Published online August 24, 2004

4512–4523 *Nucleic Acids Research, 2004, Vol. 32, No. 15*
doi:10.1093/nar/gkh782

Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences

Y. V. Sun[1,3], D. R. Boverhof[1,3,4], L. D. Burgoon[1,2,4], M. R. Fielden[1,3,4] and T. R. Zacharewski[1,3,4,*]

[1]Department of Biochemistry and Molecular Biology, [2]Department of Pharmacology and Toxicology, [3]National Food Safety and Toxicology Center and [4]Center for Integrative Toxicology, Michigan State University, East Lansing, MI 48824, USA

Received May 18, 2004; Revised June 24, 2004; Accepted August 3, 2004

ABSTRACT

Comparative approaches were used to identify human, mouse and rat dioxin response elements (DREs) in genomic sequences unambiguously assigned to a nucleotide RefSeq accession number. A total of 13 bona fide DREs, all including the substitution intolerant core sequence (GCGTG) and adjacent variable sequences, were used to establish a position weight matrix and a matrix similarity (MS) score threshold to rank identified DREs. DREs with MS scores above the threshold were disproportionately distributed in close proximity to the transcription start site in all three species. Gene expression assays in hepatic mouse tissue confirmed the responsiveness of 192 genes possessing a putative DRE. Previously identified functional DREs in well-characterized AhR-regulated genes including *Cyp1a1* and *Cyp1b1* were corroborated. Putative DREs were identified in 48 out of 2437 human–mouse–rat orthologous genes between −1500 and the transcriptional start site, of which 19 of these genes possessed positionally conserved DREs as determined by multiple sequence alignment. Seven of these nineteen genes exhibited 2,3,7,8-tetrachlorodibenzo-*p*-dioxin-mediated regulation, although there were significant discrepancies between *in vivo* and *in vitro* results. Interestingly, of the mouse–rat orthologous genes with a DRE between −1500 and +1500, only 37% had an equivalent human ortholog. These results suggest that AhR-mediated gene expression may not be well conserved across species, which could have significant implications in human risk assessment.

INTRODUCTION

Novel computational approaches are being developed to annotate genome sequence data in order to predict gene structure, function and higher order biological control processes, such as gene regulation (1). Deciphering gene regulation at the transcriptional level through the computational identification of response elements is a valuable complement to empirical approaches seeking to develop biochemical networks (2–5). Several software tools have been developed to assist with the identification of potential regulatory elements and to predict their functionality using sequence, structure, context and comparative-based methods (1,6). Sequence-based searches are the most common means of identifying putative regulatory elements *in silico* due to the availability of software (1) and databases of transcription factor binding sites (7). This approach can complement experimental genome-wide analysis of gene expression and can be used to predict gene regulatory networks (8–10).

The availability of the human, mouse and rat genomes has provided unprecedented opportunities in understanding mammalian evolution, elucidating the etiology of human disease and facilitating drug development. Comparative analysis of genomic sequence data is a powerful tool for identifying functional non-coding sequences, such as gene regulatory elements, which tend to be conserved through evolution for common responses (11–14). Putative functional regulatory elements can then be identified by searching for conserved DNA sequence motifs in orthologous genes across multiple species, such as human, mouse and rat. However, the value of computational approaches has been questioned due to supposed high false-positive rates, and therefore must be verified using empirical approaches such as genome-wide gene expression and chromatin immunoprecipitation assays.

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) is a persistent environmental contaminant that elicits a broad spectrum of aryl hydrocarbon receptor (AhR)-mediated biochemical and toxic effects. These effects are well correlated with the ability of TCDD and related compounds to bind to the AhR (15) and are manifested by inappropriate AhR-mediated regulation of gene expression. This is supported by studies demonstrating that mice with low-affinity AhR alleles are less susceptible to the effects of TCDD (16), and AhR-null mice are resistant to the prototypical toxicities elicited by TCDD and related ligands (17,18). The AhR is a basic helix–loop–helix PAS (bHLH–PAS) protein that acts as a ligand-dependent DNA-binding transcription factor (19,20).

*To whom correspondence should be addressed. Tel: +1 517 355 1607; Fax: +1 517 353 9334; Email: tzachare@msu.edu
Present address:
M. R. Fielden, Iconix Pharmaceuticals, Mountain View, CA 94043, USA

Binding of ligand to the AhR triggers nuclear translocation and subsequent dissociation of the AhR from cytosolic components. Once in the nucleus, the AhR heterodimerizes with its bHLH–PAS binding partner, the AhR nuclear translocator (ARNT). This heterodimer binds to specific genomic sequences of responsive genes to modulate their gene expression. These specific binding sequences contain the substitution intolerant 5′-GCGTG-3′ core sequence and are referred to as the dioxin response elements (DREs). Ultra violet cross-linking (21) and site selection experiments (22) indicate that the AhR occupies the 5′-TNGC half-site, while ARNT contacts the GTG-3′ half-site. However, strong evidence indicates that the 5′- and 3′-flanking nucleotides play an important role in modulating DNA-binding affinity and enhancer function (23–25).

TCDD and related compounds, including polychlorinated biphenyls and polyaromatic hydrocarbons, alter the expression of various genes involved in metabolism and detoxification (26). However, dysregulation of classical drug-metabolizing enzymes alone fails to adequately explain the tissue-, sex- and species-specific toxicity of AhR ligands (27). Consequently, the genome-wide identification of additional targets is required in order to investigate the pathological and physiological role of the AhR, and to elucidate the mechanisms of toxicity of TCDD and related compounds.

We have taken a comparative computational scanning approach to identify putative DREs in the genomic sequences of human, mouse and rat target genes. The AhR regulon provides an ideal model as most, if not all, TCDD-elicited effects are mediated through the interaction between the AhR complex with the DRE core sequence. A total of 13 bona fide DREs were used to establish a position weight matrix (PWM) and a matrix similarity (MS) score threshold to prioritize computationally identified DREs. Comparative analysis and complementary *in vitro* and *in vivo* gene expression studies validate the approach and highlight challenges in verifying the genome-wide functionality of computationally identified putative response elements.

MATERIALS AND METHODS

Computational scanning for DREs

Unambiguous genomic sequence (−5000 to +2000 bp) for 17 882 human (hg15), 11 697 mouse (mm3) and 3896 rat (rn2) genes corresponding to RefSeq accession numbers was extracted from the UCSC Genome Browser (http://genome.ucsc.edu). Genomic sequences were scanned for exact matches to the DRE core sequence, GCGTG, on both positive and negative strands. For each match, the extended 19 bp sequence was used to calculate an MS score (28), and compared to an MS score threshold of 0.85 which was based on the lowest MS score of 13 bona fide DREs (Table 1). A Java application was developed to implement the search algorithm and to calculate an MS score (available upon request). DRE frequency and location were subsequently mapped for each gene in the region of −5000 to +2000 bp in 500 bp increments. In order to investigate the chance occurrence of the DRE core sequence, a set of 10 000 DNA sequences, with each sequence having a length of 5000 bp, was compiled using a Java application that randomly selected A, C, G or T to ensure independent and identical distributions for the nucleotides. This DNA sequence set was then analyzed as described above. The Wilcoxon's rank–sum test was used to compare the DRE distributions on a per species basis to the uniform, chance distribution observed with the set of 10 000 random DNA sequences. Comparison of the number of expected (due to chance) and observed DREs per species was performed using the chi-square test. Both statistical tests were performed using R v1.8.1 (http://www.r-project.org).

Table 1. Sequence of bona fide DREs

Gene[a]	Site	DRE sequence 5′–3′	MS score	Reference
m*Cyp1a1*	Site a	caagctcGCGTGagaagcg	0.94	(24)
	Site b	cctgtgtGCGTGccaagca	0.95	(24)
	Site d	cggagttGCGTGagaagag	0.98	(24)
	Site e	ccagctaGCGTGacagcac	0.91	(24)
	Site f	cgggtttGCGTGcgatgag	0.97	(24)
m*Cyp1b1*	XRE5	ccccctGCGTGcggagct	0.96	(38)
r*Cyp1a1*	XRE1	cggagttGCGTGagaagag	0.98	(55)
	XRE2	gatcctaGCGTGacagcac	0.88	(55)
r*Aldh3*		cactaatGCGTGccccatc	0.85	(62)
r*Nqr1*		tcccctGCGTGcaaaggc	0.95	(63)
r*Sod1*		gaggcctGCGTGcgcgcct	0.89	(64)
r*Gstya*		gcatgttGCGTGcatccct	0.89	(48)
rUgt1a1		agaatgtGCGTGacaaggt	0.92	(65)

[a]Abbreviations: m*Cyp1a1*, mouse cytochrome P4501A1; m*Cyp1b1*, mouse cytochrome P4501B1; r*Cyp1a1*, rat *Cyp1a1*; r*Aldh3*, rat aldehyde dehydrogenase-3; r*Nqr1*, rat NADPH:quinone oxidoreductase; r*Sod1*, rat Cu/Zn superoxide dismutase; *rGstYa*, rat glutathione *S*-transferase Ya; and rUgt1a1, rat UDP-glucuronosyltransferase 1A1.

Identification of DREs in orthologous genes

Orthologous genes were retrieved from NCBI HomoloGene database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=homologene, Build 31). The multiple sequence alignment tool, ClustalW (http://www.ebi.ac.uk/clustalw/), was used to identify consensus regions for all genes possessing DREs between −1500 and the transcriptional start site (TSS) with an MS score above the threshold. Positionally conserved DREs were then identified by searching DREs located within conserved regions. Using indices of putative DREs, we further mapped first level conserved DREs in *Cyp1a1* to find out whether they were located within consensus regions. For *Cyp1a1*, a poorly annotated TSS required upstream regulatory sequences associated with mRNA RefSeq NM_000499 (Human *CYP1A1*), NM_009992 (Mouse *Cyp1a1*) and NM_012540 (Rat *Cyp1a1*) to be downloaded from the UCSC Genome Browser and manually curated for analysis.

