

# Structure of the SCAN Domain of Human Paternally Expressed Gene 3 Protein

Vadim Rimsa, Thomas C. Eadsforth, William N. Hunter\*

Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee, United Kingdom

## Abstract

Human paternally expressed gene 3 protein (PEG3) is a large multi-domain entity with diverse biological functions, including acting as a transcription factor. PEG3 contains twelve Cys<sub>2</sub>-His<sub>2</sub> type zinc finger domains, extended regions of predicted disorder and at the N-terminus a SCAN domain. PEG3 has been identified as partner of the E3 ubiquitin-protein ligase Siah1, an association we sought to investigate. An efficient bacterial recombinant expression system of the human PEG3-SCAN domain was prepared and crystals appeared spontaneously when the protein was being concentrated after purification. The structure was determined at 1.95 Å resolution and reveals a polypeptide fold of five helices in an extended configuration. An extensive dimerization interface, using almost a quarter of the solvent accessible surface, and key salt bridge interactions explain the stability of the dimer. Comparison with other SCAN domains reveals a high degree of conservation involving residues that contribute to the dimer interface. The PEG3-SCAN domain appears to constitute an assembly block, enabling PEG3 homo- or heterodimerization to control gene expression in a combinatorial fashion.

**Citation:** Rimsa V, Eadsforth TC, Hunter WN (2013) Structure of the SCAN Domain of Human Paternally Expressed Gene 3 Protein. PLoS ONE 8(7): e69538. doi:10.1371/journal.pone.0069538

**Editor:** Titus J. Boggon, Yale University School of Medicine, United States of America

**Received:** May 1, 2013; **Accepted:** June 10, 2013; **Published:** July 23, 2013

**Copyright:** © 2013 Rimsa et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was partly funded by AstraZeneca by way of contributing to a studentship for VR. Dr. Derek Ogg (Astra Zeneca) provided excellent support and Dr. Navratna Vajpai (Astra Zeneca) ran the NMR experiment. There are no patents, products in development or marketed products to declare. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: w.n.hunter@dundee.ac.uk

## Introduction

Human paternally expressed gene 3 protein (PEG3) is a large multi-functional protein of nearly 1600 amino acids with a mass of about 165 kDa. The protein carries twelve zinc finger DNA-binding motifs, which are an important component of transcription [1]. These zinc fingers are classed as Cys<sub>2</sub>-His<sub>2</sub> Krüppel-type motifs where metal ion coordination helps to stabilize a simple ββ $\alpha$  fold [2–4]. The Krüppel-type fingers are separated by the Krüppel-link, a conserved sequence of seven amino acids [5]. An individual zinc finger can bind three-four base pairs in the major groove of DNA, but they often occur in tandem and work collectively to allow for recognition of longer DNA sequences [6,7]. Consistent with the presence of a large number of zinc fingers, PEG3 has been shown to bind DNA in a sequence specific manner [8], and has been detected predominantly in the nucleus, supporting the notion it is a *bone fide* transcription factor [1,9]. In addition, PEG3 is reported to be present in the cytoplasm where it interacts with proteins such as tumour necrosis factor receptor-associated factor 2 (TRAF2) to regulate the NF- $\kappa$ B signal transduction pathway, and with the E3 ubiquitin-protein ligase seven-in-absentia homolog one (Siah1a) to promote p53-mediated apoptosis [10,11]. PEG3 also binds  $\beta$ -catenin to induce its degradation, thereby inhibiting the Wnt signalling pathway [12]. Whole animal studies have revealed major physiological changes in mice following targeted mutation of the *peg3* gene. When a knockout model was assessed, growth retardation, increased levels of body fat, impairment of maternal behaviour and changes in male sexual behaviour were noted [13–16]. Such

pleiotropic effects identify its importance, however the precise molecular mechanism whereby PEG3 contributes to distinct aspects of biology remains poorly understood.

