Journal of Theoretical Biology 393 (2016) 1–15

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

# An *in silico* approach to investigate the source of the controversial interpretations about the phenotypic results of the human AhR-gene G1661A polymorphism

Younes Aftabi [a], Abasalt Hosseinzadeh Colagar [a,*], Faramarz Mehrnejad [b]

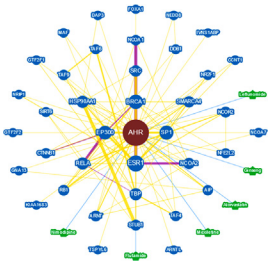
[a] Department of Molecular and Cell Biology, Faculty of Basic Sciences, University of Mazandaran, Babolsar, Post Code: 47416-95447, Mazandaran, Iran
[b] Department of Life Science Engineering, Faculty of New Sciences & Technologies, University of Tehran, P.O. Box: 14395-1561, Tehran, Iran

## HIGHLIGHTS

- AhR- G1661A causes an arginine to lysine substitution in the acidic sub-domain of AhR-TAD with controversial phenotypic results.
- We investigated the possible effects by multiple *in silico* analysis.
- Secondary structure, solvent accessibility and pattern of the binding site and post translational modification predicted to be changed in the region.
- The predicted changes may alter the interactions of TAD, especially with TATA-binding protein.
- The flexibility of TAD could act as a moderating factor and causes distinct outcomes.

## GRAPHICAL ABSTRACT

Aryl hydrocarbon receptor (AhR) is composed of modular independent, active domains. Transcriptional activation domain (TAD) locates at the C-terminal end of the receptor and includes three independent, active sub-domains. TAD is responsible for protein-–protein interactions with components of the general transcriptional machinery and chromatin remodeling co-activators. The TATA-binding protein preferentially binds to the AhR acidic sub-domain. The AhR-G1661A transition causes an arginine to lysine substitution in TAD-acidic sub-domain with controversial phenotypic results. We investigated the possible effects of the SNP by multiple *in silico* analysis. Predicted possible alterations in local secondary structure, solvent accessibility and post translational modifications sound capable of affecting AhR interactions and function. However, the controversial consequences may occur since of flexibility of AhR modular TAD.




## ARTICLE INFO

*Article history:*
Received 19 September 2015
Received in revised form
11 December 2015
Accepted 1 January 2016
Available online 9 January 2016

*Keywords:*
Aryl hydrocarbon receptor
Rs2066853 polymorphism

## ABSTRACT

Aryl hydrocarbon receptor (AhR) acts as an enhancer binding ligand-activated intracellular receptor. Chromatin remodeling components and general transcription factors such as TATA-binding protein (TBP) are evoked on AhR-target genes by interaction with its flexible transactivation domain (TAD). AhR-G1661A single nucleotide polymorphism (SNP: rs2066853) causes an arginine to lysine substitution in the acidic sub-domain of TAD at position 554 (R554K). Although, numerous studies associate the SNP with some abnormalities such as cancer, other reliable investigations refuse the associations. Consequently, the interpretation of the phenotypic results of G1661A-transition has been controversial. In this study, an in silico analysis were performed to investigate the possible effects of the transition on AhR-mRNA, protein structure, interaction properties and modifications. The analysis revealed that the R554K

*Abbreviations:* TCDD, 2,3,7,8-tetrachlorodibenzop-dioxin; FICZ, 6-formylindolo [3,2-b] carbazole; ARNT, AhR nuclear translocator; R, Arginine; AhR, Aryl hydrocarbon receptor; bHLH, Basic helix-loop-helix; K, Lysine; PAS, Per-ARNT-Sim; PTM, Post translational modification; PPI, Protein–protein interaction; SNPs, Single nucleotide polymorphisms; TBP, TATA binding protein; TAD, Transactivation domain; XREs, Xenobiotic response elements; BS, Bayesian statistics; MLR, Multiple linear regression; DT, Decision tree; NN, Neural network; SVM, Support vector machine; HumDiv, Human mutation/divergency; HumVar, Human polymorphic variants

