

## Article

# Molecular Forces Governing the Biological Function of Per-Arnt-Sim-B (PAS-B) Domains: A Comparative Computational Study

João Victor de Souza <sup>\*,†</sup>, Piotr Zaborniak <sup>†</sup>, Sylvia Reznikov, Matthew Kondal, Ruidi Zhu and Agnieszka K. Bronowska <sup>\*</sup>

Chemistry—School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK; p.zaborniak@newcastle.ac.uk (P.Z.); s.reznikov@newcastle.ac.uk (S.R.); m.kondal@newcastle.ac.uk (M.K.); r.zhu3@newcastle.ac.uk (R.Z.)

\* Correspondence: j.v.de-souza-cunha2@newcastle.ac.uk (J.V.d.S.);

agnieszka.bronowska@newcastle.ac.uk (A.K.B.); Tel.: +44-(0)-191-208-3506 (A.K.B.)

† These authors contributed equally to this work.

**Abstract:** Per-Arnt-Sim (PAS) domains are evolutionarily-conserved regions found in proteins in all living systems, involved in transcriptional regulation and the response to hypoxic and xenobiotic stress. Despite having low primary sequence similarity, they show an impressively high structural conservation. Nonetheless, understanding the underlying mechanisms that drive the biological function of the PAS domains remains elusive. In this work, we used molecular dynamics simulations and bioinformatics tools in order to investigate the molecular characteristics that govern the intrinsic dynamics of five PAS-B domains (human AhR receptor, NCOA1, HIF1 $\alpha$ , and HIF2 $\alpha$  transcription factors, and *Drosophila Suzukii* (*D. Suzukii*) juvenile hormone receptor JHR). First, we investigated the effects of different length of N and C terminal regions of the AhR PAS-B domain, showing that truncation of those segments directly affects structural stability and aggregation propensity of the domain. Secondly, using the recently annotated PAS-B located in the methoprene-tolerant protein/juvenile hormone receptor (JHR) from *D. Suzukii*, we have shown that the mutation of the highly conserved “gatekeeper” tyrosine to phenylalanine (Y322F) does not affect the stability of the domain. Finally, we investigated possible redox-regulation of the AhR PAS-B domain by focusing on the cysteine residues within PAS-B domains. The cysteines in AhR PAS-B are directly regulating the dynamics of the small molecule ligand-gating loop (residues 305 to 326). In conclusion, we comprehensively described several molecular features governing the behaviour of PAS-B domains in solution, which may lead to a better understanding of the forces driving their biological functions.

**Keywords:** Per-Arnt-Sim; molecular dynamics simulations; aryl hydrocarbon receptor; *D. Suzukii*; NCOA1; HIF1 $\alpha$ ; HIF2 $\alpha$ ; cysteinome



**Citation:** de Souza, J.V.; Zaborniak, P.; Reznikov, S.; Kondal, M.; Zhu, R.; Bronowska, A.K. Molecular Forces Governing the Biological Function of Per-Arnt-Sim-B (PAS-B) Domains: A Comparative Computational Study. *Biophysica* **2021**, *1*, 1–14. <https://doi.org/10.3390/biophysica1010001>

Received: 22 December 2020

Accepted: 12 January 2021

Published: 5 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Per-ARNT-Sim (PAS) domains can be found in proteins from all kingdoms of life. Proteins containing PAS domains in the animal kingdom are involved in cellular responses to hypoxia [1,2], hormonal signalling cascades [3], circadian cycles [2,4,5], synaptic plasticity [6] and xenobiotic responses [7]. Additionally, PAS domains are found within several transcription regulators [8,9]. Proteins containing PAS domains are involved in pathologies, such as tumourigenesis [10], excessive inflammatory responses [11] and neurodegeneration [12,13].

In animals, PAS domains are found in the basic helix-loop-helix (bHLH)-PAS transcription factors. These transcription factors usually comprise a bHLH DNA-binding domain, followed by a tandem: PAS-A and PAS-B, and a flexible C-terminal region containing transactivation (TAD) or trans-repression (TRD [14,15]) domains [8,16]. PAS-B domains may bind small molecules, and small molecule agonists and antagonists can modulate

the protein-containing PAS-B domain activity. Human PAS domains are classified into three groups. Class I includes three hypoxia-inducible factors (HIF $\alpha$ ): HIF1 $\alpha$ , HIF2 $\alpha$ , and HIF3 $\alpha$ . Class I also includes aryl hydrocarbon receptor (AhR), AhR repressor (AhRR), and two single-minded proteins (SIM1 and SIM2). Class II includes two aryl hydrocarbon receptor nuclear translocators (ARNT1 and ARNT2) and two interrelated brain and muscle ARNT-like proteins (BMAL1 and BMAL2). Class III encapsulates the nuclear receptor coactivators (NCOA1, 3 and 4) [17].

PAS domains are essential beyond the human proteome. The discovery of the methoprene-tolerant protein (Met) [18,19], a juvenile hormone receptor, has shown that it plays a crucial role in the insect life cycle and reproduction [20,21]. Met/juvenile hormone receptor (JHR) is known to have a bHLH and a PAS domain; however, sequence analysis shows that it may have a second PAS-B domain in its sequence [18,21].

Given the interest in inhibiting pathways related to PAS-related proteins, several different PAS domains had their structure solved experimentally, either in their unbound forms or complexed with ligands [22,23]. These studies show that PAS domains adopt highly similar tertiary structures, composed mainly of  $\alpha/\beta$  folds despite their low average sequence similarity (~20%) [17].

In the context of switching the proteins “on” and “off” by small molecule binding, probably the most researched PAS-B domain-containing transcription factor is the AhR receptor. This highly conserved protein is heavily involved in the cellular response to environmental pollutants and xenobiotic stress. Its translocation from the cytosol, where AhR resides in its resting state bound to the Hsp90 chaperone, to the nucleus, wherein AhR exerts its function as a transcription factor, is a multistep process, involving ligand-induced conformational changes, which remain mostly underexplored, despite years of research [24–27]. Considering a lack of its experimental structure, the complexity of this multistep process, several binding and un-binding events involve the activation of AhR, promiscuity of ligand binding, and some conflicting data on the molecular nature of the activation process, understating this process at the atomistic level of detail presents itself as a challenge. While “druggability” of the AhR PAS-B domain has been in the focus of intensive research since 2010, as the role of AhR as attractive therapeutic target emerged, the dynamics of intrinsically disordered distal segments of the domain have been overlooked. Wood and coworkers showed that the entropic force produced by the intrinsically disordered terminal regions shifts the protein’s conformational ensemble, tuning the binding affinity [28]. The function of the region did not depend on its sequence: the affinity enhancement could be accurately predicted solely based on the length of the terminal region, and was consistent with the entropic force generated by an unstructured peptide attached to the protein surface. This characteristic implies that disordered terminal segments may tune the energy landscape of proteins, which may explain the occurrence of such disordered regions in many evolutionarily conserved proteins, including transcriptional regulators containing PAS-B domains. Despite these findings, very little is currently known about the specific mechanisms of how disordered termini may modulate functions of PAS-B domains, including that of the AhR receptor.

Additionally, the reactive cysteinome of AhR’s PAS-B domain has not been addressed in detail, even though the evidence that AhR is involved in cellular redox regulation and maybe redox-regulated itself is starting to accumulate [7,29–31].

To address these gaps, we assessed the intrinsic dynamics of different ensembles of the human AhR PAS-B domain, to evaluate how changes in dynamics within intrinsically unstructured segments of the domain affect the overall stability and the misfolding and aggregation propensity of the domain. To evaluate the effect of these unstructured terminal segments, we have generated and simulated truncation mutants of human AhR PAS-B domain, where the terminal segments were truncated to various degrees. Atomistic molecular dynamics (MD) simulations showed that the entropic force generated by these unstructured termini segments controls the stability of the entire domain directly, and finely tunes the flexibility of the dimerisation interfaces and the “druggable” sites reported

in the previous studies [9]. Subsequently, we extended our AhR PAS-B simulations to other PAS-B domains, to assess whether the observations for AhR might be part of a general trend. The following PAS-B domains were investigated: NCOA1, HIF1 $\alpha$ , HIF2 $\alpha$  and the previously described Met *D. Suzukii* PAS-B. We found that the replacement of the “gatekeeper” tyrosine Y322 of AhR, by a phenylalanine (in JHR of *D. Suzukii*), did not increase the local flexibility of the loop. Finally, we evaluated the reactivity of cysteines located in the AhR PAS-B domain. We showed previously unreported oxidation of C300 and C316 residues and subsequent formation of the disulfide bond between those residues, which contributes to the AhR PAS-B stability and may therefore be functionally important.

