RESEARCH ARTICLE

[View Article Online](#)
[View Journal](#) | [View Issue](#)

Cite this: *Med. Chem. Commun.*,
2019, 10, 1126

Druggability assessment of mammalian Per-Arnt-Sim [PAS] domains using computational approaches†

João V. de Souza, Sylvia Reznikov, Ruidi Zhu and Agnieszka K. Bronowska *

Per-Arnt-Sim (PAS) domains are key regions that occur in different regulatory proteins from all kingdoms of life. PAS domains show a remarkably conserved structural scaffold, despite a highly variable primary sequence. In this study we have attempted to address some of the gaps in knowledge regarding the druggability of PAS-A domains, differences in structure and dynamics within the PAS domain family and how this affects the druggability potential, as well as give insight into the druggability of steroid receptor coactivators and putative binding modes of the NCOA1. Investigations were performed through a range of computational methods including molecular docking studies, atomistic molecular dynamics simulations, and hotspot mapping. Atomistic molecular dynamics simulations show that the function of the AhR PAS-B domain is regulated by the dynamics of the highly conserved tyrosine Y322 residue, which acts as a “gate-keeper” controlling the access to the binding cavity and finely tuning the binding affinity. Furthermore, the transition between the partially unfolded and helical conformation of the loop1 segment within PAS-B domains was shown to be essential for the generation of “druggable” sites, especially for the NCOA1 PAS-B domain. Finally, our simulations indicated the undruggability of PAS-A domains, caused by the inherent characteristics of their putative binding sites. In conclusion, this work emphasises the role of intrinsic dynamics in tuning the druggability of PAS-B domains and shows that PAS-B domains of steroid receptor coactivators, such as NCOA1, can be targeted by small molecule ligands, which highlights the potential of developing new therapeutics designed to target these coactivators using structure-based approaches.

Received 9th March 2019,
Accepted 1st May 2019

DOI: 10.1039/c9md00148d

rsc.li/medchemcomm

Introduction

Per-Arnt-Sim (PAS) domains occur in proteins from all kingdoms of life. In animals, proteins containing PAS domains are regulating responses to hypoxia,^{1,2} circadian rhythms,^{2–4} hormonal stimuli,⁵ synaptic plasticity,⁶ memory,⁷ and xenobiotic stress,⁸ exerting their activity as transcription factors and nuclear receptor coactivators. The PAS domain is also a part of the hERG potassium channel.⁹

Despite a highly variable amino acid sequence, PAS domains have a well-conserved fold with a central cavity, able to accommodate small molecule ligands. Combined with their roles in important physiological and pathophysiological processes, including tumorigenesis,¹⁰ inflammatory responses,¹¹ memory control,^{7,12} and possibly neurodegeneration,^{13,14} these make proteins containing PAS domains attractive targets for new therapeutics and tool compounds for chemical biology studies. However, PAS domains are considered chal-

lenging for drug development and a lot of aspects of PAS domain druggability by small molecule ligands and mechanisms governing their activation and inhibition by small molecules at an atomistic level of detail remain obscure.

Certain PAS domains, such as the PAS-B domain of the aryl hydrocarbon receptor (AhR), are defined through their ability to bind small molecule ligands. There are, however, proteins containing PAS domains for which no small molecule ligands have been identified to date and which are considered orphan receptors (*e.g.* Npas4). Even if the ligand binding is confirmed and essential for the biological activity of the protein containing the PAS domain, the atomistic detail of such binding may be unclear – as it is in the case of the AhR receptor.¹⁵ In addition, some PAS domains are most likely to play a mechanistic role in modulating protein–protein interactions, decoupled from the ligand binding.^{16,17}

PAS domains are important constituents of basic helix–loop–helix (bHLH)–PAS transcription factors. These transcription factors share common architectural features that include a bHLH DNA-binding domain, followed by tandem PAS domains, PAS-A and PAS-B, and a highly flexible C-terminal region containing transactivation (TAD) or trans-repression (TRD) domains.¹⁸ Class I includes the three hypoxia-

School of Natural and Environmental Sciences, Newcastle University, NE1 7RU
Newcastle, UK. E-mail: agnieszka.bronowska@ncl.ac.uk

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9md00148d

inducible factor (HIF)- α proteins (HIF-1 α , HIF-2 α and HIF-3 α), four neuronal PAS domain proteins (NPAS1–4), the aryl hydrocarbon receptor (AhR), the AhR repressor (AhRR), single-minded proteins (SIM1 and SIM2) and the clock circadian regulator (CLOCK). Npas2 is unique among class I members, as it binds heme as a cofactor.¹⁹ Class II includes the aryl hydrocarbon receptor nuclear translocator (ARNT), ARNT2, brain and muscle ARNT-like protein 1 (BMAL1) and BMAL2. Heterodimerisation between class I and class II members, which is mediated by their PAS domains, generates functional transcription factors. Yet another class of proteins containing PAS domains, denoted as p160, includes nuclear receptor coactivators (NCOA1–3), wherein dimerisation occurs between the PAS domain of one partner and the C-terminal activation domain of another partner.²⁰ In either case, small molecule ligand binding, which occurs at the PAS domains, may interfere with this process.^{18,21} This puts PAS domains at the focal point of target validation and structure-based drug design efforts.

Recent structural biology studies brought more than fifty of the experimentally solved structures of PAS domains, in complex with their small molecule ligands and protein interactors, or in their *apo* (unbound) state. Collectively, these studies show that PAS domains adopt strikingly similar α/β folds despite less than 20% average sequence identity among individual domains.²⁰ These data allow for an insight into the dimerization of class I and class II members, PAS domains interacting with other proteins, small molecule ligand binding modes, and conformational heterogeneity of the PAS domains.

In the present study, we have attempted to address some existing gaps in the understanding of the “druggability” of the PAS domains. While experimental structures exist for both PAS domains of Npas proteins (PDB: 5SY5 and 5SY7) and HIF transcription factors bound to small molecule ligands, there is no experimental structure existing for the PAS-B domain of the AhR receptor, which is a crucial ligand binding domain. Existing experimental structures (X-ray and NMR) allow for an insight into interactions between the NCOA1 and LXXLL peptide of STAT6, but its “druggability” by small molecules remains unreported. Also, despite experimental structures of several PAS-A domains being published and earlier reports on targeting the PAS-A domain of HIF-1 α with small molecule ligands,²² the “druggability” of PAS-A domains has received very little attention.

