

“Druggability” of the PAS domains of human PASK kinase, a therapeutic target for metabolic and liver disorders

Shangze Xu^{1,2±}, Lanyu Fan^{1,3±}, Piotr Zaborniak^{1±}, Ruidi Zhu^{1±}, Haoyuan Ji¹, Katrina S. Madden^{1,4†}, João V. de Souza^{1†}, Agnieszka K. Bronowska^{1,2,†*}

¹Chemistry—School of Natural and Environmental Sciences, Newcastle University, Newcastle NE1 7RU, UK

²Newcastle University Centre for Cancer, Newcastle University, Newcastle NE1 7RU, UK

³School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

⁴Translational and Clinical Research Institute, Newcastle University, Newcastle NE1 7RU UK

*Correspondence: agnieszka.bronowska@ncl.ac.uk

±Equal contribution

†Equal contribution

Abstract

Per–Arnt–Sim domain kinase (PASK) is a conserved metabolic sensor, which modulates the activation of critical proteins involved in liver metabolism and fitness. However, despite its key role in mastering the metabolic regulation, the molecular mechanism of PASK’s activity is an ongoing research, and structural information on this important protein is scarce. To investigate this, we integrated structural bioinformatics with state-of-the-art modelling and molecular simulation techniques. Our goals were to address (1) how many regulatory PAS domains PASK is likely to have, (2) how those domains modulate the kinase activity, and (3) how those interactions could be controlled by small molecules. Our results indicated the existence of three N-terminal PAS domains. Solvent mapping and fragment docking identified a consensus set of “druggable hot-spots” within all domains, as well as at domain-domain interfaces. Those “hot-spots” could be modulated with chemically diverse small molecular probes, which may serve as a starting point for rationally-designed therapeutics modulating these specific sites. Our results identified a plausible mechanism of auto-inhibition of kinase activity, suggesting that all three putative PAS domains may be required. Future work will focus on validation of the predicted PASK models, and development of small molecule inhibitors of PASK by targeting its “druggable hot-spots”.

Introduction

The protein kinase with PAS domains (PASK) is a highly conserved serine/threonine kinase containing N-terminal Per-Arnt-Sim (PAS) domains and a C-terminal kinase domain. It acts as a nutrient and energy sensor (Hurtado-Carneiro *et al.*, 2014 and 2020), and can also respond to intracellular redox state, light and oxygen (Hurtado-Carneiro *et al.*, 2021; Dongil *et al.*, 2018, Dongil *et al.*, 2021). In mammals, PASK senses nutritional status and contributes to the

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.

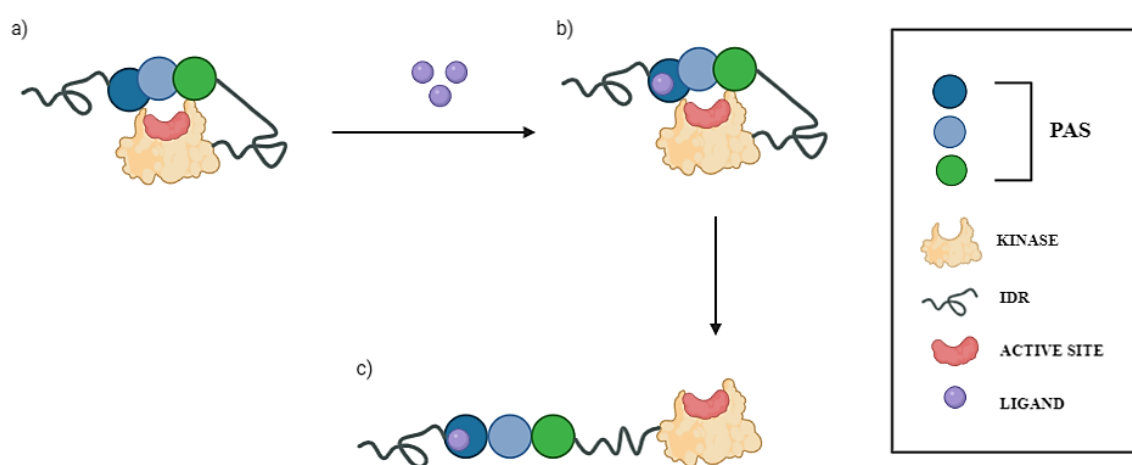
10.1017/qrd.2024.1

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

regulation of glucose homeostasis, energy metabolism, and oxidative stress (Hurtado-Carneiro *et al.*, 2021). PASK regulates glucagon and insulin secretion, and is also a critical signalling regulator of the AMPK and mTOR pathways in the liver (Zhang *et al.*, 2015). In addition, PASK involvement in protein translation, cell differentiation processes and epigenetic regulation has recently been described (Kikani *et al.*, 2019). Collectively, these make PASK an attractive therapeutic target, which has been recently validated for conditions such as non-alcoholic steatohepatitis and liver fibrosis (Zhang *et al.*, 2015; Swiatek *et al.*, 2020).

1 Despite its evolutionary conservation and established role in
 2 coordination between cellular metabolism, oxidative stress and metabolic demand, the
 3 molecular mechanism underlying PASK regulation remains an ongoing research (Hurtado-
 4 Carneiro *et al.*, 2020). The crystal structure of the kinase domain of human PASK shows that
 5 this domain adopts an active conformation and has catalytic activity in the absence of
 6 activation loop phosphorylation (Kikani *et al.*, 2010). It has been therefore suggested that
 7 the kinase domain of PASK is autoinhibited by one of its N-terminal PAS domains. The
 8 current understanding of PASK activation is that under the standard conditions PASK
 9 remains autoinhibited, and the interaction between certain small molecules (such as
 10 nutrients or metabolites) and the PAS domain terminate PASK's autoinhibited state by
 11 unblocking the kinase domain (Fig. 1). This may lead to subsequent phosphorylation of
 12 PASK's substrates and regulate its downstream effectors (Hurtado-Carneiro *et al.*, 2021).
 13



14
 15
 16 **Figure 1** The current proposed model of PASK activation. Putative PAS domains of human PASK
 17 kinase are represented as circles (blue, aqua and green). The kinase domain is represented as a
 18 yellow oval, and the active site where ATP is bound is depicted as a red circle. Intrinsically disordered
 19 regions (IDRs) are depicted as springs. **a)** Under resting conditions, PASK assumes a closed, auto-
 20 inhibited state, wherein PAS domains block the active site of the kinase domain. **b)** In the presence of
 21 small molecule ligands (e.g. nutrients, metabolites) depicted as purple circles, one of the PAS domains
 22 binds a ligand, causing conformational changes within the domain and affecting its interactions with
 23 the kinase domain. **c)** This conformational change leads to unblocking of the kinase domain, thus
 24 rendering PASK active.
 25

26 Despite ongoing efforts, the structural biology and precise mechanism of
 27 PASK regulation remain poorly understood. Likewise, the endogenous small molecules
 28 modulating PASK are still unknown, and the relationship between PASK and oxidative stress
 29 require better understanding. Since PAS domains are structurally conserved, versatile sensor
 30 domains reported to detect hypoxia, redox state, and a broad spectrum of chemically
 31 diverse small molecules (e.g. metabolites, pollutants, and drugs) (Arnaiz *et al.*, 2021; Mullard
 32 *et al.*, 2021; Poch *et al.*, 2019), it may be speculated that PAS domains of PASK are sensor
 33 domains, crucial for the regulation of this kinase.
 34

PAS domains are small (~100 residues) motifs, present in all kingdoms of life

1 (de Souza *et al.*, 2019). In eukaryotes, PAS domains are found in several classes of signalling
2 proteins, including transcriptional regulators, transmembrane channels, and kinases (de
3 Souza *et al.*, 2019). Despite their primary sequence variability, these domains have a highly
4 conserved three-dimensional structure, with a hydrophobic cavity where small molecule
5 ligands may bind. In other mammalian PAS domain sensors, such as the aryl hydrocarbon
6 receptor (AhR), such a binding event initiates a conformational change that spreads over all
7 the protein and affects protein regions distal to the binding event. This may have varied
8 results, such as dissociation of the PAS domain from its cognate binding partner (e.g. Hsp90-
9 AhR complex), formation of a new complex (e.g. AhR-ARNT) and activation (e.g.
10 transcription regulator function of AhR, NPAS and HIF proteins) (Greb-Markiewicz *et al.*,
11 2019; Janssens & Stove, 2021; Soshilov *et al.*, 2020).