## 2. Materials and Methods

### 2.1. Homology Modelling

The models created vary in the level of truncation of the terminal segments. The four models were standard length (canonical AhR denoted as AhR-Canon, residues 283–390), the N-terminal truncation (AhR N-trunc; residues 287–390), a C-truncation (AhR C-trunc; residues 283–384) and an NC-truncation (Ahr NC-trunc; residues 287–384). The length of the truncations was made based on the secondary structure assignment, where the residues predicted to be disordered were truncated. These models were created using Phyre2 web server [32], using HIF2 $\alpha$  PAS-B structure (PDB 5UFP) [22] as a template (Phyre2 confidence = 99.9, identity = 27%). The juvenile hormone receptor/methoprene-tolerant protein (Uniprot entry = LOC108005440, Uniparc sequence = UPI0007E2F3F4), PAS-B domain from *D. Suzukii* JHR PAS-B was built using Phyre2 web server, using HIF2 $\alpha$  PAS-B structure (PDB 5UFP) as a template (Phyre2 confidence = 99.9, identity = 24%). The HIF1 $\alpha$  (PDB: 4H6J), HIF2 $\alpha$  (PDB: 5UFP) and NCOA1 (PDB: 5NWM) [23] PAS-B domains were obtained from the RCSB protein databank (PDB). All solvent and co-crystallised molecules were removed when necessary.

### 2.2. Molecular Dynamics Simulations

All simulations were performed using GROMACS 2016.1 [33]. All simulations were parametrised using the AMBER99SB-ILDN force field [34], with the TIP3P water model with hydrogens added by Gromacs when necessary. Box distance was set to 1 nm, and periodic boundary conditions were applied. Each box was solvated and Na<sup>+</sup>, and Cl<sup>-</sup> ions were added to achieve a 0.1 M neutral concentration. The solvated systems were energy minimised and equilibrated. The minimisations ran using steepest descent for 1000 cycles. Energy step size was set to 0.001 nm, and the maximum number of steps was set to 50,000 in each case. The minimisation was stopped when the maximum force fell below 1000 kJ/mol/nm using the Verlet cutoff scheme. Treatment of long-range electrostatic interactions was set to Particle mesh Ewald (PME) [35], and the short-range electrostatic and van der Waals cutoff was set to 1.0 nm. After the energy minimisation, heating to 300 K was performed for 10 ps with a time step of 2 fs, and position restraints were applied to the backbone in an NVT ensemble using a V-rescaling thermostat [36]. The LINCS [37] algorithm was used to constrain the hydrogen bonds in the protein. With the Verlet cutoff scheme and the non-bonded short-range interaction, the cutoff was set to 1.0 nm. Long-range electrostatics were again set to PME. The temperature coupling was set between the protein and the non-protein entities by using a V-rescaling thermostat, with a time constant of 0.1 ps. The simulation temperature was set to reach 300 K with the pressure coupling off. Pressure equilibration was run at 300 K with a Parrinello–Rahman [38] pressure coupling on, and set to 1 bar in an NPT ensemble. The equilibration trajectories were set to 20 ns (discarded from the analysis), and the production MD simulations were performed for three replicas each. The production runs for the four different AhR PAS-B termini constructs ran for 200 ns in triplicates.

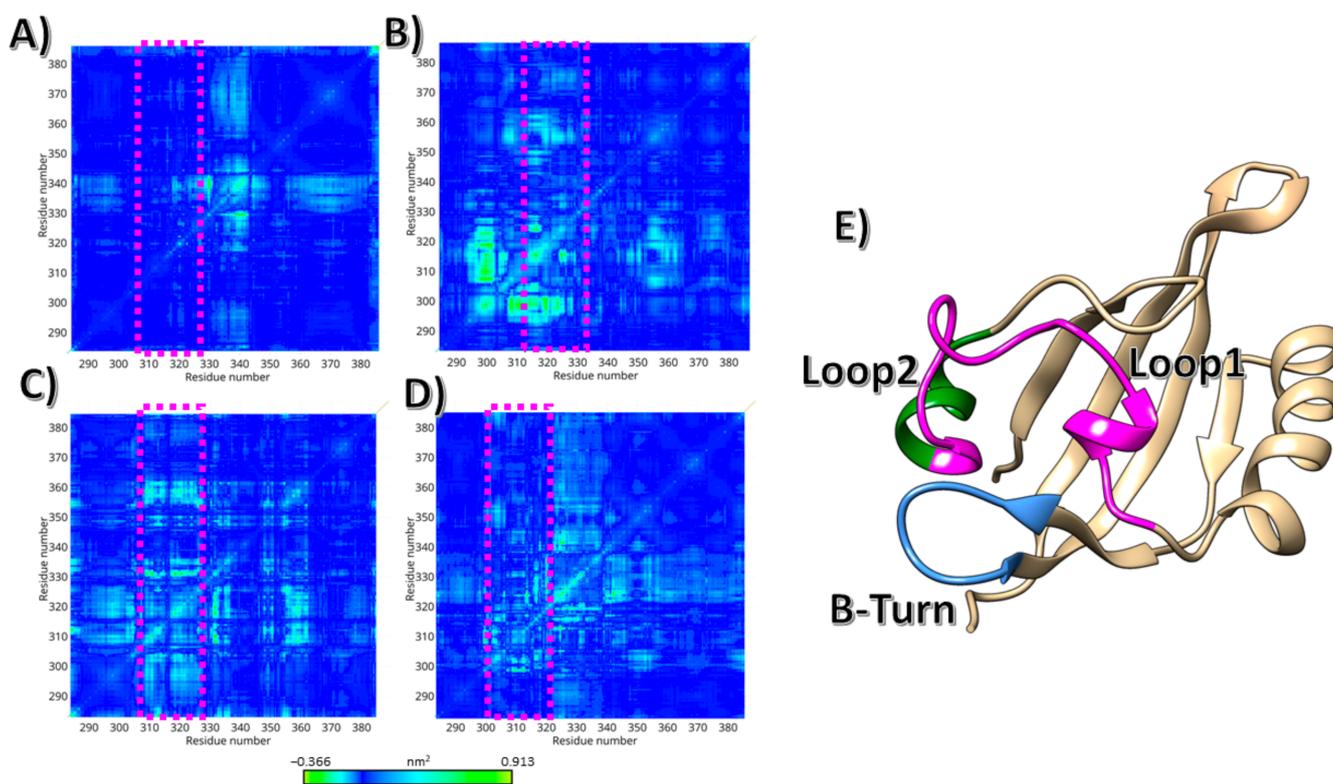
All five different PAS-B models (AhR, NCOA1, HIF1 $\alpha$ , HIF2 $\alpha$  and *D. Suzukii* JHR) were ran using the same protocol as described earlier, except for the extension of the production runs to 500 ns in all their triplicates. To evaluate the effect of the disulfide

bond on the AhR PAS-B, residues C300 and C316 were parametrised as bonded using GROMACS. The simulation procedures were the same as the ones described previously, with a production run the length of 500 ns in triplicate. All simulations converged to a stable configuration, as shown in Figures S1 and S2 (Supplementary Materials), showing the root-mean-square deviation (RMSD) for the termini length and the 500 ns runs respectively.

Analysis of the trajectories was made using GROMACS, including RMSD to assess overall stability, per-residue root-mean-square fluctuation (RMSF) to assess the local flexibility, solvent-accessible surface areas, and principal component analysis (PCA). To predict and evaluate the aggregation propensity and hotspots, we used Aggrescan3D [39,40], using the representative frame from the most populated cluster. The contact asteroid maps were created using Protein Contact Atlas [41] web server, using the representative state of the most populated cluster of each run. For the evaluation of reactivity and pKa values of the AhR cysteines, the Cy-preds webserver was used [42].