Proteins that carry Cys<sub>2</sub>-His<sub>2</sub> Krüppel-type motifs frequently contain other domains, which are implicated in subcellular localization, protein-protein and protein-DNA interactions. These include the Krüppel-associated box domain, the poxvirus and zinc finger (POZ), alternatively known as the BTB (broad-complex, tram track, and bric-a-brac), and the SCAN domain, sometimes called the leucine rich motif [17–21]. The name, SCAN, is derived from the first letters of four founding members of the family (SRE-ZBP, C<sub>tf</sub>in51, A<sub>W</sub>-1 (ZNF174), and N<sub>umber</sub> 18) [20]. PEG3 contains a SCAN domain near the N-terminus. The domain is a small, typically 80 to 90 residues, highly conserved structure. The human genome encodes 71 SCAN domains, 64 of which appear within functional genes [22]. The domain functions as a protein-protein interaction motif, mediating self-interaction and binding other proteins [23–25]. The ability of a SCAN domain to self-associate was first shown for ZNF174 using a mammalian two-hybrid system and subsequently heterodimer formation was also noted [23]. Our current understanding suggests that the presence of SCAN domains in some proteins carrying Cys<sub>2</sub>-His<sub>2</sub> Krüppel-type motifs allows them to interact in a combinatorial fashion to control gene expression [26].

Previously, two NMR and two crystal structures of SCAN domains have revealed the structural topology, which is constructed around five  $\alpha$ -helices [26–28]. The SCAN domain forms a dimer with structural homology, appearing like a domain-swapped assembly, to the C-terminal domain (CTD) of the retroviral

capsid protein [27], which is critical for capsid dimerization and viral assembly [29].

Here, we report the structure of the isolated SCAN homodimer from the human PEG3 protein (PEG3-SCAN). Comparisons with other SCAN structures in the Protein Data Bank (PDB) and the importance and contributions of residues involved in PEG3-SCAN dimerization are discussed, and provide insight into homo- and heterodimer assembly.

## Materials and Methods

### Protein Expression and Purification

The gene encoding the SCAN domain of human PEG3 (amino acids 40–130; UniProt entry Q9GZU2) was purchased in the pUC57 vector (GenScript). The gene was subsequently sub-cloned into a modified pET15b vector (Novagen), designed to express the protein of interest with an N-terminal hexa-His tag and a Tobacco Etch Virus (TEV) protease cleavage site. The sequence-verified clone (DNA Sequencing Unit, University of Dundee) was transformed into *Escherichia coli* BL21 (DE3) Gold cells (Novagen) for expression. Cells were grown in Luria-Bertani media at 37°C to an OD<sub>600</sub> of ~0.6, followed by induction with 0.2 mM IPTG. Cells were then grown overnight at 22°C and harvested by centrifugation at 3,500 g for 30 minutes at 4°C. The cell pellet was suspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole) containing DNase I (0.1 mg), with a single protease inhibitor tablet (Roche) and lysed using the French Press (16,000 psi). The extract was centrifuged at 37,500 g for 30 minutes at 4°C to pellet insoluble debris. The supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare) and a linear imidazole concentration gradient from 20 mM to 1 M was applied. The PEG3-SCAN sample eluted from the column at approximately 200 mM imidazole. Fractions containing the recombinant protein were pooled and dialyzed in buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl), prior to addition of TEV protease to remove the His-tag. TEV protease was prepared in-house by Keri Barrack (University of Dundee). 1 mg of TEV was used per 20 mg of protein in a cleavage reaction, which was performed at room temperature for three hours. After TEV digestion, three non-native residues (Gly-His-Met) remained at the N-terminus of the product. The protein was passed through a second nickel-affinity column, which removed uncleaved protein and the His-tagged TEV protease. Gel filtration (GF, Superdex 75 16/60 column; GE Healthcare) was used as a final purification step. The column had previously been calibrated with molecular weight standards, blue dextran (>2,000 kDa), thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29.5 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). PEG3-SCAN eluted from the size exclusion column as a single symmetric peak with a mass of approximately 28 kDa. The theoretical mass of PEG3-SCAN is 11,434 Da; therefore PEG3-SCAN forms a homodimer in solution. The purity of the sample was checked further by SDS-PAGE and mass spectrometry (Fingerprint Proteomics Facility, University of Dundee). The single protonated species as observed by mass spectrometry was 11,432 Da, in close agreement with the theoretical mass. Protein concentration was determined spectrophotometrically using a theoretical molar extinction coefficient of 16,960 M<sup>-1</sup> cm<sup>-1</sup> [30].