12 Amezcuca *et al.* (2002) studied the conformational changes within the first PAS
13 domain of human PASK by combining structural biology, biophysical and biochemical
14 techniques. By solving an NMR structure of this domain (PDB code: 1LL8, residues 131-237),
15 they showed that the domain contains two unusually flexible segments. The locations of
16 those segments indicated that these segments might serve as functionally relevant binding
17 interfaces. NMR-based screening of a library containing over 750 small molecules showed
18 that this domain binds a diverse range of small molecules within its binding site.
19 Furthermore, their work showed that the kinase domain of PASK interacted with the first
20 PAS domain, providing the initial direct biophysical observation of protein-protein
21 interactions (PPIs) within PASK. They concluded that the PAS and kinase domains of PASK are
22 functionally and structurally linked, suggesting a regulatory pathway for small organic
23 molecules to modulate the enzymatic activity of PASK. This study provided a robust proof-
24 of-concept for PASK allosteric regulation occurring via PAS domains. However, the N-
25 terminus of PASK is very long (~950 residues), where several PAS domains have been
26 mapped by bioinformatic approaches (Fig. 2). Even though the NMR structure of the first
27 PAS domain of human PASK has been solved (Amezcuca *et al.*, 2002), the remainder of the N-
28 terminal region of PASK, including the number, location, and potential “druggability” (i.e.
29 ability to bind small molecules, such as drugs or metabolites) of those elusive PAS domains
30 remains unexplored.

31 In this work, we have combined complementary computational
32 approaches, to address the following questions: (i) how many PAS domains are likely to span
33 the N-terminus of PASK kinase; (ii) what is their overall structure and how flexible are these
34 domains; (iii) can all of these domains bind small molecule ligands, and what is the chemical
35 nature of those putative ligands; (iv) which of those PAS domains is the most likely to
36 interact with the kinase domain, and could these putative interactions be modulated by
37 small molecules?

38 First, using several orthogonal bioinformatics methods, we mapped
39 the location of putative PAS domains. Three-dimensional models of those domains were
40 calculated using a combination of homology modelling via Phyre2 (Kelley *et al.*, 2015),
41 alongside modelling using AlphaFold (Jumper *et al.*, 2021). Next, we assessed the overall
42 stability and quality of those models by all-atom molecular dynamics simulations in explicit
43 water and excluded those models that showed abnormal energetics, had an unusual
44 number of Ramachandran outliers, or unfolded during the simulations. The resulting model
45 indicated an intrinsically disordered N-terminus, three consecutive PAS domains and a long,
46 partially disordered linker joining PAS domains and the kinase domain of PASK. The model of

1 the first two domains (PASK1 and PASK2) are consistent with structural and published data
2 (Amezcuca *et al.*, 2002; Kikani *et al.*, 2019); however, details of the architecture of the third
3 PAS domain require further investigation. Next, we assessed interactions between PAS and
4 kinase domains by protein-protein docking. These results suggest an explanation of how the
5 kinase domain may be activated by small molecule binding to N-terminal PAS domains.
6 “Druggability” mapping of individual PAS domains, free and bound to the kinase domain,
7 indicated that these PAS domains may be functionally non-redundant and serve as sensors
8 for chemically diverse ligands. Our results also indicate the presence of interfacial
9 “druggable” sites at autoinhibited PASK, which may pave the way for development of new
10 drugs to inhibit PASK via an allosteric mechanism. Finally, our results suggest that
11 metformin, a first-line drug to treat type 2 diabetes (Wang *et al.*, 2017), may inhibit PASK by
12 acting as a “molecular glue” between the PAS1-PAS2 segment and the kinase domain. This is
13 likely to shed light on metformin’s mechanism of action, which is not fully understood (Agius
14 *et al.*, 2020).

15

16 **Materials and methods**

17

18 **Structural bioinformatics**

19

20 A collection of orthogonal tools was used to predict, cross-check and annotate putative PAS
21 domains within the N-terminal portion of human PASK (UniProt ID: Q96RG2). For consensus
22 domain annotations the following tools were used: Interpro (Blum *et al.*, 2021), TIGR (Haft
23 *et al.*, 2001), SMART (Letunic *et al.*, 2021), Prosite (Sigrist *et al.*, 2013), Pfam (Mistry *et al.*,
24 2021), and the conserved domains database (CDD; (Lu *et al.*, 2020) In this work all alignment
25 parameters used their default parameters. For sequence/structural annotation, Genome3D
26 was employed (Lewis *et al.*, 2015), using its default parameter. For prediction of putative
27 domains and structured regions within PASK sequence, we used a combination of Phyre2
28 homology modelling webserver (Kelley *et al.*, 2015) , DomSerf (Buchan *et al.*, 2013), FUGUE
29 (Shi *et al.*, 2001), and pDomThreader/PSIPRED (Lobley *et al.*, 2009). All Algorithms for
30 prediction of domains used their default parameters. The outputs were manually curated,
31 analysed, and integrated into a single comprehensible consensus output.

32

33 **Molecular modelling**

34

35 All plausible domains were extracted from the full-length AlphaFold model (AFQ96RG2)
36 generated obtained from the structure available at Uniprot (Uniprot code Q96RG2) and
37 subjected to 10,000 of molecular mechanical energy minimisation using Amber99SB-ILDN
38 force field (Lindorff-Larsen *et al.*, 2010) using Gromacs (van der Spoel *et al.*, 2005).
39 Homology models were calculated using Phyre2 (Kelley *et al.*, 2015) and Swiss-Model
40 (Waterhouse *et al.*, 2018) using the following templates: PDB code 1LL8 for human PAS1
41 (Sequence similarity = 60%); PDB code 5SY5 (Sequence similarity = 30%); for human PAS2;
42 PDB code 4GW9 for human PAS3 (Sequence similarity = 29%); and PDB code 6PH4 for PAS4
43 (Sequence similarity = 31%); . For the construct consisting of both PAS1 and PAS2, PDB code:
44 5SY5 was used as a template. These templates were selected using the following criteria:
45 highest sequence similarity between the target and the template, and the lowest number of
46 outliers from the Ramachandran plot.

All-atom molecular dynamics simulations

To address structural stability, energetics and intrinsic dynamics of each putative PAS domain within human PAS kinase, all-atom molecular dynamics (MD) simulations were performed on the following distinct regions of PASK: PAS1 (residues 124-237), PAS2 (238-335), PAS1-PAS2 construct (residues 124-335); PAS3_homm (residues 337-437), PAS3_AF 3 (residues 343-400 and 885-982), and PD4 (residues 485-592).