### 3. Results

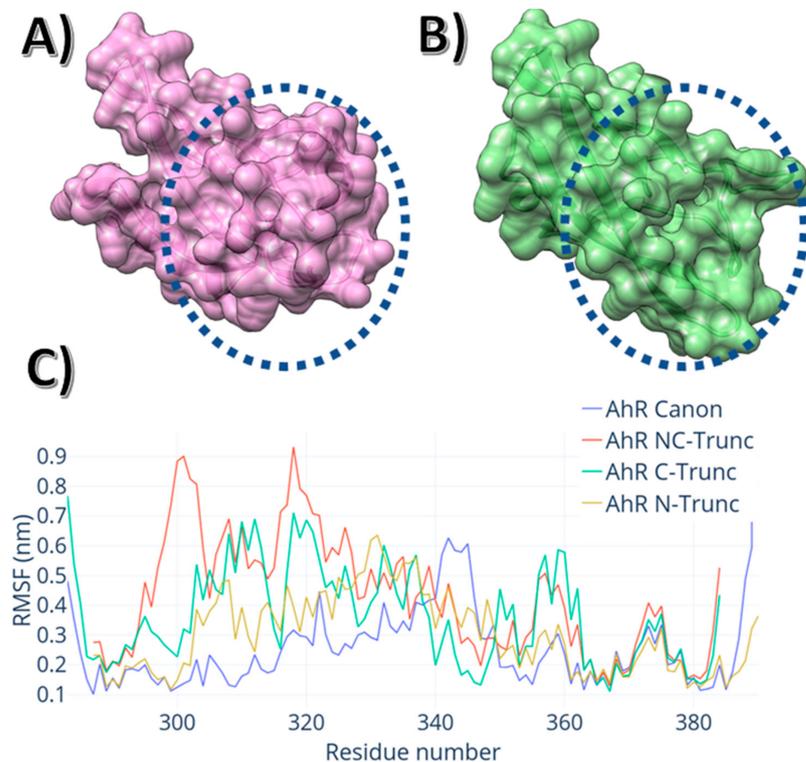
The C and N termini on AhR PAS-B directly affect its overall dynamics, particularly on loop 1 (Figure 1). Covariance analysis for these four models has shown that the tail truncations directly affected the conformations sampled for loop 1. As shown in Figure 1, the motion of loop 1 in the canonical model does not show covariation with other motions. On the other hand, the truncated models show a direct relationship between the motion of loop 1 and the remaining residues of the PAS-B domain. This effect is higher for the NC truncation (Figure 1B) and almost negligible in the canonical model.



**Figure 1.** Covariance matrices for all models. (A) AhR PAS-B Canon; (B) AhR PAS-B NC-Trunc; (C) AhR PAS-B C-Trunc; (D) AhR PAS-B N-Trunc; in purple dashed boxes, the loop 1 motif. (E) AhR Canon starting conformation with coloured specific regions, in green the loop 2, in fuchsia the loop 1 and in blue the B-Turn.

These correlation effects are related to the contacts between N and C termini and three other regions: loop 1, loop 2 and the  $\beta$ -Turn shown in Figure 1. Both of the termini are located within the  $\beta$ -sheet on the opposite facet from loop 1. They directly interact with the  $\beta$ -turn located between residues T355-R359. In turn, this  $\beta$ -turn interacts directly with

residues located in loop 1, as shown in Figure 1E. The truncation of the tail creates two distinct effects: first, it creates a cleft, as can be seen in Figure 2A,B, located within  $\beta$ -strand 2,  $\beta$ -strand 3 and loop 2. Second, there is a propensity of termini residues to integrate the  $\beta$ -sheet, reducing the overall structural fluctuation, as can be seen in the root-mean-square fluctuation shown in Figure 2B. The fluctuation on loop 1 is less pronounced for the N-terminus truncation in comparison to the C-terminus truncation.

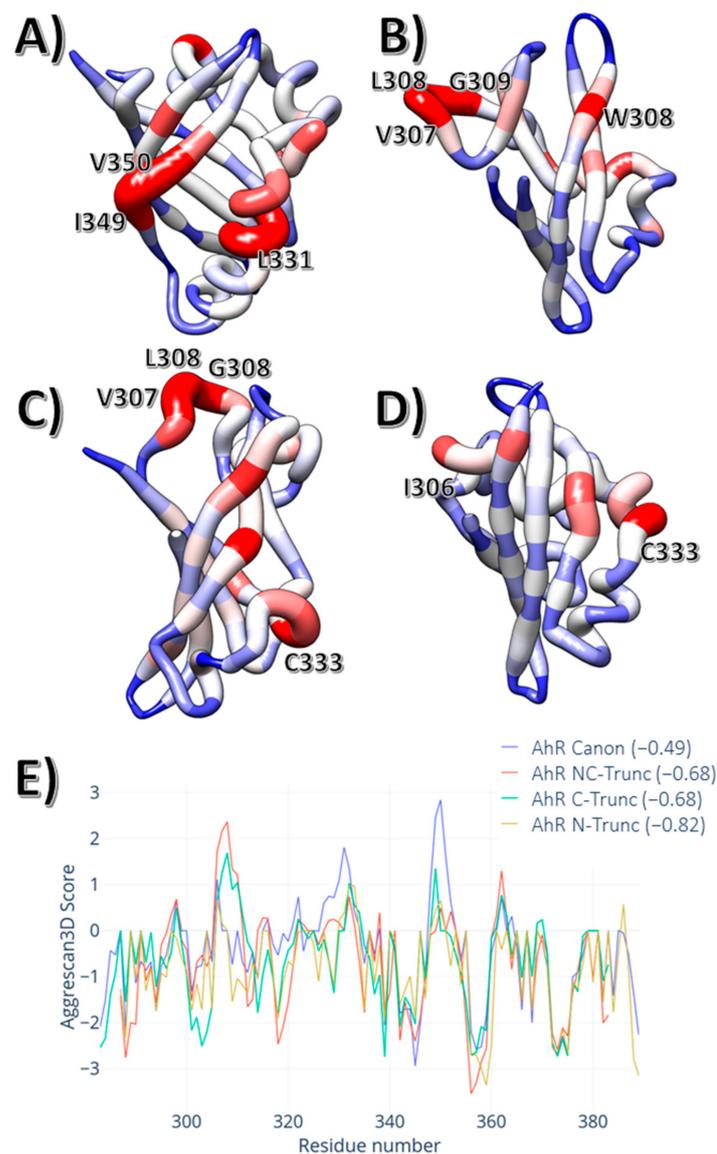


**Figure 2.** (A) AhR Canon average conformation with the surface, in dashed lines, the region occupied by the tails. (B) AhR NC-Trunc. Dashed lines indicates the clefts obtained by the termini cleavage. (C) Residual root-mean-square-fluctuation for all four models.

The deletion of N or C termini affects the aggregation propensity of AhR PAS-B. To evaluate the effect on aggregation, we used the Aggrescan3D webserver. In Figure 3, the aggregation propensity per residue and the overall average is shown for all four models. The termini regions play a significant role in stabilising the molecule against aggregation. However, the model with the highest average propensity of aggregation was the Canon (Aggrescore =  $-0.49$ ) followed by NC-Trunc and C-Trunc (Aggrescore =  $-0.68$ ). For the model with only the truncated N terminus, the aggregation scores were  $-0.82$ , the lowest of the set.

Nonetheless, this difference does not come directly from the termini truncation. As shown in Figure 3, the aggregation hotspots are not located in the chain termini. For the C-Trunc and NC-Trunc models, the primary aggregation hotspot is a trine of three hydrophobic residues (V307, L308, G309). Moreover, the central  $\beta$ -Strand also shows an aggregation hotspot, with the two central residues I349 and V350.

Four different PAS-B domains, other than AhR, were simulated in order to understand how the dynamics and the stable conformations of AhR PAS-B compare to the other PAS-B domains: HIF1 $\alpha$ , HIF2 $\alpha$ , NCOA1 and *D. Suzukii* JHR. All five molecules show a conserved primarily sequence between themselves, as shown in Figure S3 (Supplementary Materials). However, after the simulation, the structural alignment shows that the three-dimensional structures are different between the molecules, with similar areas (pink squares in Figure S4 (Supplementary Materials)), albeit with a high number of nonaligned regions.