Therefore, in this work we have developed a model of the PAS-B domain of human AhR and studied its dynamics on the microsecond time scale and also compared its “druggability” (e.g. ability to bind small molecule ligands) with several class I members, for which the experimental structures have been reported. Analysis of our data has highlighted the highly conserved tyrosine residue, which acts as a “gatekeeper” of the domain, and suggests its important role in the ligand entry into the binding cavity. Our primary focus was on the interlinking structure, dynamics, and propensity to bind small molecule ligands. We then performed

the molecular dynamics simulation for NCOA1, and for the first time we have observed the propensity of loop1 of the PAS-B domain of NCOA1 to undergo substantial conformational changes on the nanosecond time scale and adopt helical or partially unfolded conformations. This behaviour has implications for the modulation of NCOA1 and related coactivators by small molecule ligands, and we have suggested a binding mode for several confirmed NCOA1 binders. Finally, by molecular modelling coupled with atomistic MD simulations and an ultrafast “hotspot” mapping approach, we have evaluated the reason for druggability differences between PAS-A and PAS-B domains.

Experimental

Molecular modelling

The initial conformation of the human AhR PAS-B model was created by homology modelling using the Phyre2 web server²³ using the HIF-2 α crystal structure as a template (PDB code: 5UFP). The human Npas4 model has been obtained by the same procedure, using the Npas1 crystal structure as a template (PDB code: 5SY5). Both models have been checked for correctness, having any missing residues modelled using the MODELLER interface in UCSF Chimera.²⁴

For the PAS-A domain of AhR, we used the experimental structure of the domain core (PDB code: 5NJ8). Missing loops 229–231 and 251–261 were modelled and fitted using the MODELLER interface in UCSF Chimera.²⁴ The loop 173–211 has been modelled *ab initio* using i-TASSER.²⁵ The top-five models have been visually inspected, and the lowest-energy model with the conformation of the termini fitting the restraints provided by the PAS-A domain core has been selected for further studies. The modelled loop contains two cysteine residues, which might be involved in the disulfide bridge formation. Since very little is known about the redox state of these cysteines, we have decided to model both states (oxidised and reduced).

In the investigations regarding conformation transitions of loop1, alternate models of PAS-B domains of AhR, HIF-1 α , and HIF2 α , with loop1 adopting the helical conformation, were generated using the human NCOA1 NMR structure as a template (PDB code: 5NWM). The alternate model of NCOA1, with loop1 adopting the partially unfolded conformation, has been generated using the human HIF-2 α crystal structure as a template (PDB code: 5UFP). These models have been generated by SWISS-MODEL.²⁶

Equilibrium molecular dynamics (MD) simulations

All simulations were performed using GROMACS 2016.1.²⁷ The AhR simulations for druggability studies were parametrised using the AMBER99SB-ILDN force field, with the TIP3P water model.²⁸ The box distance was set to 1 nm and periodic boundary conditions were applied. The box was solvated and Na⁺ and Cl[−] ions were added to achieve a 0.1 M neutral concentration. The solvated systems were energy minimised and equilibrated. The minimisation was run

using steepest descent for 1000 cycles followed by the conjugate gradient. The energy step size was set to 0.001 nm and the maximum number of steps was set to 50 000. 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),²⁹ and the short range electrostatic and van der Waals cutoff set to 1.0 nm. After the energy minimisation, heating to 300 K was performed for 20 ps with a time step of 2 fs and position restraints applied to the backbone in an NVT ensemble. The constraint algorithm used was LINCS, which was applied to all bonds and angles in the protein.³⁰ The cut-off for non-bonded short range interaction was set to 1.0 nm with the Verlet cut-off scheme. Long range electrostatics were again set to PME. The temperature coupling was set between the protein and the non-protein entities by using a Berendsen thermostat, with a time constant of 0.1 ps and the temperature set to reach 300 K with the pressure coupling off. Pressure equilibration was run at 300 K with Parrinello–Rahman pressure coupling on and set to 1 bar (ref. 31) in a NPT ensemble. The equilibration trajectories were set to 10 ns (discarded from the analysis), and the production MD simulations were performed for 3 replicas of 100 ns (AhR PAS-A, HIF-1 α , and HIF-2 α). For AhR PAS-B, the production run has been extended to 1.5 μ s.

For any simulations involving small molecule ligands, the ligands were parametrised using ACPYPE.³² Partial atomic charges have been calculated using the AM1-BCC level of theory³³ using the DockPrep tool in Chimera.³⁴ The equilibration runs were the same as for the *apo* systems. The production trajectories for protein–ligand simulations have been set to 100 ns.

Analysis of the trajectories was performed using GROMACS tools, including root-mean square deviation (RMSD) to assess overall stability, per-residue root-mean square fluctuation (RMSF) to assess the local flexibility, solvent-accessible surface areas, dihedral angles, and principal component analysis (PCA).

Molecular docking and “hot spot” mapping

Tools for predicting binding sites are known to reliably predict novel binding areas.³⁵ The mapping of potential druggable “hot spots” was performed using FTMap³⁶ and Cryptosite³⁷ for a more reliable mapping. Protein–protein and protein–peptide docking has been performed using the ClusPro³⁸ web server.

Molecular docking was performed using the University of California, San Francisco DOCK 6.8 suite³⁹ with default grid scoring. The highest-populated structure has been selected from the MD trajectory and assigned as a receptor. The grid spacing was 0.25 Å, and the grid included 12 Å beyond the geometric centre of the PAS-B inner cavity. The energy score was the sum of electrostatic and van der Waals contributions. To check the suitability of the methodology, the docking has been run for HIF-2 α known binders, where the ligand mole-

cule has been re-docked to the protein, to check whether the docking procedure reproduced the native binding modes, as observed in the crystal structures. After the positive verification, the ligands have been docked to both AhR PAS-B and NCOA1 PAS-B sites detected by FTMap.