All simulations were carried out in GROMACS 2016.1 (van der Spoel *et al.*, 2005), using the AMBER99SB-ILDN force field with the TIP3P water model (Lindorff-Larsen *et al.*, 2010). In each simulation, the protein under investigation was placed in a centre of a cubic simulation box. The distance between the box edge and the protein was set to 1 nm. The box was filled with solvent (water) and counter-ions (Na^+ and Cl^-) were added to a concentration of 0.1 M to maintain charge neutrality of the simulation unit and to mimic their physiological concentration. Next, molecular mechanical energy minimisation was performed: first, for 1000 cycles of steepest descent, and after that, the algorithm was switched to a conjugate gradient. The minimisation step size was set to 0.01 nm, and the maximum number of steps performed was 50,000. The energy minimization stopped when the maximum force of the system was lower than $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ using the Verlet cut-off scheme (Verlet, 1967). The particle mesh-Ewald (PME) was set to long-range electrostatic interactions (Darden *et al.*, 1993), and the short-range electrostatic and van der Waals cut-off was set to 1.0 nm.

Energy-minimised systems were equilibrated, including the following steps: (1) NVT equilibration for 20 ps, with 2 fs time step, wherein the system has been heated and the target temperature was set at 300 K, by using a modified Berendsen thermostat ($\tau = 0.1 \text{ ps}$) (Berendsen *et al.*, 1984). The constraint algorithm used was LINCS (Linear Constraint Solver) for all bonds in the protein (Hess *et al.*, 1997). The Verlet cut-off scheme was set to 1.0 nm for non-bonded short interaction, long-range electrostatics were set to PME. (2) NPT equilibration (20 ps), wherein pressure was kept constant at 1 bar by Parrinello-Rahman isotropic coupling ($\tau = 2.0 \text{ ps}$) to a pressure bath (Parrinello and Rahman, 1981). (1) and (2) included positional restraints on backbone atoms, to prevent unfolding and other possible artefacts of rapid heating and equilibration. (3) 60 ps of NPT equilibration, with no positional restraints. (4) The production MD simulations, were performed for 500 ns in triplicates. The coordinates and energies were saved every 10 ps. Collectively, all production simulations amounted to 9 microseconds.

Data analysis

Analysis of MD production trajectories was performed using GROMACS tools, FTMap webserver (<http://ftmap.bu.edu/login.php>) (Kozakov *et al.*, 2015), and MDAnalysis package (Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016). GROMACS tools included RMSD (root-mean-square deviation) calculations to assess the structural stability of the domain of interest; RMSF (root-mean-square fluctuation) calculations to assess the flexibility per-residue; PCA (principal component analysis) to assess low-amplitude correlated motions, and cluster analysis to evaluate heterogeneity of produced molecular ensembles.

1 H-bond calculations was used as a metric of internal stability of the system, and its internal
2 potential energy calculated via gmx energy. For each system, all trajectories were
3 concatenated prior to the cluster analysis (Daura *et al.*, 1999) and PCA. The outcomes of PCA
4 and cluster analysis, namely top-three clusters, and clusters representative to PC1-PC3,
5 respectively, were subjected to FTMap: a computational mapping server that identifies
6 binding hotspots in proteins (Kozakov *et al.*, 2015), and a widely used tool on which our
7 group has used and tested (Sabanés *et al.*, 2019). Clustering Ensemble Similarity (CES), a
8 convergence analysis tool from MDAnalysis package (Michaud-Agrawal *et al.*, 2011) was
9 used for assessing the convergence of the systems. We monitored the cluster evolution
10 through time using MDAnalysis alongside RMSD. The tool splits the trajectory into different
11 windows, then calculate the CES value, each time a new conformation space is formed by
12 the current window plus the previous windows and compared with the conformation space
13 formed by the entire trajectory. Faster drop of CES value to 0 means the earlier convergence
14 of the trajectory. Results for Clustering Ensemble Similarity analysis is showed in Suppl. Fig.
15 S14.

Protein-protein docking

Protein-protein docking was performed with ClusPro (<https://cluspro.bu.edu/>). We used rigid docking, as described by Kozakov *et al.* (2017). Fast-Fourier Transform (FFT) employed within ClusPro enables the sampling of billions of putative complex conformations (Desta *et al.*, 2020). This makes the algorithm highly suitable for global docking without any *a priori* information on the structure of the complex. The performance validation of ClusPro is showed in Suppl. Fig. S13.

Docking of PAS and kinase domains of human PASK was carried out sequentially. First, we assigned the crystal structure of the kinase domain (PDB code: 3DLS) as the “receptor” and all PAS domains as “ligands” of binary global docking. We used the highest-populated cluster from MD simulation of each PAS domain as a representative “ligand” frame. Two alternative models of PAS3 were considered. Flexible loops and annotated intrinsically disordered regions were excluded from the calculations in order to reduce noise within the generated data. Each docking run consisted of three following stages: (1) rigid body docking by sampling billions of conformations, applying no geometric restraints; (2) root-mean-square deviation (RMSD) based clustering of the 1,000 lowest energy structures generated to find the largest clusters that will represent the most likely models of the complex, and (3) refinement of selected structures using molecular-mechanical energy minimisation. The rigid body docking step used PIPER, a docking program based on the FFT correlation approach (Kozakov *et al.*, 2006). In each run, the 10 best-ranking models were visually inspected.

Furthermore, we used the same procedure to dock: PAS1-PAS2 segment to the kinase domain; PAS2 to the PAS1-kinase domain complexes, PAS1 to the PAS2-kinase domain complexes; PAS3 to the PAS1-PAS2-kinase ternary complexes; PAS2 to the PAS1-PAS3-kinase ternary complexes; and PAS1 to the PAS2-PAS3-kinase ternary complexes. To evaluate the propensity of PAS domains binding to each other, we have also calculated PAS-PAS2-PAS3 complexes using the same procedure, and subsequently used top-five complexes for the binary docking to the kinase domain. Each calculation involving PAS3 domain included two alternative models (AlphaFold and homology). In total, 24,000 complexes were

calculated and evaluated. In each run, the 10 best-ranking models were visually inspected.

Next, we selected those models which satisfied the distance requirement of PAS1-PAS2 and PAS2-PAS3 being less than 1.2 nm apart. This is required, as the interdomain PAS1-PAS2 and PAS2-PAS3 linkers are very short. Such an approach yielded 19 models, with different relative orientation of the domains, and different numbers of PAS domains, interacting with the kinase domain simultaneously. Those complexes were energy-minimised, and the 3 best-ranking models were selected. Out of these, we prioritised the model that matched best the experimental data available (Amezcuca *et al.*, 2002; Kikani *et al.*, 2019).

Small molecule docking

The MiniFrag library (O'Reilly *et al.*, 2019) containing 80 chemically-diverse, drug-like fragments in SDF format (e.g. phenol, 2-bromopyridine, piperazine, pyrrolidin-2-ylmethanol) was obtained from Enamine (www.enamine.net). The following compounds were added to the library: glucose, fructose, n-hexane, cyclohexane, cyclohexanone, glycerol, glycerol-3-phosphate, arginine (side chain), glutamate (side chain), histidine (side chain), lysine (side chain), metformin, dimethyl fumarate, and monomethyl fumarate. The resulting set, denoted as AugMiniFrag, was used in small molecule docking calculations.

For each PAS domain, mini-ensembles derived by clustering (PCA analysis and RMSD clustering, using 0.35 nm as the cut-off; 6 distinct conformations), were subjected to the molecular docking procedure using SeeSAR version 11.2 with HYDE scoring function (Reulecke *et al.*, 2008). To evaluate potential PPI sites at kinase-PAS interface, we used the best-scoring complex from protein-protein docking, after MM energy minimisation. To map the consensus sites, we searched for the binding sites by FTMap and by the volume search tool in SeeSAR. Only those sites that were reported by both tools were considered for docking. For each consensus site, the AugMiniFrag set was docked using non-covalent docking, maximum 200 poses for each fragment, with high clash tolerance selected. Small molecules with calculated binding affinities worse than high μM range ($>500 \mu\text{M}$) and torsional strains were discarded. The remaining sets were used for mapping of the consensus "hot spots" within each predicted PAS domain and the PAS-kinase complex.