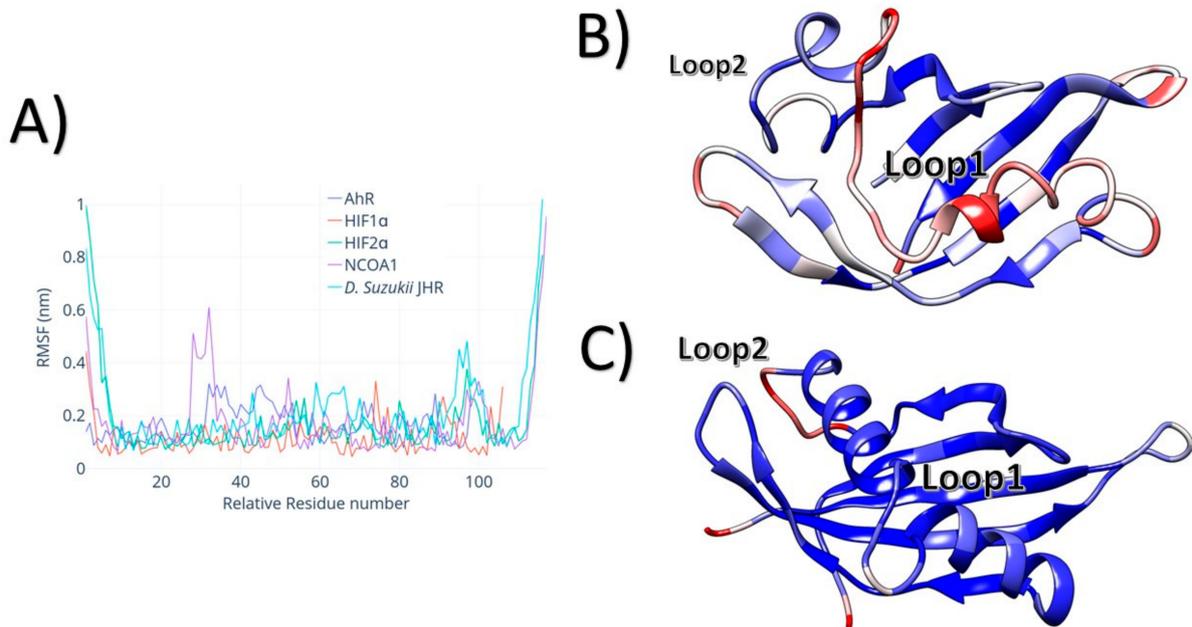


**Figure 3.** (A) AhR PAS-B Canon; (B) AhR PAS-B NC-Trunc; (C) AhR PAS-B C-Trunc; (D) AhR PAS-B N-Trunc; the red residues indicate aggregation hotspots. (E) Aggrescan score (Aggrescore) per residue for all four models. The values within parenthesis are the average aggrescore for all residues.

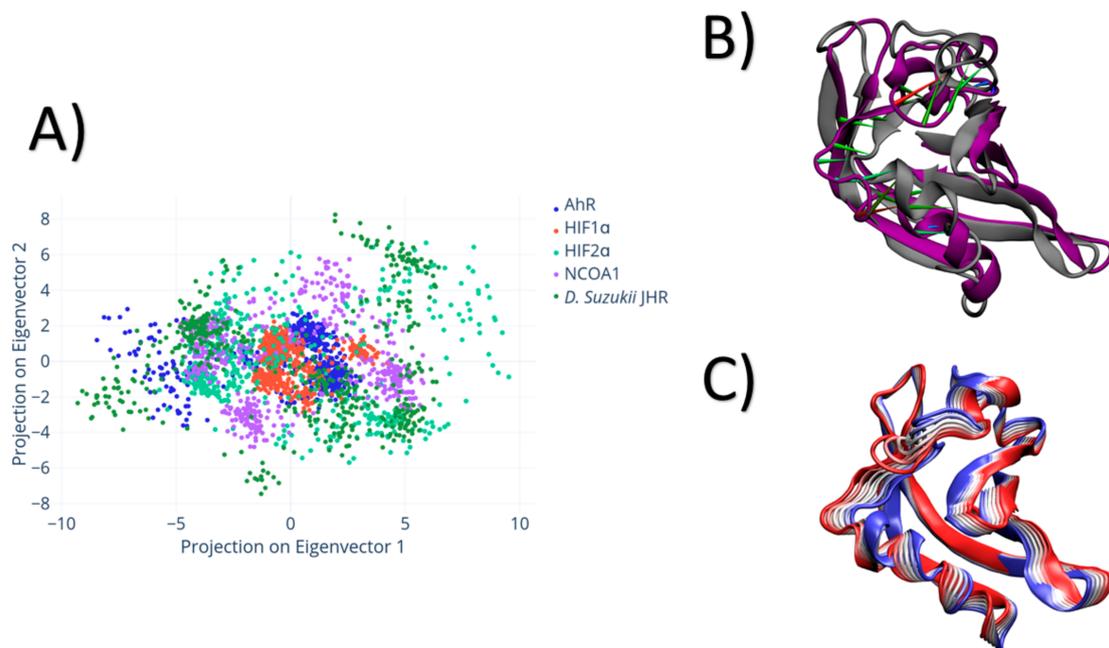
As shown in Figure 4, the residual fluctuation is significantly different between each molecule. Both NCOA1 and AhR PAS-B domains show a higher residual fluctuation for residues 317 to 325 in AhR and residues 280 to 282 in NCOA1 compared to their counterparts. The higher fluctuations may be attributed to several factors: for NCOA1, there is the formation of a short helix within the 280–282 region, as shown in Figure 4C. For AhR, the structural fluctuation within the PAS-B domain arises from a weaker interaction between loop 1 and the rest of the protein, represented by the average low total number of hydrogen bonds between these two groups ( $1.5 \pm 1$ ).

The principal component analysis (PCA) showed that the essential dynamics driving the motions are different for the different proteins considered in this study. As shown in the projection of the dynamics on principal components 1 and 2 (Figure 5A), HIF1 $\alpha$  has the most compact distribution compared to the other molecules. AhR PAS-B shows a transition between two regions, a sparse cluster representing the starting conformations and a compact cluster representing the final equilibrated ensemble. The transition between these two PCA clusters is mainly driven by a rearrangement of the residues located within loop 1, as shown in Figure 5B. Even showing high conservation on both primary sequence

and overall similar fold motif, *D. Suzukii* PAS-B domain had a significant difference from all the other four PAS-B domains discussed in this work: the conserved tyrosine (Y322 in AhR) was mutated to phenylalanine (F437).