During the docking calculations, the ligands were subjected to 5000 cycles of molecular mechanics energy minimisation. The number of maximum ligand orientations was 50 000. Solvent effects were modelled by implicit solvation (distance-dependent dielectric function). The 25 best-scoring poses (ligand–protein complexes) were further analysed by means of secondary rescoring using the SeeSAR⁴⁰ package with a more accurate HYDE scoring function.⁴¹ The best-scoring poses were subjected to all-atom MD simulations and the energetics has been assessed by PCA analysis, MM-PBSA and umbrella sampling.

For the MMPBSA end-state interaction energy calculation, the calculation was divided into 3 stages. First, the interaction energy term was calculated using the parameters used in the force-field for both the ligand and receptor. For the polar solvation term, the solute dielectric constant was set to 2, with a solvent dielectric constant of 80. The solvent probe radius was set to 1.4, with a density of grid points per Å² of 10, using the linear Poisson–Boltzmann equation. For the apolar solvation term, we used a SASA model with a surface tension of 0.0226 kJ mol^{−1} Å^{−2} and an offset constant of 3.849 kJ mol^{−1}. All calculations were performed at a temperature of 300 K. These calculations were made using the Gromacs *g_mmpbsa*^{42,43} module.

Umbrella sampling of the NCOA1 five best-scoring molecules

To calculate the binding modes and their respective energetics more accurately, umbrella sampling simulations were performed for best-scoring ligands binding to the NCOA1 PAS-B domain. From the average configuration obtained from the 100 ns simulation of the equilibrium ligand–NCOA1 complex, an umbrella sampling run was made with a harmonic force applied over the ligand, pulling in the Z direction, as shown in Fig. 1, increasing the distance from the binding site. The force constant was set as 1000 kJ mol^{−1} nm^{−2} with a

Fig. 1 The black arrow indicates the direction the umbrella pathway took to calculate the affinity for NCOA1 ligands. The spheres represent the position of the centre of mass of the ligand throughout the pull simulation.

pulling rate of 0.01 nm ps⁻¹, resulting in 30 windows per complex. The systems were parametrised with AMBER99SB-ILDN using the TIP3P solvation model, the same as the equilibrium simulations. Each umbrella window had a restrained pressure equilibration of 10 ps and a 5 ns production window. The errors were calculated using a bootstrap methodology implemented in Gromacs.

Conformational transitions of loop1 within PAS-B domains

Four pairs of initial conformations of PAS-B domains were used in the analysis of the transitions between the loop and helical conformers of loop1 (Fig. 2). The experimental structures of PAS-B domains with the partially disordered loop1 conformation of HIF1- α and HIF2- α (PDB codes 4H6J and 5UFP, respectively) were matched with alternative structures, where loop1 adopted α -helical conformations. These structures have been generated by homology modelling, as described in the previous section. The structure of NCOA1 PAS-B with loop1 adopting a partially disordered conformation was modelled using the experimental structure of HIF2- α (5UFP) as the template. Two AhR PAS-B models were considered: one based on HIF-2 α (5UFP) and another one based on NCOA1 (5WNM) as templates, as described in the previous section. All four pairs of PAS-B conformations were parametrised using the AMBER-ff03ws⁴³ force field and simulated with the TIP4P-2005 (ref. 44) water model, to improve sampling on protein disordered regions. The production trajectories were 100 ns. The coordinates were collected every 10 ps.

Per-residue root mean-square fluctuations (RMSFs) were calculated to assess the flexibility per region. To calculate the structural entropy within the protein region of interest, we used the SQuE⁴⁵ approach recently developed by our group, which integrates the principal components of the trajectory and calculates the ensemble structural entropy difference between conformers.

Results and discussion

The molecular dynamics of highly conserved tyrosine Y322 controls the solvent accessibility of the PAS-B domain of the human aryl hydrocarbon receptor (AhR)

To probe the nanosecond-to-microsecond timescale dynamics of the PAS-B domain, we have simulated the PAS-B domains of HIF-1 α , HIF-2 α , and AhR for 100 ns, in the absence of ligands or any other interactors. We have observed that loop1 (Fig. 2A) was the most flexible region for AhR (Fig. 2B), with highly conserved tyrosine controlling the access to the inner cavity (Fig. 2C). Interestingly, this tyrosine residue was highly flexible in all three proteins (Fig. 2B).

To follow this on, we have extended the simulation for the AhR PAS-B domain to 1.5 μ s and monitored the behaviour of this tyrosine residue (Y322). The results suggest that the intrinsic dynamics of Y322 affects the access to the inner cavity of the PAS-B domain and suggests tyrosine as a major “gate-keeper” for the small ligands (Fig. 3).

These results are consistent with earlier reported data. Matthews and coworkers observed that Y322 has been

Fig. 2 A Loop1 (red) of the AhR PAS-B domain, with the conserved tyrosine Y322 displayed. B RMSF plot of PAS-B domains of AhR (black), HIF-1 α (blue), and HIF-2 α (green). The loop1 region is highlighted grey, and the conserved tyrosine residue is marked by an asterisk; on the right, the aligned structures with all 3 molecules aligned, the loop region is coloured by RMSF, with yellow denoting low values and violet showing higher values; the conserved tyrosine for all three chains are highlighted with asterisks. C Structural alignment of the investigated PAS-B domains, showing the position of the conserved tyrosine residue. AhR is coloured grey with loop1 coloured red, HIF-1 α is coloured blue, and HIF-2 α is coloured green. D Sequence alignment of segments corresponding to loop1 of the PAS-B domains studied herein. Highly conserved tyrosine, corresponding to Y322 of human AhR, is marked with a red asterisk. Conserved residues are coloured red; highly similar residues are coloured green.