Results and Discussion

Annotation of putative PAS domains within human PASK

While the location and three-dimensional structure of the first PAS domain (PAS1; PDB code: 1LL8; residues 131-237) and the kinase domain (PDB code: 3DLS; residues 977-1300) of human PASK are confirmed (Amezcuca *et al.*, 2002; Kikani *et al.*, 2010), the molecular architecture of the region linking those distal domains has been largely unexplored. Due to the length of this sequence, and the presence of intrinsically disordered regions (IDRs) within, the total number of predicted domains reported and their precise location vary depending on the bioinformatics tools employed. For example, domain annotation within InterPro reports only two PAS domains (spanning residues 121-232, corresponding to PAS1, and 335-402, corresponding to PAS3, Fig. 2a), while SCOP and CATH-based classification predicts three PAS domains close together, including experimentally confirmed PAS1 (Fig.

2d).

We evaluated those predictions by (i) generating 3D models of all predicted segments, and (ii) investigating the stability of those hypothetical domains by subjecting the calculated models to 500 ns all-atom MD simulations (Suppl. Figs. S1-S8). Each run was repeated three times, and the structural features, intrinsic dynamics and stability overall were evaluated. For quality control purposes, we subjected the experimentally confirmed PAS1 domain to the same procedure (Suppl. Fig. S2).

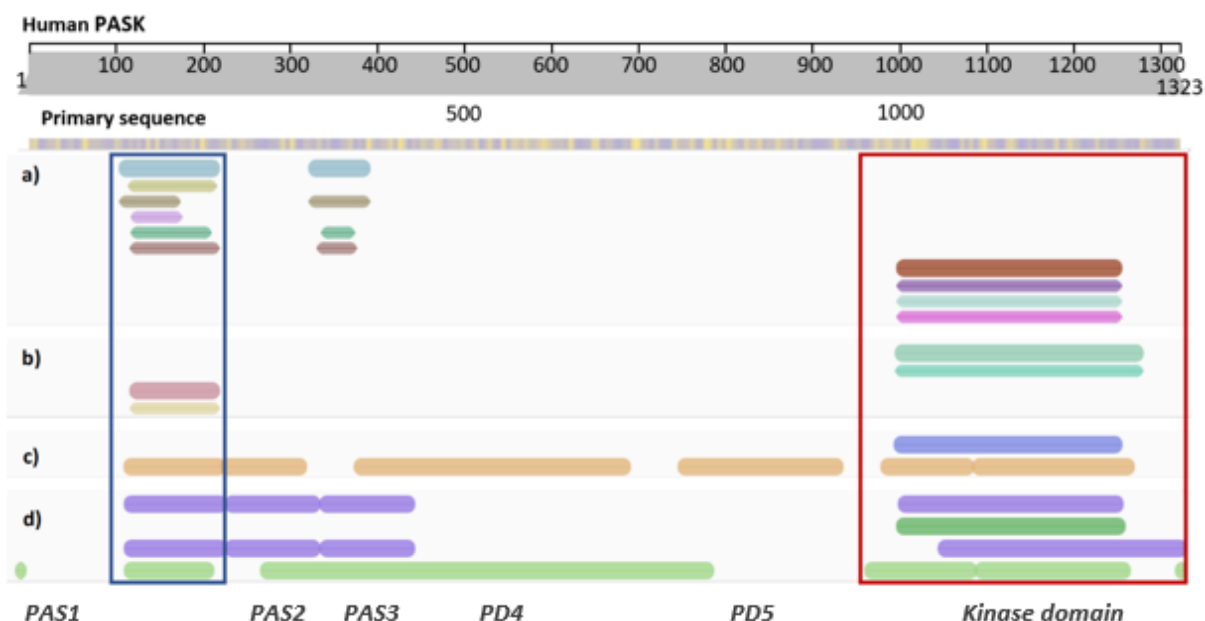
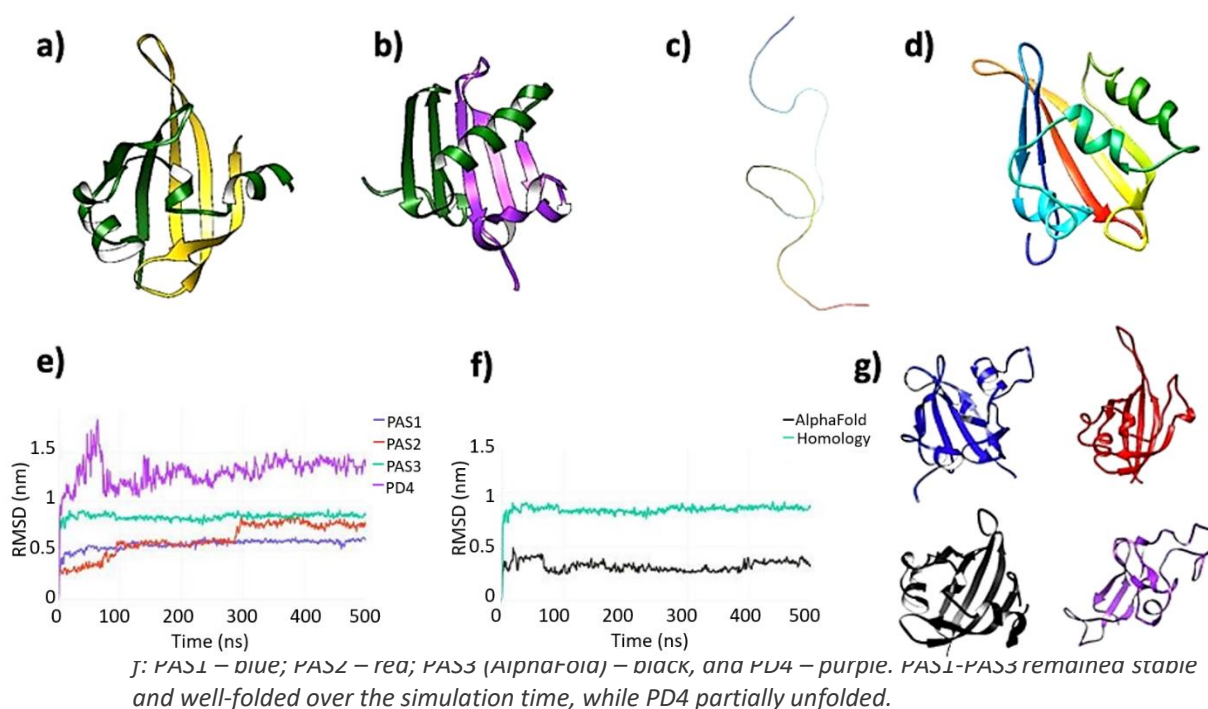


Figure 2 Diagram showing annotation of human PASK domains by different bioinformatics resources. **a)** Consensus domain annotation of N-terminal PASK segment by established structural bioinformatics databases: InterPro – blue, TIGR – khaki, SMART – brown, Prosite – lilac, Pfam – bottle green, and CDD – ochre. Kinase domain has been annotated separately and the colours correspond to dark red – InterPro, purple – SMART, aqua – ProSite, and magenta – Pfam. **b)** Consensus homologous superfamily: cyan and aqua – protein kinase-like (InterPro); red - PAS superfamily (InterPro) and orange - PYP-like sensor/PAS (InterPro). **c)** Predicted 3D structural regions annotated by Genome3D [35]: blue – PHYRE2 [36] model; brown – structural regions detected by DomSerf [37]. **d)** Predicted domains annotated by Genome3D: purple (top) – FUGUE SCOP-based classification; dark green – kinase domain reported by PHYRE2; purple (bottom) – FUGUE CATH-based classification; light green – predicted by pDomTHREADER. Two domains with resolved experimental structures are marked by rectangular selection: PAS1 domain – blue, and protein kinase domain - red. Domains PAS2-PD5 are putative, deduced from the consensus between several predictive tools.