Animal and cell treatment

Female C57BL/6 mice, ovariectomized by the vendor on postnatal day 20 and all having body weights within 10% of the average body weight, were obtained from Charles River Laboratories on postnatal day 26 (Raleigh, NC). Immature ovariectomized C67BL/6 mice provided a well-controlled responsive model to examine temporal and dose-dependent changes in gene expression that is not confounded by circulating endogenous estrogens as a result of the onset of sexual maturation and the estrus cycle. Mice were housed

in polycarbonate cages containing cellulose fiber chips (Aspen Chip Laboratory Bedding, Northeastern Products, Warrensberg, NY) in a 23°C high-efficiency particulate air (HEPA)-filtered environment with 30–40% humidity and a 12 h light/dark cycle (07.00–19.00 h). Animals were allowed free access to deionized water and Harlan Teklad 22/5 Rodent Diet 8640 (Madison, WI) and acclimatized for four days prior to dosing. Animals were treated by gavage with 0.1 ml of sesame oil for a nominal dose of 0 (vehicle control) or 10 μg/kg body weight of TCDD (provided by S. Safe, Texas A&M University, College Station, TX) for the quantitative real-time PCR (QRTPCR) time course study. Four animals were treated per dose and time point, and groups for each dose and time point were housed in separate cages. Treated mice and their time-matched vehicle controls were sacrificed 12, 24 and 72 h after dosing by gavage for QRTPCR analysis of specific genes. Animals for the dose–response microarray analysis were sacrificed 24 h following treatment by gavage with 100 or 300 μg/kg body weight of TCDD or vehicle. Animals were sacrificed by cervical dislocation and a section of the liver was snap frozen in liquid nitrogen and was stored at −80°C until further use. All procedures were performed with the approval of the Michigan State University All-University Committee on Animal Use and Care.

Mouse Hepa1c1c7 cells (obtained from Dr O. Hankinson, UCLA) were seeded (1×10^6 cells) onto 150 mm culture dishes (Corning, Acton, MA) in triplicates in phenol red-free DMEM (Invitrogen, Carlsbad, CA) supplemented with 5% fetal bovine serum (Serologicals Inc., Norcross, GA), 100 U/ml of penicillin/streptomycin (Invitrogen), 2.5 μg/ml of amphotericin B (Invitrogen) and 100 μg/ml of gentamicin reagent solution (Invitrogen) at 37°C in a 5% CO_2 humidified environment. The cells were treated with dimethyl sulfoxide (vehicle control) or TCDD (10 nM) for 2, 4, 6, 8 and 12 h, and harvested by scraping in Trizol (Invitrogen).

RNA isolation

Liver samples (~70 mg) were transferred to 1.0 ml of Trizol in a 2.0 ml microfuge tube and homogenized using a Mixer Mill 300 tissue homogenizer (Retsch, Germany). Total RNA was isolated from homogenized tissue samples and cell culture samples according to the manufacturer's protocol with an additional phenol–chloroform extraction. Isolated RNA was resuspended in RNA storage solution (Ambion Inc., Austin, TX), quantified (A_{260}) and assessed for purity by determining the A_{260}/A_{280} ratio and by visual inspection of 1.0 μg on a denaturing gel.

cDNA microarray analysis

Detailed protocols for microarray construction, labeling of the cDNA probe, sample hybridization and slide washing can be found at http://dbzach.fst.msu.edu/interfaces/microarray.html. Briefly, PCR-amplified DNA was robotically arrayed in duplicate onto epoxy-coated glass slides (Quantifoil, Germany) using an Omnigrid arrayer (GeneMachines, San Carlos, CA) equipped with 16 (4 × 4) Chipmaker 2 pins (Telechem) at the Genomics Technology Support Facility (http://www.genomics.msu.edu) at Michigan State University. Total RNA from TCDD-treated liver samples was compared to total RNA from time-matched control (vehicle-treated samples) using the 'spoke' method which involves comparing treated samples to a common vehicle control with appropriate dye swaps as described previously (29). Four replicates were performed with each hepatic sample representing a single animal. Dye-swap hybridizations were conducted to account for dye biases for a total of eight hybridizations per treatment. The animal was considered as the experimental unit and there was no pooling of samples. Total RNA (25 μg) was reverse transcribed in the presence of Cy3- or Cy5-dUTP to create fluor-labeled cDNA which was purified using a Qiagen PCR purification kit (Qiagen, Valencia, CA). The Cy3 and Cy5 samples were mixed, vacuum dried and resuspended in 32 μl of hybridization buffer (40% formamide, 4× SSC and 1% SDS) with 20 μg poly(dA) and 20 μg of mouse COT-1 DNA (Invitrogen) as competitor. This probe mixture was heated at 95°C for 3 min and was then hybridized on an array under a 22×40 mm^2 coverslip (Corning Inc., Corning NY) in a light protected and humidified hybridization chamber (Corning Inc.). Samples were hybridized for 18–24 h at 42°C in a water bath. Slides were then washed, dried by centrifugation and scanned at 635 nm (Cy5) and 535 nm (Cy3) on an Affymetrix 428 Array Scanner (Santa Clara, CA). Images were analyzed for feature and background intensities using GenePix Pro 5.0 (Axon Instruments, Union City, CA).

cDNA microarray data were normalized using a semiparametric method that combines both parametric and non-parametric methods (J. E. Eckel, C. Gennings, T. M. Therneau, D. R. Boverhof, L. D. Burgoon and T. Zacharewski, submitted for publication). A model-based *t*-test was used to calculate *t*-scores using the General Linear Mixed Model: $y_{icad} = \mu_i + C_{ic} + A_{ia} + D_{id} + \gamma_{icad}$ where y_{icad} is the normalized feature intensity for the *i*-th cDNA, μ_i is the global mean for the *i*-th gene, C_{ic} is the fixed effect of dye for the *i*-th gene and the *c*-th dye (Cy3 or Cy5), A_{ia} is the random main effect for the *a*-th microarray (1, ..., *A*), D_{id} is the fixed effect for the *d*-th dose (0, ..., *D*) and γ_{icad} is the global error term with normal mean and SD. The model-based *t*-test was calculated per cDNA and per dose-group, comparing each dose individually to the expression in vehicle-treated samples for that cDNA.

An empirical Bayes method was used to identify active genes based on the model-based *t*-scores (30). The empirical Bayes was used to filter genes for activity based on the P1(*t*)-value, the posterior probability of activity, with a P1(*t*) cutoff of 0.99.

QRTPCR analysis

QRTPCR was performed on an Applied Biosystems Prism 7000 Sequence Detection System (Applied Biosystems, Foster City, CA) using the SYBR Green PCR Core Reagents (Applied Biosystems) as described previously (29,31). Total RNA (1 μg) was used as template for a reverse transcriptase reaction in 20 μl of 1× First-Strand Synthesis buffer (Life Technologies) containing 1 μg of oligo (dT_{18}A/C/GN), 0.2 mM of dNTPs, 10 mM of DTT and 200 U of Superscript II reverse transcriptase (Life Technologies). The reaction mixture was incubated at 42°C for 60 min and was stopped by incubation at 75°C for 15 min. Amplification of cDNA was performed in a single MicroAmp Optical 96-well Reaction

Plate (Applied Biosystems). Primer pairs for each gene were designed using Primer3 (32). Gene names, accession numbers, the forward and reverse primer sequences, amplicon size, and optimal primer and Mg^{2+} concentrations are listed in Table S5. Each plate contains standards of purified PCR products with known template concentration covering at least six orders of magnitude in order to interpolate relative template concentration of the experimental samples from standard curves of log copy number versus threshold cycle (C_t). No template controls (NTC) were also included on each plate. The relative level of mRNA was standardized to the housekeeping genes β-actin in order to control the differences in RNA loading, quality and cDNA synthesis.

The relative mRNA copy number was transformed using the natural logarithm for statistical analysis, and normality assured by normal probability–probability plots. Transformed copy numbers were analyzed on a gene-by-gene basis across time, between time-matched TCDD and vehicle treatments using a General Linear Model (GLM) in SAS v8.02. Contrasts between time-matched treated and vehicle samples were performed using a GLM model-based *t*-test.

RESULTS

Establishing a DRE MS score threshold

Prior to computationally scanning for putative DREs within genomic sequence data, an MS score threshold was established that could subsequently be used to assess the putative functionality of a motif when compared to bona fide DREs empirically determined to be functional. A 19 bp PWM, which included the intolerant 5 bp core DRE sequence (GCGTG) and the adjacent variant 7 bp flanking sequences, was developed using 13 reported functional DREs (Table 1). The PWM considers most of the provided sequence information to represent the nucleotide frequency at each position (Figure 1). In addition, it has the advantage of increased specificity in distinguishing relevant binding sequences due to the ability to assign an MS score to any oligonucleotide sequence of equal length to the matrix (28). DREs with MS scores above a selected threshold are expected to have a greater probability of possessing a measurable binding affinity and presumable biological relevance. MS scores for the 13 functional DREs are listed in Table 1 with the rat aldehyde dehydrogenase-3 DRE exhibiting the lowest MS score of 0.85, which was adopted as the threshold cutoff to initially assign tentative functionality to putative DREs that were identified by computational scanning.