Fig. 3 A and B Loop1 PAS-B domain of human AhR, viewed from the deduced ligand “entrance”. Positions of the Y322 side chain, corresponding to two highest-populated X–CA–CB–X dihedral angles, are shown. The Y322 side chain is marked as a filled aromatic ring. C Solvent-accessible surface area of AhR PAS-B during the simulation. Changes correspond to the cavity being solvent-exposed, which is driven by conformational changes of Y322. D Distribution of the X–CA–CB–X dihedral angle of Y322 during 1.5 μ s of MD simulation, shown as a histogram plot. Two major conformations of Y322 are shown in panels A and B. E Distribution of the X–CA–CB–X dihedral angle of Y322 during 1.5 μ s of MD simulation, shown as a function of the simulation time.

implicated in the activation of AhR by omeprazole (Omp).⁴⁶ They found that Y322 was important for maximum activation of human AhR induced by 2,3,7,8-TCDD and 2,3,4,7,8-PeCDF, but required for 2,3,7,8-TCDF and Omp in an AhR-deficient MCF-7 human breast cancer cell line. These results provided for the first time the evidence for an important role of Y322 in AhR activation, but without explaining the mechanism. The role of Y322 was followed up later by Hegedűs and coworkers, who studied the access path to the inner cavity of human AhR using a combination of molecular dynamics simulations and molecular docking.⁴⁷ They suggested that Y322 was among the major residues defining the binding pocket for TCDD and indicated that Y322 and H337 may play a pivotal role in the different binding of molecules with high and low affinity. These results are consistent with our findings, which further elucidate the role of Y322 not only as an affinity control but as the major gatekeeper. High conservation of this tyrosine among all the PAS domains we have investigated may indicate a universal mechanism of controlling the access to the binding cavity by this residue, as well as emphasise its important role in the receptor's function. This role is likely to be a result of the balance between enthalpic factors (interactions between Y322 and other residues, such as H337, H291, and F295) and the intrinsic dynamics of Y322 (Fig. 3).

“Druggability” of PAS-B domains probed by molecular docking and “hotspot” mapping

Recent experimental and computational studies of mammalian class I bHLH-PAS proteins suggest that all the members of this protein family are likely to be able to bind to small molecules, whether *via* their PAS domains or at the interfaces between the domains.¹⁸ However, large sequence variability within the PAS-B domain and observed differences in the volumes of the binding cavities indicate that those domains may not be equally “druggable”, and the binding modes for small molecule ligands are likely to be considerably different for these domains. To assess potential differences in druggability, we have performed a rapid FTMMap scan of the domains of interest and the analysis of potential drug sites using Cryptosite (Fig. 4). While experimental structures of ligand-bound HIF-1 α and HIF-2 α show different binding modes for these domains, all the information about the binding modes of AhR ligands comes from the molecular modelling data, and very little is known about small molecules binding to NCOAs.

FTMMap and Cryptosite predicted a small “hotspot” for NCOA1, as shown in Fig. 4A and B. Although the druggability of this conformation cannot be ruled out, the dynamics of loop1, which is the major interacting part for this hotspot will make targeting it with small molecules very challenging.

Fig. 4 Binding site scans of PAS-B domains of NCOA1/FTMAP (A), NCOA1/Cryptosite (B), AhR/FTMAP (C), AhR/Cryptosite (D), HIF-2a/FTMAP (E), and HIF-2a/Cryptosite (F). The largest cluster, suggesting the most likely “druggable hotspot”, is marked with a red circle for FTMAP and a green circle for Cryptosite.

Fig. 6 Chemical structures of five confirmed ligands of human NCOA1 used in this study.

Molecular docking of a small set of confirmed NCOA1 ligands (Fig. 6) failed to obtain a binding mode due to poor energetics (data not shown).

As expected, FTMap and Cryptosite predicted the largest “hotspot” for the AhR PAS-B domain (Fig. 4). Compared to the size of the “hotspots” detected for PAS-B domains of HIF-1 α and HIF-2 α (which were consistent with experimental structures), the binding site of AhR was predicted to be able to accommodate larger and chemically diverse ligands. This is consistent with the information on the chemical structure of known AhR ligands, which include large, rigid, and highly

hydrophobic polyaromatic hydrocarbons,⁴⁸ endogenous metabolites such as kynurenine,⁴⁹ polyphenols,⁵⁰ and a lot of chemically diverse approved drug molecules⁵¹ and natural products.⁵²

Predicting the binding affinity for these compounds using virtual screening methods is nontrivial, but doable. This, however, requires the correct assignment of the relevant binding mode; the architecture of the binding site (highly hydrophobic central cavity, able to accommodate ligands of diverse sizes, due to its flexibility) makes differentiation between agonists and antagonists remains extremely challenging.

Fig. 5 A Chemical structures of two agonists, 6-formylindolo(3,2-*b*)carbazole (FICZ) and kynurenine (KYN), and two antagonists, CH223191 (CHT) and 6,2',4'-trimethoxyflavone (TMF), of human AhR. B 2D PCA plot of MD simulation trajectories of apoAhR (turquoise), AhR-FICZ (black), AhR-KYN (red), AhR-CHT (green), and AhR-TMF (blue). C Conformations of agonists (red and orange) and antagonists (blue and navy blue) at the PAS-B domain of AhR. D distributions of X–CA–CB–X dihedral angles of Y322 during MD simulations of ligand–AhR complexes, shown as a histogram plot. Agonists (red and orange), but not antagonists (blue and green) shift the ensemble towards a single conformation of Y322.

Using molecular docking methods alone on a small set of ligands with known biological activity (agonist/antagonist, Fig. 5A), we were able to rank their binding affinity, but not rationalise their activity.

Following the docked poses with MD simulations showed that agonists and antagonists, while occupying the same binding site, adopted different conformations (Fig. 5B–D). These conformations had a very different effect on the dynamics of the whole PAS domain. While both agonists, regardless of their size and the local interactions, decreased the flexibility within the domain overall, the complexes with antagonists remained flexible, counterintuitively (Fig. 5B). We have also studied the behaviour of Y322, identified as the “gatekeeper” residue (see the previous section). Antagonists restricted its mobility, quantified as changes of the X–CA–CB–X torsional angle, while its behaviour in complexes bound to antagonists was not dramatically different from that in *apo*AhR. These results indicate that intrinsic dynamics may be important, yet it is an underreported factor controlling the activation of PAS domains.

