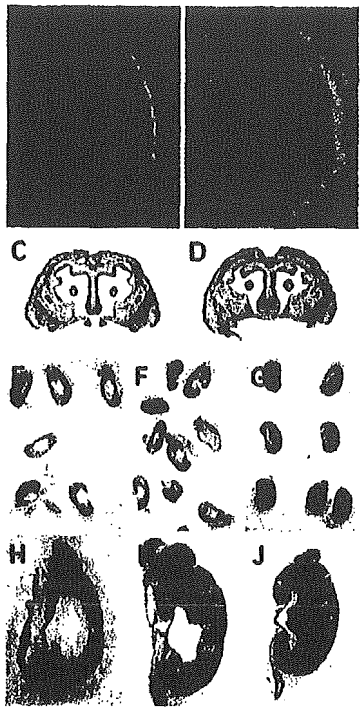



Fig. 4. Fetal teratogenesis after maternal administration of TCDD. (*A* and *C*) Cleft palate in an $Ahr^{b\text{-}1/b\text{-}1}$ fetus is shown. Filled arrowheads in *A* and open arrowheads in *C* indicate the failure of palatine shelves to fuse. Note that homozygous h*AHR* fetuses showed no cleft palates after TCDD treatment (*B* and *D*). (*E*, *F*, *H*, and *I*) Fetal hydronephrosis induced by TCDD. $Ahr^{b\text{-}1/b\text{-}1}$ (*E* and *H*) and homozygous h*AHR* (*F* and *I*) fetuses are shown. (*G* and *J*) Unaffected kidneys from untreated $Ahr^{b\text{-}1/b\text{-}1}$ fetuses are shown.

as inducibility of *CYP1A* family genes. Teratogenicity is a more integrated and complex toxicological manifestation of TCDD action. The most prominent teratogenic effects of TCDD on mouse fetus are cleft palate and hydronephrosis, both of which depend completely on AHR expression (29). The frequency and severity of these teratogenic effects of TCDD were examined in hAHR knock-in fetuses. Homozygous hAHR knock-in females were mated with males of the same genotype and given a single i.p. dose of 40 μg of TCDD per kg of body weight at GD12.5. $Ahr^{b\text{-}1/b\text{-}1}$ and $Ahr^{d/d}$ females were treated in the same way as controls. All dams were weighed to monitor the normal continuation of the pregnancy and killed at GD18.5 to remove fetuses for examination of cleft palate and hydronephrosis.

As reported previously, cleft palate was observed in 100% of the wild-type $Ahr^{b\text{-}1/b\text{-}1}$ fetuses exposed to TCDD (Fig. 4 *A* and *C* and Table 1) (22). By contrast, none of the treated homozygous h*AHR* fetuses showed abnormal palatogenesis (Fig. 4 *B* and *D* and Table 1). An intermediate frequency (30%) of cleft palate was observed in the $Ahr^{d/d}$ fetuses. Differences in the severity of cleft palate were not apparent in any of the symptomatic fetuses of any genotype. This anomaly was most frequent in $Ahr^{b\text{-}1/b\text{-}1}$, intermediate in $Ahr^{d/d}$, and least frequent in homozygous h*AHR* mice, in accordance with the transcriptional inducibility of AHR target genes, which was strongest in $Ahr^{b\text{-}1/b\text{-}1}$, intermediate in $Ahr^{d/d}$, and weakest in homozygous h*AHR* mice. Thus, a strong correlation between the incidence of cleft palate in each strain and the intrinsic transcriptional activity of their respective AHR molecules was observed.

Hydronephrosis, another teratogenic effect of TCDD, is characterized by a dilated renal pelvis. The severity of this anomaly in the fetal kidney was scored from 0 (normal) to 3 (severest)

Table 1. Incidence of anomalies caused by TCDD in homozygous h*AHR*, $Ahr^{b\text{-}1/b\text{-}1}$, and $Ahr^{d/d}$ fetuses

| Genotype of fetuses | TCDD dose, μg/kg | Dams examined, *n* | Fetuses examined, *n* | Fetuses with: Cleft palate *n* | Cleft palate %* | Hydronephrosis *n* | Hydronephrosis %* | Hydronephrosis Severity: 0–3.0† |
|---|---|---|---|---|---|---|---|---|
| $Ahr^{b\text{-}1/b\text{-}1}$ | 0 | 2 | 13 | 0 | 0 | 2 | 12.5 | 0.19 ± 0.10‡ |
| $Ahr^{b\text{-}1/b\text{-}1}$ | 40 | 5 | 29 | 29 | 100 | 26 | 89.7 | 2.54 ± 0.14§ |
| $Ahr^{d/d}$ | 0 | 2 | 15 | 0 | 0 | 2 | 13.3 | 0.20 ± 0.10‡ |
| $Ahr^{d/d}$ | 40 | 5 | 30 | 9 | 30 | 25 | 81.7 | 1.98 ± 0.14§ |
| Homo-h*AHR* | 0 | 2 | 16 | 0 | 0 | 1 | 6.3 | 0.03 ± 0.03‡ |
| Homo-h*AHR* | 40 | 5 | 37 | 0 | 0 | 30 | 81.1 | 1.19 ± 0.01§ |

*Percentage of fetuses with each anomaly of all fetuses examined.
†The criteria for severity scores are described in *Materials and Methods*. Data are expressed as mean ± SE.
‡Significant difference between TCDD-treated and -untreated fetuses of each genotype ($P < 0.0001$).
§Significant difference between TCDD-treated homozygous h*AHR* fetuses and $Ahr^{b\text{-}1/b\text{-}1}$ or $Ahr^{d/d}$ fetuses ($P < 0.0001$).

according to criteria described previously (28). When kidneys scored at 1, 2, or 3 were counted as hydronephrotic, 89.7% of the $Ahr^{b\text{-}1/b\text{-}1}$ offspring suffered from this teratogenic outcome after TCDD treatment (Fig. 4 *E* and *H*, and Table 1 for TCDD-treated animals; Fig. 4 *G* and *J*, and Table 1 for untreated animals). A similar incidence was observed in a previous study (22). $Ahr^{d/d}$ and homozygous h*AHR* fetuses also displayed this teratogenic effect with incidences of 81.7% and 81.1%, respectively (Fig. 4 *F* and *I*, and Table 1). Thus, there is no substantial difference in the incidence of hydronephrosis among the mice expressing the three distinct *Ahr* (h*AHR*) genes. When severity score values were compared among the TCDD-treated fetuses, they averaged 2.54, 1.98, and 1.19 for the $Ahr^{b\text{-}1/b\text{-}1}$, $Ahr^{d/d}$ and homozygous h*AHR* genotypes, respectively (Table 1). Therefore, hydronephrosis observed in the homozygous h*AHR* fetuses was significantly less severe compared with that in either $Ahr^{b\text{-}1/b\text{-}1}$ or $Ahr^{d/d}$ fetuses. Nonetheless, the average score of TCDD-treated homozygous h*AHR* fetuses (1.19) was still significantly higher than that of untreated homozygous h*AHR* fetuses (0.03), clearly demonstrating that the TCDD-activated hAHR mediates renal teratogenesis in mice. Although the magnitude of *CYP* gene induction is dramatically different depending on the *Ahr* genotype, the incidence of hydronephrosis is surprisingly comparable among the three strains. These results revealed that differences between human and murine AHR allowed for the emergence of discrete biological effects; e.g., hydronephrosis, but not cleft palate in homozygous h*AHR* mice.

To exclude the possibility that maternal factors affect the teratogenic manifestations on the fetuses, heterozygous h*AHR* parents were used to obtain homozygous h*AHR* and $Ahr^{b\text{-}1/b\text{-}1}$ fetuses. Heterozygous mothers were treated with TCDD as described above, and fetuses were examined for both cleft palate and hydronephrosis. As described in Table 2, the incidence of cleft palate was 100% and 0% in $Ahr^{b\text{-}1/b\text{-}1}$ and homozygous h*AHR* fetuses, respectively, which is identical to the results presented in Table 1. The incidence and severity (mean score) of hydronephrosis were 100% and 2.47 for $Ahr^{b\text{-}1/b\text{-}1}$ and 66.6% and 1.17 for homozygous h*AHR* fetuses, respectively. Again, a more moderate effect in the homozygous h*AHR* fetuses is suggested, the severity difference being statistically significant. Therefore, we conclude that the TCDD-induced teratogenic effects are independent of maternal genotypes, and that fetal AHR activity is critical for determining the outcomes.

## Discussion

One of the central issues in the uncertainty surrounding risk assessments for TCDD and its structural analogs is whether humans are relatively sensitive or resistant to the toxicities of this class of compounds. Because the pleiotropic adverse effects induced by these toxins involve multiple processes, the human response is generated by the summation and integration of the properties inherent to the human components, including expression level, ligand-binding affinity, and transcriptional activity of the AHR, as well as the variety, function and activity of the AHR target genes. Through numerous preceding studies, the primary structure of the AHR protein has been regarded as one of the most critical factors determining the susceptibility and specificity of responses of animals to various PAH/HAHs including dioxin. On the basis of several observations *in vitro*, polymorphic variation in the *Ahr* gene is considered the primary basis for differences in sensitivity to TCDD among strains of mice (9–11). In this study, we attempted to establish an *in vivo* system to evaluate the specific function of the hAHR protein to better evaluate its role in determining possible patterns of human responses to PAH/HAHs.

For this purpose, we adopted a knock-in strategy to introduce hAHR cDNA into the mouse *Ahr* genomic locus by homologous recombination. This strategy offers an obvious advantage compared with a transgenic method, because the introduced sequence is transcribed under the same regulatory mechanisms of the replaced gene (30). As desired, expression levels of the h*AHR* transcript were almost the same with those of endogenous mouse

Table 2. Incidence of anomalies caused by TCDD in fetuses from heterozygous h*AHR* parents

| Genotype of fetuses | TCDD dose, μg/kg | Dams examined, *n* | Fetuses examined, *n* | Fetuses with: Cleft palate *n* | Cleft palate %* | Hydronephrosis *n* | Hydronephrosis %* | Hydronephrosis Severity: 0–3.0 |
|---|---|---|---|---|---|---|---|---|
| $Ahr^{b\text{-}1/b\text{-}1}$ | 40 | 7 | 9 | 9 | 100 | 9 | 100.0 | 2.47 ± 0.14† |
| Hetero-h*AHR* | 40 | | 25 | 12 | 48 | 22 | 88.0 | 2.46 ± 0.13 |
| Homo-h*AHR* | 40 | | 12 | 0 | 0 | 8 | 66.6 | 1.17 ± 0.01† |

*Percentage of fetuses with each anomaly out of all fetuses examined.
†Significant difference between homozygous h*AHR* and $Ahr^{b\text{-}1/b\text{-}1}$ fetuses ($P < 0.0001$).

Ahr mRNA in multiple AHR-expressing tissues of adult mice and GD18.5 embryos. The hAHR protein was detected by immunostaining in the lungs of homozygous h*AHR* mice.

A possible explanation of the relative resistance of the hAHR knock-in mouse to TCDD lies in the qualitative difference between the human and mouse AHR molecules. Assuming that the abundance of the hAHR protein is the same as that of the endogenous mouse AHR, our results imply that the hAHR-mediated response to TCDD *in vivo* is much lower than that of DBA/2 AHR, although previous reports showed that their affinities to TCDD, as measured *in vitro*, are almost the same (9, 19). Alignment of the primary amino acid sequences of the two molecules indicates the considerable divergence in the C-terminal regions (9) and the deletion analysis differently localized the transcriptional activity within the regions (31). Such structural diversity of the C-terminal region might lead to species-specific interaction behaviors with transcriptional cofactors. TCDD-activated hAHR may not recruit coactivators as efficiently as the DBA/2 counterpart. One possibility must be noted that the incompatibility between TCDD-activated hAHR and the mouse coactivators may cause the reduced response of hAHR knock-in mouse to TCDD.

hAHR was not detectable by immunoblot analysis with the current antiserum, and its abundance relative to the constitutive level of mouse AHR protein could not be determined. Considering this lack of quantitative information, limited protein accumulation might account for the attenuated responsiveness of hAHR knock-in mice to TCDD. hAHR may have an intrinsically shorter life than mouse AHR at physiological expression levels *in vivo*.