Prior to MD simulations, we compared our generated models with predictions using AlphaFold (Jumper *et al.*, 2021). The AlphaFold model of human PASK consists of three PAS domains and the kinase domain, with PAS1 and the kinase domain matching well with the experimental structures, and the second PAS2 domain matching the model calculated by homology modelling techniques. However, the third PAS domain predicted by AlphaFold consists of two non-consecutive sequence strands (Fig. 3a): 354-397, which is part of PAS3 modelled by homology modelling and threading approaches used by Phyre2 (Fig. 3b), and residues 885-932, corresponding to the PD5 region, showed in Fig. 2. The confidence score assigned to this PAS3 model was low, particularly this non-consecutive region ($70 > \text{pLDDT} > 50$). Nevertheless, we have included this model in further analysis by all-atom MD simulations. In addition, AlphaFold did not detect the putative fourth PD4 domain (residues 485-592, Fig. 3c) which has been modelled by homology modelling (Fig. 3d). Again, we have analysed this by all-atom MD. Attempts to model PD5 as a standalone domain by homology modelling and threading approaches failed to produce a good quality model.



All-atom MD simulations indicate three PAS domains within human PASK

All initial models of individual domains were subjected to 500 ns unrestrained equilibrium MD simulations. Data analysis showed that, while PD1-PD3 were stable during the simulation time in all replica runs (Fig. 3e), PD4 partially unfolded in 2 out of 3 replicas (Fig. 3g). This is caused mainly by losing its internal polar contacts in favour to the solvent. Analysis of internal potential energies of all individual domains further suggests that PD4 is very unlikely to retain the PAS-like fold calculated by the homology modelling technique (Suppl. Fig. S5). PAS3 domain modelled by AlphaFold remained highly stable (Fig. 3f), not showing any major conformational changes during the simulation after achieving a stable configuration. PAS3 domain modelled by Phyre2 (homology modelling) remained highly stable in 2 out of 3 replicas (Fig. 3e and Suppl. Fig. S4), achieving a stable configuration during first 50 ns of the simulations. In one of the replicas, major conformational changes within the “gatekeeper” loop occurred, yet without unfolding (Suppl. Fig. S4).

To further investigate the exact location, composition, and stability of the PAS3 domain, we also simulated an AlphaFold model, as an alternative to the homology model. The comparison of these two models is shown in Fig. 3f. Despite being composed of two non-consecutive sequence strands (Fig. 3a) the model retained features of the canonical PAS domain fold, such as the L-loop with the “gatekeeper” tyrosine residue (Y389), and the core comprised of five antiparallel beta-sheets. There was a significant difference between average values for their root mean square deviation (0.2 nm for the AlphaFold model and 0.8 nm for the homology model) from the starting structure. Both models equilibrated fast and reached their stable conformations before 50 ns of MD simulation (Fig. 3f). There were several highly fluctuating residues within the AlphaFold model (Suppl. Fig. S10), but the fluctuations resulted from non-constrained terminal regions (residues L885-L982 and C343-L400), and those fluctuations are very likely

introduced by the simulation setup. However, calculations of the overall number of hydrogen bonds showed that the AlphaFold model was considerably more stabilised by intramolecular H-bonds than the homology model. On average, the analysis showed 75 ± 4 H-bonds in the AlphaFold model, compared to 60 ± 6 H-bonds within the homology model of PAS3. The analysis of potential energies for both models favoured the AlphaFold model: -5000 ± 300 kJ/mol on average, compared with -3000 ± 400 kJ/mol for the homology model. These results strongly indicate that the existence of a third PAS domain within PASK is very likely, since its prone to fold into a highly energetically favorable conformation with a high number of well organised internal hydrogen bonds.

The MD simulations of both PAS3 models show that the low-confidence model calculated by AlphaFold, comprised of two non-consecutive sequence strands, is more plausible. Nevertheless, we subjected both PAS3 models to the “druggability” assessment and subsequent further analysis.

To assess the quality of our simulated ensembles and to evaluate the extent of sampling at 500 ns time scale, we compared the simulation results of the PAS1 domain with the ensemble derived from the solution NMR data available for this domain (Amezcuca *et al.*, 2002). We found the overall trends observed in both ensembles to be consistent, with Pearson correlations between RMSF calculated from each 3 MD simulations replicas and NMR ensemble of 0.88 (replica 1), 0.71 (replica 2), and 0.77 (replica 3). Residues from helix F and loop FG (residues 175-200) were highly flexible in both ensembles (Suppl. Fig. S2). Residues 185-194 and termini showed the highest root-mean-square fluctuations in our MD simulations at 500 ns time scale. Amezcuca *et al.* (2002) showed that these residues exhibited elevated values of the spectral density function $J(\omega)$ at high frequencies [$J(0.87\omega_H)$] and correspondingly decreased values at lower frequencies [$J_{\text{eff}}(0)$], indicating that these residues are flexible on fast (ps-to-ns) timescale. Residues 175–184 and 195–200, at both ends of the FG loop, showed average $J(0.87\omega_H)$ and elevated $J_{\text{eff}}(0)$ values, which indicates that these regions undergo conformational changes on a slower (μs -ms) timescale. Interestingly, PCA analysis of our 500 ns simulations indicates slow, correlated motions within these regions of PAS1 (Suppl. Fig. S3). Collectively, this is consistent with available NMR data and validates our methodology.

As the linker connecting PAS1 and PAS2 domains is very short (3-4 residues), a segment containing both domains and the linker was subjected to the same procedure (500 ns MD simulation, followed by PCA and cluster analysis). The results are consistent with the simulations of isolated PAS1 and PAS2 domains (Suppl. Figs. S6-S8) and indicate the propensity of those two domains to change their orientation, but not to the ‘collapse’ of PAS1 on PAS2, which could be attributed to a very short linker (Suppl. Fig. S9). Furthermore, the intrinsic flexibility of a linker allows domains to re-orientate themselves upon binding to the kinase domain (Fig. 4).

Simulations of the PD4 region indicated that this region is not fully unfolded, as suggested by AlphaFold (Fig. 3c), and retains a residual fold comprised of 3 beta-sheets (residues 552-590), resembling of SH3-like domain structure. However, the biological significance of this putative domain is uncertain and requires further study, e.g. deletion of this region from the recombinant construct and assessment of the protein’s function.

In summary, our approach indicates the presence of two consensus N-terminal PAS domains in addition to the PAS1 domain, which has been experimentally

confirmed resolved by the solution NMR (Amezcuca *et al.*, 2002). The third PAS domain is likely to be composed of two non-consecutive sequence segments, as suggested by AlphaFold. Nevertheless, an alternative, consecutive model, calculated by homology modelling techniques, seemed to be stable. Both models share common structural features (Fig. 3a and b), which can be used in further analysis of interactions between this domain and other domains of PASK. In addition, established models allow for the follow-up experimental work, which may include expressing different protein constructs and evaluating the function of those domains in cell models.

Interactions between PAS and kinase domains

To explore possible interactions between all three PAS domains and the kinase domain within human PASK, we proceeded with the sequential protein-protein docking calculations, as described in Materials and Methods. Because of several constraints, such as distances between PAS1-PAS2 and PAS2-PAS3 domains, this approach narrowed down and generated 19 potential models, as shown in supplementary information (Fig. S11 and Tab. ST1). The models were cross-checked with the docked ternary complexes of three PAS domains (PAS1-PAS3) and poses that ranked highly in both calculations were prioritised for further analysis.