Defining the genomic search region

Previous studies fail to define a region to restrict the computational search for putative DREs. We took advantage of the availability of genomic sequence data to examine the distribution of DREs in 17 882 human, 11 697 mouse and 3896 rat genes. Occurrence of the DRE core sequence was scanned within 5000 bp upstream and 2000 bp downstream of the TSS for all available genes from the human (hg15), mouse (mm3) and rat (rn2) genomes. In total, 12 420 human, 7835 mouse and 3043 rat DRE core sequences were identified with an MS score greater than the threshold (0.85), which corresponds to the presence of at least one putative DRE in 8290 human, 5238 mouse and 1837 rat genes (Figure 2). The distributions of putative functional DREs in the human, mouse and rat genomes were higher in regions proximal to the TSS (Figure 3A–C). Supplementary Table S1a–c provides MS scores for each computationally identified DRE as well as the associated gene name, LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink), RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/), location and sequence.

A Wilcoxon's rank-sum test confirmed ($P < 0.0001$) an even DRE core sequence distribution pattern (Figure 3D) within 10 000 random DNA sequences, each consisting of 5000 bp, and that the DRE core sequence distribution in human, mouse and rat genomes differed significantly ($P < 0.0001$) from the uniform, random distribution (Figure 4). Moreover, the occurrence of the DRE core sequences in areas distal to the TSS is significantly ($P < 0.0001$) less than what would occur due to chance, and overall, the number of expected DRE core sequences is significantly different from the number of observed DREs for all species ($P < 0.0001$).

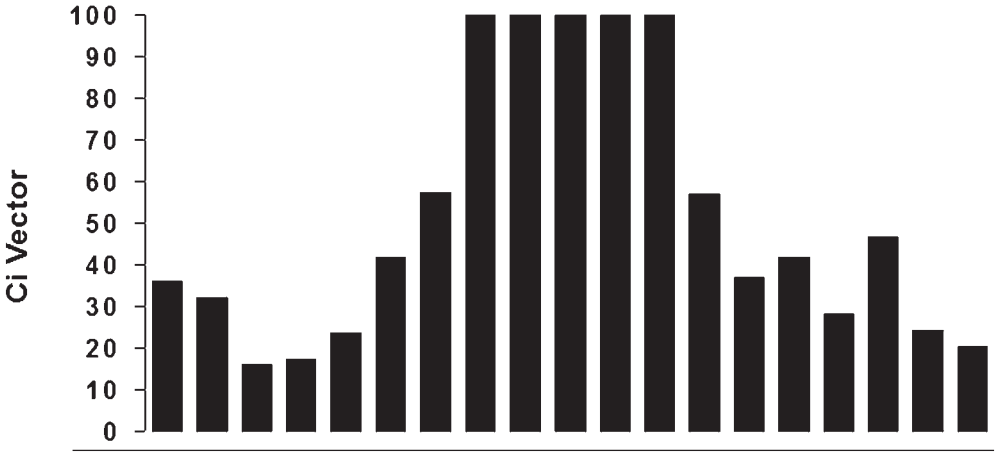
Further assessment of putative DREs

Within −1500 to +1500 bp, 5763 human, 3539 mouse and 1190 rat genes possessed at least one putative DRE. Genes annotated with the same TSS and first codon were removed, leaving 5489, 3182 and 1017 human, mouse and rat genes, respectively (Figure 2). Poor annotation of the TSS affected ~5, 10 and 15% of the human, mouse and rat genes, respectively. This has been attributed to inaccurate mapping of the TSS as a result of incomplete sequencing information for the 5′-untranslated region since expressed sequence tag sequencing is biased to the 3′ end (33).

The number of DREs within the −1500 to +1500 genomic region varied considerably between genes. We identified 1368 human, 736 mouse and 262 rat genes that contained multiple putative DREs within −1500 to +1500 bp of the TSS (Supplementary Table S2a–c). For example, the human vesicle-associated membrane protein (VAMP)-associated protein B and C gene (NM_004738) possessed 9 putative DREs, the mouse barttin gene (NM_080458) possessed 11 putative DREs and the rat insulin receptor substrate 3 gene (NM_032074) had 9 putative DREs. These genes are good candidates for AhR-mediated regulation since several of the DREs within the genomic sequence had MS scores above the threshold. Furthermore, tandem DREs are known to have a cooperative effect on gene expression (34,35), and therefore may indicate genes with a higher probability of responsiveness. Nonetheless, gene expression can be dependent on the species, sex, developmental stage, tissue, cell type and promoter context. Therefore, verification will be highly dependent on the selection of a suitable model maintained under the appropriate conditions. This may be a significant factor contributing to the high false-positive rate for identifying functional transcriptional response elements.

Microarray analysis of putative DRE containing genes

Mouse cDNA microarrays containing 13 362 features representing 8891 unique LocusLink IDs were used to identify AhR-regulated genes in C57BL/6 liver samples following treatment by gavage with 100 and 300 μg/kg body weight of TCDD



Position	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10
A	8	31	31	23	8	8	15	0	0	0	0	0	46	15	69	54	8	38	8
C	62	38	23	23	46	8	8	0	100	0	0	0	54	38	15	15	31	38	31
G	23	31	31	38	23	15	0	100	0	100	0	100	0	46	8	23	62	15	31
T	8	0	15	15	23	69	77	0	0	0	100	0	0	0	8	8	0	8	31
Ci	36	32	16	17	24	42	57	100	100	100	100	100	57	37	42	28	47	24	20

Figure 1. PWM for the DREs. The matrix was constructed from 13 experimentally verified DREs. Numbers in each column represent the percent frequency occurrence for the 4 nt at each position of the 19 bp PWM. The central positions from −2 to +3 are 100% conserved and represent the core DRE-binding site for the AhR:AhR nuclear translocator (ARNT) heterodimer. Flanking sequences show variable degrees of conservation and function to modulate the affinity of the AhR:ARNT complex and its effect on transcription. The Ci vector represents the degree of conservation of the individual nucleotide positions. Using the Ci vector, sequences were quantitatively compared with the matrix to generate a similarity score, which reflects the similarity of the query sequence to the aligned set of sequences.

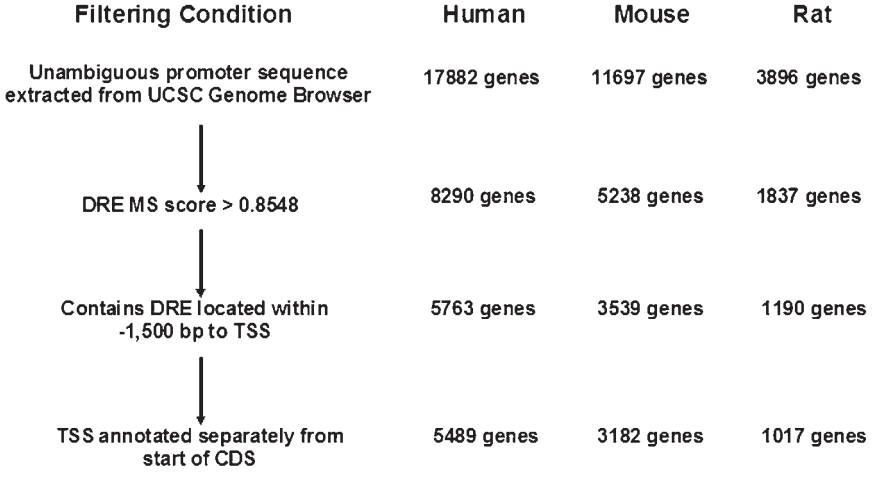


Figure 2. Genome-wide analysis of human, mouse and rat genomic sequences for putative DREs. Unambiguous genomic sequences corresponding to RefSeq accession numbers were computationally scanned for exact matches to the DRE core sequence, GCGTG. The MS score for each match was then compared to the MS score threshold. There was an increase in the occurrence of genes containing putative DREs proximal to the TSS (from −1500 to +1500).

for 24 h as part of a more comprehensive study examining temporal- and dose-dependent changes in gene expression (D.R. Boverhof, L.D. Burgoon, C. Tashiro, B. Chittim, J.R. Harkema and T.R. Zacharewski, manuscript in preparation). Of the genes represented, 6061 had available genomic sequence and a TSS annotated separately from the first codon (Figure 5). Putative DREs were computationally identified within the −1500 to +1500 bp genomic sequence for 1856 (21%) of these genes.

The hepatic expression level of 943 features representing 739 unique LocusLink IDs were found to be responsive [$P1(t) > 0.99$] following 24 h treatment with TCDD, the