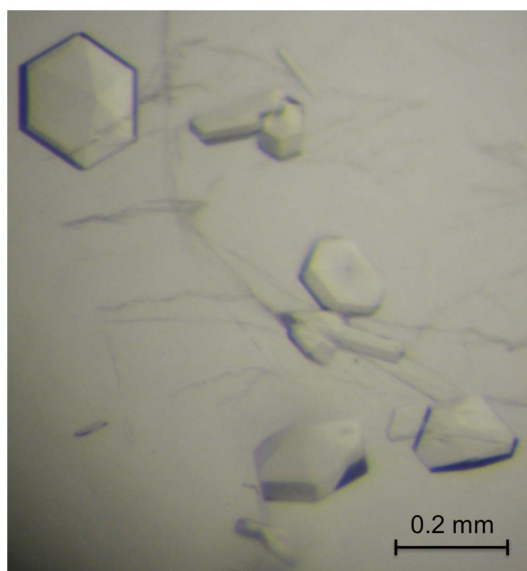
The gene coding for part of human Siah1 without the RING domain (amino acids 91–282; UniProt entry Q8IUQ4) was purchased in the pUC57 vector (GenScript). The gene fragment was transferred into the pET15b vector (Novagen) for recombinant expression. Siah1 and <sup>15</sup>N-labeled Siah1 for NMR

studies were prepared and purified using a similar protocol to that for PEG3-SCAN, except that isotopically enriched Siah1 was expressed in the minimal media. A single sharp peak was observed for Siah1 on GF column with a mass of around 39 kDa. This value matches closely to the weight of Siah1 homodimer, as the theoretical mass of a monomer is 21,897 Da. The presence of the Siah1 homodimer was confirmed by size exclusion chromatography (SEC) coupled with multi-angle light scattering. This supports previous studies showing Siah1 is a dimeric protein [31]. The purity of the sample was also analyzed by SDS-PAGE and mass spectrometry, which showed a single protonated species of 21,875, closely matching the theoretical size. Protein concentration was determined by UV spectrophotometry using a theoretical extinction coefficient of 22,960 M<sup>-1</sup> cm<sup>-1</sup> [30]. The association between PEG3-SCAN and Siah1 was tested in SEC by combining protein samples together at one-to-one stoichiometry in 50 mM Tris-HCl, pH 7.5, and 150 mM NaCl buffer. The mixture was left overnight at 4°C, before it was run on a GF column (Superdex 75 16/60 column; GE Healthcare). The NMR experiment was done under the following conditions, where 100 μM of <sup>15</sup>N-labeled Siah1 was mixed with 100 μM of unlabeled PEG3 in 50 mM HEPES, pH 7.5, 50 mM NaCl, and 5% D<sub>2</sub>O buffer. The chemical shift perturbations in the <sup>1</sup>H-<sup>15</sup>N HSQC of Siah1 were monitored upon addition of PEG3.

### Thermal Stability, Crystallization and Data Collection

Differential scanning fluorimetry (DSF) was used to investigate the influence of different buffers on the thermal stability of the samples. DSF suggested the presence of a globular SCAN domain, which displayed a melting temperature of 52°C in a number of buffers. Since no buffer appeared to enhance stability the protein was left in the GF buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl). The melting temperature of Siah1 was 64°C in the buffers tested again with a profile indicative of a folded protein.

Crystallization screening of PEG3-SCAN was carried out with the high-throughput Phoenix liquid handling system (Art Robins Instruments/Rigaku) and several commercially available screens (Hampton Research). 100 nL of protein sample at ~16.5 mg mL<sup>-1</sup> were mixed in sitting-well plates with 100 nL of reservoir solution against 70 μL of the same reservoir solution. Crystals appeared within a day in numerous conditions that contained polyethylene glycol (PEG) of different molecular weight. These crystallization conditions were scaled up from these nano-drops to micro-drops of total volume 4 μL. However, the rod-shaped crystals that were observed were small, approximately 0.1×0.02×0.02 mm, and gave poor diffraction. While preparing more protein for use in crystal optimization it was observed that larger crystals actually formed spontaneously when the protein was concentrated in the GF buffer using Vivaspin 20 concentrators with a 3,000 MW cut off (Sartorius Stedim Biotech). Hexagonal bipyramid crystals, reaching 0.2×0.2×0.2 mm dimensions, formed within minutes (Fig. 1). The average protein concentration in the centrifugal device was 5 mg mL<sup>-1</sup>, but likely to have been considerably higher near the membrane where crystal nucleation occurred. The selection of a suitable cryoprotectant required extensive screening and optimization. The use of glycerol, ethylene glycol and paratone-N produced either poor diffraction or pronounced ice rings. The most favourable cryoprotectant was PEG200. Crystals were transferred into cryo-solution of PEG200 and GF buffer at 1:1 ratio prior to flash cooling for X-ray diffraction studies.



