

SCIENTIFIC REPORTS

OPEN

Unsaturated fatty acids as high-affinity ligands of the C-terminal Per-ARNT-Sim domain from the Hypoxia-inducible factor 3 α

Received: 26 March 2015

Accepted: 07 July 2015

Published: 03 August 2015

Angela M. Fala^{1,*}, Juliana F. Oliveira^{1,*}, Douglas Adamoski¹, Juliana A. Aricetti², Marília M. Dias¹, Marcio V. B. Dias^{1,†}, Maurício L. Sforça¹, Paulo S. Lopes-de-Oliveira¹, Silvana A. Rocco¹, Camila Caldana^{2,3}, Sandra M. G. Dias^{1,†} & Andre L. B. Ambrosio^{1,†}

Hypoxia-inducible transcription factors (HIF) form heterodimeric complexes that mediate cell responses to hypoxia. The oxygen-dependent stability and activity of the HIF- α subunits is traditionally associated to post-translational modifications such as hydroxylation, acetylation, ubiquitination, and phosphorylation. Here we report novel evidence showing that unsaturated fatty acids are naturally occurring, non-covalent structural ligands of HIF-3 α , thus providing the initial framework for exploring its exceptional role as a lipid sensor under hypoxia.

HIF, closely linked to disorders of the circulatory system and cancer progression¹, consists of two multidomain subunits including the constitutively expressed HIF-1 β and an oxygen-labile HIF-1 α (or paralogs HIF-2 α and -3 α). The gene encoding for HIF-3 α (*HIF3A*) is exclusively susceptible to post-transcriptional processing in humans, with six isoforms shown to exist at mRNA levels and three already confirmed as functional protein^{2,3}. HIF-3 α 9, the longest (669 aa) and canonical isoform³, is a potent oxygen-labile transcription factor, tightly regulating the expression of a unique set of genes⁴. HIF-3 α 4 (363 aa), on the other hand, while insensitive to intracellular oxygen levels and incapable of transactivation, is a dominant-negative regulator of HIF-1 α ⁵. Due to the use of different transcription initiation sites and a combination of splicing events in *HIF3A*, the C-terminal Per-Arnt-Sim sensor domain (PASb) is the only common architectural feature among the HIF-3 α isoforms³. The PASb domains in both HIF- α and - β subunits have a crucial role in forming active HIF heterodimers and recruiting co-regulators^{6,7}.

PAS members are versatile sensory modules of stimuli such as oxygen tension, redox state and light intensity through binding to cofactors such as heme, flavin adenine dinucleotide and even divalent metal ions⁸. In the case of the HIF proteins, PAS domains-mediated protein-protein interactions have been shown independently of a cofactor and simply based on the presence of the partner PAS^{9,10}. However, recently, the identification of a 290 Å³ polar cavity in HIF-2 α PASb domain led to the proposition that endogenous ligands may exist¹¹; this cavity was later exploited for the selection of artificial small molecules that disrupt the full-length HIF-2 heterodimer^{10–12}.

Hereby, we present the crystal structure of the recombinant human HIF-3 α PASb, or PASb-3 α , at 1.15 Å maximum resolution (Table 1 and Supplementary Fig. 1). Although with a backbone virtually identical to that of the PASb domains of both HIF-1 α ⁹ and HIF-2 α ¹⁰ (Fig. 1a), PASb-3 α has a unique

¹Laboratório Nacional de Biotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil, 13083-100. ²Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil, 13083-100. ³Max Plack-partner group at the Laboratório Nacional de Ciência e Tecnologia do Bioetanol. ⁴Current address: Institute of Biomedical Science, University of São Paulo, São Paulo, SP, 05508-000, Brazil. ^{*}These authors contributed equally to this work. [†]These authors jointly supervised this work. Correspondence and requests for materials should be addressed to S.M.G.D. (email: sandra.dias@lnbio.cnpem.br) or A.L.B.A. (email: andre.ambrosio@lnbio.cnpem.br)

The C-terminal Per-Arnt-Sim (PASb) domain of human HIF-3 α bound to 18:1-1-monoacylglycerol, PDB ID 4wn5			
Data Collection		Model Refinement	
Beamline	P13 at PETRA III	Resolution range (Å)	28.5–1.15
Wavelength (Å)	0.9770	Reflections (cross-validation)	68777 (5952)
Space group	P2 ₁ 2 ₂ 1	R _{factor} /R _{free} (%)	12.2/14.1
Cell parameters a, b, c (Å)	52.7, 53.9, 67.8	Average B-factor (Å ²)	
Resolution range (Å)	28.5 – 1.15 (1.17 – 1.15)	main chain (no. of atoms)	8.8 (870)
Unique reflections	68827 (3381)	side chain (no. of atoms)	13.2 (812)
Multiplicity	9.8 (9.1)	ligands (no. of atoms)	19.9 (155)
R _{pim} (%)	1.9 (5.2)	solvent (no. of atoms)	28.0 (295)
CC (½)	0.998 (0.987)		
Completeness (%)	99.3 (99.4)	R.m.s.d. from standard geometry	
<I/σ(I)>	22.8 (11.5)	bond length (Å)	0.01
Average mosaicity (°)	0.5	bond angles (°)	1.36
B-factor Wilson Plot (Å ²)	6.7		
Monomers/AU	2	Ramachandran plot,	
Solvent content (%)	36.1	favoured (%)	100
Matthews coeff. (Å ³ /Da)	1.92		

Table 1. X-ray crystallography data collection parameters and structure refinement statistics. Data for outer shell are shown in parentheses.

extensive C-shaped hydrophobic cavity of 510 Å³ (Fig. 1b). We identified a 1-(11Z-octadecenoyl)-sn-glycerol molecule of bacterial nature buried in this cavity, as clearly suggested by the size and geometrical features of the electron density map (Fig. 1c, left panel). As such, the aliphatic tail of the fatty acid extended to eighteen carbons in length (maximum length limited by the cavity size), with a *cis* double bond between C11 and C12, interspersed by four gauche configurations (at positions C5, C7, C9 and C13, (Fig. 1c, right panel)). The ester group at the glycerol sn-1 position makes polar contact with the main chain carboxyl group of His339 and main chain amino group of Leu340. The sn-2 position is not substituted, also sharing a hydrogen bond with the Leu340 main chain amino group (Fig. 1d). Based solely on the electron density map, we could not tell whether the glycerol sn-3 position was substituted, due to absence of well-structured atoms beyond its remaining hydroxyl group. However, the presence of a well-coordinate sulfate ion (abundant in the crystallization solution), 6 Å apart from the glycerol sn-3 hydroxyl, suggested that PASb-3 α could also bind to phospholipids (Fig. 1d).