Conformational transitions within PAS-B domains

Since dynamics is very likely to play an important role in tuning the protein activity, as outlined in the previous section, we focused next on the most flexible part of the PAS domain, namely loop1. This loop adopts a fairly extended and partially unfolded conformation in all PAS-B domains but NCOA1, in which it adopts an extended α -helical conformation. Both experimental structures of this domain solved it as a complex with the STAT6 peptide, bearing an LXXLL sequence motif, which defines a conserved nuclear receptor (NR) box. This motif is a short α -helix, so it cannot be ruled out that the helical conformation of loop1 is induced by protein–protein interactions and removing the interactor (here – STAT6 peptide) may induce large conformational changes in the loop1 region, shifting the ensemble towards a more unfolded state, similar to that observed in other PAS domains.

To address this, we have modelled an alternative conformation of the NCOA1 PAS-B domain, with loop1 adopting a partially disordered helix–loop–helix conformation (Fig. 7). Both conformations were simulated for 100 ns in the absence of any protein interactors or ligands. To avoid any potential biases towards the helical conformation, the simulations were carried out using the AMBER-ff03ws force field.

The results, shown in Fig. 7, strongly indicate that NCOA1 may adopt a more extended conformation of loop1, typical for other PAS-B domains. Although the helical conformation is more stable, the difference is marginal ($0.3 \text{ kcal mol}^{-1}$, Fig. 8). This suggests potential conformational changes and the existence of NCOA1 conformations that are “druggable” by small molecules (Fig. 6).

To date, several small molecule NCOA1 ligands have been reported (Fig. 6), but their binding sites remain unknown. We have selected the compound L1, deposited in ChEMBL, reported to inhibit human NCOA1 in cell-based assay. The

Fig. 7 A Conformational change within loop1 of the NCOA1 PAS-B domain. Time-averaged structures are shown, indicating loop1 in the α -helix conformation (left panel) and in the partially disordered conformation (right panel). The backbone is coloured by the calculated *B*-factors: the most flexible parts are coloured red, while the less flexible are coloured navy blue. B FTMap scan of the NCOA1 PAS-B domain with loop1 in partially disordered conformation. A “hotspot”, detected in the centre of the PAS-B cavity, is rendered as a sphere and coloured orange; the transparent spheres are the secondary binding areas. C Cryptosite analysis of the NCOA1 PAS-B domain with loop1 in a partially disordered conformation. The highest scoring “hotspot”, detected in the centre of the PAS-B cavity, is rendered as a sphere and coloured orange.

hotspot mapping, followed by molecular docking calculations and MD simulation, showed that L1 may bind to the PAS-B domain of NCOA1 once loop1 adopts an extended conformation (Fig. 7 and Table 1). The binding event induced further conformational changes in the domain, preventing the binding to the LXXLL peptide of STAT6 (Fig. 7B). To follow up, four other confirmed small molecule NCOA1 ligands with related structures (Fig. 6) were also docked, their binding were evaluated, and the energy scores correlate with the measured IC_{50} values (Table 1).

The binding modes predicted for ligands L1–L5 explain the observed differences in binding energetics. L1 and L3 are close analogues, differing only by the heteroatom in the five-membered ring, yet their IC_{50} values, as well as computed MMPBSA energy values, are markedly different: replacing the thiophene of L1 by furan has decreased the binding affinity from the mid-nM range to $5.6 \mu\text{M}$ (Table 1). This decrease can be explained by the desolvation of the oxygen atom O1 and increased desolvation of the oxygen O2, contributing

Fig. 8 A The binding pose for the compound L1, docked to the NCOA1 PAS-B domain. Protein is coloured grey, ligand is coloured by heteroatom. Residues critical for the binding, are shown and labelled. B Docking of the LXXLL motif of STAT6 (red) to the PAS-B domain of NCOA1 with loop1 in the “druggable” partially disordered conformation (grey), compared to the docking of the same motif (green) to the PAS-B domain of NCOA1 with loop1 in the α -helical conformation (blue). The experimental structure is shown as light blue. C Predicted binding mode for the compound L1, with the per-atom contributions of the ligand mapped as the coronas, calculated by SeeSAR: green signifies favourable contribution, red signifies unfavourable contribution, and the magnitude of the contribution is proportional to the size of the corona. D Predicted binding mode for the compound L2. E Predicted binding mode for the compound L5. F Predicted binding mode for the compound L5.

Table 1 Binding data for compounds L1–L5

	IC ₅₀ ^a [nM]	Estimated K _i range ^b	Calc. ΔG^c [kcal mol ⁻¹]	Calc. ΔH^d [kcal mol ⁻¹]
L1	443	Mid/high nM	-12 ± 2	-12 ± 2
L2	1328	High nM to low μ M	-12 ± 2	-12 ± 2
L3	5695	Mid μ M	-12 ± 2	-12 ± 2
L4	3973	Low/mid μ M	-12 ± 2	-12 ± 2
L5	2442	Low/mid μ M	-12 ± 2	-12 ± 2
Pearson correlation between IC ₅₀ and calculated energy			0.83	0.97

^a ChEMBL1201862 PubChem Bioassay data set. ^b Calculated by SeeSAR. ^c Calculated by Umbrella sampling MD simulation. ^d Calculated by MM-PBSA.

unfavourably to the binding affinity (Fig. 8C). Likewise, compounds L4 and L5 are closely related analogues, and the preference of L5 ($IC_{50} = 2.4 \mu M$) over L4 ($IC_{50} = 4.0 \mu M$) can be explained by the favourable contribution of the second methoxy group in L5 (atom O2 and C2, Fig. 8D).

The free energies calculated by umbrella sampling simulations correlate with the experimental IC_{50} values, showing favourable binding energies. L3 shows, in both umbrella simulations and MMPBSA calculations, the lowest affinity. This is particularly encouraging, considering the L3 and L1 structural similarity. A reason for these differences may arise from the protein interaction after 0.5–0.7 nm (Fig. S1†). This region indicates the start of the energy increase for the systems. L1 has the lowest barrier in the region throughout its binding pathway, in comparison with the other ligands which explains the favourable IC_{50} value. This, coupled with the desolvation effect explained earlier, explains the L3–L1 difference in affinity. It was not possible to achieve statistical significance to assess the affinity difference between L1, L2, L4 and L5 in terms of the final binding free energy.