The susceptibility of embryonic kidneys of homozygous h*AHR* mice to the teratogenic effects of TCDD is noteworthy. The pathogenesis of this renal lesion induced by TCDD involves hyperplasia of the ureteric epithelium, resulting in an occlusion of the ureter and subsequent hydronephrosis (32). Adverse effects on the kidney and urinary tract have also been reported in humans exposed to TCDD (33). However, studies in Ben Tre Province in Vietnam, where defoliant containing dioxin was sprayed extensively, revealed little increase in the prevalence of cleft lip and/or palate compared with that observed in Japan (34), suggesting that hAHR is less potent to mediate the manifestation of cleft palate, and that a higher dose might be required for it. Consistent with these human reports, our analysis showed that hAHR, although expressed in mice, mediated the development of hydronephrosis induced by TCDD, but not cleft palate at our experimental dose. Thus, the knock-in animal seems to mimic some aspects of the human responses to PAH/HAHs.

An intriguing utilization of our knock-in mouse strategy would be as an *in vivo* system for the qualitative and quantitative assessment of possible human responses to various PAH/HAHs. In this study, D2N-*Ahrd* mice responded more strongly to TCDD than to 3-MC, whereas the hAHR knock-in mice responded almost equally to these two compounds. These results clearly show that the relative efficacy profiles, examined by TCDD and 3-MC, are different between D2N-*Ahrd* and our hAHR knock-in mouse. Therefore the efficacy profile specific to hAHR can be displayed by analyses of the responses of the hAHR knock-in mouse to an array of PAH/HAHs. Because environmentally relevant levels of exposure to dioxin and related compounds have garnered much concern in terms of their possible effects on reproductive, neurobehavioral, and immunological functions of humans, our hAHR knock-in mouse will serve as a humanized model mouse, exhibiting the human-specific responses to PAH/HAH congeners. This mouse should help define the range of biological and toxicological effects that could be expected to affect humans and thereby reduce some uncertainty in risk assessments of these persistent environmental contaminants.

We thank Drs. H. Ueda and S. Kimura at the Environmental Health Department of the Ministry of the Environment for valuable suggestions and support to initiate this project, and Dr. T. Kensler for the discussion and critical reading of the manuscript. We also thank Dr. N. Morito, Ms. N. Kaneko, and R. Kawai for help. This work was supported by grants from Exploratory Research for Advanced Technology Environmental Response Project (M.Y.), the Ministry of Education, Science, Sports and Culture (H.M. and M.Y.), the Ministry of Health, Labor, and Welfare (H.M., C.T., and M.Y.), Japan Society for the Promotion of Science–Research for the Future Program (M.Y.), Core Research for Evolutional Science and Technology (Y.F.-K., S.O., Y.A., C.T., and M.Y.), Program for Promotion Basic Research Activities for Innovative Biosciences (H.M. and S.T.), and Special Coordination Fund for Promoting Science and Technology (H.M.).

1. Poland, A. & Knutson, C. (1982) *Annu. Rev. Pharmacol. Toxicol.* **22**, 517–554.
2. Whitelock, J. P., Jr. (1990) *Annu. Rev. Pharmacol. Toxicol.* **30**, 251–277.
3. Landers, J. P. & Bunce, N. J. (1991) *Biochem. J.* **276**, 273–287.
4. Swanson, H. J. & Bradfield, C. A. (1993) *Pharmacogenetics* **3**, 213–230.
5. Hoffman, E. C., Reyes, H., Chu, F.-F., Sander, F., Conly, L. H., Brooks, B. A. & Hankinsin, O. (1991) *Science* **252**, 954–958.
6. Okey, A. B., Riddick, D. & Harper, P. A. (1994) *Toxicol. Lett.* **70**, 1–22.
7. Rowlands, J. & Gustafsson, J. A. (1997) *CRC Crit. Rev. Toxicol.* **27**, 109–134.
8. Poellinger, L. (2000) *Food Addit. Contam.* **17**, 261–266.
9. Ema, M., Ohe, N., Suzuki, M., Mimura, J., Sogawa, K., Ikawa, S. & Fujii-Kuriyama, Y. (1994) *J. Biol. Chem.* **269**, 27337–27343.
10. Poland, A. & Glover, E. (1990) *Mol. Pharmacol.* **38**, 306–312.
11. Poland, A., Palen, D. & Glover, E. (1994) *Mol. Pharmacol.* **46**, 915–921.
12. Ema, M., Matsushita, N., Sogawa, K., Ariyama, T., Inazawa, J., Nemoto, T., Ota, M., Oshimura, M. & Fujii-Kuriyama, Y. (1994) *J. Biochem.* **116**, 845–851.
13. Dolwick, K. M., Schmidt, J. V., Carver, L. A., Swanson, H. I. & Bradfield, C. A. (1993) *Mol. Pharmacol.* **44**, 911–917.
14. Okey, A. B., Vella, L. M. & Harper, P. A. (1989) *Mol. Pharmacol.* **35**, 823–830.
15. Pohjanvirta, R., Unkila, M. & Tuomisto, J. (1993) *Pharmacol. Toxicol.* **73**, 52–56.
16. Pohjanvirta, R., Wong, J. M. Y., Li, W., Harper, P. A., Tuomisto, J. & Okey, A. B. (1998) *Mol. Pharmacol.* **54**, 86–93.
17. Tuomisto, J. T., Viluksela, M., Pohjanvirta, R. & Tuomisto, J. (1999) *Toxicol. Appl. Pharmacol.* **155**, 71–81.
18. Mocarelli, P., Gerthoux, P. M., Ferrari, E., Patterson, D. G., Jr., Kieszak, S. M., Brambilla, P., Vincoli, N., Signorini, S., Tramacre, P., Carreri, V., *et al.* (2000) *Lancet* **355**, 1858–1863.
19. Micka, J., Milatovich, A., Menon, A., Grabowski, G. A., Puga, A. & Nebert, D. W. (1997) *Pharmacogenetics* **7**, 95–101.
20. Kohkalainen, M., Tuomisto, J. & Pohjanvirta, R. (2001) *Biochem. Biophys. Res. Commun.* **285**, 1121–1129.
21. Abnet, C. C., Tanguay, R. L., Heideman, W. & Peterson, R. E. (1999) *Toxicol. Appl. Pharmacol.* **159**, 41–51.
22. Mimura, J., Yamashita, K., Nakamura, K., Morita, M., Takagi, T. N., Nakao, K., Ema, M., Sogawa, K., Yasuda, M., Katsuki, M., *et al.* (1997) *Genes Cell.* **2**, 645–654.
23. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. (1987) *Nature* **326**, 292–295.
24. Ema, M., Sogawa, K., Watanabe, N., Chujoh, Y., Matsushita, N., Gotoh, O., Funae, Y. & Fujii-Kuriyama, Y. (1992) *Biochem. Biophys. Res. Commun.* **184**, 246–253.
25. Fernandez-Salguero, P., Pineau, T., Hilbert, D. M., McPhail, T., Lee, S. S. T., Kimura, S., Nebert, D. W., Rudikoff, S., Ward, J. M. & Gonzalez, F. J. (1995) *Science* **268**, 722–726.
26. Moriguchi, T., Sakurai, T., Takahashi, S., Goto, K. & Yamamoto, M. (2002) *J. Biol. Chem.* **277**, 16985–16992.
27. Yasuda, M., Igarashi, E., Datu, A. R. & Igawa, H. (1986) *Teratology* **34**, 454–455.
28. Birnbaum, L. S., Harris, M. W., Barnhart, E. R & Morrissey, R. E. (1987) *Toxicol. Appl. Pharmacol.* **90**, 206–216.
29. Birnbaum, L. S. & Abbott, B. (1997) in *Methods in Developmental Toxicology and Biology*, eds. Klug, S. & Thiel, R. (Blackwell Science, London), pp. 51–63.
30. Tsai, F. Y., Browne, C. P. B. & Orkin, S. H. (1998) *Dev. Biol.* **196**, 218–227.
31. Kumar, M. B., Ramadoss, P., Reen, R. K., Vanden Heuvel, J. P. & Perdew, G. H. (2001) *J. Biol. Chem.* **276**, 42302–42310.
32. Hoffman, R. E., Stehr-Green, P. A., Webb, K. B., Evans, R. G., Knutson, A. P., Schramm, W. F., Staak, J. L., Gibson, B. B. & Steinberg, K. K. (1986) *J. Am. Med. Assoc.* **255**, 2031–2038.
33. Abbott, B. D., Birnbaum, L. S. & Pratt, R. M. (1987) *Teratology* **35**, 329–334.
34. Natsume, N., Kawai, T. & Le, H. (1998) *Cleft Palate Craniofac. J.* **35**, 183–185.

# Computer-Aided Modeling of Pentachlorophenol 4-Monooxygenase and Site-Directed Mutagenesis of Its Active Site

Takashi NAKAMURA,[a] Takayuki MOTOYAMA,[a] Takatsugu HIROKAWA,[b] Shuichi HIRONO,[c] and Isamu YAMAGUCHI*,[a]

[a] *Laboratory for Remediation Research, Environmental Plant Research Group, Plant Science Center, RIKEN Institute; 1–7–22 Suehiro, Tsurumi-ku, Yokohama, Kanagawa 230–0045, Japan:* [b] *Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST); 2–41–6 Aomi, Koutou-ku, Tokyo 135–0064, Japan: and* [c] *Laboratory of Physical Chemistry for Drug Design, School of Pharmaceutical Sciences, Kitasato University; 5–9–1 Shirokane, Minato-ku, Tokyo 108–8641, Japan.*

Received August 5, 2003; accepted September 10, 2003; published online September 11, 2003

**Homology modeling was used to construct a model of the three-dimensional structure of pentachlorophenol 4-monooxygenase (PcpB). A PSI-BLAST homology search was initially performed to identify the 3D structure of proteins homologous with PcpB. The feasibility of modeled structures of PcpB was evaluated by Verify3D, which calculated structural compatibility scores based on 3D–1D profiles. The predicted structure of PcpB had an acceptable 3D–1D self-compatibility score, beyond the incorrect fold score threshold. A PcpB–pentachlorophenol (PCP) complex was then constructed utilizing the modeled PcpB structure. After energy minimization of the complex, and successive minimizations of the system that consisted of the complex and the water layer surrounding the complex, the molecular dynamics of the system were simulated. The active-site residues of PcpB were identified on the basis of the modeled structure, and PcpB mutants were then designed to change the active site residues, expressed, and purified by affinity chromatography. The mutant activity was compared with that of the wild-type to investigate the validity of the modeled structure. The experimental results suggested that Phe85, Tyr216, and Arg235 were relevant to enzyme activity, and that Tyr397 and Phe87 were important for stabilization of the structure of PcpB.**

**Key words** pentachlorophenol 4-monooxygenase (PcpB); homology modeling; molecular dynamics; site-directed mutagenesis; bioremediation

Pentachlorophenol (PCP) is a polychlorinated aromatic compound that has been used as a pesticide and herbicide for many years, and its widespread use has caused contamination of soil and groundwater.[1] In addition, paper pulp mill bleach effluent contains large amounts of chlorinated phenols, including PCP.[2] PCP is listed as a Priority Pollutant by the Environmental Protection Agency because of its toxicity and widespread distribution in the environment.[3] Bioremediation of PCP contamination has therefore become an important focus of research.

