DETECTION OF NEW POTENTIALLY ACTIVE DRE SITES IN REGULATORY REGION OF HUMAN GENES ENCODING COMPONENTS OF Ah RECEPTOR CYTOSOLIC COMPLEX

Nedosekina EA[1], Oshchepkov DY[1], Katokhin AV[1], Kuznetsova TN[1], Shamanina MY[1], Mordvinov VA[1], Tsyrlov I[2,*]

[1]Institute of Cytology and Genetics, Russian Academy of Science, Siberian Branch, Novosibirsk 630090, Russia; [2]XENOTOX Inc., Scarsdale, New York 10583, USA [*]Correspondence should be sent to xenotoxit@optonline.net

Abstract
A computational search for dioxin responsive elements (DREs) in genes of proteins comprising the Ah receptor (AhR) cytosolic core complex was performed by SITECON. This tool efficiently detected all proved functional DREs in human *CYP1A1, CYP1B1*, and rat epiregulin genes. Eventually, the following number of new DREs in 5'-flanking region was detected by SITECON: one in *AHR* gene, five in *XAP2*, eight in *HSP90AA1*, and three in *HSP90AB1* genes. Additional DREs found in genes of AhR and AhR cytosolic complex members would shed a light on potential mechanisms of expression, stoichiomentry of unliganded AhR core complex, and its degradation *vs* biosynthesis dynamics resulted from treatment of U937 cells with the AhR most potent ligand, 2,3,7,8-TCDD.

Introduction
The AhR is a ligand-activated transcription factor that regulates the expression of numerous DRE-possessing genes in a wide range of species and tissues.[1] Upon binding a ligand, typified by most potent 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), an activated AhR translocates to the nuclei, and subsequently dissociates from its cytosolic core complex partner proteins, the pair of 90 kDa heat shock proteins (HSP90s)[2], and HBV X-associated protein 2 (XAP2).[3] The nuclear translocation of the liganded AhR leads the receptor protein rapid destruction via the ubiquitin/proteasome pathway,[4] which is attributed to a multiple increase in the turnover of AhR.[5]

In the nucleus, the ligand-AhR/Arnt complex is capable of high-affinity binding within the regulatory regions of target genes to DREs, the consensus core sequence of which is 5'-TNGCGTG-3', thus modulating transcription of the target genes, extensively studied in the context of activation of *CYP1A1* promoter.[6] To this end, three DREs core sequences were computationally identified within downstream of the TSS of *AHR* human gene.[7] This indicates that *AHR* expression autoregulation could contribute to the maintenance of unliganded AhR level. However, this mechanism, which might be involved in recovery from deep depletion due to TCDD-activated AhR transcriptional signaling, and a superexpression of *AHR* gene in various malignant cells, has not been investigated.[8] Even less than genetic control of AhR level is known about what factors regulate expression of genes that encode HSP90s, and no data exist yet as regard DREs within regulatory region of *XAP2* gene. So, transcription regulation of AhR cytosolic complex key members still remains to be established. We addressed the problem using sophisticated computational approach SITECON designed to detect potentially active DREs in regulatory region of human genes encoding AhR, XAP2, and HSP90 isoforms, HSP90AA1 and HSP90AB1. SITECON was earlier proven for detection conservative conformational and physicochemical properties in transcription factor binding DNA sites.[9]

Materials and Methods
Earlier, SITECON has shown its high efficacy in recognition of SF1, SREBP (paper in publication) as well as several transcription factor binding sites other than DREs.[9] Therefore, as in[7], here a total of 13 bona fide DREs, all including the substitution intolerant core sequence (GCGTG) and adjacent variable sequences, were used here as a training sample to detect conservative conformational and physico-chemical properties for the DRE site. These data are used to construct the recognition rules and to determine conformational similarity score threshold to rank identified DREs. To this, type I errors were assessed using the jack-knife method: sequences were removed from the training sample one by one each in a series of iterations and served as controls. Type II errors were assessed based on the number of binding sites predicted to be present in a negative 500,000 bp sequence. That sequence was generated by random shuffling of the nucleotides of the sequences in the training sample; thus, the nucleotide compositions of both positive and negative samples were identical, and the search was made in both directions. Evaluation of DRE recognition errors using SITECON is shown in Table 1.