Aiming at experimental confirmation of the identity of the proposed monoacylglycerol, the hydrophobic fraction was carefully isolated from pure protein samples in solution and analyzed by thin-layer chromatography (TLC). By applying two mobile phases distinct in polarity, as well as dual staining to allow for differential detection of neutral and phosphorous-containing lipids¹³, we surprisingly observed the dominant presence of two neutral lipids and two phospholipids as ligands of PASb-3 α (Fig. 2a).

Accordingly, mass spectrometry analysis on the isolated hydrophobic fraction readily identified two differential peaks with *m/z* of 478.40 and 509.36 (Supplementary Fig. 2a,b), subsequently determined to be the lyso-phospholipids 1-(11Z-octadecenoyl)-2-hydroxy-sn-glycero-3-phosphoethanolamine, or 18:1-LPE, and 1-(11Z-octadecenoyl)-2-hydroxy-sn-glycero-3-phospho-(1'-sn-glycerol), or 18:1-LPG, respectively (Fig. 2a), in accordance with the TLC analysis. Experimental limitations related to the ionization of the neutral lipids kept us from identifying the precise nature of such.

Collectively, all complex lipids identified as PASb-3 α ligands include the 18-carbon unsaturated 11Z-octadecenoic acid (or *cis*-vaccenic acid; 18:1, *cis*-11) and an unmodified sn-2 glycerol position as common structural denominators. Given the heterologous nature of the expression system used for human PASb-3 α production, it is key to mention that *cis*-vaccenic is the sole 18-carbon unsaturated fatty acid found in *E. coli*¹⁴.

Next, in order to quantitatively check whether the degree of unsaturation influenced the binding of 18-carbon fatty acids to PASb-3 α , we performed a fluorescence-based assay in the presence of the hydrophobic probe 1,8-ANS, upon the titration of fatty acids to the delipidated protein. Intriguingly, upon fatty acid titration, no clear dose-response curve was observed for stearic acid (18:0), while sigmoid curves could be fitted in response to both oleic (18:1; *cis*-9) and linoleic acids (18:2; *cis*-9,12) (Fig. 2b). An apparent *K_d* of 39 nM for the monounsaturated oleic acid was determined, which improved 50-fold (up to 0.8 nM) for the polyunsaturated linoleic acid. Therefore, in addition to showing that 18-carbon fatty acids that occur in human cells, such as oleic and linoleic acids, are high-affinity ligands of PASb-3 α , we also establish that unsaturation is a required feature for fatty acid binding and the greater the degree of unsaturation, the stronger the interaction with the protein becomes.