Even though the predicted energy scores are approximate and thus retain only indicative values, and models are highly speculative, the mechanism in which PAS domains are likely to modulate the PASK kinase domain has emerged. The results of protein-protein docking calculations indicate three binding “hot spots” at the surface of the kinase domain, where the PAS domains are likely to bind (Fig. 5). Those “hot spots” were recurrent as highly-ranked binding sites, retrieved by all docking runs. Structural constraints within PASK, such as very short PAS1-PAS2 linker and short PAS2-PAS3 loop, favoured certain complexes over others.

Our results suggest that all three PAS domains may be involved in interactions with the kinase domain simultaneously (Fig. 4). This model is supported by available mutagenesis data on PASK residues regulated by autophosphorylation: S1149, T1161, T1165, S1287, S1289, S1273, S1277, and S1280 (Kikani *et al.*, 2010). PAS1 interacts with T1161, PAS2 interacts with T1165 and S1149, while PAS3 interacts with residues downstream R1267. This is shown in Fig. 4. Residues in PAS domains predicted to be important in maintaining interactions with the kinase domain are listed in Supp. Tab. ST2. In future work, we plan to mutate those residues to further validate our model.

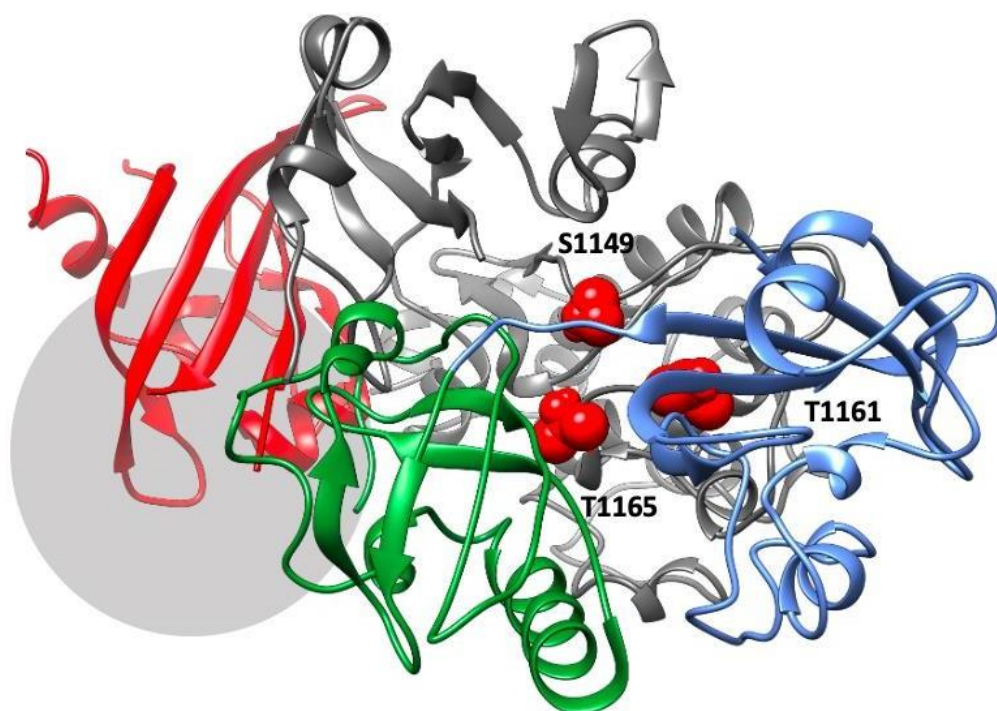


Figure 4 The best-ranking model of PAS domains interacting with the kinase domain of PASK. PASK domains are coloured as follows: PAS1 – blue; PAS2 – dark green; PAS3 – red, kinase domain – dark grey. Residues S1149, T1161 and T1165, which are autophosphorylated in the activated kinase, are coloured red and highlighted as spheres. The region downstream R1267, which contains several phosphorylated serine residues including S1280, S1287 and S1289, is marked by a light grey circle.

Modulation of PASK by small molecule ligands

A novel crystallographic screening method has been developed and reported recently by O'Reilly *et al.* (2019). In this approach, the high-concentration aqueous soaks were made with a chemically diverse and ultra-low-molecular-weight fragment library denoted as “MiniFrag” (heavy atom count 5–7). This allowed the identification of “druggable” sites on proteins of interest. High screening hit rates reflected enhanced sampling of chemical space. MiniFrag screening can represent thus a highly effective method for “druggability” assessment and guiding optimisation of fragment-derived lead compounds. Such an approach is applicable for *in silico* screening. In this work, we tested it by employing fast ensemble docking of an enriched version of the MiniFrag library (AugMiniFrag; 94 compounds). The original MiniFrag library, which contains a set of chemically-diverse fragments obtained from Enamine (www.enamine.net), was been augmented by fragments derived from amino acids known to modulate proteins (e.g. arginine), sugars, free fatty acids, and small drug-like molecules (see Materials and Methods).

The set was used in molecular docking against all binding sites detected by FTMap and validated by the binding site mapping mode within SeeSAR (www.biosolveit.de). First, we used all individual domains as targets, using ensembles derived from PCA and cluster analysis of MD simulation trajectories as inputs. Next, we probed interfacial pockets, using outputs from MD simulations of the PAS1-PAS2 segment and results of protein-protein

docking (PAS-kinase domains) as inputs.

Results are shown in Fig. 5 and Suppl. Tab. ST3. All PAS domains under investigation were found to be “druggable” by small molecules using FTMap and SeeSAR, including both alternative models of PAS3. The major difference between the two models of PAS3 is the nature of the binding site at the inner cavity of the PAS domain: while the binding site within PAS3 predicted by AlphaFold is very hydrophobic (all side chains but Y894, lining the central cavity, are hydrophobic), the model calculated by homology modelling has a more polar binding cavity (residues Y396, S399, N413, R420, Q428, D432, and K473; Suppl. Fig. S12), able to bind a broad spectrum of chemically diverse small molecules, including phosphorylated sugars and side chains of free amino acids (Suppl. Tab. ST3). We suggest that future studies involving mutation of those residues may contribute to validation of the PASK model and further our understanding of its regulation by small molecules.

Inner pockets of PAS1 and PAS2 have a propensity to bind hydrophobic ligands, however, additional sites were detected on those domains, including the interfacial site at the protein-protein interaction (PPI) interface between both domains. Amezcua *et al.* (2002) reported nine compounds that bound PAS1 with dissociation constants smaller than 100 μM . Those compounds were hydrophobic in character, containing one or two aromatic rings substituted with polar groups around their peripheries, and separated by a short linker (0–2 heavy atoms). Molecular docking of those six compounds into the central cavity of PAS1 reproduced those measured K_d values well (p -value = 0.0029; Suppl. Fig. S16 and Tab. ST4). Furthermore, molecular docking of the enriched MiniFrag set identified 2 distinct binding sites on PAS1, additional to that central cavity: one of those sites was modulated by simple sugars (Suppl. Tab. ST3). Collectively, we identified >10 consensus “druggable hotspots” involving all three PAS domains of PASK. Interfacial binding sites were found on each PAS domain of PASK, which may be explored in the development of allosteric inhibitors of PASK in the future.