**Figure 1. Crystals of PEG3-SCAN.** Crystals grown in 50 mM Tris-HCl pH 7.5 and 150 mM NaCl.

doi:10.1371/journal.pone.0069538.g001

### Structure Solution and Refinement

Diffraction data were collected in-house with a Micromax-007 rotating anode generator using CuK $\alpha$  ( $\lambda = 1.5414 \text{ \AA}$ ) radiation and an AFC11 Saturn 944+ CCD detector (Rigaku). The data were indexed and integrated with *iMOSFLM* [32] and scaled with *AIMLESS* from the *CCP4* program suite [33]. The structure was solved by molecular replacement with *PHASER* [34] using a poly-Ala model of the SCAN domain dimer from the mouse zinc finger protein 206 (Zfp206, PDB code 4E6S [26]) that shares 38% sequence identity with the PEG3-SCAN domain. The output model was subjected to a round of rigid body and restrained refinement using *REFMAC5* [35]. The poly-Ala model was modified to the sequence of human PEG3-SCAN based on inspection of electron and difference density maps in *COOT* [36]. Several residues and side chains for which there was no convincing electron density were deleted. Additional rounds of restrained least-squares refinement followed, interspersed with map inspection and model manipulation. The refinement used the automatic geometry and *B*-factor restraint weights. Neither non-crystallographic symmetry (NCS) restraints nor TLS (Translation/Libration/Screw) were used in refinement. A number of ligands (ethylene glycol, diethylene glycol and triethylene glycol) were included in the model on the basis of the difference density and chemical environment, and refined successfully. These molecules are likely to be decomposition products or impurities of the PEG200 cryoprotectant. The final model also includes multiple side chain conformers and water molecules. The stereochemistry of the model was checked using *MOLPROBITY* [37]. *PISA* (Protein Interfaces, Surfaces and Assemblies [38]) was used to calculate surface and dimer interface areas. Fig. 2 was prepared with *ALINE* [39] and the others were produced using *PyMOL* [40]. Data collection and structure refinement statistics are shown in Table 1. The atomic coordinates and structure factors have been deposited in the PDB with accession code 4BHX.

**Table 1. Crystallographic statistics.**

PEG3-SCAN	
Space group	$P6_5$
Unit cell dimensions: <i>a</i> , <i>b</i> , <i>c</i> (Å)	83.61, 83.61, 55.23
Resolution <sup>a</sup> (Å)	13.8–1.95 (2.00–1.95)
No. of reflections	453776 (23734)
Unique reflections	16090 (1143)
Completeness (%)	99.7 (99.8)
Multiplicity	28.2 (20.8)
$\langle I/\sigma \rangle$	38.2 (9.7)
Wilson <i>B</i> (Å <sup>2</sup> )	20.6
Mosaicity (°)	0.5
<b>Residues</b>	
Chain A	40–127
Chain B	40–129
Water/ethylene glycol/diethylene glycol/triethylene glycol	155/21/4/2
$R_{\text{merge}}^b$ (%)	7.0 (31.8)
$R_{\text{work}}^c$ (%)	17.15 (19.0)
$R_{\text{free}}^d$ (%)	22.38 (24.4)
<b>Mean B-factors (Å<sup>2</sup>)</b>	
Chain A	27.4
Chain B	25.9
Waters	38.4
Other ligands	43.2
R.m.s.d. bond lengths (Å)	0.02
R.m.s.d. bond angles (°)	2.02
<b>Ramachandran plot (%)</b>	
Most favoured	98.4
Additional allowed	1.6
Outliers	0.0

<sup>a</sup>Values in parentheses refer to the highest resolution shell (2.00–1.95 Å).

<sup>b</sup> $R_{\text{merge}} = \sum_h \sum_i |I(h,i) - \langle I(h) \rangle| / \sum_h \sum_i I(h,i)$ ; where  $I(h,i)$  is the intensity of the  $i$ th measurement of reflection  $h$  and  $\langle I(h) \rangle$  is the mean value of  $I(h,i)$  for all  $i$  measurements.