Table 1.

Recognition threshold	Type I errors (sites missed during recognition)	Type II errors (sites recognized accidentally)	
0.95	0.0000	1.11e-003	(1 / 901)
0.96	0.0769	9.10e-004	(1 / 1099)
0.97	0.2308	6.70e-004	(1 / 1492)
0.98	0.3077	1.50e-004	(1 / 6665)

As it follows from the table, at the recognition threshold of 0.95, SITECON employed conformational similarity score[9] allowing recognize all sites in the training sample. In other words, at a 0.95 threshold, type I error was equal to zero. Here we used the same approach Michigan University group used for mammalian DREs,[7] and the threshold

value for which type I error equal to zero. Generally, computational approaches utilizing conformation and physico-chemical properties of DNA sites, and thus considering dependence of interactions with nucleotides at site positions, can be more accurate in depicting the true binding site specificities, since the methods like weight matrix rely upon largely unproven assumption that the nucleotides of binding sites exert independent effects on binding affinity.[10,11]

Furthermore, an efficacy of SITECON in detection of potentially functionally active DREs was tested using sequences of genes encoding human CYP1A1[12] and CYP1B1,[13] and rat epiregulin,[14] which were shown to be regulated by DRE binding AhR-Arnt transcription complex. When compared to the experimentally proven data, it was demonstrated that at the recognition threshold of 0.95, SITECON efficiently identified all tested functionally active DREs. Those include DREs of the rat epiregulin gene,[14] absent in the training sample,[7] and recognized with a 0.8357 MS score by Michigan University group[7], i.e., lower than accepted threshold (0.85) for this study. It is noteworthy that all of the known DREs were recognized by SITECON with a high score of 0.974-0.987. Altogether, we proved SITECON as effective tool for searching potentially active DREs in the genes of interest.

Results and Discussion

As Table 2 shows, three DREs were identified at +8, +406, and +696 positions of the *AHR* gene. Similar positioned DREs have been already predicted within a downstream region of the gene,[7] however, two sites out of three have been filtered by threshold accepted (0.85). In contrast, SITECON recognized that DREs with a high score of 0.975-0.984. It is known, however, that MS scores not necessarily correlate with transcription factor binding affinity.[16,17]

Table 2. DRE positions computationally determined in regulatory regions of human *AHR, HSP90* and *XAP2* genes

Human genes	Sequence of DRE elements	SITECON (threshold of 0.95)		Sun et al.[7] (threshold of 0.85)		Boutros et al.[20]
		Position	Score	Position	Score	Position
AHR (NM_001621)	CGAGAGCGTGCCCC	**+8**	**0.975**	**+69**	**0.825**	ND
	AGCCTGCGTGAGCC	**+406**	**0.984**	**+467**	**0.895**	ND
	CACCCGCGTGCCTG	**+696**	**0.984**	**+756**	**0.818**	ND
	GAATCGCGTGAACC	**-7392**	**0.971**	ND	-	ND
HSP90AA1 (NM_001017963)	TTCCTGCGTGTGAT	**+550**	**0.959**	ND	-	ND
	ACCCCGCGTGCTGG	**+256**	**0.973**	ND	-	ND
	ACCTGGCGTGCTCC	**-39**	**0.972**	ND	-	ND
	CGCAGGCGTGCTCA	**-52**	**0.970**	ND	-	ND
	GGTGGGCGTGATCC	**-197**	**0.968**	ND	-	ND
	GCCCTGCGTGCCGA	**-993**	**0.988**	ND	-	ND
	CGGTGGCGTGCTCA	**-2830**	**0.972**	ND	-	ND
	TACAGGCGTGAGCC	**-4138**	**0.968**	ND	-	ND
HSP90AA1 (NM_005348)	TGTTCGCGTGCGGC	**+306**	**0.973**	ND	-	ND
	CGGGGGCGTGCGAG	**-241**	**0.970**	ND	-	**-234**
	CCCGGGCGTGCCCT	**-364**	**0.971**	ND	-	ND
	GTCGTGCGTGGACG	**-465**	**0.955**	ND	-	ND
	CGGGCGCGTGAGAC	**-491**	**0.970**	ND	-	**-484**
	TACAGGCGTGAGCC	**-2157**	**0.968**	ND	-	ND
	GAATGGCGTGAACC	**-3506**	**0.968**	ND	-	ND
	TAGAGGCGTGAGCC	**-3687**	**0.968**	ND	-	ND
HSP90AB1 (NM_007355)	CTACTGCGTGCCCC	**-22**	**0.985**	ND	-	ND
	AGGCCGCGTGACGA	**-1132**	**0.971**	ND	-	ND
	TGGGAGCGTGATCC	**-1701**	**0.973**	ND	-	ND
	CAGTGGCGTGATCT	**-2618**	**0.970**	ND	-	ND
	TACAGGCGTGAGCC	**-3175**	**0.968**	ND	-	ND
XAP2 (NM_003977)	GAATGGCGTGAACC	**-858**	**0.968**	ND	-	ND
	TACAGGCGTGAGCC	**-1037**	**0.968**	ND	-	ND
	CACAGGCGTGCACC	**-1293**	**0.971**	ND	-	ND
	TACAGGCGTGAGCC	**-2092**	**0.968**	ND	-	ND
	TACAGGCGTGAGCC	**-2404**	**0.968**	ND	-	ND