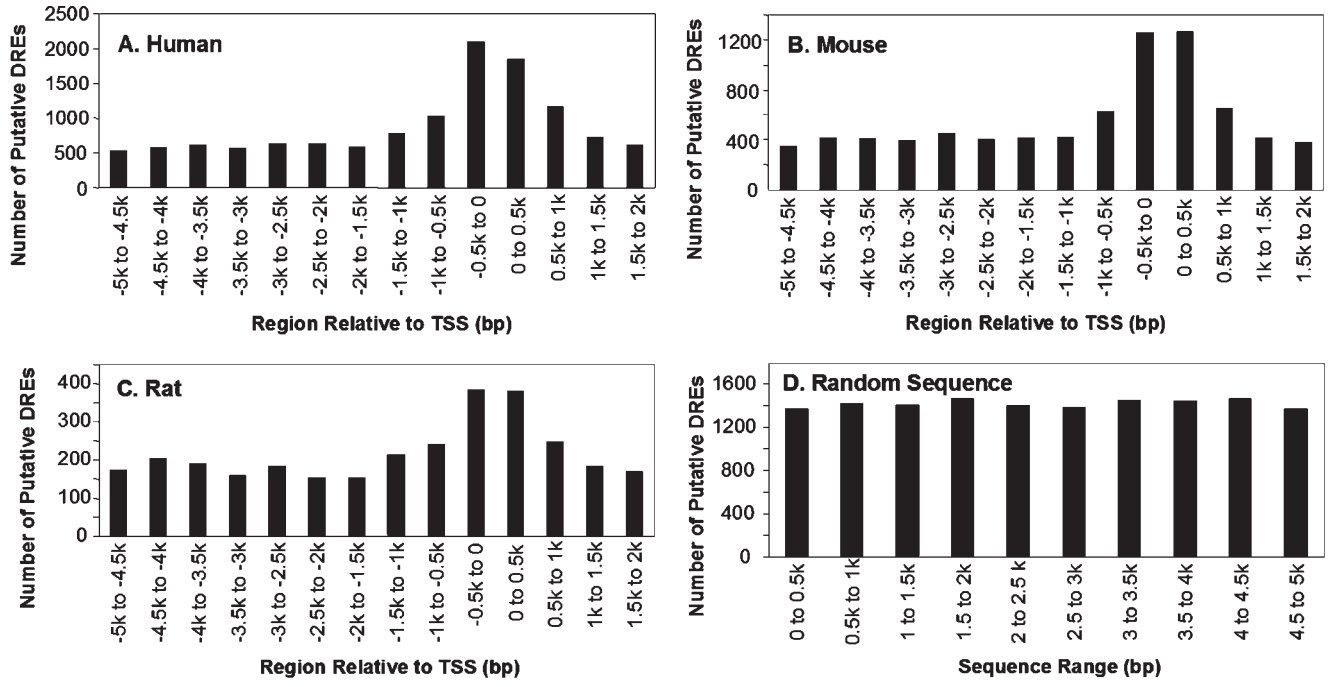


Figure 3. DRE distribution in 17 882 human (**A**), 11 697 mouse (**B**) and 3896 rat (**C**) genes relative to random sequence (**D**). Genomic sequence (from −5000 to +2000 bp) corresponding to RefSeq accession numbers was scanned for exact matches to the DRE core sequence, GCGTG, on both positive and negative strands. For each match, the extended 19 bp sequence was used to calculate an MS score. The occurrence of DREs with MS scores above the MS score threshold is preferentially distributed proximal to the TSS (i.e. from −1500 to +1500 bp). In order to investigate the chance occurrence of the DRE core sequence, a set of 10 000 random DNA sequences was computationally scanned. An equal distribution of DREs across the 5000 bp random sequences was found.

prototypical AhR ligand (Supplemental Table S3). Genomic sequence was available for 585 responsive genes with well-annotated TSSs. A total of 192 genes exhibited significant [$P1(t) > 0.99$] induction or repression and had putative DREs within −1500 and +1500 bp with 81 of these responsive genes exhibiting a significant change in expression of at least 1.5-fold. If limited to −1500 to the TSS, the region where all known functional DREs have been reported, 118 responsive [$P1(t) > 0.99$] genes possessed a putative DRE. It is important to note that preliminary time course studies indicate that 24 h may not be the optimal time to capture TCDD-elicited expression changes for all responsive genes (data not shown). It is expected that as other models and time points are examined, the number of responsive genes will increase, although some responses may be due to secondary effects.

No evidence of overt toxicity was observed for 24 h at any of the doses examined. However, histopathological examination using hematoxylin and eosin staining indicated some evidence of very mild hydropic degeneration at 100 and 300 μg/kg TCDD. This likely contributes to the high-false-positive rate for TCDD responsive genes that did not possess a putative DRE as identified by computational scanning.

Comparative analysis and verification of positionally-conserved DREs in orthologous genes

The identification of evolutionary conserved regions has proven to be a powerful strategy to identify putative functional elements within genomic sequence (11–14). To increase the probability of computationally identifying functional response elements, the −1500 to TSS genomic region of orthologous human, mouse and rat genes were comparatively scanned for the presence of putative DREs. Coding regions (i.e. TSS to +1500 bp) were not included in this analysis as no previously characterized DREs have been located within these regions, and homology biases inherent in coding regions are likely to increase the false-positive rate. Orthologous human, mouse and rat genes were obtained from HomoloGene (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=homologene, Build 31) based on a reciprocal best match. Among 3087 human, 1745 mouse and 554 rat DRE-containing genes within −1500 bp and the TSS, 365 human–mouse, 140 human–rat and 133 mouse–rat orthologous genes containing DREs were identified (Figure 6 and Supplementary Table S4a–c). In addition, 48 human–mouse–rat orthologous genes were identified that contain a putative DRE in the −1500 to TSS genomic region (Supplementary Table S4d).

ClustalW, a multiple sequence alignment tool, was then used to identify 19 human, mouse and rat orthologous genes containing positionally conserved DREs in the −1500 to TSS genomic region (Table 2). Another 21 mouse–rat orthologous genes were identified to contain positionally conserved DREs but were not conserved in human. Surprisingly, no additional positionally conserved DREs were identified between human–mouse and human–rat orthologs.

AhR-mediated regulation for the 19 orthologous genes containing positionally conserved DREs was examined using QRTPCR in C57BL/6 mice liver and mouse Hepa1c1c7

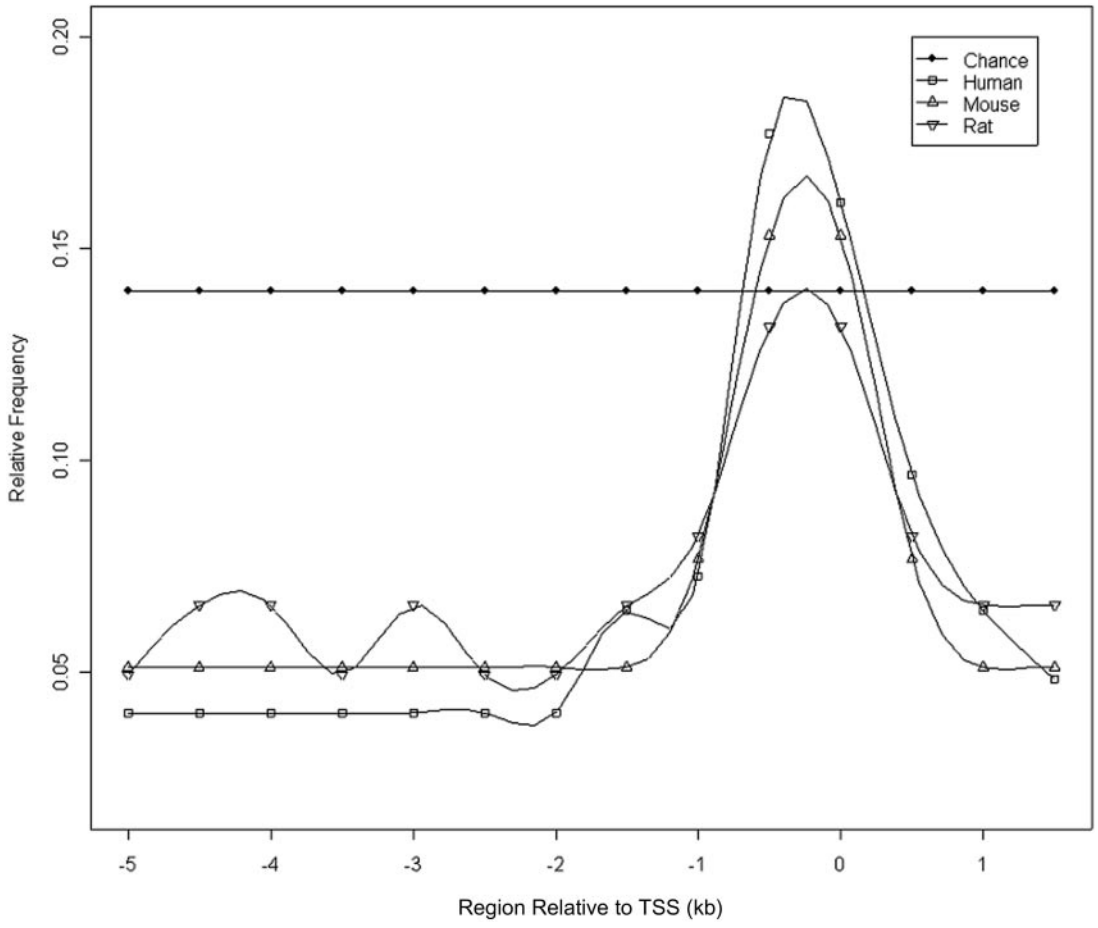


Figure 4. Comparison of DRE core sequence distribution in random DNA sequence relative to human, mouse and rat genomic sequence. The Wilcoxon's rank-sum test was used to compare DRE core sequence distributions on a per species basis to the uniform, chance distribution observed with the set of 10 000 random DNA sequences. DRE core sequence distribution in human, mouse and rat genomes differed significantly from the uniform, random distribution ($P < 0.0001$). DRE core sequences in areas distal to the TSS occur less frequently than would occur due to chance.

hepatoma cell models following TCDD treatment. Six of the nineteen genes exhibited *in vivo* hepatic induction with marked differences in kinetics and levels of expression (Table 3). For example, *Cyp1a1* and *Cyp1b1*, two well-characterized responsive genes, were significantly induced by TCDD at 12 h while reaching 1420- and 82-fold induction, respectively, at 72 h in the mouse liver. Four other genes, *Ugdh*, *Stc2*, *Znf148* and *Rab10*, were also significantly induced. In contrast, only two of the six *in vivo* responsive genes, *Cyp1a1* and *Cyp1b1*, were significantly induced in Hepa1c1c7 cells. Moreover, *Khsnpd1* repression (2.6-fold at 2 h) was specific to Hepa1c1c7 cells. Gene expression levels for the other 12 genes were not significantly affected by TCDD in either model.