Given the plateau achieved by L3, it can be concluded that a significant part of the binding free energy arises from intrinsic conformational effects of the protein. This is supported further by the protein RMSD calculations. For L1, the RMSD of the ligand–protein complex converges to 0.25 nm and the RMSD of loop1 converges to 0.3 nm, while for the L3 complex RMSD converges to 0.3 nm and the loop1 RMSD converges to 0.35 nm.

The difference in their respective final configurations results from two effects: the thiophene moiety from L1 is more favourable at the binding site (as shown by the MM-PBSA), allowing loop1 to acquire a more stable conformation. This configuration is achieved given the thiophene moiety interaction with loop1 being unfavourable, which is compensated by the favourable side chain–side chain interactions.

Analysis of the pathway differences between L4 and L5 shows that the additional methoxyl group of L5 affects how the overall ligand is solvated in the unbound state, which can explain the more favourable IC_{50} value.

Regardless, all five ligands show favourable values of binding energy to the unstructured loop1 conformation in all three scoring modes: SeeSAR/HYDE, MMPBSA, and umbrella sampling. This indicates that NCOA1 is druggable upon configurational transition, since the favourable binding site is found only in the canonical loop–helix–loop conformation. The structural entropy for NCOA1 loop1 can be easily compensated by the binding energetics. This is shown in Fig. 7, as the entropic barrier is significantly lower ($0.6 \text{ kcal mol}^{-1}$) in comparison with the binding energetics. Therefore, our results indicate that loop1 of NCOA1 may adopt the “canonical” (disordered helix–loop–helix) conformation, which may be targeted by small molecules.

These findings are very encouraging for structure-based drug discovery efforts targeting steroid receptor coactivators for therapeutic interventions. The PAS-B domain of NCOA1 has been shown to interact with several proteins linked to

cancer, including the transactivation domain of STAT6.⁵⁰ As such, disruption of the STAT6–NCOA1 complex by targeting the PAS-B domain of NCOA1 could be an attractive therapeutic strategy. Also, NR boxes of NCOA1 and other coactivators interact with the PAS-B domain of NCOA1, promoting homo- and heterodimerisation of NCOA1, and this interaction may also be exploited therapeutically.⁵²

To assess the generality of the conformational behaviour observed for NCOA1, we took the same approach to other PAS-B domains, HIF-1 α , HIF-2 α , and AhR, inducing the α -helical conformation of loop1, resembling that of NCOA1. The results, shown in Fig. 7 and Table 2, indicate that the extended conformation of loop1 is more entropically favourable for all three proteins. Therefore, the conformational transitions within these regions, in order to occur, must be driven by a very strong enthalpic contribution (*e.g.* arising from protein–protein interactions). While this cannot be ruled out completely, particularly for AhR (Fig. S2†), such transitions are less likely for unbound PAS class I members, such as HIF-2 α than for p160 transcription factors, such as NCOA1 (Table 2).

“Druggability” of PAS-A domains

Finally, we observed whether PAS-A domains can also be targeted by small molecule ligands, in a manner similar to PAS-B domains which has been suggested for HIF-1 α .²² Experimental structures of PAS-A domains of AhR, HIF-1 α , HIF-2 α , Npas1, and Npas3 have been subjected to the analysis. As these structures miss several flexible loops (Fig. 9), attempts were made to model these loops *via* the MODELLER interface in UCSF Chimera. While this approach was adequate for shorter segments (4–15 residues), some proteins, including the AhR and Npas family, are missing larger segments of about 35–40 residues; *e.g.* in human AhR the 38 amino acid segment, corresponding to residues 173–211, is missing. Those segments are not amenable to conventional homology modelling approaches, due to lack of suitable templates. This not only requires the use of more sophisticated modelling approaches for these missing loops (see Materials and methods), but also raises concerns about the potential impact of those loops on the “druggability” of the domains.

To address these, we have modelled the missing loop of AhR using the *ab initio* modelling approach. The best-scoring model of the loop has been fitted in the PAS-A domain using protein–protein docking, and the complete structure of the PAS-A domain of the AhR receptor has been subjected to 100 ns of all-atom MD simulation in explicit water. During the

Table 2 Entropy values from unstructured to structured conformations

Protein	$T\Delta S$ [kcal mol^{-1}]
AhR	7.6
HIF-1 α	11.3
HIF-2 α	11.9
NCOA1	0.3

Fig. 9 A RMSD plots of AhR PAS-A with two conformations of the *ab initio* modelled loop (residues 173–211): the plot for the structure with the loop with the C183–C202 disulphide bridge is coloured red, and the plot for the model where cysteines C183 and C202 are reduced is coloured black. The plots show the production phase (100 ns) of the MD simulation. B PAS-A (orange and red) and PAS-B (cyan) domains of AhR are superimposed, to compare the relative sizes of the inner cavities, which affect druggability. The 173–211 loop is coloured red and its molecular surface is displayed (mesh). Another loop, 251–261, is coloured orange, and its surface is displayed to signify the size of these loops relative to the PAS-A core. The side chain of the “gatekeeper” residue, corresponding to Y322 in the PAS-B domain, is displayed. C FTMap results for the averaged structure of AhR PAS-A. The protein backbone is coloured grey, loops and all side chains but “gatekeeper” (green) are not shown. FTMap clusters, highlighting potential druggable “hotspots”, are coloured red. D Hydrophobicity of the PAS-A domain. Hydrophobic residues lining the PAS-A inner cavity are highlighted as “balls and sticks”.

simulation, the structure was stable, with no substantial conformational changes occurring (Fig. 9A). The results indicate that due to its size, the loop is likely to obstruct the entrance of any small ligand to the internal cavity of the PAS-A domain (Fig. 9B), thus rendering PAS-A of AhR “undruggable”, despite the cavity size allowing for binding of a small molecule. However, this is unlikely to be the case, as indicated by FTMap (Fig. 9C). The inner cavity of the PAS-A domain is smaller, compared to that of PAS-B, while the number of hydrophobic residues is higher than that of the PAS-B domain, with not a single H-bonding residue present (Fig. 9D). This is very likely to be the case for the Npas family of proteins and NCOA transcriptional coactivators, considering all of them have a long loop within their PAS-A domains (PDB codes: 5SY5 and 5SY7), as well as being lined with hydrophobic residues.