Therefore, the usage of the approaches, which assess a conformation and physico-chemical properties of DNA sites, might increase a value of prediction, as regards DRE functional activity. In regard to the orientation of the above sites, according to model studies conducted with DREs,[15] their sequences localized in the positions downstream of

TSS might manifest enhancing effect. Notably, two DREs situated at +8 and +696 positions are shown conserved among mammals (Table 3). Supposedly, these DREs are involved in *AHR* gene induction resulted from activation of AhR by its high affinity ligands, epitomized by TCDD.

At the same time, SITECON recognized one additional DRE sequence in human *AHR* gene, particularly at positions –7398 to –7385, which has not been previously detected by any other computational tools.[7] The newly DRE is characterized by high recognition score (0.971), and therefore could be functionally active, despite its distant location. Some DREs, located as far from TSS as in a region between - 4000 and -7000, have been suggested functionally active.[12,15] After detecting these four unusually located DREs, we now develop human cell model of autoregulatory loop that seems to function in maintaining *AHR* own expression. We also study to what extent DREs are involved in retaining the level of the receptor mRNA in the course of ligand-activated AhR destruction-restitution cycle in human pro-monocytic U937 cell line treated with 0.1nM-3.0 nM TCDD.

Table 3. DRE-containing sequences identified in homologous mammalian genes

Human genes	DRE-containing sequence	Existence of DRE sequence in homologous mammalian gene	DRE position
AHR (NM_001621)	CGAGAGCGTGCCCC	Chimp; Mouse; Rat; Dog	**+8**
	AGCCTGCGTGAGCC	Chimp	**+406**
	CACCCGCGTGCCTG	Macaques; Mouse; Rat	**+696**
HSP90AA1 (NM_0010179)	TACAGGCGTGAGCC	Chimp; Macaques; Cow	**-4138**
	CGGTGGCGTGCTCA	Chimp; Macaques	**-2830**
	GCCCTGCGTGCCGA	Chimp; Macaques; Mouse; Rat; Rabbit; Dog; Elephant; Cow	**-993**
	CGCAGGCGTGCTCA	Chimp	**-52**
	ACCTGGCGTGCTCC	Chimp; Macaques	**-39**
	TTCCTGCGTGTGAT	Chimp;	**+550**
HSP90AA1 (NM_005348)	TAGAGGCGTGAGCC	Chimp; Macaques	**-3687**
	GAATGGCGTGAACC	Chimp; Macaques	**-3506**
	TACAGGCGTGAGCC	Chimp;	**-2157**
	CGGGCGCGTGAGAC	Chimp; Macaques	**-491**
	GTCGTGCGTGGACG	Macaques	**-465**
	CCCGGGCGTGCCCT	Macaques	**-364**
	CGGGGGCGTGCGAG	Macaques	**-241**
	TGTTCGCGTGCGGC	Chimp; Macaques	**+306**
HSP90AB1 (NM_007355)	TACAGGCGTGAGCC	Chimp	**-3175**
	CAGTGGCGTGATCT	Chimp	**-2618**
	AGGCCGCGTGACGA	Macaques	**-1132**
	TGGGAGCGTGATCC	Chimp; Macaques	**-1701**
	CTACTGCGTGCCCC	Chimp	**-22**
XAP2 (NM_003977)	TACAGGCGTGAGCC	Chimp	**-2404**
	TACAGGCGTGAGCC	Chimp	**-2092**
	CACAGGCGTGCACC	Chimp	**-1293**
	TACAGGCGTGAGCC	Chimp	**-1037**
	GAATGGCGTGAACC	Chimp	**-858**