<sup>c</sup> $R_{\text{work}} = \sum_h k | |F_o| - |F_c| | / \sum_h |F_o|$ , where  $F_o$  is the observed structure factor amplitude and the  $F_c$  is the structure-factor amplitude calculated from the model.

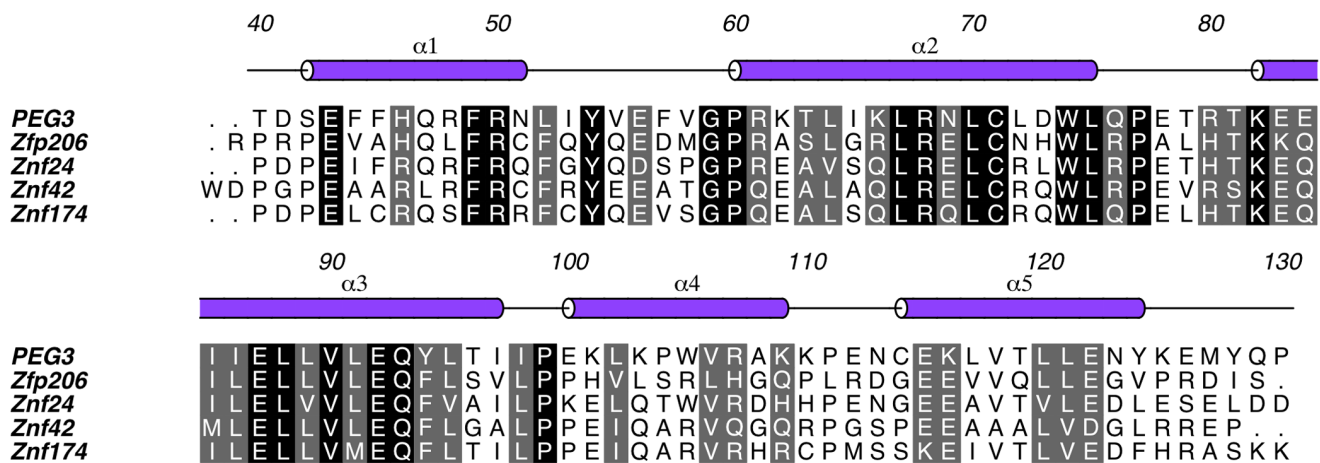
<sup>d</sup> $R_{\text{free}}$  is the same as  $R_{\text{work}}$  except calculated with a subset, 5%, of data that are excluded from refinement calculations.

doi:10.1371/journal.pone.0069538.t001

## Results and Discussion

### Structure Quality

Crystals of human PEG3-SCAN belong to space group  $P6_5$ , with a  $V_M$  value of  $2.44 \text{ \AA}^3 \text{ Da}^{-1}$  and solvent content of approximately 50% for an asymmetric unit comprising two polypeptide chains. The polypeptides are arranged as a symmetrical dimer consistent with the GF results obtained during protein purification and also with previously solved structures of SCAN domains. The crystals diffract to a resolution of 1.95 Å and the majority of the residues are located within well-defined electron density, apart from a few residues at the C-terminus. In addition, the final model contains two extra residues (His and Met) at the N-terminus, which are remnants from proteolytic cleavage of the histidine tag. A Ramachandran plot indicates that 98.4% of the



**Figure 2. The primary and secondary structure of PEG3-SCAN.** Five  $\alpha$ -helices are shown as cylinders (purple) and are numbered accordingly. Multiple sequence alignment of PEG3-SCAN with other SCAN proteins from PDB was performed with ClustalW2 [48]. PEG3-SCAN residues that are strictly conserved in Zfp206 (PDB: 4E65), Znf24 (PDB: 3LHR), Znf42 (PDB: 2F12) and Znf174 (PDB: 1Y7Q) are encased in black, while residues sharing similar properties in five proteins are encased in grey. The numbers that are shown above the secondary structure mark residues in the full length PEG3 protein (UniProt: Q9GZU2).