Pentachlorophenol 4-monooxygenase (PcpB) has been purified from the PCP-degrading bacterium *Sphingobium chlorophenolicum* (formerly *Sphingomonas chlorophenolica*) strain ATCC9723.[4—6] This enzyme is a flavin monooxygenase that converts PCP to tetrachlorobenzoquinone (TCBQ) in the presence of oxygen and a reduced form of nicotinamide adenine dinucleotide phosphate (NADPH).[7] The nucleotide sequence of the *pcpB* gene has already been determined and its recombinant protein has been produced,[8,9] but its three-dimensional (3D) structure has not yet been solved. In fact, little information is currently available regarding the structure and catalytic mechanism of action of this enzyme.

Computational chemistry and molecular modeling offer effective tools to study the mechanisms of biodegradation reactions at the molecular level,[10—12] and can provide guidelines for mutagenesis experiments. We have therefore used computer modeling for the construction of 3D-structures and the analysis of active-site residues in PcpB. PcpB was selected, first, because it appears to catalyze the rate-limiting step in the biodegradation of PCP.[3,13] Thus, with a PcpB mutant it may be possible to degrade PCP more efficiently than with the wild-type. Second, PcpB is strikingly non-specific, unlike most flavin monooxygenases. It can turn over a variety of substituted phenols and can replace hydrogen, nitro, amino, and cyano groups, as well as halogens, with a hydroxyl group.[14] Consequently, it seems that PcpB can be altered to break down synthetic organic pollutants that natural microorganisms currently are unable to degrade.

In the present study, the 3D structure of PcpB was derived from homology modeling, molecular mechanics (MM) calculations, and molecular dynamics (MD) calculations, and the active-site residues of PcpB were analyzed. Protein expression and the activities of purified PcpB mutants, designed on the basis of the modeled structure, were compared with those of wild-type PcpB to investigate the model's validity.

## Experimental

**Homology Modeling of PcpB** The amino acid sequence of PcpB was used as the search entry in the NR sequence database, using the position-specific iterated basic local alignment search tool (PSI-BLAST),[15] iterating five times. The 3D structures of phenol hydroxylase (PH) chain D (Protein Data Bank (PDB)[16] ID: 1FOH, PSI-BLAST E-value $1\times10^{-59}$)[17] and *p*-hydroxybenzoate hydroxylase (PHBH) (PDB ID: 1PBE, PSI-BLAST E-value $2\times10^{-54}$)[18] were extracted from the PDB and used as templates. The crystal structure of 'closed' PH (chain D) was selected from among four crystal structures, because the closed conformation is favorable for hydroxylation. The putative PcpB domains were determined by multiple sequence alignment using ClustalW[19] and visual inspection. The alignment is shown in Fig. 1. The structure of variable regions between domains was generated by finding suitable peptide segments in other proteins from the PDB. The conformations of complicated and 3D amino acid groups were revised using the rotational isomer library, and this 3D structure was finally optimized by the molecular mechanics method with the constant valence force field (CVFF). The homology modeling and energy optimizations of PcpB were carried out

* To whom correspondence should be addressed. e-mail: nakamura.takashi@nies.go.jp

using Insight II Homology and Discover 3 modules (v2000, Accelrys Inc., San Diego, CA, U.S.A.) on a SGI workstation (Silicon Graphics Inc., Mountain View, CA, U.S.A.). The final PcpB homology model was evaluated by Verify3D Structure Evaluation Server.[20]

**Construction of the PcpB–PCP Complex** The modeled structure of PcpB was superimposed on the 3D structure of PH (PDB ID: 1FOH), containing FAD and phenol (substrate). PCP was substituted for the substrate, and then docked manually into the binding site without significant conflict. The CVFF parameters of Discover 3 were used for the MM and MD simulations of the PcpB–PCP complex. Due to the lack of hydrogen atoms in the PDB structure, the initial positions of the hydrogen atoms were generated by the Modify/Hydrogens command in the Biopolymer module of the Insight II package. The energy-minimized complex was placed in a sphere consisting of 21992 TIP3P water molecules.[21] The radius of the sphere was 59 Å and the 7-Å outer layer was restrained (tethering, force constant 500) because of the prevention of water vaporization during MM and MD simulations. The 10 amino acid residues from the N-terminal were deleted, and the N-terminal was capped with an acetyl group so that the aqueous sphere could sufficiently hold the complex. Finally, the entire system was energy minimized until the value of final convergence was below 0.4. In the potential energy minimization, the steepest descent (SD) method was used for the early cycles, until the value of final convergence was below 8.0, and then the conjugate gradient (CG) method was used. The MD simulation of the energy-minimized system was run on a SGI Octane-workstation using Discover 3 modules in Insight II. This simulation was carried out at a constant temperature of 298 K and time step of 1 fs. The 600-ps MD calculation was performed. Every 10-ps structure, from 550 to 600 ps, was extracted from the MD trajectory file and the structures were then energy minimized under the same conditions as described above, with the exception of CG final convergence, which was 0.8. The resultant structures of the PcpB–PCP complex were inspected and the putative active site residues of PcpB were identified. Secondary structure elements were calculated using the Kabsch–Sander method.[22]

**Bacterial Strains, Plasmids, and Culture Conditions** *S. chlorophenolicum* ATCC39723 was purchased from Sumitomo Pharmaceuticals International Co. (SPI, Tokyo, Japan). The vector pET-19b (Novagen, Madison, WI, U.S.A.), which carries an N-terminal histidine tag sequence, followed by an enterokinase site and three cloning sites, was used for the cloning and expression of recombinant PcpB. The vector pKF18K (TaKaRa Shuzo, Kyoto, Japan) was used for construction of PcpB mutants. *S. chlorophenolicum* ATCC39723 was cultured in mineral salt medium at 25 °C[3] with shaking. *Escherichia coli* strain DH5$\alpha$ (Toyobo, Osaka, Japan) was cultured in Luria-Bertani (LB) medium[23] and *E. coli* strain BL21 (DE3) pLysS (Novagen) was cultured in Terrific Broth (TB)[23] at 37 °C with vigorous shaking. Media were solidified with 1.5% Bacto agar (Difco Laboratories, Detroit, MI, U.S.A.) and, where appropriate, supplemented with kanamycin or carbenicillin at final concentrations of 200 and 50 $\mu$g/ml, respectively. Restriction enzymes were purchased from TaKaRa Shuzo. Primers for site-directed mutagenesis were purchased from Sigma Genosys Japan (Hokkaido, Japan).

**Site-Directed Mutagenesis** The pET-19b vector, described above, was digested with *Bam*HI and *Xba*I. The obtained *pcpB* fragment was ligated into the pKF18K vector. Site-directed mutagenesis of PcpB was performed with the pKF18K vector and a Mutan-Express Km kit (TaKaRa Shuzo). The provided protocol was followed with one exception; the kanamycin concentration was changed to 200 $\mu$g/ml. All of the nucleotide sequences of mutants were confirmed by the dideoxy-chain termination method with an automated DNA sequencer (ABI PRISM 3700 genetic analyzer, Applied Biosystems, Foster City, CA, U.S.A.). The primers for site-directed mutagenesis were phosphorylated by T4 polynucleotide kinase (TakaRa Shuzo) and ATP prior to site-directed mutagenesis and had the following sequences (the underlined codons include the italicized substitutions that encoded the mutation): Phe85Ala (5′-GTCAGGAGCAACGGG*GCG*ACGTTCAACTTCGAG-3′), Phe87Ala (5′-GCAACGGGTTCACG*GCG*AACTTCGAG-3′), Tyr216Ala (5′-GACGACTGGATTCAC*GCG*TTCATCGGTCAGGAC-3′), Tyr216Phe (5′-CGACTGGATTCAC*TTC*TTCATCGGTCAGG-3′), Phe223Ala (5′-GGACAAA*GCG*GTCTTCGTGACGAAGC-3′), Phe225Ala (5′-CAGGACAAATTCGTC*GCG*GTGAC-3′), Arg235Ala (5′-GCCGGGTTCCAATTAT*GCG*GTGATTATCAGCGACC-3′), Arg235Glu (5′-CGGGTTCCAATTAT*GAA*GTGATTATCAGCGAC-3′).

Mutants Tyr397Ala, Tyr397Phe, and Arg235Lys were constructed by the Long and Accurate PCR method with the same procedure as mentioned above for the Mutan-Express Km kit. These mutants were also phosphorylated and used as PCR primers with the selection primer of the kit. The sequences were: Arg235Lys (5′-GCCAAGGTCGCTGATAATCAC*CTT*ATAATTGGAAC-3′), Tyr397Ala (5′-GCTGACGAGCTGGTCGCG*CGC*ATATAGG-3′), Tyr397Phe (5′-GCTGACGAGCTGGTCGCG*GAA*ATTATAGG-3′).

PCR was performed, using these primers, for 20 cycles of denaturation at 95 °C for 30 s, annealing at 50 °C for 30 s, and extension at 68 °C for 2 min.

**Enzyme Expression and Purification** The modified pET-19b vectors, as constructed above, were transformed into *E. coli* BL21 (DE3) pLysS strain. The cells were allowed to grow in TB medium in the presence of carbenicillin (200 $\mu$g/ml) at 37 °C until the optical density ($OD_{600}$) of the medium reached 0.6. The expression of the desired protein was induced by adding isopropyl $\beta$-D-thiogalactopyranoside and carbenicillin to final concentrations of 100 $\mu$M and 500 $\mu$M, respectively. The cells were incubated at 20 °C overnight, and all purification steps were performed at 4 °C. All buffers contained 20 mM $\beta$-mercaptoethanol and 15% (v/v) glycerol. The cells were collected by centrifugation, suspended in 50 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris)–HCl (pH 8.0) and 0.09% (v/v) Triton X-100, and broken by sonication (TOMY Ultrasonic disruptor UD-201, TOMY Digital Biology, Tokyo, Japan). The suspension was centrifuged and the soluble fraction of the enzyme was mixed with the same amount of bind buffer containing 800 mM NaCl and 1.6 mM imidazole. The solution was loaded into an Econo Column (1.5×15 cm) (BioRad, Hercules, CA, U.S.A.) packed with Ni-NTA agarose (Qiagen K. K., Tokyo, Japan) equilibrated in buffer containing 50 mM Tris–HCl (pH 8.0) and 400 mM NaCl. The solution was treated with two kinds of wash buffer. Buffer 1 contained 50 mM Tris–HCl (pH 8.0), 50 mM NaCl, 0.8 mM imidazole, 5 mM $MgCl_2$, and 0.1 mM EDTA–Na (pH 8.0). Buffer 2 was the same as buffer 1, except that the concentration of imidazole was 40 mM instead of 0.8 mM. PcpB activity was then eluted with the elution buffer, which was the same as the wash buffers except that the imidazole concentration was 200 mM. The eluate was concentrated by ultrafiltration (Amicon Centriprep YM-30, Millipore, Billerica, MA, U.S.A.), loaded onto NAP10 columns (Amersham Biosciences, Upsalla, Sweden), and equilibrated in 50 mM Tris–HCl (pH 8.0) and 15% (v/v) glycerol. The imidazole was removed from the buffer using the NAP 10 column. The eluate was again concentrated by ultrafiltration (Amicon Centricon YM-30, Millipore) and stored at −80 °C. In the case of Y397 mutants, the enzyme was purified and concentrated in the presence of 25% glycerol, to stabilize the enzyme.