FTMap and Cryptosite scans of PAS-A domains with relatively shorter loops, *e.g.* HIF-1 α and HIF-2 α , failed to detect any “druggable” cavities inside the domain of HIF-1 α (Fig. 10A). This is again very likely due to the relatively small size of the PAS-A inner pocket compared to that of PAS-B. The only PAS-A domain which has a “druggable” binding site inside the domain predicted by FTMap, was HIF-2 α (Fig. 10B). The inner cavities of these PAS-A domains are lined with hydrophobic and bulky residues (particularly Met, Phe, and Leu); there is only one H-bonding side chain (Thr135/136). This makes the binding site significantly smaller than that found in the PAS-B domain (Fig. 10B). These differences may limit the size of the small molecule ligand, while the hydrophobicity is likely to make the development of any potential selective ligand very challenging. To

Fig. 10 A FTMap results for the highest-populated cluster of the HIF-1 α PAS-A domain. The protein backbone is coloured grey; F133 corresponding to the Y322 “gatekeeper” of the AhR PAS-B domain, is highlighted. FTMap clusters, highlighting potential druggable “hotspots”, are coloured red. B FTMap results for the highest-populated cluster of the HIF-2 α PAS-A domain. The protein backbone is coloured dark blue, FTMap clusters, highlighting potential druggable “hotspots”, are coloured red, and the “hotspot” located inside the PAS-A cavity is coloured green.

conclude, a combination of steric factors and the microenvironment of PAS-A inner cavities make them unlikely to be “druggable” by small molecules.

Conclusions

In the present study, we employed computational approaches to assess the “druggability” of several mammalian PAS-A and PAS-B domains to address some existing gaps in understanding their intrinsic dynamics. Our molecular dynamics simulations show that the function of the AhR PAS-B domain, and, very likely, other PAS-B domains, is regulated by the dynamics of the highly conserved tyrosine residue, which acts as a “gatekeeper” controlling the access to the binding cavity and finely tuning the binding affinity. Investigation of the PAS-B domain of AhR showed differences in dynamics depending on either agonist or antagonist binding. This indicates the role of entropic factors in fine-tuning the biological activity of AhR.

Considering the lack of data regarding the druggability of PAS-A domains, we have investigated the reasons for the druggability differences between PAS-A and PAS-B domains. The lack of reported druggability of PAS-A domains can be assigned to their smaller cavity size and the presence of bulky hydrophobic residues within their cavity in comparison with PAS-B domains.

While investigating PAS-B domains, we have observed two possible conformations these domains can adopt – the partially unfolded conformation of loop1 found in most reported structures, including HIF-2 α , and the α -helical conformation observed in NCOA1. Both conformations have been studied, and marginal stability differences are reported for NCOA1. This implies the different druggability potential of the two conformers, as the structural changes affect the cavity size and the access of the ligand/drug molecule. Importantly, our results indicate the presence of the NCOA1 PAS-B conformation amenable to binding small molecule ligands. Considering an important role of NCOA1 in hormone-dependent cancers, our findings highlight the potential for structure-based development of new anticancer drugs targeting NCOA1.

Ethical statement

All works were conducted in accordance with the “Newcastle University Ethics Policy for Research, Teaching and Consultancy” guidelines, and the study was approved by the ethics committee at Newcastle University.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We are grateful to Dr. M. Garner for technical assistance and to R. Cairns and P. Lanzoni Mangutchi for critical review of the manuscript and constructive feedback. This work was funded by the School of Natural and Environmental Sciences (scholarship to J. V. S.) and by the Faculty of Science, Agriculture, and Engineering, Newcastle University (scholarship to S. R.).

Notes and references

- 1 Z. Feng, X. Zou, Y. Chen, H. Wang, Y. Duan and R. K. Bruick, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**(52), 13240–13245.
- 2 R. Depping and H. Oster, *FEBS J.*, 2017, **284**(22), 3801–3803.
- 3 M. Rojas-Pirela, D. J. Rigden, P. A. Michels, A. J. Cáceres, J. L. Concepción and W. Quiñones, *Mol. Biochem. Parasitol.*, 2018, **219**, 52–66.
- 4 Y. Zhang, M. J. Markert, S. C. Groves, P. E. Hardin and C. Merlin, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**(36), E7516–E7525.
- 5 Z. Culig and F. R. Santer, *Methods Mol. Biol.*, 2018, **1786**, 259–262.
- 6 A. L. Hartzell, K. M. Martyniuk, G. S. Brigidi, D. A. Heinz, N. A. Djaja, A. Payne and B. L. Bloodgood, *eLife*, 2018, **7**, pii: e35927.
- 7 X. Sun and Y. Lin, *Trends Neurosci.*, 2016, **39**(4), 264–275.
- 8 A. C. Harmon, V. Y. Hebert, S. A. Cormier, B. Subramanian, J. R. Reed, W. L. Backes and T. R. Dugas, *PLoS One*, 2018, **13**(10), e0205412.
- 9 X. Tang, J. Shao and X. Qin, *Acta Crystallogr., Sect. F: Struct. Biol. Commun.*, 2016, **72**(Pt 8), 578–585.
- 10 X. Che and W. Dai, *Curr. Drug Targets*, 2019, **20**(6), 625–634.
- 11 R. Shinde and T. L. McGaha, *Trends Immunol.*, 2018, **39**(12), 1005–1020.
- 12 F. J. Weng, R. I. Garcia, S. Lutz, K. Alviña, Y. Zhang, M. Dushko, T. Ku, K. Zemoura, D. Rich, D. Garcia-Dominguez, M. Hung, T. D. Yelhekar, A. T. Sørensen, W. Xu, K. Chung, P. E. Castillo and Y. Lin, *Neuron*, 2018, **97**(5), 1137–1152, e5.
- 13 S. Adesso, I. Paterniti, S. Cuzzocrea, M. Fujioka, G. Autore, T. Magnus, A. Pinto and S. Marzocco, *J. Clin. Med.*, 2018, **7**(10), pii:E365.
- 14 S. Adesso, T. Magnus, S. Cuzzocrea, M. Campolo, B. Rissiek, O. Paciello, G. Autore, A. Pinto and S. Marzocco, *Front. Pharmacol.*, 2017, **8**, 370.
- 15 A. A. Soshilov and M. S. Denison, *Mol. Cell. Biol.*, 2014, **34**(9), 1707–1719.