doi:10.1371/journal.pone.0069538.g002

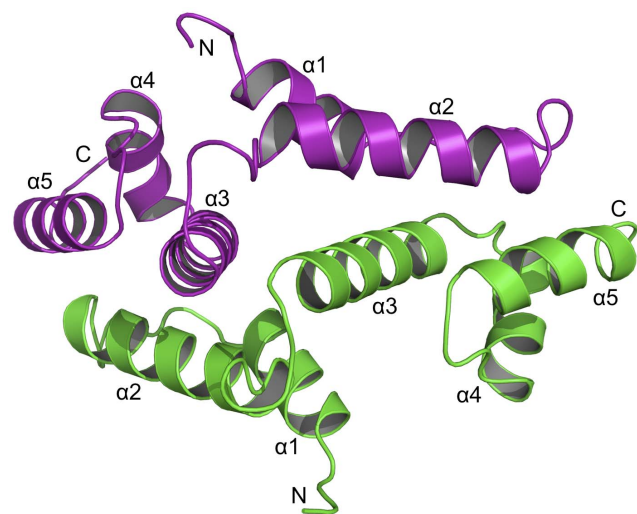
amino acids are located in the most favoured region with no outliers.

### Overall Structure

The human PEG3-SCAN domain folds as an extended V-shaped structure, with approximate overall dimensions of  $50 \times 25 \times 25$  Å. Each arm of the V-shape is approximately 35 Å in length. The subunit comprises five  $\alpha$  helices and the assignment of secondary structure onto the sequence is presented in Fig. 2 with the fold depicted in Fig. 3. Helices  $\alpha 1$  and  $\alpha 2$  which form an N-terminal sub-domain are aligned antiparallel to create one half of the V. A C-terminal sub-domain, which forms the other half, results from  $\alpha 3$ ,  $\alpha 4$  and  $\alpha 5$  being packed together. The domain aligns with the N-terminal sub-domains interacting with partner C-terminal sub-domains (Fig. 3) to form a dimer with approximate dimensions of  $50 \times 37 \times 30$  Å. A least-squares superposition of subunits with *LSQKAB* [41] gives an r.m.s.d. (root-mean-square deviation) of 0.57 Å for 90 C $\alpha$  atoms, which shows there are no major conformational differences between the two subunits. It is noteworthy that such a low value was obtained in the absence of NCS restraints.

The total surface area of a subunit, calculated with *PISA* [38], is approximately 7400 Å<sup>2</sup> of which 1700 Å<sup>2</sup> are buried within the dimer. Therefore, about 23% of the surface area of each monomer is involved in dimerization. The free energy of dissociation ( $\Delta G^{\text{diss}}$ ) is estimated as 19.4 kcal mol<sup>-1</sup>, and suggests that this assembly is thermodynamically stable, consistent with the observation of a stable dimer in solution. Similar values are observed for other SCAN structures. For example, the interface area and  $\Delta G^{\text{diss}}$  for the Znf24 dimer (PDB code 3LHR) are 23% and 21.8 kcal mol<sup>-1</sup>, respectively.

At present there are four SCAN domain structures in the PDB, two crystal structures and two determined by solution NMR. Sequence conservation of these four with human PEG3-SCAN is presented in Fig. 2. The superposition of the PEG3-SCAN dimer onto these other dimers reveals an overall structural conservation (Fig. 4), with calculated r.m.s.d. values presented in Table 2. The largest deviations among SCAN structures occur at the N- and C-terminal ends, which show higher flexibility than the core, and  $\alpha 4$ , which is positioned away from the dimer interface. The r.m.s.d.



**Figure 3. Overall structure of PEG3-SCAN.** The homodimer is shown as ribbons with one subunit green, the partner purple. The N- and C-termini as well as the five  $\alpha$ -helices of each monomer are labeled accordingly.

doi:10.1371/journal.pone.0069538.g003

values for alignments with the SCAN domain dimers of Znf42 and Znf174 show higher variation, more than 1.0 Å greater, than for the X-ray structures, because of the greater uncertainties associated with the NMR structures and that the fit involves an average of 20 conformers that represent their NMR derived structures.