Enzyme purity and the molecular weights of the subunits were determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), with 5—20% gradient gels (PAGEL NPG520L, ATTO, Tokyo, Japan) and molecular weight markers (Daiichi III, Daiichi Pure Chemicals, Tokyo, Japan). The proteins were stained with Bio-Safe CBB G-250 Stain (Bio-Rad). The protein concentration was determined by the method of Bradford,[24] with bovine serum albumin (BSA) as the standard.

**Comparison of the Activity of PcpB and PcpB Mutants** The activity of PcpB and its mutants was assayed by measuring the production of tetrachlorohydroquinone (TCHQ), which is probably produced by NADPH-reduction of the reaction product (TCBQ) of PCP and PcpB.[7] A 200-$\mu$l enzyme solution was made from PcpB (124.2 $\mu$g, 10 $\mu$M), 25% glycerol (v/v), and 10 mg/ml BSA (Fraction V, crystalline, Calbiochem-Novabiochem Co., La Jolla, CA, U.S.A.), because of enzyme stability. The assay system contained 80 $\mu$l of 100 mM potassium phosphate (KPi) buffer (pH 7.0), 5 $\mu$l of 2 mM NADPH, 1 $\mu$l of 2 mM PCP, and 5 $\mu$l of enzyme solution in a volume of 100 $\mu$l. The PCP degradation reaction by PcpB and its mutants was carried out for 90 s at a constant temperature of 25 °C. The extent of reaction by wild-type PcpB was approximately 5% under these conditions. The reaction mixtures were acidified with 20 $\mu$l of 2 mol/l HCl, and extracted 3 times with 100 $\mu$l of ethyl acetate after the reaction. The ethyl acetate was taken up into glass inserts in screw-capped vials for GC-MS. Aliquots of ethyl acetate solution containing 200 ng of pentachlorothioanisole (PCTA) used as an external standard were also taken up into the glass inserts. Following evaporation of the ethyl acetate under a stream of nitrogen gas, the residue was dissolved in 100 $\mu$l of a 2:1 solution of pyridine and acetic anhydride, for acetylation of PCP and TCHQ, and allowed to stand at room temperature for several hours. The production of acetylated TCHQ was analyzed by GC-MS. Activities of PcpB mutants were calculated relative to that of wild-type. Experiments were repeated 4 times.

**Analytical Methods** GC-MS analysis was performed with a model HP 6890 series GC system and 5973 Mass Selective Detector system (Agilent Technologies, Palo Alto, CA, U.S.A.) fitted with a fused silica chemically bonded capillary column (DB-5W; inside diameter, 0.25 mm; length, 30 m; film thickness, 0.25 $\mu$m: J&W Scientific, Inc., Folsom, CA, U.S.A.). Each 1-$\mu$l sample was injected into the column at 130 °C in splitless mode, using an HP 7863 series injector and auto sampler. After 1 min at 130 °C, the column

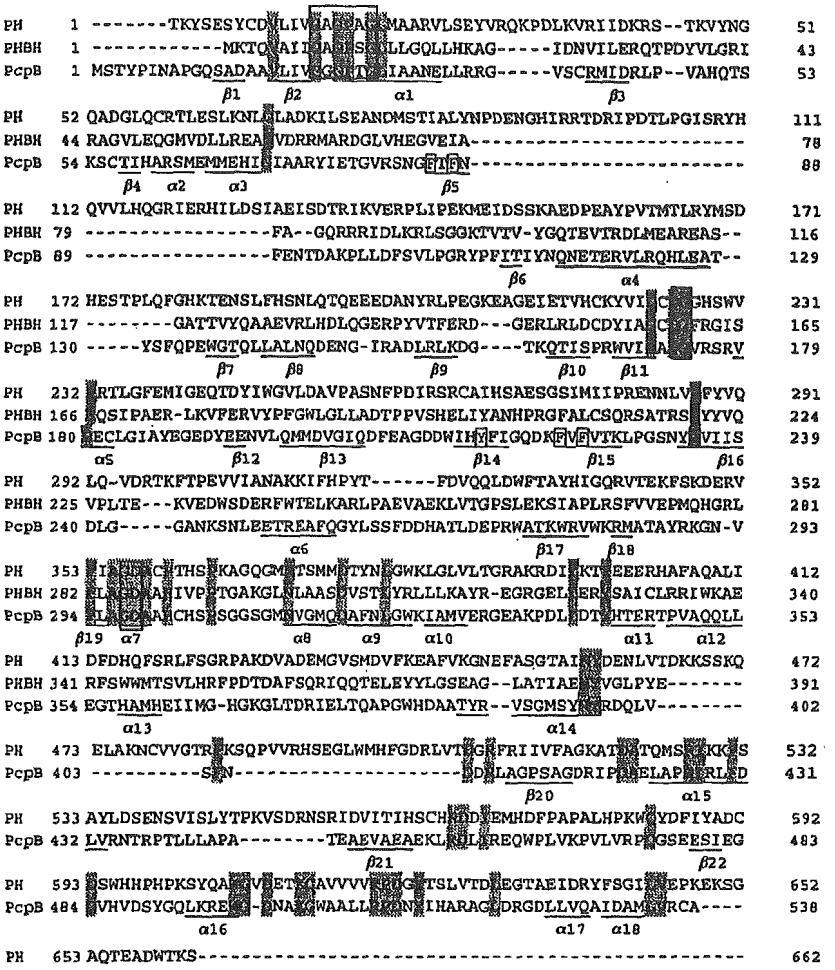


Fig. 1. The Protein Sequence Alignment of PcpB and Two Types of Hydroxylase Used in Homology Modeling

Sequence alignment was performed using ClustalW and further modified manually. The sequences used for the alignment were phenol hydroxylase (PH) (PDB ID: 1FOH) and *p*-hydroxybenzoate hydroxylase (PHBH) (PDB ID: 1PBE). Conserved residues are shaded. The residues mutated in the experimental section are in squares. In PDB data of 1FOH, residues from Ala294 to Arg301 have been deleted, as indicated by '~'. The three consensus sequences (GXGXXG, DG, GD) in the flavoprotein hydroxylases are boxed. The positions of the secondary elements in PcpB are indicated by underlined sequences. The secondary structure elements of PcpB are numbered according to the predictions of the Kabsch–Sander method.

temperature was increased by 10 °C/min to 250 °C, and held at 250 °C for 4 min. The amounts of PCTA were calculated based on the peak area, determined by the selected ion monitoring method. The amounts of acetylated PCP and TCHQ were calculated based on the calibration curves of these compounds relative to PCTA. The selected *m/z* of these compounds were as follows: PCTA 246, 296; PCP 266, 308; and TCHQ 248, 290, 332.

**Chemicals** PCP, PCTA, and ethyl acetate were purchased from Wako Pure Chemical Ind., Ltd. (Osaka, Japan). These chemicals were of residual pesticide analysis grade, and the purities of PCP and PCTA were more than 99%. All other chemicals were of analytical grade and of the highest purity available.

## Results and Discussions

**Construction of 3D Structure of PcpB by Homology Modeling** The PSI-BLAST homology search[15] was initially performed to identify the 3D structures of proteins homologous to PcpB, because no template proteins for PcpB were found on the BLAST search. PH and PHBH were selected as templates, with significant E-values, for construction of the PcpB homology model. PcpB showed 21.0% and 16.5% sequence identity to PH and PHBH, respectively. PSI-BLAST methods have been well benchmarked, with low sequence similarity (less than 30%),[25–27] and both enzymes are well-studied flavoproteins that catalyze hydroxylation of the *ortho* position to the hydroxyl group of a substrate (phenol or *p*-hydroxybenzoate) in the presence of oxygen and NADPH. The same substances, FAD, NADPH, and oxygen, were also used in PcpB. Moreover, the three consensus sequences in the hydroxylases were also conserved in PcpB (Fig. 1).[28] Thus, it is likely that PcpB and these two proteins have similar enzymatic functions. The construction of a structural model for PcpB was therefore attempted with the Homology module of Insight II, as described in the Experimental section. The feasibility of the modeled structure of PcpB was evaluated by Verify3D, which calculated the structural compatibility score, based on the 3D–1D profile.[20] The predicted structure of PcpB had an acceptable 3D–1D self-compatibility score (159.05), beyond the incorrect fold score threshold (111.01).

**PcpB–PCP Complex and the Prediction of Active Site Residues** The PcpB–PCP complex was constructed utilizing the above PcpB structure, as described in the Experimental section. After energy minimization of the complex, and successive minimization of the system that consisted of the

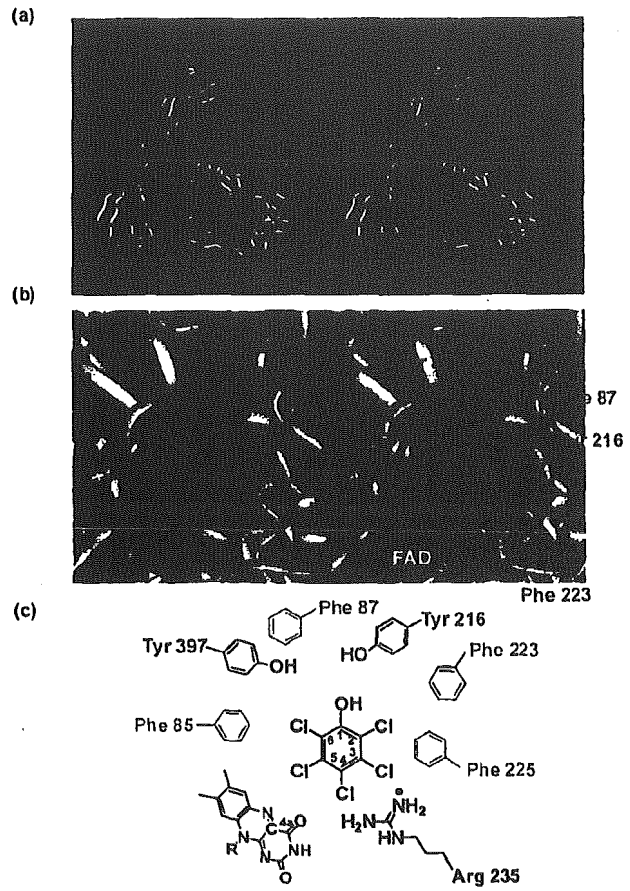



Fig. 2. The 3D Structure of PcpB, Constructed by Molecular Modeling

(a) Stereo view of the entire structure of PcpB. The color of the ribbon corresponds to the secondary structure elements calculated by the Kabsch–Sander method. The regions of red, yellow, and light blue refer to $\alpha$-helix, $\beta$-sheet, and random coil, respectively. The N-terminal and C-terminal are shown in capital letters. (b) Stereo view of the active site of PcpB. Important residues, predicted from inspection, are shown and distinguished with different colors: Tyrosine, phenylalanine, and arginine are shown in pink, light green, and blue, respectively. PCP and FAD are shown in orange and yellow CPK models, respectively. The red atom in FAD shows the $C_{4a}$ atom to which molecular oxygen is likely added. (c) A schematic of the active site of PcpB shown in (b). The blue atoms in PCP are important for PcpB recognition.