Earlier, the regulation of cytosolic AhR has been examined primarily from the context of the role of protein-protein interactions, namely the role of chaperone proteins hsp90 and XAP2 interaction with unliganded AhR.[18]
In this study, we examined the problem from the context of genetic factors regulating cellular level of these proteins. Eventually, five earlier unknown DREs were detected by SITECON within regulatory region of human *XAP2* gene, namely at positions –858, -1037, -1293, -2092, and -2404 (Table 2). Every DREs were recognized with scores ranged from 0.968 to 0.971, which are high enough to be close to SITECON scores characteristic for experimentally proven functional DREs in *CYP1A1*, *CYP1B1* and epiregulin genes. In case DRE in *XAP2* gene are determined functionally active, such numerous enhancer DREs might reflect high sensitivity of *XAP2* gene to activation by AhR-Arnt transcription complex, comparable to such sensitivity of *CYP1A1* gene. DRE detected in human *XAP2* gene could be recognized functionally active, because that five DRE sequences are also found in homologous gene of chimp. If a notable elevation of *XAP2* gene expression by TCDD or other AhR ligand is demonstrated, this chaperone protein could be recognized a regulatory factor for unliganded AhR, because *XAP2* overexpression has been shown to enhance cytosolic AhR levels.[19]

As essential molecular chaperons, HSP90 proteins are involved in stress management, signal transduction, cell cycle control, and folding, degradation, and transport of proteins. With regard to the current subject, cytoplasmic unliganded AhR exists as a heterotetrameric complex composed of AhR (a proper ligand binding subunit), the immunophilin-like XAP2, and a dimer of HSP90s.[19] It is shown that two distinct domains of AhR, bHLH and PAS, are capable of binding to HSP90, and supposedly interact with both proteins in the HSP90 dimer.[21] The PAS of AhR is considered an integral part of the ligand binding domain, and a number of reports suggest that HSP90 dimer maintains the AhR in a conformation required for ligand binding.[22] Cytosolic HSP90 belongs to Group A, which occurs across all eukaryotes, and is produced from *HSP90AA1* and *HSP90AB1* paralogs that have arisen in vertebrates. It remains to be determined whether two subunits of cytosolic HSP90 dimer relate to HSP90AA1 and HSP90AB1, or just to one of these isoforms. That is why in this study we used SITECON to detect potentially active DREs in genes encoding both HSP90AA1 and HSP90AB1. Additionally, according to NCBI GeneID 3320, the *HSP90AA1* gene can produce two mRNA isoforms: NM_001017963.1→NP_001017963.1 HSP90kDa alpha (cytosolic), class A member 1, isoform 1, and NM_005348.2→NP_005339.2 HSP90kDa alpha (cytosolic), class A, member 1, isoform 2. These isoforms differ in gene TSS position and promoter structure, which predict different pattern of gene expression. So, both HSP90AA1 promoters were studied by SITECON. In terms of DRE data, it is noteworthy that RefSeq sequence isoform 2 was generated in 2003, whereas that of isoform 1 in 2007.

As it follows from Table 2, for the first time, SITECON mapped eight DREs in regulatory region of *HSP90AA1* gene, of those two (+256, +550) downstream of TSS, and six (-39, -52, -197, -993, -4138) upstream of TSS. It is important that all discovered DREs characterized by high recognition scores, ranged from 0.959 to 0.988.