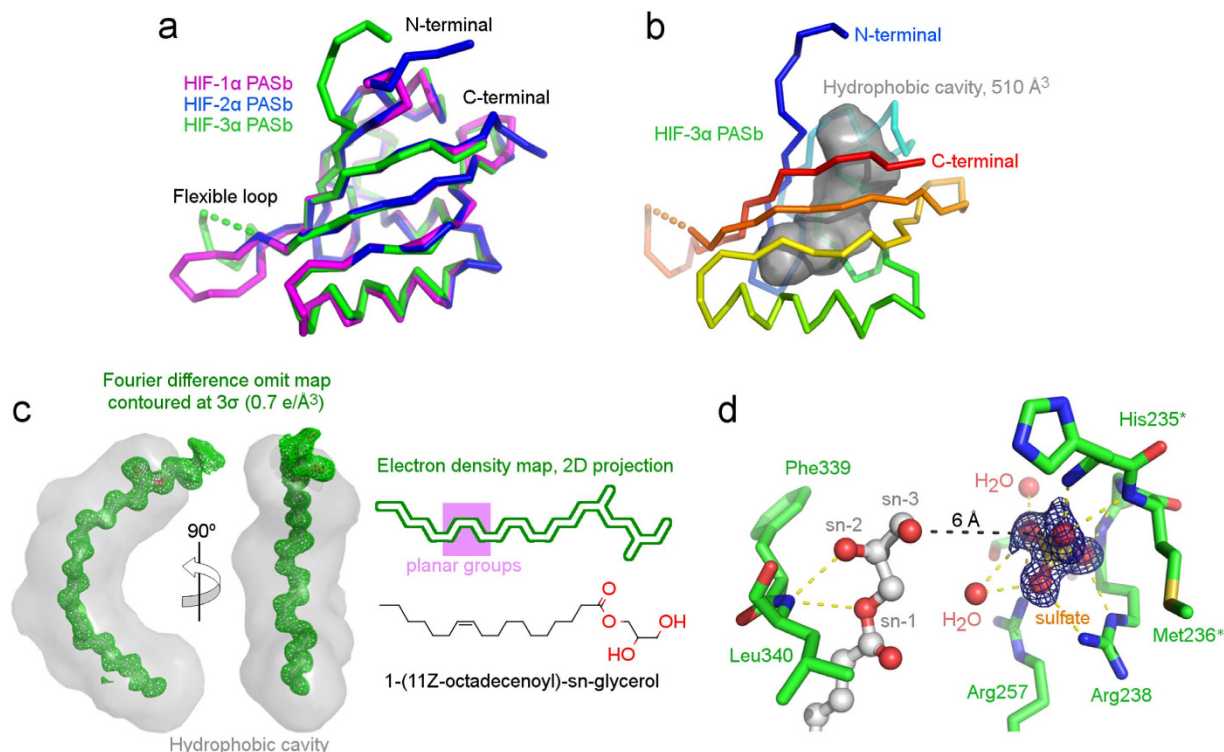


Figure 1. Structure of lipid-bound PASb-3α. (a) High structural identity between the main chain tracing of PASb-3α and the PASb domains of both HIF-1α and HIF-2α (similar core r.m.s.d. of 0.6 Å). (b) However, a 510 Å³ hydrophobic cavity was exclusively identified in the core of PASb-3α (delimited by gray surface), and (c) found to enclose a monoacylglycerol molecule. (d) Extensive network of polar contacts (yellow dashed lines) shared by specific residues in PASb-3α with the lipid and a neighbor sulfate ion, suggesting that minor rearrangements in the protein may allow for binding of phospholipids.

The mechanism by which the lipid molecule enters the PASb-3α pocket is unclear. However, the outcomes of the binding assay above can be reasoned from the perspective of the free fatty acids, since the ligand conformations in the bound state are dependent on the frequency of such microstates as part of the unbound ensemble in vacuum (thus simulating lipid desolvation). Upon ensemble recapitulation, it is curious to observe that, while the putative bound conformation of stearic acid has the lowest conformational enthalpy inside the pocket (−3 kcal/mol), such a conformation has very low probability of happening upon desolvation (Fig. 2c). On the other hand, the bound conformations of oleic and linoleic acids, although with higher enthalpies (4 and 37 kcal/mol, respectively), are more frequent and easily achieved since the estimated ensemble entropies are significantly smaller than that of stearic acid (Fig. 2c).

NMR and fluorescence polarization data showed that purified, lipid-associated PASb-3α, has stronger association with HIF-1β PASb, albeit with slower binding kinetics, when compared to the well-characterized HIF-1α:HIF-1β PASb dimer (Supplementary Fig. 3a–e). In hopes of characterizing exactly how the absence of lipids could affect the heterodimerization of PASb-3α to HIF-1β PASb, we performed extensive attempts to remove the accompanying bacterial lipids from the purified protein by conventional hydrophobic interaction chromatography, reversible unfolding and detergents, all of which failed. Limited success was attained after treatment with the LIPIDEX-1000 matrix allowing, however, the recovery of only minor amounts of protein¹⁵. This observation may point to the role of the lipids as a structural cofactors for PASb-3α, which is not without precedents, since some ligand-free PAS have already been shown to be predominantly unstructured in solution and prone to precipitation^{16,17}. A fitting example is the association between the ligand-free PASb domain of the Aryl Hydrocarbon Receptor (AHR), with a predicted similarly large hydrophobic cavity (500 Å³) used for ligand binding, and the chaperone *Heat shock protein 90* (Hsp90). The presence of compounds known to activate the AHR response pathway displaces Hsp90 from this complex, suggesting that these chemicals stabilize a folded form of the PASb domain¹⁸. Therefore, we suggest that the binding of lipids to PASb is mandatory for the proper folding and stability of HIF-3α, inducing its translocation to the nucleus and subsequent activation of HIF function. This hypothesis is strongly supported by the direct observation that HIF-3α

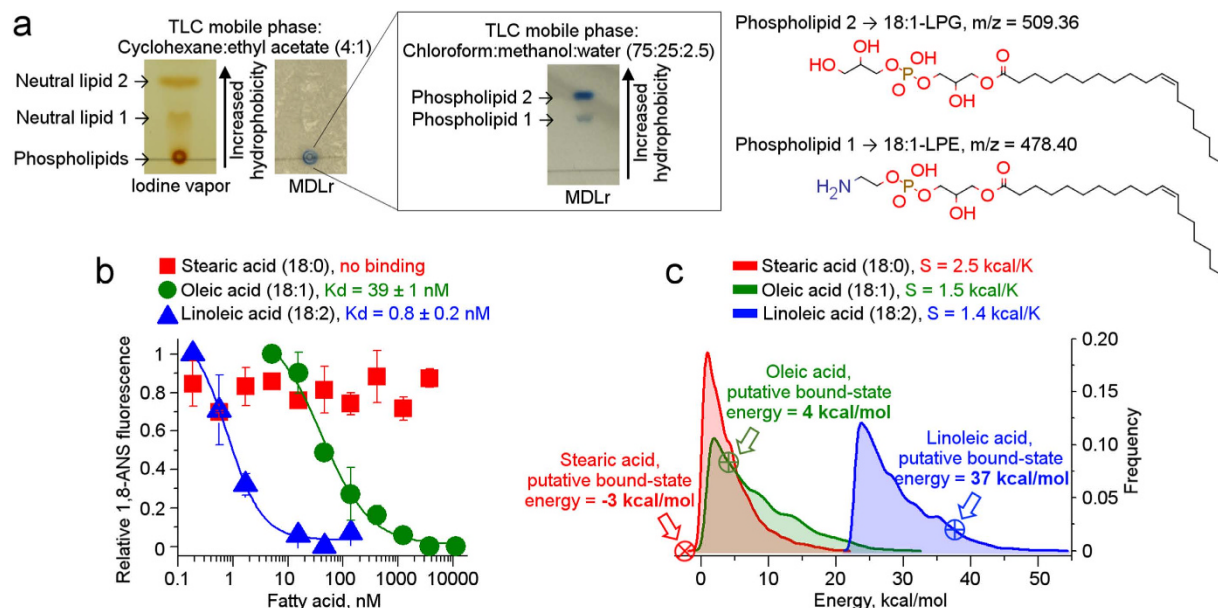


Figure 2. Characterization of the lipids bound to PASb-3 α . (a) Two neutral lipids (yellow bands on the iodine vapor stained plate) and two phosphorous-containing lipids (blue bands over white background) were carried over by the purified protein samples. The two phospholipids were identified as 18:1-lysophospholipids. (b) Free oleic and linoleic acids bind with low-nanomolar affinity to delipidated PASb-3 α , suggesting that unsaturation is a mandatory feature for the partner fatty acid and, as unsaturation accumulates, the higher the affinity for the protein. (c) Boltzmann distributions representing the canonical ensembles of fatty acid desolvation. The estimated entropies decrease as the degree of unsaturation increase. Crossed-circles over the curves point out the frequency of putative bound-state enthalpy into the distribution.

indeed co-precipitates with HSP90 in cell models, likely via its PASb domain¹⁹. However, the dependence of this interaction on the presence of lipids has yet to be tested.