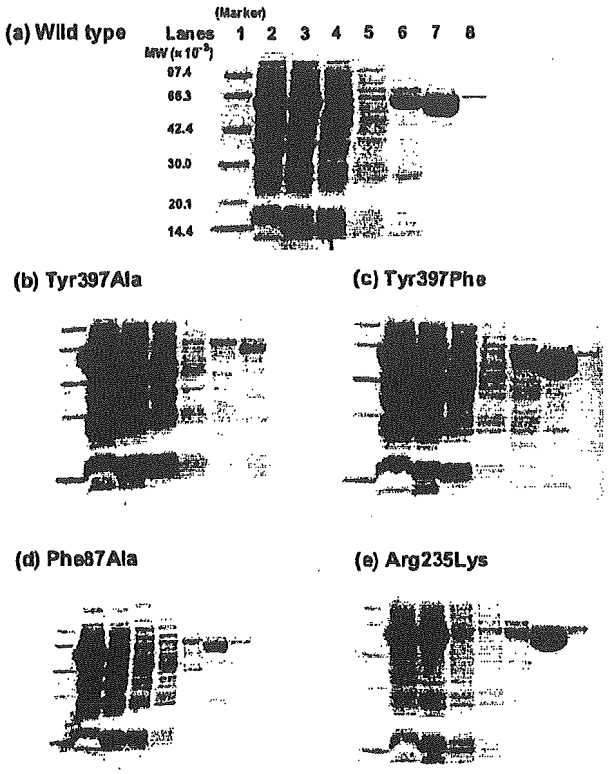



Fig. 3. Purification Steps of Recombinant PcpB (a) and Its Mutants (b—e), Traced by SDS-PAGE

The type of mutant is shown above each figure. Lanes: 1, molecular size markers; 2, cell suspension after sonication; 3, supernatant after centrifugation of the cell suspension; 4, the fraction after the supernatant was passed through the column; 5, the fraction of wash1 buffer; 6, the fraction of wash 2 buffer; 7, the fraction of elution1 buffer; 8, the fraction of elution 2 buffer.

complex and the water layer surrounding the complex, MD simulation of the system was performed. The resultant structure of the PcpB–PCP complex is shown in Fig. 2a.

From inspection of the active site residues in PcpB, 7 residues seemed to be in direct contact with either the substrate molecule or with each other, as shown in Fig. 2b and 2c. Tyr216 and Tyr397 were selected because these tyrosines had the potential to make hydrogen bonds between the enzyme and the hydroxyl group of the substrate in both PH (Tyr289)[17] and PHBH (Tyr201).[18] The residues Tyr216 and Tyr397 were conserved in the template enzyme (Fig. 1). Moreover, the residues Tyr201 and Tyr385 in PHBH, which corresponded to Tyr216 and Tyr397 in PcpB, respectively, played important roles in recognition and substrate specificity of PHBH.[29] Arg235 was also conserved, and its positively charged guanidinium group seemed to stabilize the chloride ion released from PCP during the enzymatic reaction. To our knowledge, such a role of arginine in other enzymes has not been reported, but the chloride–arginine interaction has been reported in some proteins such as halorhodopsin[30] and deoxyhemoglobin Rothschild.[31] Phe225 possibly made a $\pi$–$\pi$ interaction with PCP. The phenyl ring edge of Phe87 seemed to interact with the $\pi$ face of Tyr397. Finally, Phe85 and Phe223 seemed to be involved in recognition of the *ortho* chloride of PCP, and was therefore selected, because PcpB could convert the substrate for both dichlorinated and trichlorinated phenols as long as one of the *ortho* positions was chlorinated.[14]

**Expression and Purification of the PcpB Mutants** Recombinant PcpBs with mutations in the seven residues were expressed in *E. coli* and purified by affinity chromatography. The process of purification of PcpB and some of its mutants is shown in Fig. 3. The purified protein was manifested as a major band on SDS-PAGE, with an apparent molecular mass of approximately 60000, as reported previously.[5] PcpB was readily purified because the enzyme moved to the soluble fraction (Fig. 3a, lane 3 and lane 7). This result is consistent with those of a previous report, which indicated that PcpB was translocated from the cytoplasm, *via* the inner membrane, to the periplasm of *S. chlorophenolicum*, and that the majority of PcpB was soluble at all times.[32] Most PcpB mutants, such as Tyr397Phe and Arg235Lys, showed similar behavior (Figs. 3c, e). Only small amounts of Tyr397Ala and Phe87Ala mutants moved to the soluble fraction, however, presumably due to the structural change of the protein (Figs. 3b, d). The results of SDS-PAGE for other mutants are shown in the supporting information (Fig. 5).

**Comparison of the Activities of PcpB and PcpB Mu-**

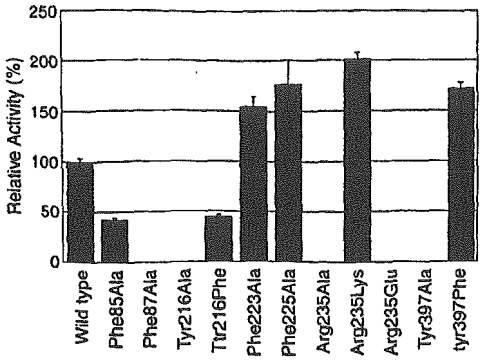



Fig. 4. Relative Activities of the PcpB Mutant in Comparison with That of Wild-Type PcpB

The amount of TCHQ made from PCP with wild-type PcpB (Wild type) was used as a standard. The mutant type is shown below the histogram. Data are shown as means±standard deviation, with 4 replicates per treatment.

**tants** The amounts of TCHQ produced by PcpB and NADPH *via* TCBQ were subsequently analyzed by GC-MS, and the activities of PcpB mutants were calculated relative to that of the wild type. The results are shown in Fig. 4. Replacement of either Tyr216 or Tyr397 with alanine greatly decreased the activity. Replacement of either Tyr216 or Tyr397 with phenylalanine, however, decreased activity by 60% in Tyr216Phe and increased activity by 60% in Tyr397Phe. Thus, Tyr216 appears to be more important for the mechanism of action of PCP than Tyr397.

Replacement of Arg235 with alanine and glutamic acid markedly decreased the activity, but replacement by lysine increased activity. The positive charge of the guanidinium group in arginine could be important for the reaction of PcpB. The proposed role of Arg235 for the stabilization of the released chloride ion is interesting and specific to PcpB, because such a function was not obvious in either template enzyme, even though the sequence was conserved (Fig. 1). Thus, the function of Arg235 might come from the evolution of a metabolic pathway of the template enzymes. Replacement of Phe87 with alanine greatly decreased enzyme activity. As shown in Figs. 2b and d, the benzene rings of Phe87 and Tyr397 likely play an important role in stabilizing the active site of PcpB. Interestingly, establishing the 3D-structure revealed that Phe87 is important in stabilizing the active site of PcpB (Figs. 2c, 3d). Moreover, replacement of Phe85 with alanine decreased activity by 60%, suggesting the importance of the Phe85 benzene ring.

The reaction mechanism of PcpB is not obvious, but the reaction likely proceeds *via* electrophilic attack of an intermediate hydroperoxide form of the flavin cofactor ($C_{4a}$-hydroperoxyflavin) on the $C_4$ substrate as in PHBH and PH, which have been well studied.[33—35] Thus, the distance between the $C_{4a}$ atom of FAD and the substrate $C_4$ atom in the modeled structure of PcpB was compared to the distance between the $C_{4a}$ atom of FAD and the substrate $C_2$ atom in the crystal structures of PH and PHBH. The result was indicative of the propriety of the modeled structure, because the distance in PcpB, 4.73 Å, was between those of PHBH, 4.37 Å, and PH, 5.37 Å. Moreover, the flavin cofactors, present in the model structure of PcpB or in the crystal structure of PH and PHBH, were modified to the active $C_{4a}$-hydroperoxyflavin intermediate using the coordinates reported previously[36] on the

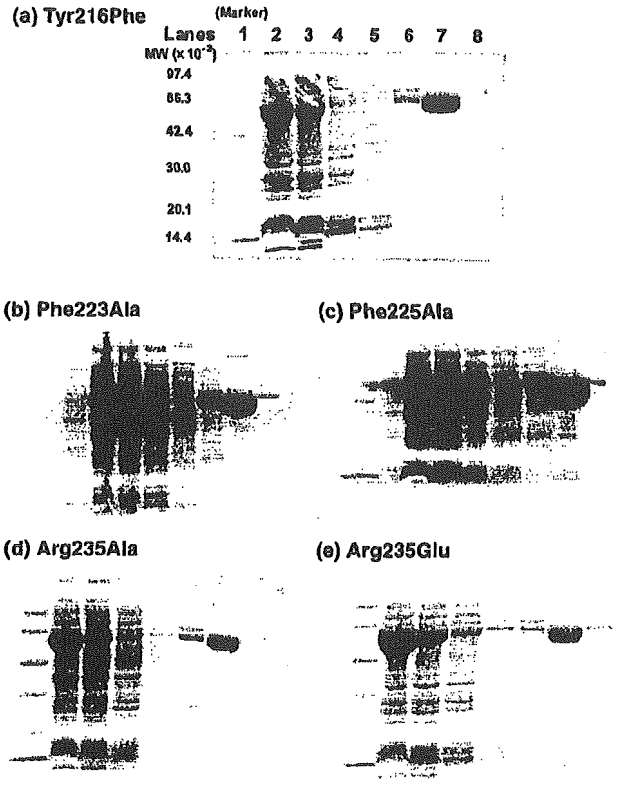



Fig. 5. Purification Steps of PcpB Mutants Traced by SDS-PAGE

(a) Tyr216Phe, (b) Phe223Ala, (c) Phe225Ala, (d) Arg235Ala, and (e) Arg235Glu. Conditions are as same as that of Fig. 3.

basis of molecular modeling studies. Schreuder *et al.*[37] suggested in the previous report that the almost ideal position of the distal oxygen, at an approximate angle of 60° with the plane of the aromatic ring of the substrate, is certainly important in enhancing the reaction rate. Thus, the distance between the distal oxygen from FAD ($O_d$) and the substrate carbon atom ($C_r$), which is attacked to the $O_d$ atom, and the angle formed by the line $O_d$–$C_r$ with the plane of the aromatic ring were examined. The distance $O_d$–$C_4$ and angle formed by the line $O_d$–$C_4$ with the plane of the aromatic ring were 3.5 Å and 81°, respectively in the case of PcpB. This result was comparable with those of PH, 3.2 Å and 74°.

The model did not explain all the results of site-directed mutagenesis; for example, the increased activities of Phe223Ala and Phe225Ala mutants relative to the wild-type were not explained. The 3D-structure of PcpB optimized by further MD simulation or partial modification might be able to account for the experimental result. The measurement of kinetic parameters, such as $K_m$ and $k_{cat}$, and the analysis of structural change induced by site-directed mutagenesis are still required. This study provided the first analysis of the 3D structure of PcpB. PcpB is an attractive enzyme for bioremediation, because PCP should be a model compound of recalcitrant polychlorinated aromatic compounds, such as PCBs and dioxin. Thus, it is very important that PcpB be designed for improved degradation and substrate specificity.

**Acknowledgments** We are grateful to the Research Resources Center at the Riken Brain Science Institute for DNA sequencing. We thank Drs. Noriaki Okimoto from RIKEN in Japan and Jiri Damborsky from the National Center for Biomolecular Research in the Czech Republic for helpful discus-

sion. This work was supported by The Ministry of Education, Culture, Sports, Science, and Technology (MEXT) through a Grant-in-Aid for Young Scientists (B) (grant no. 14760067).