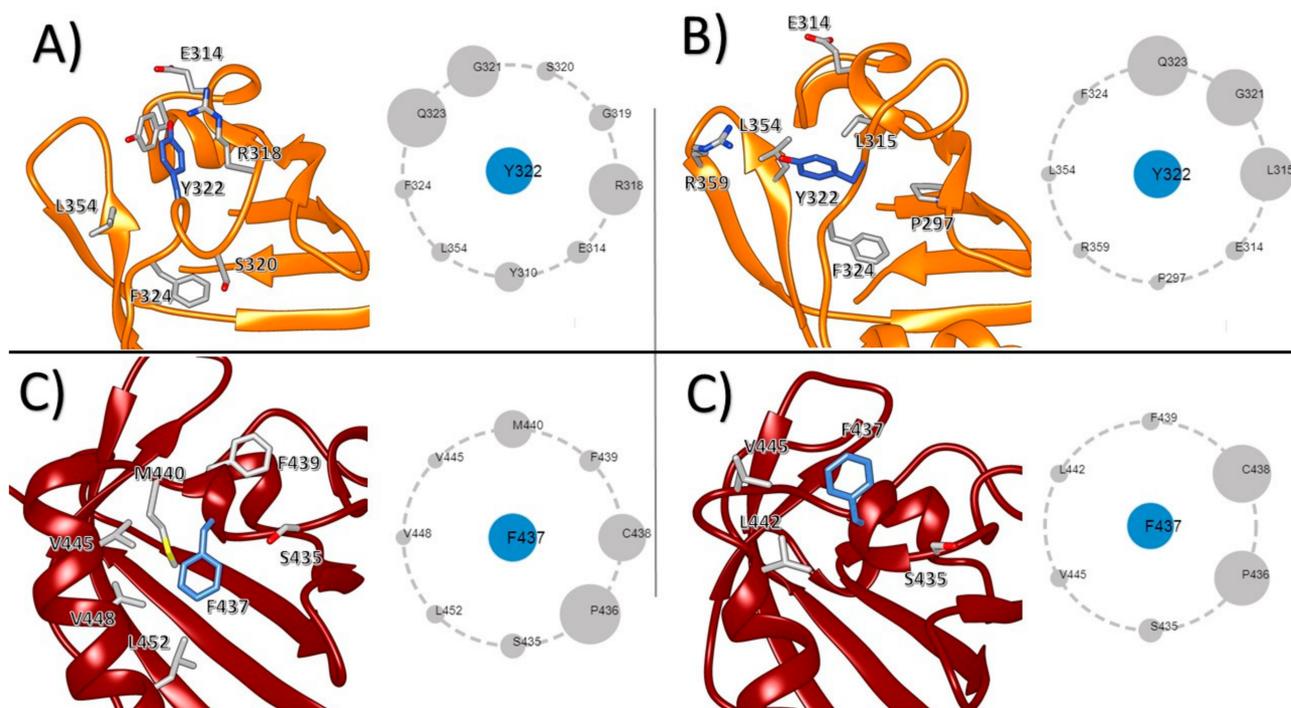


**Figure 4.** (A) Root-mean-square-fluctuation for all five models. (B) AhR PAS-B domain average structure after 500 ns; (C) NCOA1 PAS-B domain average structure after 500 ns. Red, residues with high spatial fluctuation; in blue, residues with a low spatial fluctuation.



**Figure 5.** Principal component decomposition for all five simulated PAS-B domains. (A) Principal component 2D scatter plot for different five different PAS-B models. (B) Porcupine plot of the largest principal component: in grey, the molecular representation of the low extreme of the largest principal component for AhR PAS-B; in purple, the molecular representation for the higher extreme. (C) Principal component pathway: in blue, a representative of conformations projected on low values of the PC; in red, a representative conformation for the high-value projections, with a projection difference of 7 nm between conformations.

To evaluate the effect of Y322 and F437 on the dynamics of their respective PAS-B, we focused on the dynamics of said residues in relation to their neighbouring atoms. In AhR, Y322 has interactions primarily with its neighbours Q23 and G21 and shows contact points with R318 (Figure 6A) in its starting configuration. In the final configurations, loop 1 experiences a conformational change, which flips Y322, which results in contacts being created between L315 and R359 (Figure 6B). *D. Suzukii* JHR F437 keeps most of the interactions throughout the simulation. The residues with the highest amount of contact points with F437 besides its adjacent residues are M440, S435 and V445 (Figure 6C). F437 flips and breaks its contact with M440 and V445, but it does not change its overall loop 1 configuration (Figure 6D).

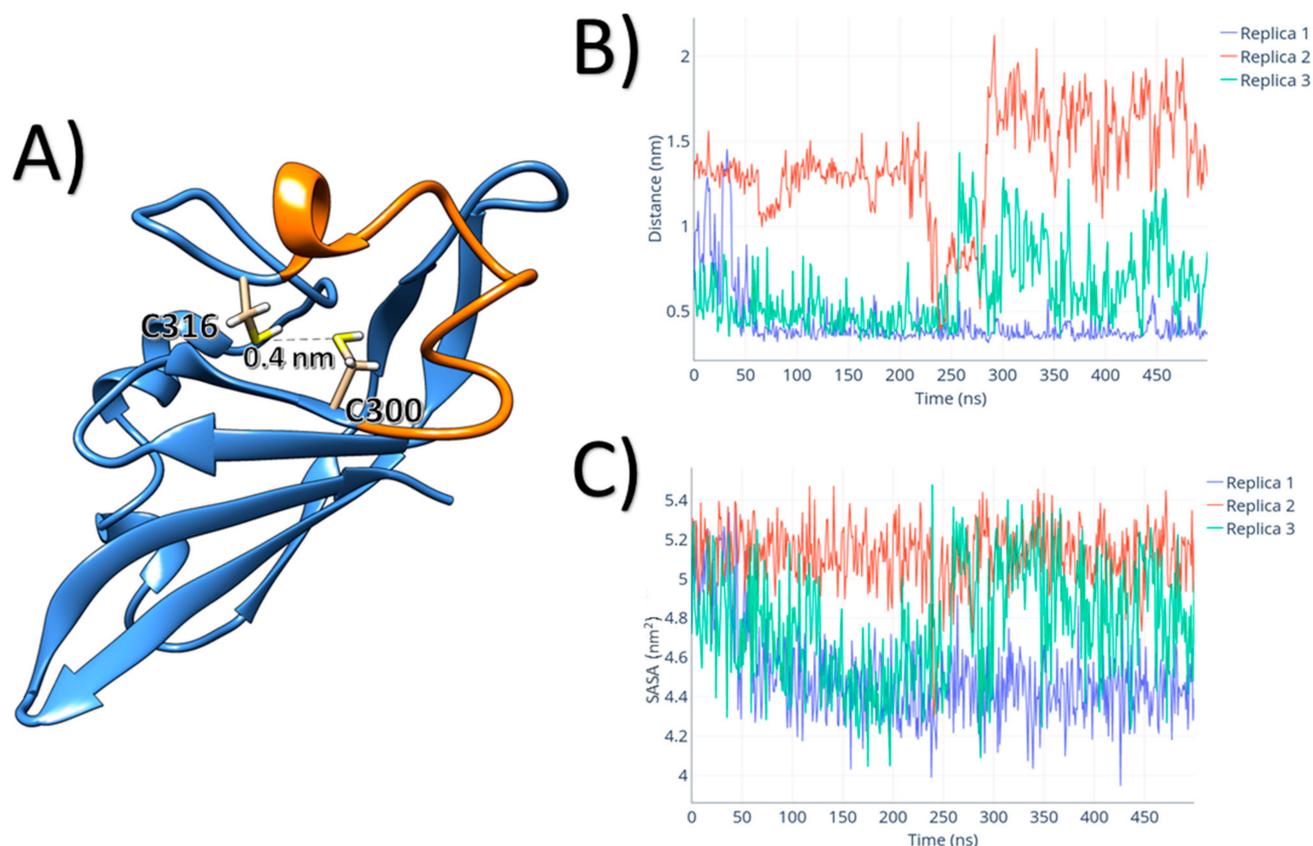


**Figure 6.** Molecular interactions of AhR PAS-B Y322 and *D. Suzukii* JHR F437. (A) AhR PAS-B Y322 interactions in the starting conformation; (B) AhR PAS-B Y322 interactions in the final conformation; (C) *D. Suzukii* JHR F437 interactions found on the starting conformation; (D) *D. Suzukii* JHR F437 interactions found on the final conformation. The size of the circle represents the amount of contact points between the central residue and the respective neighbour.

The Y322 counterparts within HIF1 $\alpha$  (Y278), HIF2 $\alpha$  (Y276) and NCOA1 (Y297) are less flexible than AhR. The Y322 average RMSF is 0.3 nm compared to 0.14 nm for NCOA1, 0.2 for HIF1 $\alpha$ , and 0.16 nm from HIF2 $\alpha$ . The main factor that changes the fluctuation on these residues in contrast to AhR is the microenvironment. NCOA1 Y297 is located in helix 2, which is non-existent in all the other four models. HIF1 $\alpha$  Y278 stability comes from the fact that it remains buried within the whole simulation structure (solvent-accessible surface area (SASA) =  $0.1 \pm 0.1$  nm<sup>2</sup>). Finally, the stability of HIF2 $\alpha$  Y278 is explained by the backbone hydrogen bond with the buried Y281 (Figure S5 (Supplementary Materials)).