Computational scanning of the human, mouse and rat *Cyp1b1* gene, identified 7, 10 and 10 DREs, respectively, between −1500 and the TSS region (Table 4). ClustalW alignment identified four positionally conserved DREs in highly homologous regions between −1100 and −800 bp relative to the TSS (Figure 7). Three of the four positionally conserved DREs (−1026, −855 and −835 for human; −1033, −891 and −872 for mouse; −1001, −859 and −840 for rat) had MS scores above the 0.85 threshold. The four positionally conserved DREs have an average MS score of 0.91, well above the average MS score of 0.80 for the 15 non-conserved identified DREs. Gel-mobility shift assays have demonstrated that the AhR/ARNT heterodimers bind to DREs at positions −1026, −855 and −835 in the human *CYP1B1* gene (36). Moreover, the positionally conserved DREs located at −872 and −891 in the mouse *Cyp1b1* promoter have also proved functional (37,38). Extrapolation of these results also facilitates the ranking and prioritization of putative DREs in the rat promoter for functional analysis.

DISCUSSION

Following completion of the human, mouse and rat genomes, attention has focused on identifying and elucidating the precise location of all sequence-based functional elements. While significant progress has been made identifying protein-coding sequences, most other sequence-based functional elements within the genome, including transcriptional regulatory elements, remain largely uncharacterized. The AhR, a phylogenetically ancient protein with homologs in nearly all living vertebrates including early chordates such as the sea

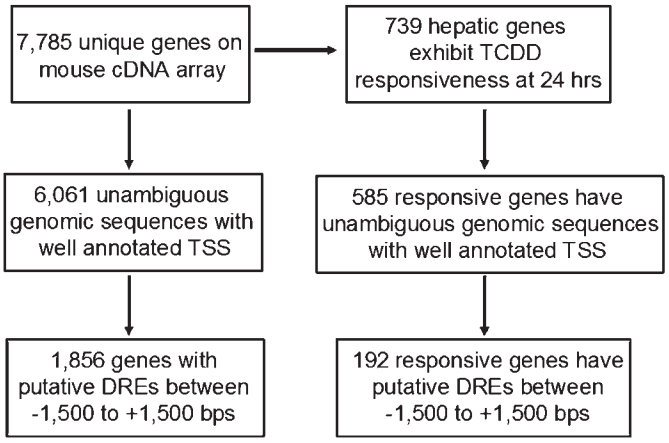


Figure 5. cDNA Microarray analysis of TCDD responsive genes. Immature, female ovariectomized C57BL/6 mice were treated by gavage with 100 and 300 μg/kg body weight of TCDD or vehicle for 24 h. Total RNA was extracted from hepatic tissue and reverse transcribed in the presence of Cy3- or Cy5-dUTP to create fluor-labeled cDNA as described previously (29). The probe mixture was hybridized to a 13 362 mouse cDNA array representing 8891 unique LocusLink IDs. Four independent replicates were performed and dye-swap hybridizations for a total of eight hybridizations per dose of TCDD. The animal was the experimental unit and there was no pooling of samples. Raw intensity data were normalized using a semiparametric method that combines both parametric and non-parametric methods (J. E. Eckel, C. Gennings, T. M. Therneau, D. R. Boverhof, L. D. Burgoon and T. Zacharewski, submitted for publication), and significant changes in gene expression were identified using an empirical Bayes method (30). Only 192 of the 1856 well-annotated genes with putative DREs between −1500 to +1500 bp exhibited a significant change in expression.

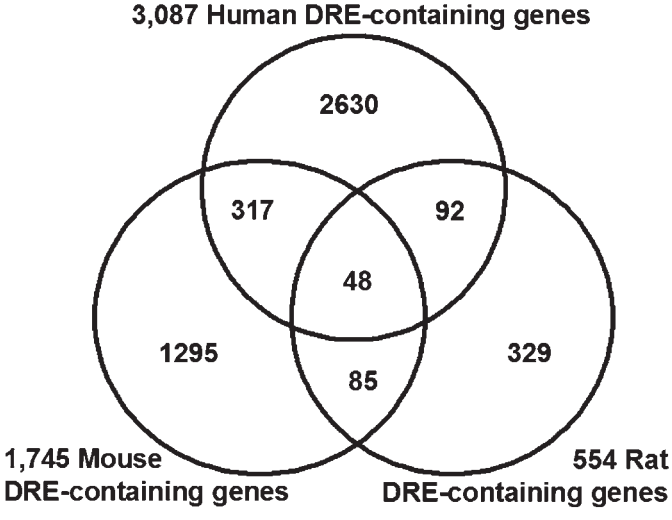


Figure 6. Venn diagram of human, mouse and rat DRE-containing genes. Well-annotated human, mouse and rat genes with DREs between −1500 bp and the TSS and an MS score greater than 0.8548 were identified. Overlapped regions indicate the number of orthologous genes as defined by the HomoloGene database (Build 31).

lamprey (39), provides an ideal model to develop computational scanning approaches that can be complemented with genome-wide empirical studies to identify functional transcriptional elements. The structure and mode of action of the AhR is well conserved, and most, if not all, effects elicited by TCDD and related compounds result from dysregulation of gene expression following AhR:ARNT complex interactions with DREs in the regulatory region of responsive genes. Previous efforts to computationally identify DREs had limited success, due to incomplete sequence data for mammalian genomes (40).

Defining the genomic region to search for *cis*-regulatory elements, particularly in vertebrate organisms such as human and rodents, is challenging. Within well-characterized mouse genes, 85% of the regulatory elements are in the immediate vicinity (<2 kb) of promoters, 10% in introns and 5% are more distally located (41). For example, functional estrogen response elements (EREs) have been characterized in the 5′ promoter (42), coding (43) or 3′ untranscribed (44) regions. Recently, a genome-wide computational scanning study confirmed that EREs are enriched in genomic sequences proximal to the TSS, but could also be identified distal to the TSS (45). Interestingly, distal EREs occasionally were conserved between species, and exhibited greater *in vitro* binding affinity (45). In this study, DRE core sequence distribution was significantly greater in the immediate vicinity (−1500 to +1500 bp) of the TSS (Figure 3A–D) when compared to more distal regions and random DNA sequences suggesting a greater likelihood of functionality for putative DREs in this region.

Note that the substitution intolerant DRE core sequence, GCGTG, would be expected to occur every 512 bp by random chance alone (i.e. 4^5 within the +/− strands). Based on the distribution of DREs proximal to the TSS, a PWM approach followed by a complementary comparative analysis strategy was used to increase the probability of identifying functional response elements. Position weighting allows mismatches at less conserved positions to be considered, since the quality of a match is based on the frequency of a given symbol found at the position in the sequence alignment. The utility of PWM-based scanning to identify functional elements is heavily dependent on the sequence data from which the matrix was constructed. Although constructed from 13 functional DREs, the PWM used in this study may be biased against human AhR-regulated genes due to the fact that the characterized elements were from rodent responsive genes and all mediated induction as opposed to repression. Moreover, putative DREs above and below the MS score may prove to be functional due to context, and therefore the threshold should be used more for guidance when selecting putative DREs that warrant further investigation. Studies have also shown that DRE enhancer and repressor activity is dependent on promoter context as well as on other proximal elements (23,46–50). Therefore, in addition to the MS score, the number of DREs within a specified region may also be a factor in ranking the probability of a functional element (34,35). Secondary search programs, such as FastM (6), could also aid in the identification of composite elements involving DREs (47). It will be interesting to see how the PWM changes between species for enhancer and repressor elements as additional functional DREs are empirically characterized.

Putative functional DREs with MS scores above the threshold were identified in 5238 mouse genes within the −5000 to +2000 bp genomic region. About 1856 of these genes are represented in our in-house developed mouse cDNA array. All have well-annotated TSSs and putative DREs between

Table 2. Orthologous human, mouse and rat genes containing positionally conserved DREs

Human RefSeq	Locuslink	Gene abbreviation	Mouse RefSeq	Locuslink	Gene abbreviation	Rat RefSeq	Locuslink	Gene abbreviation
NM_000475	190	*NR0B1*	NM_007430	11614	*Nr0b1*	NM_053317	58850	*Nr0b1*
NM_001609	36	*ACADSB*	NM_025826	66885	*Acadsb*	NM_013084	25618	*Acadsb*
NM_002156	3329	*HSPD1*	NM_010477	15510	*Hspd1*	NM_022229	63868	N/A
NM_003359	7358	*UGDH*	NM_009466	22235	*Ugdh*	NM_031325	83472	N/A
NM_004501	3192	*HNRPU*	NM_016805	51810	*Hnrpu*	NM_057139	117280	N/A
NM_005389	5110	*PCMT1*	NM_008786	18537	*Pcmt1*	NM_013073	25604	*Pcmt1*
NM_005662	7419	*VDAC3*	NM_011696	22335	*Vdac3*	NM_031355	83532	N/A
NM_006559	10657	*KHDRBS1*	NM_011317	20218	*Khdrbs1*	NM_130405	117268	N/A
NM_012479	7532	*YWHAG*	NM_018871	22628	*Ywhag*	NM_019376	56010	*Ywhag*
NM_021964	7707	*ZNF148*	NM_011749	22661	*Zfp148*	NM_031615	58820	*Znf148*
NM_053024	5217	*PFN2*	NM_019410	18645	*Pfn2*	NM_030873	81531	N/A
NM_145805	64843	*ISL2*	NM_027397	104360	*Isl2*	NM_020471	57233	N/A
NM_000722	781	*CACNA2D1*	NM_009784	12293	*Cacna2d1*	NM_012919	25399	*Cacna2d1*
NM_003040	6522	*SLC4A2*	NM_009207	20535	*Slc4a2*	NM_017048	24780	*Slc4a2*
NM_003129	6713	*SQLE*	NM_009270	20775	*Sqle*	NM_017136	29230	*Sqle*
NM_016131	10890	*RAB10*	NM_016676	19325	*Rab10*	NM_017359	50993	N/A
NM_000104	1545	*CYP1B1*	NM_009994	13078	*Cyp1b1*	NM_012940	25426	*Cyp1b1*
NM_003714	8614	*STC2*	NM_011491	20856	*Stc2*	NM_022230	63878	N/A
NM_000499	1543	*CYP1A1*	NM_009992	13076	*Cyp1a1*	NM_012540	24296	*Cyp1a1*

N/A, gene abbreviation not available.