Curiously, the splicing event that gives rise to the HIF-3 α 4 isoform completely changes the amino acid sequence for the second half of its PASb domain³ (Supplementary Fig. 4a). *Ab initio* homology modelling of PASb-3 α 4 strongly suggest that amino acid changes relative to PASb-3 α (Supplementary Fig. 4a), result in substantial deformation of the topography of the hydrophobic cavity as a consequence of the presence of longer, bulkier and polar side chains inside the cavity (Supplementary Fig. 4b). These observations strongly suggest that, out of all HIF-3 α isoforms, the splicing variant HIF-3 α 4 - which is uniquely insensitive to cellular oxygen levels⁵ - may as well be exclusively insensitive to lipids. Of note, HIF-3 α 4 is a negative regulator of HIF-1⁵. Extensive attempts to produce PASb-3 α 4 domain in bacterial system only resulted in insoluble protein products, hindering the testing of our hypothesis.

With all the possible scenarios described above several outstanding questions arise. These include the identification of which lipids may in fact regulate HIF-3 α function, for instance in cancer (Supplementary Figs. 5a–b and 6a–d) or obesity²⁰, under what exact circumstances lipids must be sensed and what are the possible outcomes of lipid binding. However, by defining PASb-3 α as a lipid-binding domain, our findings lay the groundwork to explore the role of HIF-3 α as a link coupling lipid sensing under hypoxia, thus adding an unanticipated layer of complexity to the already intricate mechanism for HIF-mediated gene expression. Finally, similarly to HIF-2 α , the description of an exclusive hydrophobic pocket yield unique implications for the future development of small molecule-oriented therapeutics targeting specifically HIF-3 α .

Collectively, our structural and biochemical observations also carry implications for the overall molecular aspects of the PAS fold. For the first time, we document that a eukaryotic PAS domain may have lipids as high-affinity cofactors, thus introducing a new class of PAS ligands and expanding its already vast functional portfolio⁸.

Methods

Cloning. PASb-1 α (Asp238-His348, UniProt Q16665), PASb-1 β (Pro349-Ser467, UniProt P27540) and PASb-3 α (Gly237-Thr345, UniProt Q9Y2N7-1) coding sequences were amplified by Polymerase Chain Reaction from a cDNA library from the breast cancer cell line SKBR3. PASb-1 α and PASb-3 α constructs were cloned into the pET28a plasmid (Novagen) using the *NdeI* and *XhoI* restriction sites. PASb-1 β was cloned into a modified version of pETSUMO vector (Invitrogen) using *BamHI* and *HindIII*

restriction sites. The following pairs of oligonucleotides were used: PASb-1 α – Forward: 5'-AGTCTCA TATGGATAGCAAGACTTTCCTCAGTCGA-3' and Reverse: 5'-CGTACCTCGAGTTAGTGCTGAATA ATACCACTCAC-3'; PASb-1 β – Forward: 5'-TTGGTGGATCCCCCACTGTACAGACATGAGTAAT-3' and Reverse: 5'-GCCGCAAGCTTTTAAGAGTTCTTCACATTGGTGT-3'; PASb-3 α – Forward: 5'-AGTCTCATATGGGCCGAGGGGCCTTCCTCAGCCGC-3' and Reverse: 5'-CGTACCTCGAGTTAGG TCTCTTCCACCTGGCTGAT-3'.

Expression and purification of recombinant proteins. The following expression and purification protocol was applied to the three Pasb constructs, transformed into *Escherichia coli* Rosetta-2 thermocompetent cells (Merck). Overnight cultures, grown in LB medium supplemented by 50 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin and 50 $\mu\text{g}\cdot\text{ml}^{-1}$ chloramphenicol, were inoculated in a ratio of 1:100 in 5 liter cultures supplemented with the same antibiotics and left shaking at 200 rpm for 5 hours at 37 °C. The cultures were then down-tempered to 18 °C for 1 hour before induction with 0.2 mM IPTG (isopropyl β -D-1-thiogalactopyranoside) for additional 16 hours at 18 °C. Cells were harvested by centrifugation (5,000 $\times g$) for 10 min at 4 °C and resuspended in buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 2 mM BME (β -mercaptoethanol) and 1 mM phenylmethylsulfonyl fluoride. Cell lysis was performed chemically by incubation with hen egg-white lysozyme, DNase I and deoxycholate (all three reagents from Sigma-Aldrich) for about 1 hour, incubated on ice. The soluble fractions were separated from the debris by high speed centrifugation (23,000 $\times g$) and subsequently loaded, by gravity, on Co²⁺-charged resin TALON (Clontech), previously equilibrated with the running buffer 50 mM Tris-HCl pH 7.5, 30 mM NaCl and 2 mM BME. The constructs, bearing a 6xHis-tag fused to their N-terminus, were eluted step-wise using running buffer to which up to 500 mM imidazole (v/v) was added. The tags were removed by overnight digestion with either bovine thrombin (Sigma-Aldrich) or ULP-1 protease and the samples loaded into a MonoQ 5/50 GL anion exchange chromatography column (GE Healthcare). Elution was done by performing a linear gradient with a buffer containing 1 M NaCl, 30 mM Tris-HCl pH 7.5 and 2 mM BME. The fractions containing the protein of interest were loaded into a HiLoad 16/600 Superdex 75 size-exclusion column (GE Healthcare), equilibrated with 30 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.5 mM TCEP (tris(2-carboxyethyl)phosphine). Protein concentrations were determined by UV 280 nm using calculated extinction coefficients.