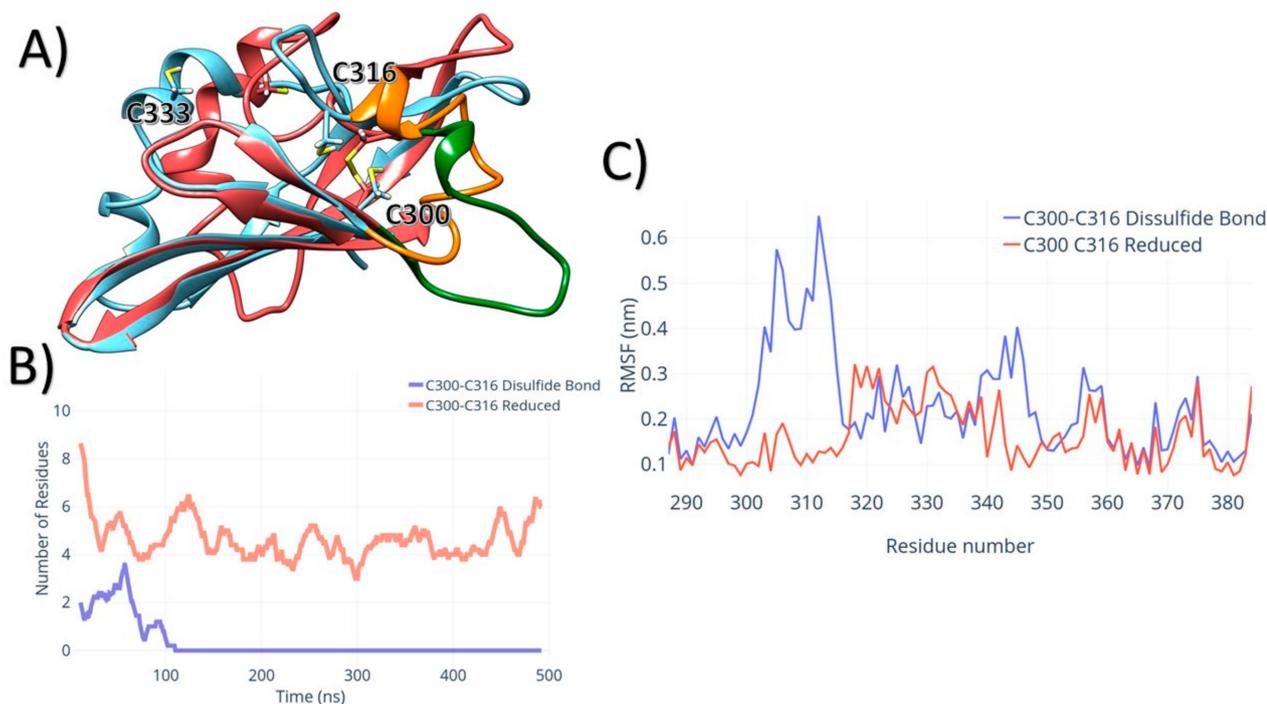
AhR PAS-B contains a unique cysteine in position 316, which is not found in any other PAS-B domain studied in this work. In the starting homology model, the distance between C316 S<sub>thiol</sub> and the thiol group of another, more conserved cysteine C300 (S<sub>thiol</sub>) was 0.35 nm. Several conformers sampled throughout the replicas had those residues close (distances below 0.4 nm), without any restraints. As shown in Figure 7, one replica was able to sample approximately 450 ns with both thiol groups within an average S–S distance of 0.4 nm. To achieve a configuration with both residues in this distance, C300 and C316 were buried within loop 2. This short distance reduced the accessible area for both residues

from 5 nm<sup>2</sup> to an average 4.2 nm<sup>2</sup> for the replica. A third cysteine C333 is located on the loop 1 region, which should be close to the previously predicted areas to be a druggable binding site [9]. Cypreds were used to evaluate the reactivity and pKa of C300, C316 and C333. C300 and C316 obtained predicted pKa values of  $9 \pm 1$ , indicating none of these cysteines to be reactive with external molecules. However, C333, a cysteine located on the other helix 1 of the AhR PAS-B domain, has a predicted pKa of  $8 \pm 0.5$ , which indicates a slightly higher predicted reactivity than its other two counterparts.



**Figure 7.** (A) Average conformation of the AhR PAS-B, highlighted is the interatomic distance between C316 S<sub>thiol</sub>–C300 S<sub>thiol</sub>, with loop 2 shown in orange; (B) C316 S<sub>thiol</sub>–C300 S<sub>thiol</sub> distances vs. Time for all three replicas; (C) Solvent-accessible surface area (SASA) vs. Time for all three AhR PAS-B replicas.

For a further study of the dynamic effect of the disulfide bond formation between C300 and C316, both residues were connected, as shown in Figure 8, and subsequently simulated for 500 ns in triplicate. The region comprised between C300–C316, which encapsulates loop 2 and a great part of loop 1, has a significant difference in dynamics regarding its spatial fluctuation, as shown in Figure 8C. The average RMSF of the region for the disulfide bond model is 0.5 nm, in contrast to 0.15 nm for the reduced system. This difference comes from the transition between a partially  $\alpha$ -helical structural to a fully disordered structure. As shown in Figure 8B, the average number of residues characterised as  $\alpha$ -helical decreases to 0 through the disulfide bond trajectory. In contrast, the total starting number of residues considered as  $\alpha$ -helical is eight, and it steadily decreases in time, stabilising in an average value of five.



**Figure 8.** (A) Overlaid average conformations of AhR PAS-B Domain: In red, the disulfide-bonded model with its C300-C316 loop highlight in green; In Blue, the reduced model with its C300-C316 loop highlighted in orange; (B) Number of residues in an  $\alpha$ -Helical conformation vs Time for C300-C316 loop. (C) Root mean square fluctuation for both bonded and reduced AhR PAS-B domains.

#### 4. Discussion

Using our previously developed homology model of the AhR PAS-B domain, we outlined properties of interest regarding its stability and intrinsic dynamics. Primarily, we studied the effect that different AhR PAS-B termini constructs had on its stability. Secondly, we compared AhR PAS-B dynamics with several different PAS-B domains over a longer timescale. Finally, we studied and reported how AhR PAS-B dynamics are affected by different states of its cysteines.

The termini directly affect the loop 1 dynamics, and hence, should be essential to control heterodimerisation. As previously shown by Soshilov and colleagues [15], the N and C termini are critical for the normal function and its resulting AhR DNA transformation. Hence, the heterodimer comprising AhR and hsp90 appears to be critical not only for the AhR cytosolic survival but also for its translocation to the nucleus. This overall process of the AhR PAS-B translocation and function remains elusive; however, several works have indicated that the opposite interface of loop 1 is the area which modulates the hsp90–PAS-B heterodimerisation [15,43,44]. As shown in this work, there is a direct relationship in dynamics between the termini and loop 1 interfaces.

The results found in this work shows that the length of these termini may directly affect not only AhR PAS-B stability in solution but also may also help to regulate its heterodimerisation and its function. The truncation of the termini directly affected the dynamics of the ligand binding region. The covariance matrices showed a direct relationship between the dynamics of loop 1 and the termini. Therefore, ligand binding to the PAS-B domain, modulated by loop 1 [9], also affects the dimerisation interface to Hsp90. This relationship is shown by how the motions of loop 1 directly covariates with the dimerisation interface. The correlated effect between loop 1 ligand binding and the dimerisation interface can explain how the hsp90–AhR dimerisation is signalled to occur.

The termini's length also plays a role in applying an entropic force to modulate and stabilise the binding regions. Some studies have shown that free disordered loops affect the dynamics through the means of an entropic force [28]: the existence of a disorder

region, regardless of its residual composition. This idea of “entropic rectifier” can be applied to the termini of AhR PAS-B to be one of the two factors which can explain the difference in dynamics for the binding region. As shown by our simulations, the residual fluctuation of loop 1 is lower for the canonical PAS-B model than its counterparts. The higher fluctuation on its termini can balance out the entropic cost for the lower RMSF. This characteristic should also have an opposite effect after the heterodimerisation with Hsp90, because its interconnecting spring between PAS-A and PAS-B would be rigidified, affecting the structural entropy of the molecule [16]. This entropic effect should also modify the internal dynamics of other domain that comprise the AhR, such as the PAS-A domain. As previously discussed in De Souza et al. [9], The AhR PAS-A should be able to bind small molecules, and future research will clarify whether the termini length affects its dynamics and ability to bind small molecules. The second reason for the related motion between this to the areas is the chain of contacts and the cleft formed, which is substantially more extensive on the NC-Trunc model.