**References**


1) McAllister K. A., Lee H., Trevors J. T., *Biodegradation*, **7**, 1—40 (1996).
2) Thakur I. S., Verma P. K., Upadhaya K. C., *Biochem. Biophys. Res. Commun.*, **286**, 109—113 (2001).
3) McCarthy D. L., Claude A. A., Copley S. D., *Appl. Environ. Microbiol.*, **63**, 1883—1888 (1997).
4) Saber D. L., Crawford R. L., *Appl. Environ. Microbiol.*, **50**, 1512—1518 (1985).
5) Xun L., Orser C. S., *J. Bacteriol.*, **173**, 4447—4453 (1991).
6) Takeuchi M., Hamana K., Hiraishi A., *Int. J. Syst. Evol. Microbiol.*, **51**, 1405—1417 (2001).
7) Dai M. H., Rogers J. B., Warner J. R., Copley, S. D., *J. Bacteriol.*, **185**, 302—310 (2003).
8) Orser C. S., Lange C. C., Xun L., Zahrt T. C., Schneider B. J., *J. Bacteriol.*, **175**, 411—416 (1993).
9) Wang H., Tiirola M. A., Puhakka, J. A., Kulomaa, M. S., *Biochem. Biophys. Res. Commun.*, **289**, 161—166 (2001).
10) Bohac M., Nagata Y., Prokop Z., Prokop M., Monincova M., Tsuda M., Koča J., Damborsky J., *Biochemistry*, **41**, 14272—14280 (2002).
11) Ridder L., Mulholland A. J., Rietjens I. M. C. M., Vervoort J., *J. Am. Chem. Soc.*, **122**, 8728—8738 (2000).
12) Ridder L., Palfey B. A., Vervoort J., Rietjens I. M. C. M., *FEBS Lett.*, **478**, 197—201 (2000).
13) Copley S. D., *TIBS*, **25**, 261—265 (2000).
14) Xun L., Topp E., Orser C. S., *J. Bacteriol.*, **174**, 2898—2902 (1992).
15) Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J., *Nucleic Acids Res.*, **25**, 3389—3402 (1997).
16) Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E., *Nucleic Acids Res.*, **28**, 235—242 (2000).
17) Enroth C., Neujahr H., Schneider G., Lindqvist Y., *Structure*, **6**, 605—617 (1998).
18) Schreuder H. A., Prick P. A. J., Wierenga R. K., Vriend G., Wilson K. S., Hol W. G. J., Drenth J., *J. Mol. Biol.*, **208**, 679—696 (1989).
19) Thompson J. D., Higgins D. G., Gibson T. J., *Nucleic Acids Res.*, **22**, 4673—4680 (1994).
20) Luthy R., Bowie J. U., Eisenberg D., *Nature* (London), **356**, 83—85 (1992).
21) Jorgensen W. L., Chandrasekhar J., Madura J. D., *J. Chem. Phys.*, **79**, 926—935 (1983).
22) Kabsch W., Sander C., *Biopolymers*, **22**, 2577—2637 (1983).
23) Sambrook J., Fritsch E. F., Maniatis T., "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, New York, 1989.
24) Bradford M. M., *Anal. Biochem.*, **72**, 248—254 (1976).
25) Park J., Karplus K., Barret C., Hughey R., Haussler D., Hubbard T., Chothia C., *J. Mol. Biol.*, **284**, 1201—1210 (1998).
26) Salamov A. A., Suwa M., Orengo C. A., Swindells M. B., *Protein Sci.*, **8**, 771—777 (1999).
27) Friedberg I., Kaplan T., Margalit H., *Protein Sci.*, **9**, 2278—2284 (2000).
28) Eppink M. H. M., Schreuder H. A., van Berkel W. J. H., *Protein. Sci.*, **6**, 2454—2458 (1997).
29) van der Bolt F. J. T., van den Heuvel R. H. H., Vervoort J., van Berkel W. J. H., *Biochemistry*, **36**, 14192—14201 (1997).
30) Braiman M. S., Walter T. J., Briercheck D. M., *Biochemistry*, **33**, 1629—1635 (1994).
31) Kavanaugh J. S., Rogers P. H., Case D. A., Arnone A., *Biochemistry*, **31**, 4111—4121 (1992).
32) Wang H., Marjomaki V., Ovod V., Kulomaa M. S., *Biochem. Biophys. Res. Commun.*, **299**, 703—709 (2002).
33) Husain M., Entsch B., Ballou D. P., Massey V., Chapman P. J., *J. Biol. Chem.*, **255**, 4189—4197 (1980).
34) Maeda-Yorita K., Massey V., *J. Biol. Chem.*, **268**, 4134—4144 (1993).
35) Entsch B., Berkel W. J. H., *FASEB J.*, **9**, 476—483 (1995).
36) Ridder L., Mullholland A. J., Rietjens I. M. C. M., Vervoort J., *J. Mol. Graph. Mod.*, **17**, 163—175 (1999).
37) Schreuder H. A., Hol W. G. J., Drenth J., *Biochemistry*, **29**, 3101—3108 (1990).

ELSEVIER

Journal of Molecular Structure (Theochem) 634 (2003) 159–170

THEO CHEM

www.elsevier.com/locate/theochem

# Geometry-dependent atomic charge calculations using charge equilibration method with empirical two-center Coulombic terms

Akifumi Oda[a,b,*], Shuichi Hirono[b]

[a]*Discovery Laboratories, Toyama Chemical Co. Ltd, 2-4-1 Shimookui, Toyama 930-8508, Japan*
[b]*School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan*

Received 28 February 2003; accepted 30 April 2003

**Abstract**

The calculation of atomic charges plays an important role in the non-bonded interaction terms of energy functions for molecular mechanics and molecular dynamics methods. Since the electronic states vary with changing molecular conformations, different atomic charges are required for different conformers. Although the charge equilibration (QEq) method was proposed by Rappé and Goddard as a technique for easily calculating atomic charges that depend on the molecular conformation, this method utilizes the two-center Coulombic integrals of Slater type orbitals, which have unrealistic values for small interatomic distances. In this study, we instead calculated the atomic charges for 87 compounds using QEq methods in combination with five types of empirical equations to calculate the two-center Coulombic terms, and investigated which equation gives the most appropriate charges. As a result, it was found that the DasGupta–Huzinaga equation and the Ohno–Klopman equation are effective for the calculations of QEq charges and that the Ohno–Klopman equation with QEq/PD parameters derived by Bakowies and Thiel are best for calculating dipole moments. It was also discovered that both the quantitative and qualitative derivation of appropriate charges for compounds that contain formal charges or aromatic five-membered hetero-rings is not always possible using QEq methods.


*Keywords:* Atomic charge; Charge equilibrium method; Two-center Coulombic terms; Non-bonded interaction

## 1. Introduction

For energy calculations in molecular mechanics (MM) and molecular dynamics (MD) methods, non-bonded interaction terms play important roles. In particular, accurate estimations of these terms are indispensable for calculating the binding free energy of protein–ligand complexes by using MD trajectories, which is one of the most important methods in computer-aided drug design. In most MM and MD methods, van der Waals and electrostatic interaction terms are included in the non-bonded terms, the electrostatic interaction term requiring the calculation of every interatomic Coulombic interaction. Therefore, it is essential to break down the molecular charge as a whole into components for each atom in order to reproduce the molecular electrostatic properties.

* Corresponding author. Address: Discovery Laboratories, Toyama Chemical Co. Ltd, 2-4-1 Shimookui, Toyama 930-8508, Japan.
*E-mail address:* odaa@pharm.kitasato-u.ac.jp (A. Oda).

0166-1280/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0166-1280(03)00338-5

For the calculation of atomic charges, methods using the electrostatic potentials obtained by quantum chemical calculations [1] and the partial equalization of orbital electronegativity, which depends on the molecular connectivity, [2,3] have been proposed. However, because quantum chemical methods are very costly in terms of computational resources, it is unfeasible to recalculate atomic charges for every change in molecular conformation, and only fixed charges can be used for representing the relaxation of charge distribution that is dependent on changing molecular structure. On the other hand, although the partial equalization of orbital electronegativity methods are very fast, the electronegativities of only covalently bonded atom pairs are used; the atomic charges do not vary with conformational changes when the Gasteiger–Marsili method [2] is used, and even if the method developed by Cho et al. [3] is used they vary only with changes in bond lengths. Because variations of the torsional angle are the most important conformational changes in molecular modeling problems such as protein folding and molecular recognition, the above methods yield inadequate atomic charges for these purposes. Thus, new calculational methods for atomic charges are required that depend on molecular geometry, i.e. bond lengths, bond angles and torsional angles; these must be undemanding enough, in terms of computation, that the charges at every step of the MM and MD calculations can be determined.

In 1991, Rappé and Goddard proposed the charge equilibration (QEq) method, which allows the charges to vary with changes in the chemical environment. In this method, atomic charges are calculated by using the electronegativities, self-Coulomb integrals and two-center Coulombic terms, and the dependence on interatomic distance of the two-center Coulombic terms allows the geometry-dependent atomic charges for conformers to be determined. Since the atomic charges can be obtained simply by solving sets of simultaneous equations, these calculations can be carried out with lower computational costs than quantum chemical methods. On the other hand, it has been reported [4,5] that the two-center Coulombic terms derived by the Rappé and Goddard method, which are two-center Coulomb integrals of Slater-type orbitals, have unrealistic values for atomic pairs with short interatomic distances.

In order to overcome this problem in the QEq method, the introduction of empirical treatments of the two-center Coulombic terms that are used in semi-empirical molecular orbital calculations was proposed [5,6]. Furthermore, Bakowies and Thiel [5] parameterized electronegativities and self-Coulomb integrals that are used in the QEq method and obtained reasonable atomic charges. These treatments can yield realistic values of the terms with low computational cost. However, there has been no systematic investigation of the QEq approach using two-center Coulombic terms empirically calculated by different methods, and thus far it has remained uncertain which empirical equation is most adequate for atomic charge calculations.

In this study, the proposed QEq methods utilize empirically obtained two-center Coulombic integrals from not only the Ohno–Klopman [7,8] and Nishimoto–Mataga [9] approximations, which are already incorporated into the QEq methods [5,6], but also from the Ohno [7], Nishimoto–Mataga–Weiss [10] and DasGupta–Huzinaga [11] approximations. The calculated atomic charges and dipole moments are then compared with each other. Furthermore, atomic charges and dipole moments are calculated using the Ohno–Klopman approximation with new parameters reported by Bakowies and Thiel [5] and compared with those obtained using the Rappé and Goddard parameters.