* Corresponding author. Tel./fax: +98 11 35242161/35302452.
*E-mail address:* ahcolagar@umz.ac.ir (A.H. Colagar).

http://dx.doi.org/10.1016/j.jtbi.2016.01.001

Protein–protein interaction
Post translational modification

substitution affects secondary structure and solvent accessibility of adjacent residues. Also, it causes to decreasing of the AhR stability; altering the hydropathy features of the local sequence and changing the pattern of the residues at the binding site of the TAD-acidic sub-domain. Generating of new sites for ubiquitination and acetylation for AhR-K554 variant respectively at positions 544 and 560 was predicted. Our findings intensify the idea that the AhR-G1661A transition may affects AhR-TAD interactions, especially with the TBP, which influence AhR-target genes expression. However, the previously reported flexibility of the modular TAD could act as an intervening factor, moderate the SNP effects and causes distinct outcomes in different individuals and tissues.



## 1. Introduction

The aryl hydrocarbon receptor (AhR), a member of the basic helix-loop-helix (bHLH)/Per-ARNT-Sim (PAS) family of transcription factors, is a well conserved protein during vertebrate species evolution that binds enhancer sequences after activation by ligand (Hankinson, 1995; Hahn, 2002). Along exogenous ligands such as 2,3,7,8-tetrachlorodibenzop-dioxin (TCDD), it is suggested that chemical agents such as tryptophan derived kynurenine and 6-formylindolo [3,2-b] carbazole (FICZ) act as AhR endogenous activators (Stockinger et al., 2014; Hubbard et al., 2015). Upon ligand binding in cytosol, AhR interacts with the importin; enters into the nucleus and forms a complex with the AhR nuclear translocator (ARNT) (Petrulis et al., 2003). Afterwards, the AhR-ARNT heterodimer binds specifically to xenobiotic response elements (XREs) in target genes and activates transcription of them via chromatin remodeling over the promoter and the assembly of the general transcription machinery (Reisz-Porszasz et al., 1994, Gu et al., 2000; Hestermann and Brown, 2003; Wang et al., 2004). AHR pathway is involved in xenobiotic induced carcinogenesis, immunosuppression, and tumor development (Andersson et al., 2002; Opitz et al., 2011; Safe et al., 2013) and plays several roles in cell proliferation and differentiation (Apetoh et al., 2010; Latchney et al., 2010) and organ development (Sahlberg et al., 2002; Walisser et al., 2004; Bock and Köhle, 2006).

Structurally, the AhR is composed of modular independent, active domains: Two PAS domains and one bHLH domain occur at its highly conserved N terminal end and transcriptional activation domain (TAD) locates at the C-terminal of the receptor (Fukunaga et al., 1995; Gu et al., 2000). AhR-TAD, itself includes three independent, active sub-domains: the acidic sub-domain (residues: 500–600), the glutamine (Q) rich sub-domain (residues: 600–713) and the proline/serine/threonine (PST) rich sub-domain with residues 713 up to 848 (Kumar et al., 2001; Hankinson, 2005). TAD is responsible for protein-protein interactions (PPIs) with components of the general transcriptional machinery and remodeling co-activators (Rowlands et al., 1996; Kumar et al., 2001; Wang et al., 2004; Watt et al., 2005). The acidic and Q rich sub-domains of TAD have been shown to interact directly with co-regulators and transcription factors (Kumar et al., 1999; Matthews et al., 2005). Especially a selective interaction and preferentially bound described between the acidic sub-domain of AhR-TAD and TATA binding protein (TBP) (Rowlands et al., 1996; Watt et al., 2005).