Crystallization. Size-exclusion purified PASb-3 α was concentrated down with an Amicon ultrafiltration device (10 KDa cutoff; Millipore) to a final concentration of approximately 300 μM . Crystallization experiments were performed at 291 K using the conventional sitting drop vapor diffusion technique. Drops were made by mixing equal parts of protein and well solution, containing 3% PEG 400, 1.9 M Amonium Sulphate, 0.1 M Sodium Acetate pH 5.7. Small bar-shaped crystals started growing two days after the setting up of the crystallization drops, reaching about 50 μM in longest dimension on average usually after 5 days. Before data collection at cryogenic temperature (100 K), harvested crystals were cryoprotected with 10% ethylene glycol added to the mother liquor (v/v).

X-ray crystallography. X-ray diffraction data sets were obtained at the P13 beamline at PETRA III, Hamburg, Germany. Data was processed using Mosflm²¹ and merged and scaled with Aimless²². The first set of phases was obtained by the molecular replacement as implemented in the program Phaser²³, using the PASb domain of HIF-1 α (PDB ID 4h6j⁹) as a search model. Positional and anisotropic B-factor refinement cycles were carried out with Phenix²⁴. Manual building of the extra portions and real space refinement, including Fourier electron density map inspection, were performed with Coot²⁵. Solvent water molecules, treated as oxygen atoms, were added using the appropriate Coot routine. The survey for cavities was done using KVFinder²⁶. During model refinement, the inspection of Fourier difference maps indicated the presence of very strong non-protein electron densities (over 3 σ in height in the $F_{\text{obs}} - F_{\text{calc}}$ Fourier maps) that were interpreted - due to their sizes and shapes and in respect with the crystallization condition - as two 1-11Z-vaccenoyl-sn-glycerol molecules, five PEG 400 molecules and two sulfate ions, all subsequently refined as such. Two additional strong electron density distributions also resembling PEG400 molecules were found on the vicinities of Tyr264 in chain A and Trp315 in chain B. The refinement of such molecules, however, did not converge after many modeling configurations and the densities were then left as unexplained. The overall stereochemical quality of the final models and the agreements between them and experimental data were assessed by the program Molprobity²⁷ and the appropriate Phaser and Coot routines.

Thin-layer chromatography. Chromatographic separations were carried out on 4 \times 2 cm TLC Silica gel 60 F₂₅₄ aluminum plates (Merck). The complex lipid mixture was applied on the edge of the plates (at the sample spotting line) using a glass capillary. The samples were allowed to run for 4 cm in a 8 \times 4 cm glass chamber, previously saturated with the mobile phase. Two distinct mobile phases were used - cyclohexane:ethyl acetate (4:1) and chloroform:methanol:water (75:25:2.5) - which respectively allow for the separation of neutral lipids and the separation of phospholipids by head group polarity. The plates were then left exposed to air to allow for evaporation of the mobile phase residual. For differential detection of neutral lipids and phosphorous-containing lipids, iodine vapor and a modified Dittmer-Lester reagent¹³ were employed for dual staining. The plates were scanned immediately after drying.

Mass spectrometry. Lipids were extracted from 400 μL of purified protein ($\sim 1 \text{ mg} \cdot \text{mL}^{-1}$) using 1 mL of a precooled (-15°C) mixture of MTBE: methanol: water 3:1:1 (v/v/v) as previously described^{28,29}. Five hundred microliters of the upper phase, containing the lipids, were transferred to a fresh 1.5 mL Eppendorf tube and dried down in a speed-vac. The extracts that were reconstituted in acetonitrile: isopropanol 7:3 (v/v) (final concentration $25 \mu\text{g} \cdot \text{mL}^{-1}$) were gradient-eluted on an UPLC-system (Accela, Thermo) using a reverse phase C8 column ACQUITY UPLC BEH $1.7 \mu\text{m}$, $2.1 \times 150 \text{ mm}$ (Waters, 186003377). The gradient of separation was done at $400 \mu\text{L} \cdot \text{min}^{-1}$ using (A) 0.1% acetic acid in water and (B) 0.1% acetic acid in methanol (60% B during 4 min, 60% B to 80% B in 4 min, 80–99% B in 0.5 min, 99% B for 3.5 min). The samples were measured in negative ion mode and mass spectra were acquired using an LTQXL mass spectrometer (MS) (ThermoFisher Corporation) with electrospray ionization (ESI). The MS interface capillary was maintained at 325°C , with a sheath gas flow of 45 (arbitrary units) and auxiliary gas flow of 2 (arbitrary units) for negative injections. The spray voltage ion injection was 4 kV, and the instrument scanned 350–600 m/z. MS/MS normalized collision energy was set to 41, activation Q 0.25, activation time 30 ms, and a 2 m/z isolation window. MS/MS scans were collected using dynamic exclusion with an exclusion time of 30 s. Obtained raw-chromatograms were further processed using Excalibur software version 2.10 (Thermo Fisher). The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the standard library spectrum from the Lipid Mass Database^{30,31}.