Table 3. *In vivo* and *in vitro* QRTPCR examination of orthologous genes with positionally conserved DREs

Mm RefSeq	Locuslink	Gene	Maximum fold change (time point)		*P*-value[a]	
			In vivo[b]	*In vitro*[c]	*In vivo*	*In vitro*
NM_009466	22235	*Ugdh*	+4.0 (12 h)	No change	2.29E−06	N/A
NM_009992	13076	*Cyp1a1*	+1420 (72 h)	28 (12 h)	2.89E−12	1.09E−11
NM_009994	13078	*Cyp1b1*	+86 (72 h)	1.7 (4 h)	1.10E−12	0.0698
NM_011317	20218	*Khsnpd1*	No change	−2.6 (2 h)	—	0.00332
NM_011491	20856	*Stc2*	+2.2 (24 h)	No change	0.009	—
NM_011749	22661	*Znf148*	+1.9 (24 h)	No change	0.105	—
NM_016676	19325	*Rab10*	−2.1 (72 h)	No change	0.068	—

[a]*P*-values not reported for genes with no change in expression following treatment with TCDD.
[b]Hepatic samples from immature ovariectomized female C57BL/6 mice gavaged with 10 μg/kg TCDD for 12, 24 and 72 h.
[c]Mouse Hepa1c1c7 hepatoma cells treated with 10 nM TCDD for 2, 4, 6, 8 and 12 h.

Table 4. Orthologous *Cyp1b1* DRE locations, sequences and MS scores

DRE	Human Location[a]	Sequence[b]	MS score	Mouse Location	Sequence	MS score	Rat Location	Sequence	MS score
Conserved	−1026	cgactgtGCGTGcgcagcc	0.93	−1033	cagctgtGCGTGcgccgcc	0.91	−1001	cagctgtGCGTGcgccgcc	0.91
	−942	cgcctccGCGTGtcaggtg	0.84	−978	cgcctcgGCGTGtcaggtg	0.83	−946	cgcctcgGCGTGtcaggtg	0.83
	−855	gggctttGCGTGcgccgct	0.92	−891	gtgccttGCGTGcgccgct	0.92	−859	gtgccttGCGTGcgccgct	0.92
	−836	ccccttttGCGTGcggagct	0.95	−872	ccccttGCGTGcggagct	0.96	−840	ccccattGCGTGcggagct	0.95
Non-conserved	−1494	agaggttGCGTGacttaac	0.87	−712	tggacacGCGTGacagctc	0.83	−677	tggacatGCGTGacagttc	0.87
	−991	acccttgGCGTGgggtgcc	0.82	−709	ctgtcacGCGTGtccagca	0.81	−590	cgcaaagGCGTGcaccatc	0.79
	−265	gtaccgaGCGTGgttctgg	0.73	−625	ccccaagGCGTGcaccatc	0.79	−351	gaggaggGCGTGctcttgg	0.77
				−397	gaggaggGCGTGctcctgg	0.77	−293	ggggaaaGCGTGggggtct	0.76
				−322	gaggaaaGCGTGggggtct	0.76	−79	gggtgggGCGTGaggggcg	0.83
				−94	gggtgggGCGTGaggggcg	0.83	−18	agggatgGCGTGtcgcctc	0.78

[a]Location of DREs was relative from TSS. The negative numbers indicate the upstream region of TSS.
[b]The DRE sequences included the core region and flanking sequences comparable to the DRE PWM regardless of the strand located.

−1500 and +1500 bp (Figure 4). Microarray analysis using hepatic tissue from mice treated with TCDD for 24 h identified 739 genes that exhibited a significant change in expression, with 192 of these genes possessing at least one DRE between −1500 and +1500 with a MS score above the threshold. Collectively, these results provide compelling indirect evidence for the presence of a functional DRE, and support hypotheses regarding the elucidation of the DRE(s)

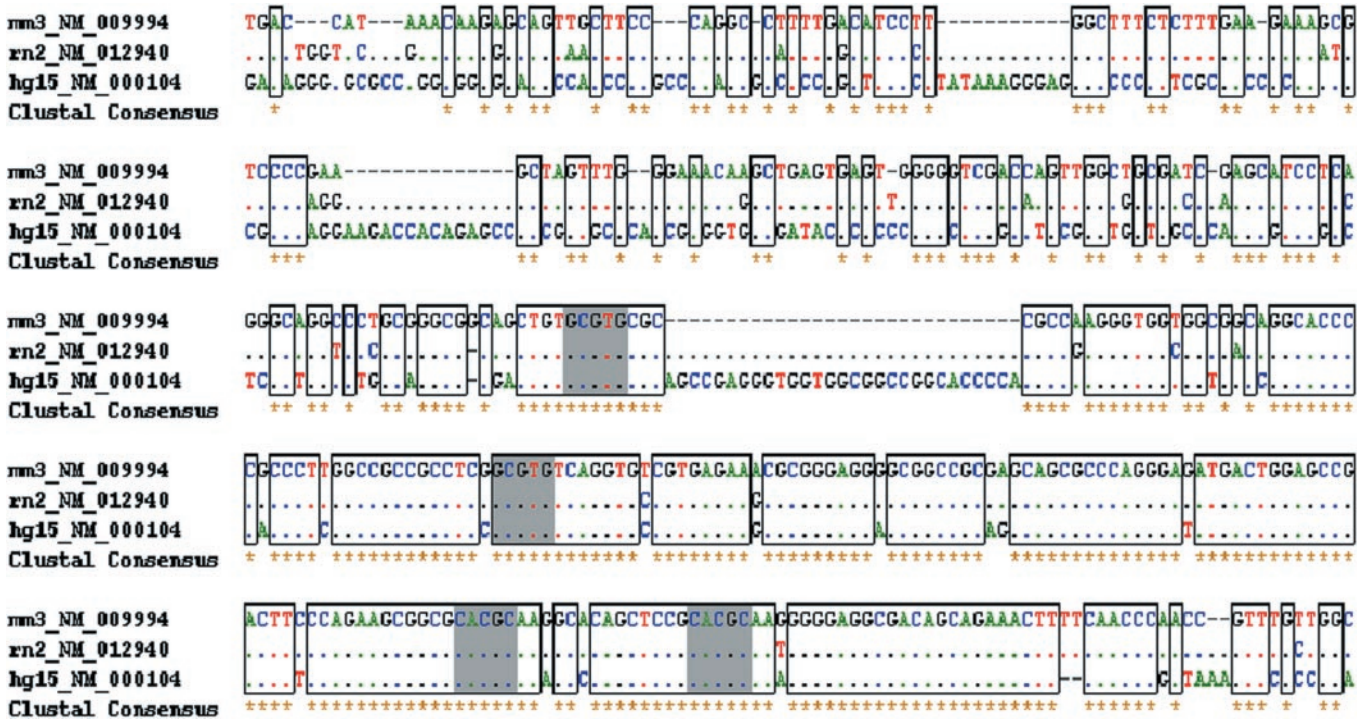


Figure 7. Positionally conserved DREs in human, mouse and rat *Cyp1b1* genomic sequence. Alignment of orthologous human, mouse and rat *Cyp1b1* genomic sequences extracted from the UCSC Genome Browser. Four positionally conserved DREs [human *CYP1B1* (NM_000104), −1123 to −787; mouse *Cyp1b1* (NM_009994), −1133 to −823; and rat *Cyp1b1* (NM_012940), −1101 to −791], shaded in gray, were identified. Three of these positionally conserved DREs have been shown to be functional in gel–mobility shift assays (36).

responsible for mediating the response when considered in combination with MS score data. Although only ∼10% of the genes possessing DREs within −1500 to +1500 bp of the TSS that were represented on the array exhibited a significant change in mRNA transcript levels (either induction or repression) following exposure to TCDD, this percentage will increase as other models and additional time points are examined. However, given the exposure time, some of the responses may be due to secondary effects, and therefore must be verified by further empirical studies. In addition, some responsive genes may have functional DREs that are below the MS score threshold. Nevertheless, gene expression and computational scanning data have provided valuable supporting data for subsequent verification by chromatin immunoprecipitation procedures including genome-wide approaches such as ChIP on Chip assays (51,52).

Comparative human, mouse and rat genome analysis was pursued using ClustalW, a multiple sequence-alignment tool, in order to further enhance the probability of identifying functional DREs by computational scanning. Sequence alignment of homologous regions has proven to be a superior approach to identify functional motifs when compared to simple alignment approaches. *Cis*-regulatory elements often cannot be simply aligned in promoter sequences in order to identify putative functional motifs. Empirical studies have shown that the distance of some functionally conserved response elements from the TSS varies significantly due to gap region length (53,54). For example, comparative analysis of the human *CYP1B1* from −1500 bp to the TSS identified a total of seven DREs. Four of these are positionally conserved in the mouse and rat, and three have been shown to be functional in gel-mobility shift assays (36). Other studies have also verified the functionality of positionally conserved *Cyp1b1* and *Cyp1a1* DREs in the mouse (24,37,38,55–57).