It is described profoundly that single-nucleotide polymorphisms (SNPs) are the most common type of genetic variation in the human genome (Sachidanandam et al., 2001). In the human AhR gene, SNPs occur predominantly in a region that encodes a major part of the TAD (Harper et al., 2002). As a non-synonymous SNP (nsSNP) G1661A transition (dbSNP ID: rs2066853) causes an arginine (R) to lysine (K) exchange within the acidic sub-domain of AhR-TAD at position 554 (R554K). This transition is the first discovered SNP of human AhR gene and the most widely studied mutation of it (Kawajiri et al., 1995). However, the interpretation of the phenotypic effects of the rs2066853 SNP has been complicated and controversial. While numerous studies reported a significant association of it with some abnormalities such as a variety of cancers (Berwick et al., 2004; Long et al., 2006; Chen et al., 2009; Shin et al., 2008; Ng et al., 2010; Wang et al., 2010); nevertheless, other reliable investigations refuse the associations (Kim et al., 2007; Cotterchio et al., 2008; Kobayashi et al., 2013; Luo et al., 2013).

The most comprehensive surveys on related literature does not result in finding any convincing explanation about the probable or proven effects of AhR-G1661A transition on its structure and function (Cauchi et al., 2001; Harper et al., 2002; Connor and Aylward, 2006; Merisalu et al., 2007; Shin et al., 2008; Chen et al., 2009; Celius & Matthews, 2010; Ng et al., 2010; Gu et al., 2011; Wang et al., 2012; Luo et al., 2013; Brokken et al., 2014; Cannavo et al., 2014; Ridolfi et al., 2014; Huang et al., 2015). The only thing that is discussed in the articles is the residue change place in AhR protein or restatement of the former reports, that how contradictory the phenotypic results and effects of the rs2066853 SNP has been.

SNPs can cause different structural folds of mRNA and then affect translation processes (Ng and Henikoff, 2006). Also, nsSNPs change the sequence of proteins and can affect their stability, folding, ligand binding characteristics, catalysis and post translational modification (PTM) and PPI-properties (Wang and Moult, 2001; Johnson et al., 2011; David et al., 2012). Hydrophobicity content, secondary structure and solvent accessibility of interacting sequences are major factors in PPI-forming (Lesk and Chothia, 1980; Baud and Karlin, 1999) that could be modified by nsSNPs (Ng and Henikoff, 2006).

Protein PTMs including phosphorylation, acetylation, sumoylation and methylation and ubiquitination can regulate almost entire transcription factor functions rapidly and reversibly (Filtz et al., 2014). AhR-TAD is the target of PTMs such as phosphorylation (Mahon and Gasiewicz, 1995; Long et al., 1998), sumoylation (Xing et al., 2012) and ubiquitination (Ma and Baldwin, 2000). Ubiquitination of the AhR-TAD that manages proteolysis events plays a crucial role in its stability (Roberts and Whitelaw, 1999; Salghetti et al., 2001; Chen et al., 2005). The G1661A transition causes an R to K exchange in AhR-protein and the lysine-specific PTMs described to be major regulators of gene expression, PPI and protein degradation (Zencheck et al., 2012). Indeed, both R and K residues are important target sites for crucial PTMs in proteins.

The current importance and efficacy of bioinformatics tools and in silico approaches in biomedical studies and other spheres of life sciences has been reviewed recently (Mehmood et al., 2014; Chou 2015). Meanwhile, understanding and predicting the downstream effects of the genetic variation via bioinformatics is getting increasingly important. Indeed, the approach is very efficient in conducting studies about the SNP and the molecular causes of the related diseases (Mooney, 2005). During the last decade a large number of deleterious and disease-associated nsSNP detection tools and servers have been developed (Bromberg and Rost, 2007; Calabrese et al., 2009; Adzhubei et al., 2013; Kumar et al., 2014; Choi and Chan, 2015) but, there is still a need for development of methods with improved prediction accuracies (Acharya and Nagarajaram, 2012).