### Residues Forming the SCAN Dimer Interface

The human PEG3-SCAN homodimer is held together by an extensive network of hydrogen-bonding, salt-bridge interactions and van der Waals forces. Even though the overall sequence identity among the five known SCAN structures is only 40–50% (Fig. 2), the key residues located at the dimer interface and that contribute to inter-subunit associations are conserved. The

**Table 2.** Structure and sequence similarity between PEG3-SCAN and other SCAN domains.

Protein name	PDB codes	R.m.s.d (Å)	R.m.s.d alignment length	Sequence identity (%)
Zfp206	4E6S	1.57	157	38
Znf24	3LHR	1.51	164	48
Znf42	2F12	2.85	155	35
Znf174	1Y7Q	2.87	167	43

These included crystal structures of Zfp206 and Znf24, and solution NMR structures of Znf42 and Znf174. R.m.s.d. calculations were carried out with PDBeFold using secondary structure matching [49] with the PEG3-SCAN dimer in the superposition. Sequence alignment was performed with ClustalW2 using residues 40–130 of the full-length PEG3 against the core of the SCAN domain, as well as 2–5 flanking residues, of other proteins.  
doi:10.1371/journal.pone.0069538.t002

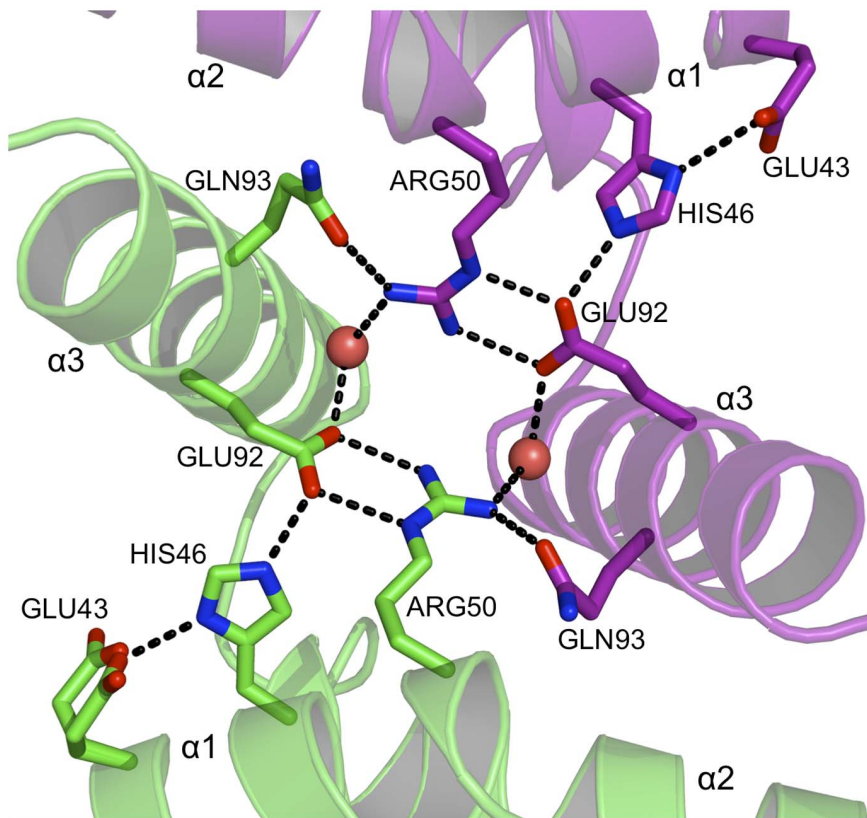
majority of these intermolecular contacts are formed between  $\alpha 1$  and  $\alpha 2$  (the N-terminal sub-domain) of one subunit and  $\alpha 3$  on the C-terminal sub-domain of the partner. Helices  $\alpha 2$  and  $\alpha 3$  show the highest amino acid conservation when comparing the sequences of these known SCAN domain structures and the conserved residues contribute significantly to either direct inter-subunit associations or organising the side chain positions to facilitate such interactions. For example, a hydrogen bond linking Glu43 with His46 helps to position the imidazole to also interact with Glu92 (Fig. 5). Glu92, close to the position of the NCS two-fold axis, forms a salt bridge interaction with Arg50 and also a well ordered and buried water molecule. Arg50 then donates a hydrogen bond to the partner subunit Gln93 and the water molecule mediates contact between Arg50 and the partner subunit Glu92. Arg68 forms an inter-subunit salt bridge with Glu87, so linking  $\alpha 2$  of one subunit with  $\alpha 3$  of the partner subunit (Fig. 6). These residues are highly

conserved within SCAN domains and observed to form similar hydrogen bonding patterns where the structures are known. Furthermore, mutation of both the equivalent Arg50 and Arg61 residues to alanines in Zfp206 destabilizes heterodimerization with Zfp110 [26]. Thus, these invariant residues seem to play an important role in both homo- and heterodimerization. Other residues establishing inter-subunit contacts include Lys82, which interacts with Pro77 and Arg61 with Glu115 (Fig. 6). The former makes a hydrogen bond, while the later makes a salt-bridge interaction, which links  $\alpha 2$  with the partner subunit  $\alpha 5$ . The later pair of residues is substituted to a similar lysine glutamine pair in some SCAN domain sequences.