Examination of the 48 human–mouse–rat orthologs (Figure 6), in addition to the manually annotated *Cyp1a1* gene, resulted in the identification of 19 genes with positionally conserved DREs. Only two of these nineteen genes, *Cyp1a1* and *Cyp1b1*, have been reported to be AhR regulated. Seven of these nineteen genes were found to be TCDD responsive with six (i.e. *Ugdh*, *Cyp1a1*, *Cyp1b1*, *Stc2*, *Znfl48*, *Rab10*) exhibiting a change in mouse liver and three in mouse Hepa1c1c7 cells (i.e. *Cyp1a1*, *Cyp1b1*, *Khsnpd1*). These results (Table 3) illustrate the importance of selecting an appropriate model when verifying putative response elements, and support our earlier argument that additional putative DREs will likely be verified as other models and time points are investigated. For example, maximum induction of *Cyp1b1* mRNA was 87-fold in mouse liver but only 1.7-fold in Hepa1c1c7 cells while 18-fold induction was reported in Hepa1c1c7 cells using different RT–PCR primers and conditions (58). These studies also demonstrate that different probes for different regions of *Cyp1b1* can significantly affect gene expression results which may be a contributing factor in the discrepancy between the reported *Cyp1b1* microarray and QRTPCR results in this study. Although the clone representing *Cyp1b1* was sequenced verified in-house, and our records indicate that it amplified well, basic local alignment sequence tool (BLAST) analysis indicates that regions of the *Cyp1b1* are homologous to other mouse BAC clones. In addition, examination of *Cyp1b1* mRNA expression in other ongoing microarray studies using the same representative clone also

provided equivocal results (D.R. Boverhof, L.D. Burgoon, C. Tashiro, B. Chittin, J.R. Harkema and T.R. Zacharewski, manuscript in preparation). Collectively, this suggests that although the clone is representative of *Cyp1b1*, it probes a poor region of the transcript which may be a contributing factor in the discrepancy between the *Cyp1b1* microarray and RT–PCR results.

Gene product functions associated with these orthologs are also consistent with biological activities elicited by TCDD and related compounds that are conserved between species. Among the 48 genes, 7 could be classified as being involved in oxidative stress, hypoxia and detoxification, 5 are associated with calcium homeostasis, and 5 are localized to the endoplasmic reticulum. Associations between oxidative stress, hypoxia and detoxification (59), calcium homeostasis (60,61) and TCDD exposure are well-established effects.

In addition to examining similarities, the availability of the human, mouse and rat genomes also facilitates identifying potential differences in DRE distributions that may reflect species-specific mechanisms of toxicity and interactions with other signaling pathways that could affect AhR-mediated gene expression. On a genome-wide scale, only 37% (49 out of 134 when the manually annotated Cyp1A1 is included) of mouse–rat orthologs with a DRE between −1500 and +1500 had an equivalent human ortholog. About 82% (40 out of 49) of the mouse–rat orthologs shared positionally conserved DREs. In contrast, only 39% (19 out of 49) of the human orthologs with positionally conserved DREs had a rodent counterpart with a positionally conserved DRE. Although biases may have been introduced through the predominant use of bona fide rodent DREs in the development of the PWM, these statistics fuel the debate regarding the suitability of rodent models to assess the potential human health risks associated with exposure to AhR ligands. However, TCDD and related compounds are known to elicit species-specific effects and therefore, not all regulatory elements will be identified using comparative approaches. Nevertheless, these searches rank and prioritize the most promising DREs for further investigation which will facilitate the development of AhR regulons for physiological and toxic responses when integrated with other genome-wide technologies.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Jeremy Burt, Cora Fong, Josh Kwekel and Ed Dere for their constructive comments. D.R.B. is supported by a Michigan Agricultural Experiment Station Competitive Grant. T.R.Z. is partially supported by the Michigan Agricultural Experiment Station. This work was funded by NIH grant ES012245.

REFERENCES

1. Bork,P., Dandekar,T., Diaz-Lazcoz,Y., Eisenhaber,F., Huynen,M. and Yuan,Y. (1998) Predicting function: from genes to genomes and back. *J. Mol. Biol.*, **283**, 707–725.
2. Brazhnik,P., de la Fuente,A. and Mendes,P. (2002) Gene networks: how to put the function in genomics. *Trends Biotechnol.*, **20**, 467–472.
3. Pugh,B.F. and Gilmour,D.S. (2001) Genome-wide analysis of protein–DNA interactions in living cells. *Genome Biol.*, **2**, REVIEWS1013.
4. Wang,W., Cherry,J.M., Botstein,D. and Li,H. (2002) A systematic approach to reconstructing transcription networks in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **99**, 16893–16898.
5. Wyrick,J.J. and Young,R.A. (2002) Deciphering gene expression regulatory networks. *Curr. Opin. Genet. Dev.*, **12**, 130–136.
6. Frech,K., Quandt,K. and Werner,T. (1997) Software for the analysis of DNA sequence elements of transcription. *Comput. Appl. Biosci.*, **13**, 89–97.
7. Wingender,E., Chen,X., Hehl,R., Karas,H., Liebich,I., Matys,V., Meinhardt,T., Pruss,M., Reuter,I. and Schacherer,F. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.*, **28**, 316–319.
8. Brazma,A., Jonassen,I., Vilo,J. and Ukkonen,E. (1998) Predicting gene regulatory elements *in silico* on a genomic scale. *Genome Res.*, **8**, 1202–1215.
9. Pilpel,Y., Sudarsanam,P. and Church,G.M. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nature Genet.*, **29**, 153–159.
10. Werner,T. (2001) Target gene identification from expression array data by promoter analysis. *Biomol. Eng.*, **17**, 87–94.
11. Frazer,K.A., Elnitski,L., Church,D.M., Dubchak,I. and Hardison,R.C. (2003) Cross-species sequence comparisons: a review of methods and available resources. *Genome Res.*, **13**, 1–12.
12. Hardison,R., Slightom,J.L., Gumucio,D.L., Goodman,M., Stojanovic,N. and Miller,W. (1997) Locus control regions of mammalian β-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene*, **205**, 73–94.
13. McGuire,A.M., Hughes,J.D. and Church,G.M. (2000) Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res.*, **10**, 744–757.
14. Tagle,D.A., Koop,B.F., Goodman,M., Slightom,J.L., Hess,D.L. and Jones,R.T. (1988) Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J. Mol. Biol.*, **203**, 439–455.
15. Safe,S. (1990) Polychlorinated biphenyls (PCBs), dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs) and related compounds: environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). *CRC Crit. Rev. Toxicol.*, **21**, 51–84.
16. Okey,A.B., Vella,L.M. and Harper,P.A. (1989) Detection and characterization of a low affinity form of the Ah receptor in livers of mice nonresponsive to induction to cytochrome P1-450 by 3-methylcholanthrene. *Mol. Pharmacol.*, **35**, 823–830.
17. Mimura,J., Yamashita,K., Nakamura,K., Morita,M., Takagi,T.N., Nakao,K., Ema,M., Sogawa,K., Yasuda,M., Katsuki,M. *et al.* (1997) Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor. *Genes Cells*, **2**, 645–654.
18. Vorderstrasse,B.A., Steppan,L.B., Silverstone,A.E. and Kerkvliet,N.I. (2001) Aryl hydrocarbon receptor-deficient mice generate normal immune responses to model antigens and are resistant to TCDD-induced immune suppression. *Toxicol. Appl. Pharmacol.*, **171**, 157–164.
19. Hankinson,O. (1995) The aryl hydrocarbon receptor complex. *Ann. Rev. Pharmacol. Toxicol.*, **35**, 307–340.
20. Swanson,H.I. and Bradfield,C.A. (1993) The Ah receptor: genetics, structure and function. *Pharmacogenetics*, **3**, 213–230.
21. Bacsi,S.G., Reisz-Porszasz,S. and Hankinson,O. (1995) Orientation of the heterodimeric aryl hydrocarbon (dioxin) receptor complex on its asymmetric DNA recognition sequence. *Mol. Pharmacol.*, **47**, 432–438.
22. Swanson,H.I., Chan,W.K. and Bradfield,C.A. (1995) DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *J. Biol. Chem.*, **270**, 26292–26302.
23. Gillesby,B.E., Stanostefano,M., Porter,W., Safe,S., Wu,Z.F. and Zacharewski,T.R. (1997) Identification of a motif within the 5′ regulatory region of pS2 which is responsible for AP-1 binding and TCDD-mediated suppression. *Biochemistry*, **36**, 6080–6089.
24. Lusska,A., Shen,E. and Whitlock,J.P.,Jr (1993) Protein–DNA interactions at a dioxin-responsive enhancer. Analysis of six bona fide