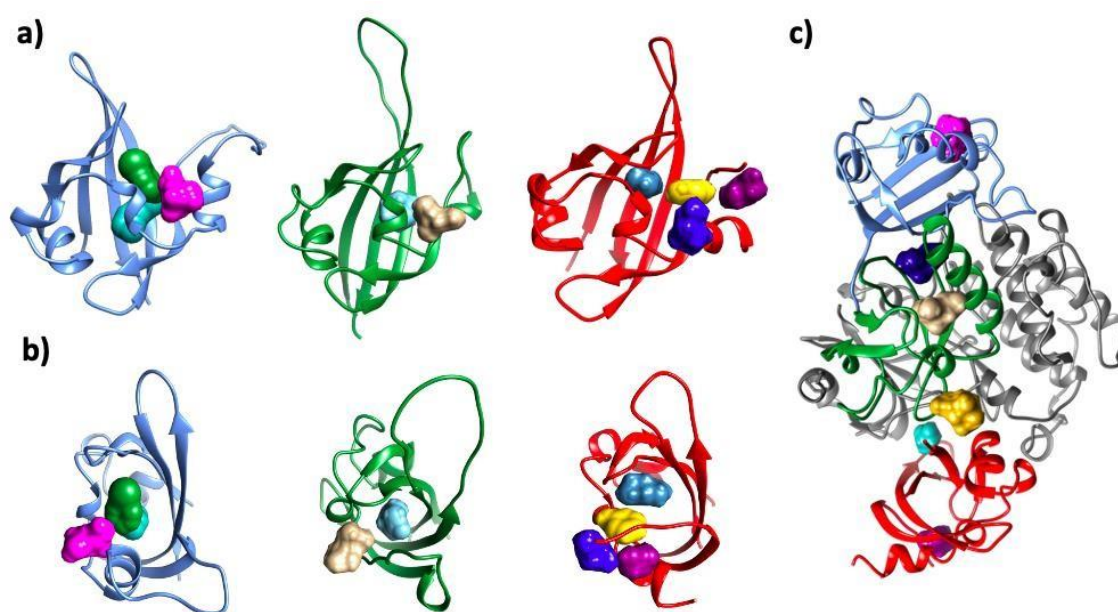


Figure 5 Consensus “hot-spots” identified on all three putative PAS domains and the interface between PAS and kinase domains. The following colours were used for each domain: PAS1 – blue; PAS2 – dark green; PAS3 – red; kinase domain – dim grey. For each PAS domain, two orientations are shown: side (a) and top (b). For the hotspots, the following colours were used: PAS1: H1 – bottle green; H2 – aqua; H3 – pink; PAS2: H4 – tan; H5 – cyan; PAS3: H6 – marine blue; H7 – yellow; H8 – electric blue; H9 – dark magenta. Kinase-PAS domains (c): H13 – navy blue; H14 – gold; H15 – turquoise. In panel c), location of probes H3 in PAS1 (pink); H4 in PAS2 (tan), and H9 in PAS3 (dark magenta) are shown for reference. Nomenclature of hotspots is consistent with Suppl. Tab. ST3.

We inspected the predicted interface between the kinase domain and PAS domains. Three “hot-spots” were detected at the interface between PAS and kinase domains, and the site comprised of three domains: PAS1, PAS2 and kinase domain (Fig. 5). These two sites were found to bind arginine and metformin. Metformin is a first-line therapy for the treatment of diabetes type 2 due to its robust glucose-lowering effects and well-established safety and tolerability (Diabetes Prevention Program Research Group, 2012). However, its mechanism of action is poorly understood. Its putative mechanisms of action include indirect activation of AMPK, inhibition of cAMP production, modulation of mTOR, blocking the action of glucagon, and activation of TAK1- $\text{IKK}\alpha/\beta$ signalling (Agius *et al.*, 2020). However, the main protein target of metformin remains unknown, and metformin’s activity may rely on modulating multiple targets. Since PASK is regulated by the activity of AMPK and mTOR (Hurtado-Carneiro *et al.*, 2014), it is plausible that it can be among the targets modulated by metformin, and its direct interactions should be followed up experimentally. If it were determined that PASK is indeed directly modulated by metformin, it would be beneficial to evaluate whether metformin would indeed the kinase as well as PAS domains of PASK, and whether such binding would disrupt the PPI interface, or make the interactions between the domains stronger.

Conclusions

In this work, we present results of an integrative computational study on human PASK kinase, combining structural bioinformatics, molecular modelling, all-atom MD simulations,

molecular docking and “hot-spot” mapping. Our results strongly suggest the presence of three consensus N-terminal PAS domains in human PASK kinase. The experimental structure of one of those domains has resolved by solution NMR and two other domains are predicted; future work will focus on experimental confirmation of those domains. All domains are predicted to be “druggable” by small molecules, albeit with different domains likely to be modulated by different classes of ligands. This is likely to contribute to PASK’s role as a versatile metabolic sensor. Predicted binding sites for diverse small molecules include sites within individual PAS domains as well as “druggable hot-spots” at domain-domain interaction interfaces. Proposed three-dimensional model of interactions between PAS domains and the kinase domain of PASK fits with experimental data available and consolidates the model of the kinase domain being inhibited by direct interaction with PAS domain; these interactions may be modulated by small molecules. Future work will focus on those small molecules and their binding sites, which may pave the way for future drug discovery in not only liver and metabolic disorders, but also other diseases influenced by oxidative stress, such as neurodegenerative disease. Increased understanding of PASK and its modulation by small molecules has the potential to profoundly impact healthcare research.

Acknowledgements

We are grateful to Dr S. Kometa (HPC Rocket) and Dr C. Wills (Newcastle University NMR Facility) for their technical assistance.

Author contributions

P.Z. and A.K.B. outlined the problem and conceptualised the study; J.V.d.S., K.S.M. and A.K.B. designed the study; L.F., A.K.B. and J.V.d.S. performed the bioinformatic analysis; R.Z.; H.J. and L.F. carried out the equilibrium atomistic molecular dynamics simulations of PAS1 and PAS2 domains; S.X., and P.Z. carried out the equilibrium atomistic molecular dynamics simulations of PAS3 and PD4 domains; J.V.d.S. and S.X. wrote scripts for data analysis; S.X., L.F., P.Z., R.Z., and H.J. extracted and analysed data from the simulations; R.Z., H.J., A.K.B., and K.S.M. performed molecular docking calculations; A.K.B. and K.S.M. supervised the study; all authors produced the figures, plots, and supplementary data. J.V.d.S. and A.K.B. wrote the first draft of the manuscript. K.S.M. edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by a EPSRC Ph.D. studentship to L.F. (EPR51209X1) and EPSRC funding to A.K.B. (EP/S022791/1).

Conflict of interest

The authors declare no conflict of interest.

References

Agius L, Ford BE and Chachra SS (2020) The Metformin Mechanism on Gluconeogenesis and

AMPK Activation: The Metabolite Perspective. *International Journal of Molecular Sciences* **21**, 3240.

Amezcuca CA, Harper SM, Rutter J and Gardner KH (2002) Structure and interactions of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation. *Structure* **10**, 1349-1361.

Arnaiz E, Miar A, Bridges E, Prasad N, Hatch SB, Ebner D, Lawrie CH and Harris AL (2021) Differential effects of HIF2 α antagonist and HIF2 α silencing in renal cancer and sensitivity to repurposed drugs. *BMC Cancer* **21**, 896.

Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A and Haak JR (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690.

Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A and Finn RD (2021) The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* **49**, 344-354.

Buchan DW, Minneci F, Nugent TC, Bryson K and Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**, 349-357.

Darden T, York D and Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092.

Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE (1999) *Angew. Chem. Int. Ed* **38**, 236-240.

de Souza JV, Reznikov S, Zhu R and Bronowska AK (2019) Druggability assessment of mammalian Per–Arnt–Sim [PAS] domains using computational approaches. *MedChemComm* **10**, 1126-1137.

Desta IT, Porter KA, Xia B, Kozakov D and Vajda S (2020) Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure* **28**, 1071-1081.e1073.

Diabetes Prevention Program Research Group (2012) Long-term safety, tolerability, and weight loss associated with metformin in the Diabetes Prevention Program Outcomes Study. *Diabetes care* **35**, 731-737.