- 16 Y. Guo, T. H. Scheuermann, C. L. Partch, D. R. Tomchick and K. H. Gardner, *J. Biol. Chem.*, 2015, **290**(12), 7707–7721.
- 17 Y. Guo, C. L. Partch, J. Key, P. B. Card, V. Pashkov, A. Patel, R. K. Bruick, H. Wurdak and K. H. Gardner, *ACS Chem. Biol.*, 2013, **8**(3), 626–635.
- 18 D. Wu, X. Su, N. Potluri, Y. Kim and F. Rastinejad, *eLife*, 2016, **5**, pii: e18790.
- 19 T. Uchida, I. Sagami, T. Shimizu, K. Ishimori and T. Kitagawa, *J. Inorg. Biochem.*, 2012, **108**, 188–195.
- 20 C. L. Partch and K. H. Gardner, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**(19), 7739–7744.
- 21 D. Wu, N. Potluri, J. Lu, Y. Kim and F. Rastinejad, *Nature*, 2015, **524**(7565), 303–308.
- 22 E. J. Park, D. Kong, R. Fisher, J. Cardellina, R. H. Shoemaker and G. Melillo, *Cell Cycle*, 2006, **5**(16), 1847–1853.
- 23 L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg, *Nat. Protoc.*, 2015, **10**(6), 845–858.
- 24 B. Webb and A. Sali, *Methods Mol. Biol.*, 2017, **1654**, 39–54.
- 25 J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, *Nat. Methods*, 2015, **12**, 7–8.
- 26 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nucleic Acids Res.*, 2018, **46**(W1), W296–W303.
- 27 D. van der Spoel, E. Lindahl and H. J. Berendsen, *et al.*, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
- 28 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, I. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**(8), 1950–1958.
- 29 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 30 B. Hess, H. Bekker and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 31 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182.
- 32 A. W. Sousa da Silva and W. F. Vranken, *BMC Res. Notes*, 2012, **5**, 367.
- 33 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**(16), 1623–1641.
- 34 C. C. Huang, E. C. Meng, J. H. Morris, E. F. Pettersen and T. E. Ferrin, *Nucleic Acids Res.*, 2014, **42**(Web Server issue), W478–W484.
- 35 H. C. Chan, Y. Li, T. Dahoun, H. Vogel and S. Yuan, *Trends Biochem. Sci.*, 2019, **44**(4), 312–330.
- 36 P. Cimermancic, P. Weinkam, T. J. Rettenmaier, L. Bichmann, D. A. Keedy, R. A. Woldeyes, D. Schneidman-Duhovny, O. N. Demerdash, J. C. Mitchell, J. A. Wells, J. S. Fraser and A. Sali, *J. Mol. Biol.*, 2016, **428**(4), 709–719.
- 37 D. Kozakov, L. E. Grove, D. R. Hall, T. Bohnuud, S. E. Mottarella, L. Luo, B. Xia, D. Beglov and S. Vajda, *Nat. Protoc.*, 2015, **10**(5), 733–755.
- 38 D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov and S. Vajda, *Nat. Protoc.*, 2017, **12**(2), 255–278.
- 39 W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz and R. C. Rizzo, *J. Comput. Chem.*, 2015, **36**(15), 1132–1156.
- 40 <https://www.biosolveit.de/SeeSAR/>.
- 41 N. Schneider, G. Lange, S. Hindle, R. Klein and M. Rarey, *J. Comput.-Aided Mol. Des.*, 2013, **27**(1), 15–29.
- 42 R. Kumari, R. Kumar, Open Source Drug Discovery Consortium and A. Lynn, *J. Chem. Inf. Model.*, 2014, **54**(7), 1951–1962.
- 43 R. B. Best, D. de Sancho and J. Mittal, *Biophys. J.*, 2012, **102**(6), 1462–1467.
- 44 J. L. Abascal and C. Vega, *J. Chem. Phys.*, 2005, **123**(23), 234505.
- 45 J. V. de Souza and A. K. Bronowska, available on <https://chemrxiv.org/s/b83db2a6e9342fe9edc3>.
- 46 M. Powis, T. Celius and J. Matthews, *Biochem. Biophys. Res. Commun.*, 2011, **410**(4), 859.
- 47 D. Szöllősi, Á. Erdei, G. Gyimesi, C. Magyar and T. Hegedűs, *PLoS One*, 2016, **11**(1), e0146066.
- 48 A. Kühnert, C. Vogs, B. Seiwert, S. Aulhorn, R. Altenburger, H. Hollert, E. Küster and W. Busch, *Environ. Pollut.*, 2017, **230**, 1–11.
- 49 B. W. Labadie, R. Bao and J. J. Luke, *Clin. Cancer Res.*, 2019, **25**(5), 1462–1471.
- 50 J. E. B. Koper, L. M. P. Loonen, J. M. Wells, A. D. Troise, E. Capuano and V. Fogliano, *Mol. Nutr. Food Res.*, 2018, e1800722.
- 51 E. F. O'Donnell, D. C. Koch, W. H. Bisson, H. S. Jang and S. K. Kolluri, *Cell Death Dis.*, 2014, **5**, e1038.
- 52 L. Ge, Y. Cui, K. Cheng and J. Han, *Molecules*, 2018, **23**(10), pii: E2600.