AhR PAS-B has been known to be a challenging protein to be studied in vitro. As previously discussed, it requires a dimerisation partner for a longer cytosolic half-life. The modelling of the N and C termini may be used on an approach of generating a rationale for in vitro stabilisation of the AhR PAS-B domain. Our models predicted that truncating the termini may improve the propensity of aggregation. The main beneficial effect comes from the fact that truncating the N-terminus creates a partial cleft that buries the hydrophobic trine located between residues 307–309. Interestingly, the canonical AhR PAS-B has shown a highly exposed hydrophobic patch located in one of the strands of the central  $\beta$ -Sheet. Burying this patch within Hsp90 may be one factor to stabilise dimerisation, not only the polar interactions created with the N-terminal segment, as previous works suggested [15].

From the five PAS-B molecules simulated for 500 ns, AhR and NCOA1 PAS-B domains are the ones with the most mobile regions. Specifically for the NCOA1 PAS-B, the molecular structures used in this work were obtained via solution NMR bound to a STAT6 LXXLL motif [23]. Hence, the loop 1 region from NCOA1 was structured in a different conformation, showing a structured  $\alpha$ -loop- $\alpha$  conformation compared to a more unstructured loop found on HIF2 based models. Nonetheless, in our apo simulations, residues 280 to 282 on NCOA1 changed from a loop to a short  $\alpha$ -helical configuration, which showed a high motion to acquire said conformation to maintain the binding area for STAT6. This motion is different from the one shown in the AhR PAS-B.

The conservation of the “gatekeeper” tyrosine residue through the PAS-B sequences is not required to maintain a stable configuration. The *D. Suzukii* JHR PAS-B, although highly conserved in its entirety, has the tyrosine replace phenylalanine. However, this mutation did not increase the fluctuations within the loop. The change from an aromatic polar residue to a hydrophobic residue mainly affected which residues will interact with *D. Suzukii* JHR PAS-B F437 compared to AhR PAS-B Y322. This mutation indicates a change in dimerisation partner and function, enabling a novel mechanism to target the *D. Suzukii* life cycle.

The motions related to the three cysteines in AhR indicate their importance to the proper function of AhR molecule binding mechanisms. The residues located between CYS300 and CYS316 showed a mechanical transition, which allowed for several conformations sampled with both of these cysteines within a distance to form a disulfide bond. Simulations of the bonded cysteines resulted in higher local mobility for loop 1, which is directly related to the unfolding of the C300–C316 loop. The dynamics of loop 1 on AhR PAS-B domain are directly related to the ligand binding event; therefore, the formation of this disulfide bridge may be relevant for the AhR biological function. Additionally, C333 is located within reach of the previously proposed binding site located within the central cavity of the PAS-B. C333 is unique in comparison to its other four counterparts studied in this work, and covalently targeting this residue may lead to the development of specific AhR PAS-B inhibitors.

## 5. Conclusions

In conclusion, this work shows how the different characteristics of AhR PAS-B affect its dynamics and how it compares to four other counterparts. The data shown here for AhR indicate that modulating the lengths of the C and N termini may result in favourable conditions for in vitro studies (expression of the recombinant protein for biophysical binding assays and structural biology studies). The termini control also indicates that the entropic effects within this domain might be far from negligible and could lead to a better understanding of the mechanisms which regulate PAS-B heterodimerisation events. The studies of other PAS-B domains over a longer timescale show structural changes which can be crucial for its functions, alongside the redox state, which could couple the oxidation of the cysteines, the intrinsic dynamics, protein stability, and its biological activity. Generating mutants of these cysteine residues may shed light on the role of these residues in ligand binding and protein stability.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2673-4125/1/1/1/s1>.

**Author Contributions:** J.V.d.S., S.R. and A.K.B. designed the study. J.V.d.S., S.R., R.Z. and P.Z. carried out the equilibrium molecular dynamics simulations. J.V.d.S., S.R. and P.Z. and M.K. extracted and analysed data from the MD simulations. M.K. performed the cysteine analysis. A.K.B. supervised this work and the project. J.V.d.S. produced the figures and plots and wrote the first draft of the manuscript. All authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the School of Natural and Environmental Sciences (scholarship to J.V.d.S.) and by the Faculty of Science, Agriculture, and Engineering, Newcastle University (scholarship to S.R.). IAFRI-FERA scholarship to P.Z., EPSRC-AstraZeneca iCASE scholarship to M.K., and EPSRC funding to A.K.B. (EP/S022791/1).

**Data Availability Statement:** Data available on request.

**Acknowledgments:** We are grateful to M. Garner and K. Bower for technical assistance.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Feng, Z.; Zou, X.; Chen, Y.; Wang, H.; Duan, Y.; Bruick, R.K. Modulation of HIF-2 $\alpha$  PAS-B domain contributes to physiological responses. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 13240–13245. [[CrossRef](#)] [[PubMed](#)]
2. Depping, R.; Oster, H. Interplay between environmentally modulated feedback loops—hypoxia and circadian rhythms—two sides of the same coin? *FEBS J.* **2017**, *284*, 3801–3803. [[CrossRef](#)] [[PubMed](#)]
3. Culig, Z.; Santer, F.R. Studies on Steroid Receptor Coactivators in Prostate Cancer. In *Methods in Molecular Biology*; Humana Press Inc.: New York, NY, USA, 2018; Volume 1786, pp. 259–262.
4. Zhang, Y.; Markert, M.J.; Groves, S.C.; Hardin, P.E.; Merlin, C. Vertebrate-like CRYPTOCHROME 2 from monarch regulates circadian transcription via independent repression of CLOCK and BMAL1 activity. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7516–E7525. [[CrossRef](#)] [[PubMed](#)]
5. Rojas-Pirela, M.; Rigden, D.J.; Michels, P.A.; Cáceres, A.J.; Concepción, J.L.; Quiñones, W. Structure and function of Per-ARNT-Sim domains and their possible role in the life-cycle biology of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **2018**, *219*, 52–66. [[CrossRef](#)]
6. Hartzell, A.L.; Martyniuk, K.M.; Brigidi, G.S.; Heinz, D.A.; Djaja, N.A.; Payne, A.; Bloodgood, B.L. NPAS4 recruits CCK basket cell synapses and enhances cannabinoid-sensitive inhibition in the mouse hippocampus. *Elife* **2018**, *7*. [[CrossRef](#)]
7. Harmon, A.C.; Hebert, V.Y.; Cormier, S.A.; Subramanian, B.; Reed, J.R.; Backes, W.L.; Dugas, T.R. Particulate matter containing environmentally persistent free radicals induces AhR-dependent cytokine and reactive oxygen species production in human bronchial epithelial cells. *PLoS ONE* **2018**, *13*. [[CrossRef](#)]
8. Wu, D.; Su, X.; Potluri, N.; Kim, Y.; Rastinejad, F. NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. *Elife* **2016**, *5*. [[CrossRef](#)]
9. De Souza, J.V.; Reznikov, S.; Zhu, R.; Bronowska, A.K. Druggability assessment of mammalian Per-Arnt-Sim [PAS] domains using computational approaches. *Medchemcomm* **2019**, *10*, 1126–1137. [[CrossRef](#)]
10. Weng, F.J.; Garcia, R.I.; Lutz, S.; Alviña, K.; Zhang, Y.; Dushko, M.; Ku, T.; Zemoura, K.; Rich, D.; Garcia-Dominguez, D.; et al. Npas4 Is a Critical Regulator of Learning-Induced Plasticity at Mossy Fiber-CA3 Synapses during Contextual Memory Formation. *Neuron* **2018**, *97*, 1137–1152.e5. [[CrossRef](#)]