## 2. Methods

### 2.1. QEq methods

For QEq methods, the total energy of a molecule X is described as the sum of one-center and two-center terms:

$$E(X) = \sum_{I}^{X} E_I + \sum_{I<}^{X} \sum_{J}^{X} E_{IJ} \tag{1}$$

The atomic electrostatic energy, i.e. the one-center term $E_I$ in Eq. (1), is assumed to be a differentiable function of the atomic charge $q_I$, and $E_I$ can written as

a Taylor series up to second order around $q_I = 0$:

$$E_I = E_I^0 + q_I\left(\frac{\partial E}{\partial q_I}\right)_0 + \frac{1}{2}q_I^2\left(\frac{\partial^2 E}{\partial q_I^2}\right)_0 \quad (2)$$

The first derivative of the energy is obtained from the atomic electronegativity $\chi_I^0$ defined by Mulliken [12]

$$\chi_I^0 = \frac{1}{2}(\mathrm{IP}_I + \mathrm{EA}_I) = \left(\frac{\partial E}{\partial q_I}\right)_0 \quad (3)$$

where $\mathrm{IP}_I$ and $\mathrm{EA}_I$ are the ionization potential and the electron affinity of atom $I$, respectively. The difference between $\mathrm{IP}_I$ and $\mathrm{EA}_I$ is given by the second derivative of the energy

$$J_{II} = \mathrm{IP}_I - \mathrm{EA}_I = \left(\frac{\partial^2 E}{\partial q_I^2}\right)_0 \quad (4)$$

where $J_{II}$ is the Coulombic repulsion between two electrons in one atom (the one-center Coulombic integral). By using $\chi_I^0$ and $J_{II}$, $E_I$ can be written as follows:

$$E_I = E_I^0 + \chi_I^0 q_I + \tfrac{1}{2}J_{II}q_I^2 \quad (5)$$

If the two-center term is assumed to consist of simple Coulombic interactions, $E_{IJ}$ is written as follows:

$$E_{IJ} = J_{IJ}q_I q_J \quad (6)$$

Therefore, the molecular energy is the sum of $E_I$ as described in Eq. (5) for all atoms and $E_{IJ}$ as described in Eq. (6) for all diatomic pairs:

$$E(X) = \sum_I^X \left(E_I^0 + \chi_I^0 q_I + \frac{1}{2}J_{II}q_I^2\right) + \sum_{I<}^X \sum_J^X J_{IJ}q_I q_J \quad (7)$$

In analogy with Eq. (3), taking the first derivative of $E(X)$ with respect to $q_I$ allows the electronegativity $\chi_I$ of an atom $I$ in a molecule to be calculated:

$$\chi_I = \frac{\partial E}{\partial q_I} = \chi_I^0 + J_{II}q_I + \sum_J^X J_{IJ}q_J \quad (8)$$

According to the principle of electronegativity equalization

$$\chi_1 = \chi_2 = \cdots = \chi_N \quad (9)$$

and the condition on the total charge

$$Q_{\mathrm{total}} = \sum_I^N Q_I \quad (10)$$

a total of $N$ simultaneous equations are obtained

$$\mathbf{CQ} = \mathbb{D} \quad (11)$$

where the vector $\mathbf{Q}$ contains all the atomic charges $(q_1, q_2, \ldots, q_N)$, the components of $\mathbf{C}$ are

$$C_{1J} = 1 \quad (12)$$

$$C_{IJ} = J_{1J} - J_{IJ} \qquad \text{for } I \geq 2$$

and the elements of $\mathbb{D}$ are

$$D_I = Q_{\mathrm{total}} \quad (13)$$

$$D_I = \chi_I^0 - \chi_1^0 \qquad \text{for } I \geq 2$$

For the calculation of the matrix elements, terms $\chi_I^0$, $J_{II}$, and $J_{IJ}$ are necessary. There are two types of $\chi_I^0$ and $J_{II}$, derived either by Rappé and Goddard [4] (hereafter abbreviated as RG parameters) or Bakowies and Thiel [5] (hereafter abbreviated as BT parameters); we use both parameter sets. The latter set can be used for the Ohno–Klopman equation, and only QEq/PD parameters are adopted in this study. Although the two-center terms $J_{IJ}$ in the original QEq method are obtained using integrals of Slater type orbitals, we calculate these terms by using empirical equations as described in Section 2.2.

### 2.2. Calculations of $J_{IJ}$

In calculations of the terms $J_{IJ}$, which are essential for evaluation of the components of matrix $\boldsymbol{C}$, integrals of Slater-type orbitals proposed by Rappé and Goddard lead to unrealistic results for small interatomic distances. Thus, in this study we instead use five types of empirical approximations in order to calculate $J_{IJ}$. In these approximations, which are listed below, $J_{IJ}$ is expressed as a function of an interatomic distance.

(i) Nishimoto–Mataga equation (NM)

$$J_{IJ} = \frac{1}{R_{IJ} + \dfrac{2}{J_{II} + J_{JJ}}} \quad (14)$$

(ii) Nishimoto–Mataga–Weiss equation (NMW)

$$J_{IJ} = \frac{f_x}{R_{IJ} + \frac{2f_x}{J_{II} + J_{JJ}}} \quad (15)$$

where $f_x$ is a constant with a value of 1.2 in this study.

(iii) Ohno equation (O)

$$J_{IJ} = \frac{1}{\sqrt{R_{IJ}^2 + \left(\frac{2}{J_{II} + J_{JJ}}\right)^2}} \quad (16)$$

(iv) Ohno–Klopman equation (OK)

$$J_{IJ} = \frac{1}{\sqrt{R_{IJ}^2 + \left(\frac{1}{2J_{II}} + \frac{1}{2J_{JJ}}\right)^2}} \quad (17)$$

(v) DasGupta–Huzinaga equation (DH)

$$J_{IJ} = \frac{1}{R_{IJ} + \frac{1}{\frac{J_{II}}{2}e^{k_I R_{IJ}} + \frac{J_{JJ}}{2}e^{k_J R_{IJ}}}} \quad (18)$$

where $k$ is the Klondike parameter ($0.4 \leq k \leq 0.8$). In this study, $k = 0.4$ is adopted for all atoms.

In the above equations, $J_{II}$ and $J_{JJ}$ are one-center Coulombic integrals, and $R_{IJ}$ is the distance between atoms $I$ and $J$. The aim of this study is to compare the atomic charges obtained from QEq methods with several empirically calculated $J_{IJ}$, and to elucidate which approximation and parameters are most appropriate for calculations of the atomic charges. The combinations of the parameter set and the empirical approximation for the different methods of calculating $J_{IJ}$ are shown in Table 1.

### 2.3. Compounds

The compounds used for testing the QEq methods are shown in Table 2. In this table, the reference molecules used by Bakowies and Thiel for their parameterizations [5] and compounds possessing formal charges and five-membered aromatic hetero-rings are indicated. Empirically determined dipole moments are also listed in the cases where they have been observed. We generated four sets of compounds from those in the table: all compounds (set **i**), compounds without formal charges and five-membered aromatic hetero-rings (set **ii**), compounds belonging to set **i** but excluding the reference molecules in Ref. [5] (set **iii**), and compounds belonging to set **ii** but excluding the reference molecules (set **iv**). The number of compounds in each set is listed in Table 3. The structures of these compounds were optimized using ab initio Hartree–Fock (HF) calculations, with 6-31G** basis sets. For comparison with results obtained using the QEq methods, atomic charges derived from the electrostatic potentials which were obtained by the ab initio HF/6-31G** method (hereafter abbreviated to HF-ESP charges) [1] were calculated. Original QEq charges were calculated using the ArgusLab 3.0 software [13]. Gasteiger–Hückel charges were also computed. Dipole moments were evaluated from the atomic charges calculated using different methods, and were then compared with the experimentally observed values in Refs. [14,15]. Atomic charge

Table 1
Computational methods for two-center Coulombic terms used in this paper

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Two-center Coulombic terms | NM | NMW | O | OK | DH | OK |
| Parameter | Rappé and Goddard | | | | | Bakowies and Thiel |

The combinations of the empirical equations and the parameters, which are self-Coulomb integrals and electron negativities, to calculate Coulombic two-center terms are shown. Abbreviation is as follows; NM: Nishimoto–Mataga, NMW: Nishimoto–Mataga–Weiss, O: Ohno, OK: Ohno–Klopman, DH: DasGupta–Huzinaga.

Table 2
Compounds used in this study

| | Compound | Dipole moment (D) | | Compound | Dipole moment (D) |
|---|---|---|---|---|---|
| 1 | 1,2-Butadiene | 0.401 | 45 | Hydrazine | 1.75 |
| 2 | 1,3,5-Trioxane | 2.08 | 46 | Hydrogenperoxide | 1.5728 |
| 3 | Acetaldehyde | 2.75 | *47* | *Indole* | 2.08 |
| 4 | Acetic acid | 1.7 | 48 | Isobutane | 0.132 |
| 5 | Acetone | 2.90 | *49* | *Isooxazole* | 2.95 |
| 6 | Acetonitrile | 3.925191 | 50 | Isoprene | 0.25 |
| 7 | Acetophenone | 3.00 | 51 | Isoquinoline | 2.53 |
| 8 | Acetylene | – | 52 | Ketene | 1.42215 |
| 9 | Acrylic aldehyde | 3.117 | 53 | Malononitrile | 3.735 |
| 10 | Acrylonitrile | 3.89 | 54 | Methane | – |
| 11 | Ammonia | 1.471772 | 55 | Methanol | 1.66 |
| 12 | Aniline | 1.018 | 56 | Methyl acetate | 1.67 |
| 13 | Anisole | 1.35 | 57 | Methyl isocyanate | 3.18 |
| 14 | Aziridine | 1.89 | **58** | **Methyl isocyanide** | 3.83 |
| 15 | Benzaldehyde | 2.76 | 59 | Methyl nitrite | 2.05 |
| 16 | Benzene | – | 60 | Methylamine | 1.27 |
| 17 | Benzonitrile | 4.14 | 61 | Methyleneamine | 1.97 |
| 18 | β-Propiolacton | 4.18 | **62** | **Nitro-benzene** | 4.21 |
| 19 | Camphor | 2.95 | **63** | **Nitroethylene** | 3.70 |
| 20 | Carbondioxide | – | 64 | Nitromethane | 3.46 |
| 21 | Carbonmonoxide | 0.10980 | **65** | **Nitrooxymethane** | 3.081 |
| 22 | *cis*-2-Butene | 0.253 | 66 | Oxetan | 1.93 |
| 23 | *cis*-2-Butenenitrile | 4.08 | 67 | Oxirane | 1.88 |
| 24 | Cyanoacetylene | 3.7225 | 68 | Phenol | 1.224 |
| 25 | Cyclohexane | 0.332 | 69 | Propane | 0.0841 |
| 26 | Cyclopentanone | 3.30 | 70 | Propionaldehyde | 2.46 |
| 27 | Cyclopropene | 0.454 | 71 | Propionitrile | 4.02 |
| **28** | **Diazomethane** | 1.5 | 72 | Propylene | 0.364 |
| 29 | Diethyl ether | 1.061 | 73 | Propylene oxide | 2.00 |
| 30 | Diketene | 3.51 | 74 | Propyne | 0.7839 |
| 31 | Dimethyl ether | 1.302 | 75 | Pyridazine | 4.221 |
| 32 | Dimethyl-diazene | – | 76 | Pyridine | 2.15 |
| 33 | Divinyl ether | 0.782 | 77 | Pyrimidine | 2.334 |
| 34 | Ethane | – | *78* | *Pyrrole* | 1.74 |
| 35 | Ethanol | 1.441 | 79 | Pyruvonitrile | 3.45 |
| 36 | Ethyl formate | 1.98 | 80 | Quinoline | 2.22 |
| 37 | Ethylene | – | 81 | *s-cis*-Formylcyclopropane | 2.74 |
| 38 | Formaldehyde | 2.3315 | 82 | *s-trans*-Formylcyclopropane | 3.26 |
| 39 | Formaldehyde oxime | 0.44 | 83 | *t*-Butylacetylene | 0.661 |
| 40 | Formamide | 3.711 | 84 | Tetrahydrofuran | 1.75 |
| 41 | Formicacid | 1.4214 | 85 | Toluene | 0.375 |
| 42 | Formicacidmethylester | 1.77 | 86 | Trimethylamine | 0.612 |
| 43 | Fulvene | 0.4236 | 87 | Water | 1.8546 |
| *44* | *Furan* | 0.661 | | | |

Compounds used for parameterizations by Bakowies and Thiel [5] are underlined. Compounds which have formal charges are described by bold and italic letters mean the compounds which include the five-membered aromatic hetero-ring.

calculations for two additional compounds tested by Bakowies and Thiel, α- and β-D-idose, were also carried out. These molecules are shown in Fig. 1, where the numbering conventions for both molecules are the same as in the previous study on MNDO/PD charges by Merz [16]. The molecular geometries of these compounds were also optimized using the ab initio HF/6-31G** method.