Dongil P, Pérez-García A, Hurtado-Carneiro V, Herrero-De-Dios C, Álvarez E and Sanz C (2020) PAS kinase deficiency reduces aging effects in mice. *Aging* **12**, 2275-2301.

Dongil P, Pérez-García A, Hurtado-Carneiro V, Herrero-De-Dios C, Blazquez E, Alvarez E and Sanz C (2018) Pas Kinase Deficiency Triggers Antioxidant Mechanisms in the Liver. *Scientific reports* **8**, 13810-13810.

Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, Dotson DL, Domanski J, Buchoux S, Kenney IM and Beckstein O (2016) MDAAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. *Python in Science Conference* **15**, 98-105.

- Greb-Markiewicz B and Kolonko M (2019) Subcellular Localization Signals of bHLH-PAS Proteins: Their Significance, Current State of Knowledge and Future Perspectives. *International Journal of Molecular Sciences* **20**, 4746.
- Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT and White O (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* **29**, 41-43.
- Hess B, Bekker H, Berendsen HJC and Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472.
- Hurtado-Carneiro V, Dongil P, Pérez-García A, Álvarez E and Sanz C (2021) Preventing Oxidative Stress in the Liver: An Opportunity for GLP-1 and/or PASK. *Antioxidants (Basel, Switzerland)* **10**, 2028.
- Hurtado-Carneiro V, Pérez-García A, Alvarez E and Sanz C (2020) PAS Kinase: A Nutrient and Energy Sensor “Master Key” in the Response to Fasting/Feeding Conditions. *Frontiers in endocrinology* **11**, 594053-594053.
- Hurtado-Carneiro V, Roncero I, Egger SS, Wenger RH, Blazquez E, Sanz C and Alvarez E (2014) PAS Kinase Is a Nutrient and Energy Sensor in Hypothalamic Areas Required for the Normal Function of AMPK and mTOR/S6K1. *Molecular Neurobiology* **50**, 314-326.
- Janssens LK and Stove CP (2021) Sensing an Oxygen Sensor: Development and Application of Activity-Based Assays Directly Monitoring HIF Heterodimerization. *Analytical Chemistry* **93**, 14462-14470.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589.
- Kelley LA, Mezulis S, Yates CM, Wass MN and Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845-858.
- Kikani CK, Antonysamy SA, Bonanno JB, Romero R, Zhang FF, Russell M, Gheyi T, Iizuka M, Emtage S, Sauder JM, Turk BE, Burley SK and Rutter J (2010) Structural bases of PAS domain-regulated kinase (PASK) activation in the absence of activation loop phosphorylation. *J Biol Chem* **285**, 41034-41043.
- Kikani CK, Wu X, Fogarty S, Kang SAW, Dephore N, Gygi SP, Sabatini DM and Rutter J (2019) Activation of PASK by mTORC1 is required for the onset of the terminal differentiation program. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 10382-10391.
- Kozakov D, Brenke R, Comeau SR and Vajda S (2006) PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure* **65**, 392-406.

Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D and Vajda S (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **10**, 733-755.

Kozakov D, Hall DR, Xia B, Porter KA, Padhorney D, Yueh C, Beglov D and Vajda S (2017) The ClusPro web server for protein–protein docking. *Nature Protocols* **12**, 255-278.

Letunic I, Khedkar S and Bork P (2021) SMART: recent updates, new developments and status in 2020. *Nucleic Acids Research* **49**, 458-460.

Lewis TE, Sillitoe I, Andreeva A, Blundell TL, Buchan DW, Chothia C, Cozzetto D, Dana JM, Filippis I, Gough J, Jones DT, Kelley LA, Kleywegt GJ, Minneci F, Mistry J, Murzin AG, Ochoa-Montaño B, Oates ME, Punta M, Rackham OJ, Stahlhacker J, Sternberg MJ, Velankar S and Orengo C (2015) Genome3D: exploiting structure to help users understand their sequences. *Nucleic Acids Res* **43**, 382-386.

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO and Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958.

Lobley A, Sadowski MI and Jones DT (2009) pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* **25**, 1761–1767.

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ and Marchler-Bauer A (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* **48**, 265-268.

Michaud-Agrawal N, Denning EJ, Woolf TB and Beckstein O (2011) MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem* **32**, 2319-2327.

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD and Bateman A (2021) Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, 412-419.

Mullard A (2021) The Nobel Prize-winning path from oxygen sensing research to new drugs. *Nat Rev Drug Discov* **20**, 732-733.

O'Reilly M, Cleasby A, Davies TG, Hall RJ, Ludlow RF, Murray CW, Tisi D and Jhoti H (2019) Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design. *Drug Discov Today* **24**, 1081-1086.

Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182.

Poch A, Villanelo F, Henriquez S, Kohen P, Muñoz A, Strauss JF3RD and Devoto L (2019) Molecular modelling predicts that 2-methoxyestradiol disrupts HIF function by binding to the PAS-B domain. *Steroids* **144**, 21-29.

Reulecke I, Lange G, Albrecht J, Klein R and Rarey M (2008) Towards an Integrated Description

of Hydrogen Bonding and Dehydration: Decreasing False Positives in Virtual Screening with the HYDE Scoring Function. *ChemMedChem* **3**, 885-897.

Sabanes FZ, De Souza JV and Bronowska AK (2019) Cosolvent Analysis Toolkit (CAT): a robust hotspot identification platform for cosolvent simulations of proteins to expand the druggable proteome. *Scientific Reports* **9**, 19118.

Shi J, Blundell TL and Mizuguchi K (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* **310**, 243-257.

Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L and Xenarios I (2013) New and continuing developments at PROSITE. *Nucleic Acids Res* **41**, D344-7.

Smith DA and Jones RM (2008) The sulfonamide group as a structural alert: A distorted story? *Current Opinion in Drug Discovery & Development* **11**, 72-79.

Soshilov AA, Motta S, Bonati L and Denison MS (2020) Transitional States in Ligand-Dependent Transformation of the Aryl Hydrocarbon Receptor into Its DNA-Binding Form. *International Journal of Molecular Sciences* **21**, 2474.

Swiatek W, Parnell KM, Nickols GA, Scharschmidt BF and Rutter J (2020) Validation of PAS Kinase, a Regulator of Hepatic Fatty Acid and Triglyceride Synthesis, as a Therapeutic Target for Nonalcoholic Steatohepatitis. *Hepatol Commun* **4**, 696-707.

Uniprot-Q96RG2. PAS domain-containing serine/threonine-protein kinase, Homo sapiens (Human). Available at <https://www.uniprot.org/uniprot/Q96RG2>.

Verlet L (1967) Computer "Experiments" on Classical Fluids. I. *Thermodynamical Properties of Lennard-Jones Molecules*. *Physical Review* **159**, 98-103.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJC (2005) GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701-1718.

Wang YW, He SJ, Feng X, Cheng J, Luo YT, Tian L and Huang Q (2017) Metformin: a review of its potential indications. *Drug Des Devel Ther* **11**, 2421-2429.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R and Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303.

Zhang DD, Zhang JG, Wang YZ, Liu Y, Liu GL and Li XY (2015) Per-Arnt-Sim Kinase (PASK): An Emerging Regulator of Mammalian Glucose and Lipid Metabolism. *Nutrients* **7**, 7437-7450.

Zhang DD, Zhang JG, Wu X, Liu Y, Gu SY, Zhu GH, Wang YZ, Liu GL and Li XY (2015) Nuciferine downregulates Per-Arnt-Sim kinase expression during its alleviation of lipogenesis and inflammation on oleic acid-induced hepatic steatosis in HepG2 cells. *Frontiers in Pharmacology* **6**, 238.