11. Shinde, R.; McGaha, T.L. The Aryl Hydrocarbon Receptor: Connecting Immunity to the Microenvironment. *Trends Immunol.* **2018**, *39*, 1005–1020. [[CrossRef](#)]
12. Adesso, S.; Paterniti, I.; Cuzzocrea, S.; Fujioka, M.; Autore, G.; Magnus, T.; Pinto, A.; Marzocco, S. AST-120 Reduces Neuroinflammation Induced by Indoxyl Sulfate in Glial Cells. *J. Clin. Med.* **2018**, *7*, 365. [[CrossRef](#)] [[PubMed](#)]
13. Adesso, S.; Magnus, T.; Cuzzocrea, S.; Campolo, M.; Rissiek, B.; Paciello, O.; Autore, G.; Pinto, A.; Marzocco, S. Indoxyl sulfate affects glial function increasing oxidative stress and neuroinflammation in chronic kidney disease: Interaction between astrocytes and microglia. *Front. Pharmacol.* **2017**, *8*. [[CrossRef](#)] [[PubMed](#)]
14. Soshilov, A.A.; Denison, M.S. Ligand Promiscuity of Aryl Hydrocarbon Receptor Agonists and Antagonists Revealed by Site-Directed Mutagenesis. *Mol. Cell. Biol.* **2014**, *34*, 1707–1719. [[CrossRef](#)] [[PubMed](#)]
15. Soshilov, A.A.; Motta, S.; Bonati, L.; Denison, M.S. Transitional States in Ligand-Dependent Transformation of the Aryl Hydrocarbon Receptor into Its DNA-Binding Form. *Int. J. Mol. Sci.* **2020**, *21*, 2474. [[CrossRef](#)] [[PubMed](#)]
16. Kolonko, M.; Greb-Markiewicz, B. BHLH–PAS proteins: Their structure and intrinsic disorder. *Int. J. Mol. Sci.* **2019**, *20*, 3653. [[CrossRef](#)]
17. Partch, C.L.; Gardner, K.H. Coactivators necessary for transcriptional output of the hypoxia inducible factor, HIF, are directly recruited by ARNT PAS-B. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7739–7744. [[CrossRef](#)]
18. Hirano, M.; Toyota, K.; Ishibashi, H.; Tominaga, N.; Sato, T.; Tatarazako, N.; Iguchi, T. Molecular Insights into Structural and Ligand Binding Features of Methoprene-Tolerant in Daphnids. *Chem. Res. Toxicol.* **2020**. [[CrossRef](#)]
19. Charles, J.P.; Iwema, T.; Epa, V.C.; Takaki, K.; Rynes, J.; Jindra, M. Ligand-binding properties of a juvenile hormone receptor, Methoprene-tolerant. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 21128–21133. [[CrossRef](#)]
20. Dahm, K.H.; Bhaskaran, G.; Peter, M.G.; Shirk, P.D.; Seshan, K.R.; Röller, H. On the Identity of the Juvenile Hormone in Insects. In *The Juvenile Hormones*; Springer: New York, NY, USA, 1976; pp. 19–47.
21. Jindra, M.; Uhlířová, M.; Charles, J.-P.; Smykal, V.; Hill, R.J. Genetic Evidence for Function of the bHLH-PAS Protein Gce/Met As a Juvenile Hormone Receptor. *PLoS Genet.* **2015**, *11*, e1005394. [[CrossRef](#)]
22. Cho, H.; Du, X.; Rizzi, J.P.; Liberzon, E.; Chakraborty, A.A.; Gao, W.; Carvo, I.; Signoretti, S.; Bruick, R.K.; Josey, J.A.; et al. On-target efficacy of a HIF-2 $\alpha$  antagonist in preclinical kidney cancer models. *Nature* **2016**, *539*, 107–111. [[CrossRef](#)]
23. Russo, L.; Giller, K.; Pfitzner, E.; Griesinger, C.; Becker, S. Insight into the molecular recognition mechanism of the coactivator NCoA1 by STAT6. *Sci. Rep.* **2017**, *7*. [[CrossRef](#)] [[PubMed](#)]
24. Bell, D.R.; Poland, A. Binding of aryl hydrocarbon receptor (AhR) to AhR-interacting protein: The role of hsp90. *J. Biol. Chem.* **2000**, *275*, 36407–36414. [[CrossRef](#)]
25. Tsuji, N.; Fukuda, K.; Nagata, Y.; Okada, H.; Haga, A.; Hatakeyama, S.; Yoshida, S.; Okamoto, T.; Hosaka, M.; Sekine, K.; et al. The activation mechanism of the aryl hydrocarbon receptor (AhR) by molecular chaperone HSP90. *FEBS Open Bio* **2014**, *4*, 796–803. [[CrossRef](#)] [[PubMed](#)]
26. Henry, E.C.; Gasiewicz, T.A. Transformation of the aryl hydrocarbon receptor to a DNA-binding form is accompanied by release of the 90 kDa heat-shock protein and increased affinity for 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Biochem. J.* **1993**, *294*, 95–101. [[CrossRef](#)] [[PubMed](#)]
27. Fukunaga, B.N.; Probst, M.R.; Reisz-Porszasz, S.; Hankinson, O. Identification of functional domains of the aryl hydrocarbon receptor. *J. Biol. Chem.* **1995**, *270*, 29270–29278. [[CrossRef](#)] [[PubMed](#)]
28. Keul, N.D.; Oruganty, K.; Bergman, E.T.S.; Beattie, N.R.; McDonald, W.E.; Kadirvelraj, R.; Gross, M.L.; Phillips, R.S.; Harvey, S.C.; Wood, Z.A. The entropic force generated by intrinsically disordered segments tunes protein function. *Nature* **2018**, *563*, 584–588. [[CrossRef](#)]
29. Jin, H.; Ji, C.; Ren, F.; Aniagu, S.; Tong, J.; Jiang, Y.; Chen, T. AHR-mediated oxidative stress contributes to the cardiac developmental toxicity of trichloroethylene in zebrafish embryos. *J. Hazard. Mater.* **2020**, *385*. [[CrossRef](#)] [[PubMed](#)]
30. Omid, M.; Ghafarian-Bahraman, A.; Mohammadi-Bardbori, A. GSH/GSSG redox couple plays central role in aryl hydrocarbon receptor-dependent modulation of cytochrome P450 1A1. *J. Biochem. Mol. Toxicol.* **2018**, *32*. [[CrossRef](#)] [[PubMed](#)]
31. Kubli, S.P.; Bassi, C.; Roux, C.; Wakeham, A.; Göbl, C.; Zhou, W.; Jafari, S.M.; Snow, B.; Jones, L.; Palomero, L.; et al. AhR controls redox homeostasis and shapes the tumor microenvironment in BRCA1-associated breast cancer. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 3604–3613. [[CrossRef](#)] [[PubMed](#)]
32. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Pyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)]
33. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. [[CrossRef](#)]
34. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 1950–1958. [[CrossRef](#)] [[PubMed](#)]
35. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [[CrossRef](#)]
36. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [[CrossRef](#)] [[PubMed](#)]
37. Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122. [[CrossRef](#)]

38. Parrinello, M.; Rahman, A. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.* **1980**, *45*, 1196–1199. [[CrossRef](#)]
39. Zambrano, R.; Jamroz, M.; Szczasiuk, A.; Pujols, J.; Kmiecik, S.; Ventura, S. AGGRESCAN3D (A3D): Server for prediction of aggregation properties of protein structures. *Nucleic Acids Res.* **2015**, *43*, W306–W313. [[CrossRef](#)]
40. Kuriata, A.; Iglesias, V.; Pujols, J.; Kurcinski, M.; Kmiecik, S.; Ventura, S. Aggrescan3D (A3D) 2.0: Prediction and engineering of protein solubility. *Nucleic Acids Res.* **2019**, *47*, W300–W307. [[CrossRef](#)]
41. Kayikci, M.; Venkatakrishnan, A.J.; Scott-Brown, J.; Ravarani, C.N.J.; Flock, T.; Babu, M.M. Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. *Nat. Struct. Mol. Biol.* **2018**, *25*, 185–194. [[CrossRef](#)]
42. Soylu, İ.; Marino, S.M. *Cy-preds*: An algorithm and a web service for the analysis and prediction of cysteine reactivity. *Proteins Struct. Funct. Bioinform.* **2016**, *84*, 278–291. [[CrossRef](#)]
43. Soshilov, A.; Denison, M.S. Role of the Per/Arnt/Sim domains in ligand-dependent transformation of the aryl hydrocarbon receptor. *J. Biol. Chem.* **2008**, *283*, 32995–33005. [[CrossRef](#)] [[PubMed](#)]
44. Corrada, D.; Denison, M.S.; Bonati, L. Structural modeling of the AhR:ARNT complex in the bHLH-PASA-PASB region elucidates the key determinants of dimerization. *Mol. Biosyst.* **2017**, *13*, 981–990. [[CrossRef](#)] [[PubMed](#)]