Table 3
The number of compounds

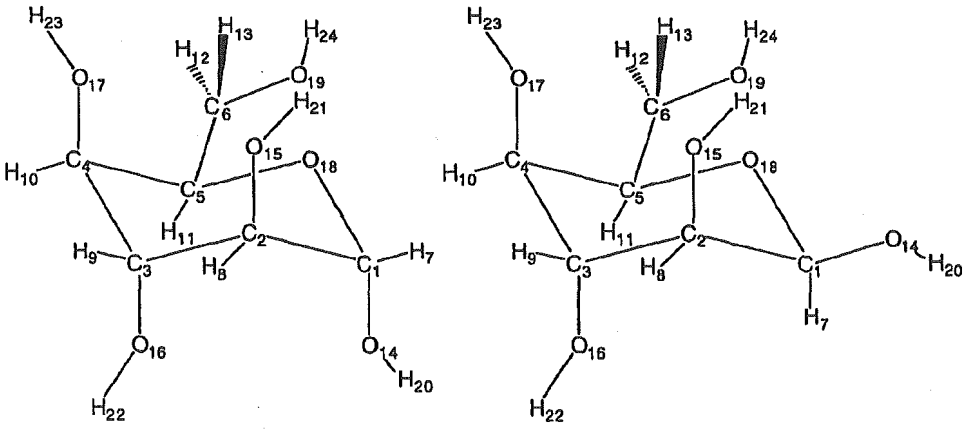
| | With Bakowies' compounds | | Without Bakowies' compounds | |
|---|---|---|---|---|
| | Set | Number | Set | Number |
| All | i | 87 | iii | 63 |
| Without formal charges and five-membered aromatic hetero-ring | ii | 76 | iv | 54 |

The number of compounds which belong to each set are shown.

The geometry optimizations and evaluations of HF-ESP charges by ab initio HF/6-31G** calculations were carried out using the GAMESS US program package [17]. The calculations of Gasteiger–Hückel charges and dipole moments were performed using the SYBYL 6.8 package [18].

## 3. Results and discussion

The atomic charges obtained using the QEq methods are shown and compared with the HF-ESP charges in Fig. 2. Root mean square deviations (RMSDs) and correlation coefficients quantifying the level of agreement between QEq and HF-ESP charges for compound sets i and iii are summarized in Table 4. Gasteiger–Hückel charges for carbon monoxide cannot be evaluated by the SYBYL 6.8 program package and the results in this table do not include them. It is apparent that atomic charges calculated using methods 3–6 are more consistent with the HF-ESP charges than methods 1 and 2. Although methods 3–6 give better results than the original QEq methods using ArgusLab 3.0, methods 1 and 2 give worse results; as shown in Fig. 2, they can reproduce the HF-ESP data qualitatively but there are large differences in the absolute values of the atomic charges. These results suggest that the two-center Coulombic terms given by the Nishimoto–Mataga and Nishimoto–Mataga–Weiss approximations are inappropriate for atomic charge calculations using the QEq method. The Gasteiger–Hückel method always gave the worst agreement with the HF-ESP method, indicating that QEq methods are more desirable for atomic charge calculations, not only because geometry-dependence is taken into account, but also because better numerical values are obtained. For compound set iii, from which the compounds in Ref. [5] are excluded, roughly similar results were obtained. Method 6, in which the Ohno–Klopman approximation in combination with BT parameters was used, gave slightly worse agreement with HF-ESP for set i than for set iii. On the other hand method 5, in which the DasGupta–Huzinaga approximation was used, yielded the best correlation coefficients for all compound sets and the best RMSDs for set iii. For sets i and iii, the statistical results from method 4 are better than those from method 6, and hence we conclude that parameterized values of $J_{II}$




(a) α-D-idose (b) β-D-idose

Fig. 1. α- and β-D-Idose with numbers.

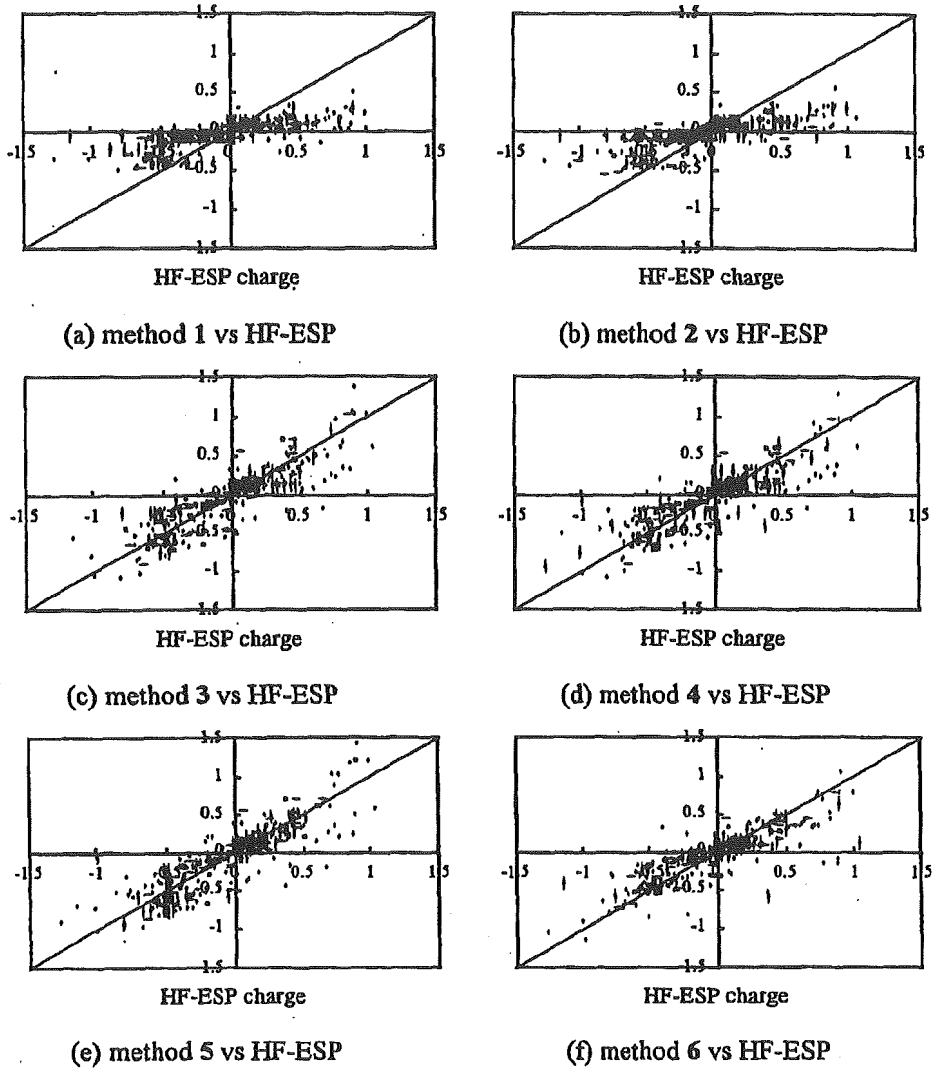



Fig. 2. Correlation between QEq and HF-ESP charges. The HF-ESP charges are taken in a horizontal axis and the QEq charges are in the vertical axis.

Table 4
RMSD and correlation between QEq and HF-ESP charges

| | 1 | 2 | 3 | 4 | 5 | 6 | QEq | G–H |
|---|---|---|---|---|---|---|---|---|
| Set i | | | | | | | | |
| RMSD | 0.239 | 0.231 | 0.167 | 0.162 | 0.165 | 0.169 | 0.195 | 0.251 |
| Correlation | 0.751 | 0.763 | 0.866 | 0.868 | 0.878 | 0.857 | 0.805 | 0.718 |
| Set iii | | | | | | | | |
| RMSD | 0.241 | 0.233 | 0.167 | 0.164 | 0.164 | 0.174 | 0.198 | 0.253 |
| Correlation | 0.719 | 0.732 | 0.853 | 0.855 | 0.868 | 0.838 | 0.784 | 0.671 |

The Gasteiger–Hückel charges are calculated for compounds except for carbon monoxide in set i. G–H means Gasteiger–Hückel methods.

Table 5
RMSD and correlation between QEq and HF-ESP charges for compounds which do not include formal charges and five-membered aromatic rings with hetero-atoms

| | 1 | 2 | 3 | 4 | 5 | 6 | QEq | G–H |
|---|---|---|---|---|---|---|---|---|
| Set ii | | | | | | | | |
| RMSD | 0.233 | 0.224 | 0.150 | 0.145 | 0.144 | 0.130 | 0.181 | 0.243 |
| Correlation | 0.780 | 0.791 | 0.893 | 0.897 | 0.909 | 0.921 | 0.835 | 0.757 |
| Set iv | | | | | | | | |
| RMSD | 0.234 | 0.226 | 0.151 | 0.147 | 0.143 | 0.136 | 0.186 | 0.243 |
| Correlation | 0.744 | 0.757 | 0.880 | 0.883 | 0.901 | 0.909 | 0.809 | 0.713 |

G–H means Gasteiger–Hückel methods.

and $\chi_I^0$ are not always adequate for atomic charge calculations using the QEq method.

The RMSDs and correlation coefficients for sets ii and iv are shown in Table 5. These results are better than those for sets i and iii for all of the methods, but especially for method 6. The RMSDs and correlation coefficients obtained using method 6 are the best not only for set ii, but also for set iv in which the compounds in Ref. [5] are excluded, contrary to the results of the analogous calculations for sets i and iii. The effectiveness of this method for sets ii and iv is greater than that for sets i and iii due to differences in the parameterization schemes. The RG parameters used in methods 1–5 were obtained by using experimental results for atoms. On the other hand, the BT parameters used in method 6 were determined using quantum chemical results for molecules that did not possess formal charges or five-membered aromatic hetero-rings. The above results indicate that method 6 (and to a lesser extent method 5) is appropriate for atomic charge calculations of compounds which do not possess formal charges and five-membered aromatic hetero-rings. The observation that QEq methods do not give reasonable atomic charges for formally charged molecules is consistent with previous studies [5]. In five-membered aromatic hetero-rings, the non-covalent electron pair on the hetero-atom is delocalized across the aromatic ring, and hence cannot be described as a lone pair in these compounds. The electronic states around hetero-atoms in such compounds are very different from those in atoms and, therefore, QEq charges are inadequate in comparison with the HF-ESP charges.

In Table 6, correlation coefficients among atomic charges derived by each of the QEq methods are shown. The correlation coefficient between atomic charges obtained using the Nishimoto–Mataga approximation (method 1) and its modification, the Nishimoto–Mataga–Weiss equation (method 2), is higher than 0.95. Correlation coefficients among charges obtained by the approaches of Ohno (method 3), Ohno–Klopman (method 4), DasGupta–Huzinaga (method 5) and Ohno–Klopman with BT parameters (method 6) are as high as that between methods 1 and 2. Because two-center Coulombic integrals calculated by the DasGupta–Huzinaga approximation with Klondike parameter $k = 0.4$ tend to resemble those evaluated by the Ohno equation [11], the six types of QEq charges can be classified in two groups that are determined not by the parameters $J_{II}$ and $\chi_I^0$ but by the similarities in approximations of $J_{IJ}$.

We calculated the dipole moments derived from the atomic charges, and Fig. 3 shows the comparison between calculated and experimental

Table 6
Correlation coefficient among six types of QEq charges

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | *1.000* | | | | | |
| 2 | *0.999* | *1.000* | | | | |
| 3 | 0.900 | 0.911 | *1.000* | | | |
| 4 | 0.915 | 0.926 | *0.998* | *1.000* | | |
| 5 | 0.871 | 0.883 | *0.993* | *0.990* | *1.000* | |
| 6 | 0.872 | 0.883 | *0.958* | *0.963* | *0.959* | *1.000* |

Correlation coefficients which are 0.95 or more are indicated in italics.