Fluorescence polarization. Purified PASb-1 β was labeled with the FITC fluorophore (fluorescein isothiocyanate, Thermo Scientific) by mixing 1 mg of protein with 20-fold molar excess of FITC in 50 mM borate buffer pH 8.5 for 1 hour at room temperature on the dark. The excess FITC was removed by overnight dialysis at 4°C with 30 mM Hepes pH 7.5, 150 mM NaCl and 0.5 mM TCEP. Unlabeled PASb-1 α and PASb-3 α , initially at a concentration of approximately 1 mM, were serially diluted 1:2 nineteen consecutive times into buffer 30 mM Hepes pH 7.5, 150 mM NaCl and 0.5 mM TCEP. Then, 18.75 μL from each protein dilution was mixed with 6.25 μL of PASb-1 β labeled with FITC at 80 nM (final concentration of labeled protein was fixed at 20 nM). A final reaction volume of 25 μL was pipetted into a 384-well black-walled microplate (Greiner Bio-one), in triplicate. Protein:protein complex samples were incubated at room temperature in the dark and fluorescence polarization (FP) was measured in a CLARIOstar microplate reader (BMG Labtech) with excitation and emission wavelengths of 480 nm and 520 nm, respectively. Experimental values were output as mean \pm standard deviation and logistic sigmoidal curves fitted to determine dissociation constants (K_d) for each protein:protein complex interaction.

Lipid binding assay. Removal of bacterial lipids from purified PASb-3 α was reached by protein incubation (1.7 mL at 10 μM) with 100 mg of LIPIDEX 1000 (Sigma Aldrich) for 1 h at 37°C in buffer containing 30 mM Tris pH 7.5, 150 mM NaCl and 0.5 mM TCEP, for three to five consecutive times. After delipidation the protein was incubated with different fatty acids for binding assay using the fluorescent probe 8-Anilino-1-naphthalenesulfonic acid (1,8-ANS). The final protein concentration used in the measurements was around 1 μM , with 10 μM of ANS. The fatty acid solutions, at initial concentration of 33 μM in pure ethanol, were diluted in ethanol by serially diluted 1:3 twelve consecutive times. The concentration of ethanol in the final reaction never exceeded 1%. The fluorescence measurements were recorded on a EnSpire multimode plate reader (Perkin Elmer) using a 96-well black-walled plate (Perkin Elmer), with excitation wavelength of 380 nm and emission wavelength range of 400–600 nm, at 25°C .

Isotopic labeling of PASb-1 β . To prepare ^{15}N - and $^{13}\text{C}/^{15}\text{N}$ -labelled PASb-1 β , *Escherichia coli* Rosetta II cells were allowed to grow in a modified M9 minimal medium³² supplemented with 1 g $\cdot \text{L}^{-1}$ $^{15}\text{NH}_4\text{Cl}$ (Cambridge Isotope Laboratories) combined with 4 g $\cdot \text{L}^{-1}$ D(+)-glucose (Sigma) to obtain only ^{15}N -labelled proteins or with [^{13}C 6]-D(+)-glucose (Cambridge Isotope Laboratories) to obtain double-labeled proteins. Expression and purification of labeled PASb-1 β was performed using the same procedure as described for the unlabeled protein.

NMR spectroscopy. ^{15}N - and $^{13}\text{C}/^{15}\text{N}$ -labeled PASb-1 β was produced and purified in a final buffer containing 30 mM $\text{Na}_2\text{HPO}_4\text{-NaH}_2\text{PO}_4$ pH 7.2, 70 mM NaCl, 0.5 mM TCEP and 10% D_2O . NMR experiments were performed using an Agilent Inova spectrometer, operating at a ^1H Larmor frequency of 599.887 MHz and temperature of 298 K. The spectrometer is equipped with a triple resonance cryogenic probe and a Z pulse-field gradient unit. The spectra were processed with NMRPipe/NMRDraw³³ and analyzed with NMRView³⁴. The ^{15}N HSQC spectra of PASb-1 β was collected and analyzed based on a previously deposited set of assigned chemical shifts into a public database at BioMagResBank (BMRB Entry 6597) and confirmed using the three-dimensional experiments HNCACB, CBCA(CO)NH, HNCO, and HN(CA)CO^{35–38}. PASb-1 α and PASb-3 α , both initially at 900 μM , were titrated in ^{15}N -labeled PASb-1 β (200 μM), followed by the recording of the two-dimensional ^{15}N HSQC spectra. Binding of proteins was characterized by changes in ^{15}N HSQC signals, intensities and chemical shift values, as a function of the concentration of unlabeled protein. The chemical shift perturbation of the ^{15}N -HSQC spectra was normalized according to the equation $\Delta\delta(^{15}\text{N} + ^1\text{H}) = [(\Delta\delta^{15}\text{N}/10)^2 + (\Delta\delta^1\text{H})^2]^{1/2}$, in ppm units³⁹ and the relative peak intensity changes calculated using the equation $(I - I_0)/I_0$, where I is the peak intensity

measured in the presence of the maximum amount of unlabeled protein and I_0 is the peak intensity measured in the free labeled protein.

Molecular Modelling of PASb-3 α 4. *Ab initio* homology modelling of PASb domain of HIF-3 α 4, or PASb-3 α 4 (Gly237-Pro351, Uniprot Q9Y2N7-4) was done using methodology implemented by I-TASSER⁴⁰. The top output model was used for comparison against the crystallographic model of lipid-bound PASb-3 α .