DNA-binding sites for the liganded Ah receptor. *J. Biol. Chem.*, **268**, 6575–6580.
25. Shen,E.S. and Whitlock,J.P.,Jr (1992) Protein–DNA interactions at a dioxin-responsive enhancer. Mutational analysis of the DNA-binding site for the liganded Ah receptor. *J. Biol. Chem.*, **267**, 6815–6819.
26. Nebert,D.W., Puga,A. and Vasiliou,V. (1993) Role of the Ah receptor and the dioxin-inducible [Ah] gene battery in toxicity, cancer, and signal transduction. *Ann. NY Acad. Sci.*, **685**, 624–640.
27. Peterson,R.E., Theobald,H.M. and Kimmel,G.L. (1993) Developmental and reproductive toxicity of dioxins and related compounds: cross species comparisons. *Crit. Rev. Toxicol.*, **23**, 283–335.
28. Quandt,K., Frech,K., Karas,H., Wingender,E. and Werner,T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.*, **23**, 4878–4884.
29. Boverhof,D.R., Fertuck,K.C., Burgoon,L.D., Eckel,J.E., Gennings,C. and Zacharewski,T.R. (2004) Temporal- and dose-dependent hepatic gene expression changes in immature ovariectomized mice following exposure to ethynyl estradiol. *Carcinogenesis*, **25**, 1277–1291.
30. Efron,B. and Tibshirani,R. (2002) Empirical bayes methods and false discovery rates for microarrays. *Genet. Epidemiol.*, **23**, 70–86.
31. Fertuck,K.C., Eckel,J.E., Gennings,C. and Zacharewski,T.R. (2003) Identification of temporal patterns of gene expression in the uteri of immature, ovariectomized mice following exposure to ethynylestradiol. *Physiol. Genomics*, **15**, 127–141.
32. Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.
33. Suzuki,Y., Yamashita,R., Nakai,K. and Sugano,S. (2002) DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. *Nucleic Acids Res.*, **30**, 328–331.
34. Rushmore,T. and Pickett,C. (1990) Transcriptional regulation of the rat glutathione *S*-transferase Ya subunit gene. Characterization of a xenobiotic-responsive element controlling inducible expression by phenolic antioxidants. *J. Biol. Chem.*, **265**, 14648–14653.
35. Xu,C. and Pasco,D.S. (1998) Suppression of CYP1A1 transcription by H_2O_2 is mediated by xenobiotic-response element. *Arch. Biochem. Biophys.*, **356**, 142–150.
36. Tsuchiya,Y., Nakajima,M. and Yokoi,T. (2003) Critical enhancer region to which AhR/ARNT and Sp1 bind in the human CYP1B1 gene. *J. Biochem. (Tokyo)*, **133**, 583–592.
37. Zhang,L., Zheng,W. and Jefcoate,C.R. (2003) Ah receptor regulation of mouse Cyp1B1 is additionally modulated by a second novel complex that forms at two AhR response elements. *Toxicol. Appl. Pharmacol.*, **192**, 174–190.
38. Zhang,L., Savas,U., Alexander,D.L. and Jefcoate,C.R. (1998) Characterization of the mouse Cyp1B1 gene. Identification of an enhancer region that directs aryl hydrocarbon receptor-mediated constitutive and induced expression. *J. Biol. Chem.*, **273**, 5174–5183.
39. Hahn,M.E., Karchner,S.I., Shapiro,M.A. and Perera,S.A. (1997) Molecular evolution of two vertebrate aryl hydrocarbon (dioxin) receptors (AHR1 and AHR2) and the PAS family. *Proc. Natl Acad. Sci. USA*, **94**, 13743–13748.
40. Lai,Z.W., Pineau,T. and Esser,C. (1996) Identification of dioxin-responsive elements (DREs) in the 5′ regions of putative dioxin-inducible genes. *Chem. Biol. Interact.*, **100**, 97–112.
41. Waterston,R.H., Lindblad-Toh,K., Birney,E., Rogers,J., Abril,J.F., Agarwal,P., Agarwala,R., Ainscough,R., Alexandersson,M., An,P. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
42. Shapiro,R.A., Xu,C. and Dorsa,D.M. (2000) Differential transcriptional regulation of rat vasopressin gene expression by estrogen receptor alpha and beta. *Endocrinology*, **141**, 4056–4064.
43. Hyder,S.M., Chiappetta,C. and Stancel,G.M. (1999) Synthetic estrogen 17alpha-ethinyl estradiol induces pattern of uterine gene expression similar to endogenous estrogen 17beta-estradiol. *J. Pharmacol. Exp. Ther.*, **290**, 740–747.
44. Ikeda,K., Orimo,A., Higashi,Y., Muramatsu,M. and Inoue,S. (2000) Efp as a primary estrogen-responsive gene in human breast cancer. *FEBS Lett.*, **472**, 9–13.
45. Bourdeau,V., Deschenes,J., Metivier,R., Nagai,Y., Nguyen,D., Bretschneider,N., Gannon,F., White,J.H. and Mader,S. (2004) Genome-wide identification of high affinity estrogen response elements in human and mouse. *Mol. Endocrinol.*, **18**, 1411–1427.
46. Duan,R., Porter,W., Samudio,I., Vyhlidal,C., Kladde,M. and Safe,S. (1999) Transcriptional activation of c-fos protooncogene by 17beta-estradiol: mechanism of aryl hydrocarbon receptor-mediated inhibition. *Mol. Endocrinol.*, **13**, 1511–1521.
47. Kobayashi,A., Sogawa,K. and Fujii-Kuriyama,Y. (1996) Cooperative interaction between AhR●Arnt and Sp1 for the drug-inducible expression of CYP1A1 gene. *J. Biol. Chem.*, **271**, 12310–12316.
48. Pimental,R.A., Liang,B., Yee,G.K., Wilhelmsson,A., Poellinger,L. and Paulson,K.E. (1993) Dioxin receptor and C/EBP regulate the function of the glutathione *S*-transferase Ya gene xenobiotic response element. *Mol. Cell. Biol.*, **13**, 4365–4373.
49. Takahashi,Y., Nakayama,K., Itoh,S., Fujii-Kuriyama,Y. and Kamataki,T. (1997) Inhibition of the transcription of CYP1A1 gene by the upstream stimulatory factor 1 in rabbits. competitive binding of USF1 with Ahr·Arnt complex. *J. Biol. Chem.*, **272**, 30025–30031.
50. Wang,F., Hoivik,D., Pollenz,R. and Safe,S. (1998) Functional and physical interactions between the estrogen receptor Sp1 and nuclear aryl hydrocarbon receptor complexes. *Nucleic Acids Res.*, **26**, 3044–3052.
51. Ren,B., Robert,F., Wyrick,J.J., Aparicio,O., Jennings,E.G., Simon,I., Zeitlinger,J., Schreiber,J., Hannett,N., Kanin,E. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science*, **290**, 2306–2309.
52. Weinmann,A.S. and Farnham,P.J. (2002) Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods*, **26**, 37–47.
53. Conkright,M.D., Guzman,E., Flechner,L., Su,A.I., Hogenesch,J.B. and Montminy,M. (2003) Genome-wide analysis of CREB target genes reveals a core promoter requirement for cAMP responsiveness. *Mol. Cell*, **11**, 1101–1108.
54. Liu,Y., Liu,X.S., Wei,L., Altman,R.B. and Batzoglou,S. (2004) Eukaryotic regulatory element conservation analysis and identification using comparative genomics. *Genome Res.*, **14**, 451–458.
55. Fujisawa-Sehara,A., Sogawa,K., Yamane,M. and Fujii-Kuriyama,Y. (1987) Characterization of xenobiotic responsive elements upstream from the drug-metabolizing cytochrome P-450c gene: a similarity to glucocorticoid regulatory elements. *Nucleic Acids Res.*, **15**, 4179–4191.
56. Fujisawa-Sehara,A., Sogawa,K., Nishi,C. and Fujii-Kuriyama,Y. (1986) Regulatory DNA elements localized remotely upstream from the drug-metabolizing cytochrome P-450c gene. *Nucleic Acids Res.*, **14**, 1465–1477.
57. Sogawa,K., Fujisawa-Sehara,A., Yamane,M. and Fujii-Kuriyama,Y. (1986) Location of regulatory elements responsible for drug induction in the rat cytochrome P-450c gene. *Proc. Natl Acad. Sci. USA*, **83**, 8044–8048.
58. Eltom,S.E., Zhang,L. and Jefcoate,C.R. (1999) Regulation of cytochrome P-450 (CYP) 1B1 in mouse Hepa-1 variant cell lines: a possible role for aryl hydrocarbon receptor nuclear translocator (ARNT) as a suppressor of CYP1B1 gene expression. *Mol. Pharmacol.*, **55**, 594–604.
59. Nebert,D.W., Roe,A.L., Dieter,M.Z., Solis,W.A., Yang,Y. and Dalton,T.P. (2000) Role of the aromatic hydrocarbon receptor and [Ah] gene battery in the oxidative stress response, cell cycle control, and apoptosis. *Biochem. Pharmacol.*, **59**, 65–85.
60. Holsapple,M.P., Karras,J.G., Ledbetter,J.A., Schieven,G.L., Burchiel,S.W., Davila,D.R., Schatz,A.R. and Kaminski,N.E. (1996) Molecular mechanisms of toxicant-induced immunosuppression: role of second messengers. *Annu. Rev. Pharmacol. Toxicol.*, **36**, 131–159.
61. Tilson,H.A. and Kodavanti,P.R. (1997) Neurochemical effects of polychlorinated biphenyls: an overview and identification of research needs. *Neurotoxicology*, **18**, 727–743.
62. Boesch,J.S., Miskimins,R., Miskimins,W.K. and Lindahl,R. (1999) The same xenobiotic response element is required for constitutive and inducible expression of the mammalian aldehyde dehydrogenase-3 gene. *Arch. Biochem. Biophys.*, **361**, 223–230.
63. Favreau,L. and Pickett,C. (1991) Transcriptional regulation of the rat NAD(P)H:quinone reductase gene. Identification of regulatory elements controlling basal level expression and inducible expression by planar aromatic compounds and phenolic antioxidants. *J. Biol. Chem.*, **266**, 4556–4561.
64. Yoo,H.Y., Chang,M.S. and Rho,H.M. (1999) Xenobiotic-responsive element for the transcriptional activation of the rat Cu/Zn superoxide dismutase gene. *Biochem. Biophys. Res. Commun.*, **256**, 133–137.
65. Emi,Y., Ikushiro,S. and Iyanagi,T. (1996) Xenobiotic responsive element-mediated transcriptional activation in the UDP-glucuronosyltransferase family 1 gene complex. *J. Biol. Chem.*, **271**, 3952–3958.