According to information on a model utilizing different number of DREs inserted into *SV40* enhancer[15], and our comparative data on human cells infected with HIV-1 (gene of which contains one promoter DRE)[23], and CMV (ten DREs are determined in this viral gene promoter)[24], the level of AhR-mediated expression of the target gene by TCDD or dioxin-like compounds is dependent on how many DREs located in its enhancer region.[25] To this end, genes possessing five or more DREs in their regulatory region, such as *HSP90AA1* (eight DREs), *XAP2* (five DREs), and *HSP90AB1* (five DREs) are presumably target genes of environmental toxicants, and expression could be caused by lower cellular concentration of TCDD, than *AHR* expression. Our new suggestion that cytosolic AhR-XAP2-HSP90 core complex assembly is dependent on the level of each particular gene expression coheres with existing data that HSP90 is required for XAP2-AhR binding, and that AhR is required to stabilize XAP2-HSP90 binding.[18,19]

Comprehending molecular interrelationships of *de novo* synthesized proteins within the AhR core cytosolic complex, mechanisms of TCDD-activated *AHR, HSP90, and XAP2* genes induction, and functional role in it of DREs detected in the regulatory regions will provide new insights into AhR signaling trunsduction in mammals.

References

1. Sogawa K, Fujii-Kuriyama Y. *J Biochem (Tokyo)* 1997; 122:1075.
2. Kazlauskas A, Sundstrom S, Poellinger L, Pongrantz I. *Mol Cell Biol* 2001; 21:2594.
3. Meyer BK, Perdew GH. *Biochemistry* 1999; 38:8907.
4. Roberts BJ, Whitelaw ML. *J Biol Chem* 1999; 274:36351.
5. Ma Q, Baldwin KT. *J. Biol Chem* 2000; 275:8432.
6. Whitlock JP. *Ann Rev Pharmacol Toxicol* 1999; 39:103.
7. Sun YV, Boverhof DR, Burgoon LD, Fielden MR, Zacharewski TR. *Nucleic Acids Research* 2004; 32:4512.
8. Harper PA, Riddick DS, Okey AB. *Biochem Pharmacol* 2006; 72:267.
9. Oshchepkov D, Viyaev E, Grigorovich D, Ignatieva E, Khlebodarova TM. *Nucleic Acids Res* 2004; 32:W208.
10. Barash Y, Kaplan T, Friedman F, Elidan G. *RECOMB* (Berlin) 2003; p.28.
11. Benos PV, Bulyk ML, Stormo GD. *Nucleic Acids Res* 2002; 30:4442.
12. Kress S, Reichert J, Schwarz M. *Eur J. Biochem* 1998; 258:803.
13. Tsuchiya Y, Nakajima M, Yokoi T. *J. Biochem (Tokyo*) 2003; 133:583.
14. Patel RD, Kim DJ, Peters JM, Perdew GH. *Toxicol Sci* 2006; 89:75.
15. Kawajiri K, Watanabe J, Gotoh O, Tagashira Y, Sogawa K, Fujii-Kuriyama Y. *Eur J Biochem* 1986; 159:219.
16. Benos PV, Bulyk ML, Stormo GD. *Nucleic Acids Res* 2002; 30:4442.
17. Roulet E, Fisch I, Junier T, Bucher P, Mermod N. *In Silico Biol* 1998; 1:21.
18. Carver LA, LaPres JJ, Jain S, Dunham EE, Bradfield CA. *J Biol Chem* 1998; 273:33580.
19. Petrius JR, Perdew GH. *Chem-Biol Interact* 2002; 141:25.
20. Boutros PC, Moffat ID, Franc MA, Tuomisto J, Okey AB. *Bichem Biphys Res Commun* 2004; 321:707.
21. Perdew GH, Brradfield CA. *Biochem Mol Biol Int* 1996; 39:589.
22. Whitelaw M, McGuire J, Picard D, Gustafsson J, Poellinger L. *Proc Natl Acad Sci USA* 1995; 92:4437
23. Tsyrlov IB, Pokrovsky AG. *Xenobiotica* 1993; 23:457.
24. Murayama T, Inoue M, Nomura T, Mori S, Eizuru Y *Biochem Biophys Res Commun* 2002; 296:651.
25. Tsyrlov IB. *Organohalogen Comp* 2006; 68:552.