Tissue array. Human normal and tumor tissue arrays containing 78 normal tissue spots (FDA805-1, U.S. Biomax) or 18 normal tissues and 54 tumors (FDA805-2, U.S. Biomax) from multiple organs were subjected to immunohistochemical analysis with Anti-HIF3 alpha antibody (ab10134, Abcam). The tissue array slides were deparaffinized by two xylene washes (10 min each) and then rehydrated in ethanol series. After washing with water, the slides were microwaved for 10 rounds of 1 min in 10 mM sodium citrate buffer (pH 6.0) and allowed to cool to room temperature to retrieve antigens. Endogenous peroxidase activity was extinguished after 30 min incubation in 0.3% H₂O₂ in PBS. The slides were then blocked with Avidin and Biotin blocking solutions (Avidin/Biotin Blocking Kit; Vector Labs) and 1% goat serum in PBS (Vectastain Elite ABC Kit, Vector Labs). The tissue arrays were incubated overnight with the primary antibody (1:100 dilution in 1X PBS, 0.2% BSA, 0.02% Triton X-100, 10 mM Glycine), washed in same buffer, incubated with biotinylated secondary antibody 1:200 and treated for 30 min with the Elite ABC reagent (Universal R.T.U. Vectastain Elite ABC Kit; Vector Labs). After washing with PBS, the slides were processed with chromogen solution (ImmPACT DAB Substrate; Vector Labs). The slides were then dehydrated by ethanol series and mounted using Entellan New (Merck). The resulting staining obtained were captured using Operetta (PerkinElmer) 2X magnification objective, stitched by Harmony software (PerkinElmer) and quantified using ImageJ software. To allow comparison between slides, all data was normalized first to Skin Malignant melanoma spot (present in both slides). A second normalization was done to the highest expression in each specific tissue type. Samples were selected based on $n \geq 2$ and the simultaneous presence of both normal and cancer representatives. Statistical significance was tested using Two-sample t-test, considering $p < 0.1$.

TCGA data analysis. All the gene-expression data derived from RNA-seq were downloaded from The Cancer Genome Atlas data portal (<http://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm>). For the evaluation of HIF-3 α 9 expression levels across tumor types, the normalized RSEM count from uc002peh.2 were used. For tumors with mean expression above 5, the 2-component Gaussian mixture distribution model with fixed variance for definition of expressing and non-expressing patients^{41,42} was implemented using mclust⁴³ package (R statistical software⁴⁴) and fitted independently for each tumor dataset using the uc002peh.2 expression levels. Patients were categorized in each class and gene differential expression analysis was performed using EBSeq⁴⁵ package (R statistical software) using as input raw RSEM values and performing a 75h quantile normalization⁴⁶. Genes up- (>1) and down- regulated (<1) with posterior probability of being differentially expressed above 0.95 were used as input for KOBAS2.0⁴⁷ pathway enrichment analysis for REACTOME database. The Venn plot were made with VennDiagram⁴⁸ package (R statistical software).

Canonical ensemble simulation of lipid in vacuum. A Metropolis-Monte Carlo simulation was performed to recapitulate a canonical ensemble for the free stearic, oleic and linoleic acids. One-million-steps simulation, as implemented in Python, was performed, with temperature set to 300 K. The associated energies of randomly generated microstates was calculated using YASARA's function Energy("All")⁴⁹ and the AMBER03 force field⁵⁰ for the parametrization of the lipid molecules. Probability of micro-states $P(s)$ were calculated by discretization of Boltzmann distributions using $\text{hist}(\text{Energy})/\text{density}$ function of the statistical software R⁴⁴. Entropy was estimated with $S(\text{lipid}, \text{vacuum}) = -\sum P(s) \log(P(s))$.

References

1. Semenza, G. L. Oxygen sensing, hypoxia-inducible factors, and disease pathophysiology. *Annu Rev Pathol* **9**, 47–71 (2014).
2. Maynard, M. A. *et al.* Multiple splice variants of the human HIF-3 α locus are targets of the von Hippel-Lindau E3 ubiquitin ligase complex. *J Biol Chem* **278**, 11032–11040 (2003).
3. Pasanen, A. *et al.* Hypoxia-inducible factor (HIF)-3 α is subject to extensive alternative splicing in human tissues and cancer cells and is regulated by HIF-1 but not HIF-2. *Int J Biochem Cell Biol* **42**, 1189–1200 (2010).
4. Zhang, P. *et al.* Hypoxia-inducible factor 3 is an oxygen-dependent transcription activator and regulates a distinct transcriptional response to hypoxia. *Cell Rep* **6**, 1110–1121 (2014).
5. Maynard, M. A. *et al.* Human HIF-3 α 4 is a dominant-negative regulator of HIF-1 and is down-regulated in renal cell carcinoma. *FASEB J* **19**, 1396–1406 (2005).
6. Partch, C. L., Card, P. B., Amezcua, C. A. & Gardner, K. H. Molecular basis of coiled coil coactivator recruitment by the aryl hydrocarbon receptor nuclear translocator (ARNT). *J Biol Chem* **284**, 15184–15192 (2009).
7. Partch, C. L. & Gardner, K. H. Coactivators necessary for transcriptional output of the hypoxia inducible factor, HIF, are directly recruited by ARNT PAS-B. *Proc Natl Acad Sci USA* **108**, 7739–7744 (2011).
8. Henry, J. T. & Crosson, S. Ligand-binding PAS domains in a genomic, cellular, and structural context. *Annu Rev Microbiol* **65**, 261–286 (2011).
9. Cardoso, R. *et al.* Identification of Cys255 in HIF-1 α as a novel site for development of covalent inhibitors of HIF-1 α /ARNT PasB domain protein-protein interaction. *Protein Sci* **21**, 1885–1896 (2012).

10. Scheuermann, T. H. *et al.* Artificial ligand binding within the HIF2 α PAS-B domain of the HIF2 transcription factor. *Proc Natl Acad Sci USA* **106**, 450–455 (2009).
11. Rogers, J. L. *et al.* Development of inhibitors of the PAS-B domain of the HIF-2 α transcription factor. *J. Med. Chem.* **56**, 1739–1747 (2013).
12. Scheuermann, T. H. *et al.* Allosteric inhibition of hypoxia inducible factor-2 with small molecules. *Nat Chem Biol* **9**, 271–276 (2013).
13. Dynńska-Kukulska, K., Ciesielski, W. & Zakrzewski, R. The use of a new, modified Dittmer-Lester spray reagent for phospholipid determination by the TLC image analysis technique. *Biomed Chromatogr* **27**, 458–465 (2013).
14. Cronan, J. E. & Vagelos, P. R. Metabolism and function of the membrane phospholipids of *Escherichia coli*. *Biochim Biophys Acta* **265**, 25–60 (1972).
15. Glatz, J. F. & Veerkamp, J. H. Removal of fatty acids from serum albumin by Lipidex 1000 chromatography. *J Biochem Biophys Methods* **8**, 57–61 (1983).
16. Hoff, W. D. *et al.* Global conformational changes upon receptor stimulation in photoactive yellow protein. *Biochemistry* **38**, 1009–1017 (1999).
17. Kroon, A. R. *et al.* Spectral tuning, fluorescence, and photoactivity in hybrids of photoactive yellow protein, reconstituted with native or modified chromophores. *J Biol Chem* **271**, 31949–31956 (1996).
18. Pandini, A., Denison, M. S., Song, Y., Soshilov, A. A. & Bonati, L. Structural and functional characterization of the aryl hydrocarbon receptor ligand binding domain by homology modeling and mutational analysis. *Biochemistry* **46**, 696–708 (2007).
19. Katschinski, D. M. *et al.* Interaction of the PAS B domain with HSP90 accelerates hypoxia-inducible factor-1 α stabilization. *Cell Physiol Biochem* **15**, 351–360 (2004).
20. Dick, K. J. *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383**, 1990–1998 (2014).
21. Leslie, A. G. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4+ESF-EAMCB Newsletter on Protein Crystallography* **26**, 27–33 (1992).
22. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214 (2013).
23. McCoy, A. J. *et al.* Phaser Crystallographic software. *J Appl Cryst* **40**, 658–674 (2007).
24. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).
25. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. P. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
26. Oliveira, S. H. *et al.* KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinformatics* Jun **15**, 197 (2014).
27. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).
28. Giavalisco, P. *et al.* Elemental formula annotation of polar and lipophilic metabolites using (13)C, (15)N and (34)S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J* **68**, 364–376 (2011).
29. Hummel, J. *et al.* Ultra performance liquid chromatography and high resolution mass spectrometry for the analysis of plant lipids. *Front Plant Sci* **2**, 54 (2011).
30. Fahy, E., Sud, M., Cotter, D. & Subramaniam S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res* **35**, W606–12 (2007).
31. Fahy, E. *et al.* LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* **50**, S9–S14 (2009).
32. Li, M. X., Corson, D. C. & Sykes B. D. Structure determination by NMR: isotope labeling. *Methods Mol Biol* **173**, 255–265 (2002).
33. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293 (1995).
34. Johnson, B. & Blevins, R. NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* **4**, 603–614 (1994).
35. Ikura, M., Kay, L. & Bax, A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of larger proteins: Heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659–4667 (1990).
36. Grzesiek, S. & Bax, A. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR* **3**, 185–204 (1993).
37. Wittekind, M. & Mueller, L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J Magn Reson Ser B* **101**, 201–205 (1993).
38. Yamazaki, T., Lee, W., Arrowsmith, C., Muhandiram, D. R. & Kay, L. E. A suite of triple resonance NMR experiments for the backbone assignment of 15N, 13C, 2H labeled proteins with high sensitivity. *J Am Chem Soc* **116**, 11655–11666 (1994).
39. Cavanagh, J., Fairbrother, W., Palmer, A., III, Rance, M. & Skelton, N. *Protein NMR Spectroscopy: Principles and Practice* (Academic Press, San Diego, 2007).
40. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Prot* **5**, 725–738 (2010).
41. Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* **121**, 2750–2767 (2011).
42. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–52 (2012).
43. Fraley, C. & Raftery, A. E. Model-based Clustering, Discriminant Analysis and Density Estimation. *J Am Stat Assoc* **97**, 611–631 (2002).
44. Ihaka, R. & Gentleman, R. R. A language for data analysis and graphics. *J Comput Graph Stat* **5**, 299–314 (1996).
45. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010) *Genome Biol* **11**, R106 (2010).
46. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010). *BMC Bioinformatics* **11**, 94 (2010).
47. Xie, C. *et al.* KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* **39**, 316–322 (2011).
48. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**, 35 (2011).
49. Krieger, E., Darden, T., Nabuurs, S., Finkelstein, A. & Vriend, G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* **57**, 678–683 (2004).
50. Duan, Y. *et al.* Point-Charge Force Field for Molecular Mechanics Simulations of Proteins. *J Comput Chem* **24**, 1999–2012 (2003).

Acknowledgements

This work was supported by the São Paulo State Research Foundation, FAPESP, under grants 2012/14298-9 and 2014/20673-2 (to ALBA) and 2014/15968-3 (to SMGD), and fellowships 2010/13739-6 and 2014/04927-4 (to AMF). We thank LNBio for financial support and access to facilities (LPP, MAS, LBE, NMR and Robolab). We thank Dr. Alessandra Girasole for expert technical and secretarial support. We are very grateful to Dr. Chi Van Dang and Zandra Walton (Abramson Cancer Center, University of Pennsylvania, USA) for expert comments and suggestions as well as stimulating discussions. We are also grateful to Dr. Richard Charles Garratt (IFSC/USP, Brazil) for detailed and critical reading of the manuscript. The results published here are part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

Author Contributions

A.M.F., J.F.O., S.M.G.D. and A.L.B.A. designed research; A.M.F., J.F.O., D.A., J.A., S.A.R., M.M.D., M.V.B.D., M.L.S. and P.S.L.O. performed research; A.M.F., J.F.O., D.A., J.A., S.A.R., M.M.D., M.V.B.D., M.L.S., P.S.L.O., C.C., S.M.G.D. and A.L.B.A. analyzed and discussed data; S.M.G.D. and A.L.B.A. wrote the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Fala, A.M. *et al.* Unsaturated fatty acids as high-affinity ligands of the C-terminal Per-ARNT-Sim domain from the Hypoxia-inducible factor 3 α . *Sci. Rep.* **5**, 12698; doi: 10.1038/srep12698 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>