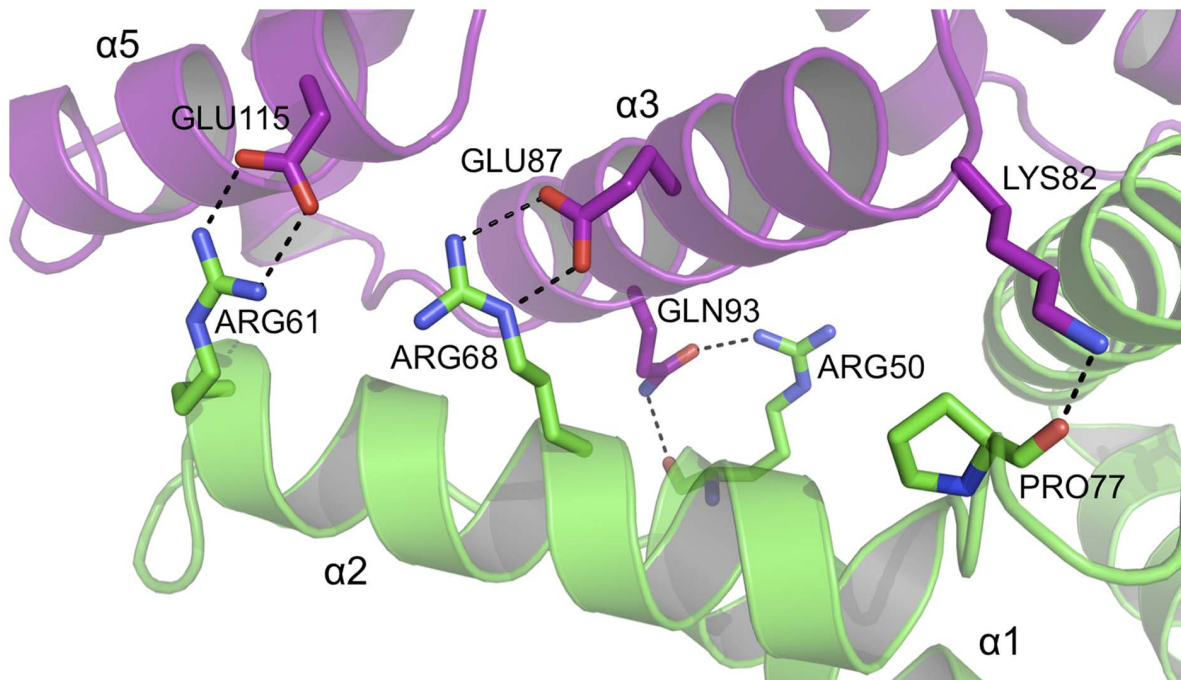
Some of the interactions noted in the PEG3-SCAN dimer are absent from the other structures. For example, a hydrogen bond donated from the side chain of Tyr94 on one subunit to the carbonyl of Pro60 (Fig. 7) on the partner cannot occur in other



**Figure 4. Superposition of the PEG3-SCAN homodimer (purple) with other SCAN structures.** Zfp206 (PDB 4E6S), Znf24 (PDB 3LHR), Znf42 (PDB 2F12) and Znf174 (PDB 1Y7Q) are shown in cyan, green, yellow and grey, respectively. Superposition was calculated using secondary structure matching [49].  
doi:10.1371/journal.pone.0069538.g004



**Figure 5. The dimer interface of PEG3-SCAN (I).** A hydrogen-bonding network is formed between conserved residues lining the subunit-subunit interface. Water molecules are shown as red spheres, N and O positions are colored blue and red respectively, C positions are purple or green depending on the subunit to which they belong, hydrogen bonds are depicted as dashed lines. The same color scheme is used in Figures 6, 7 and 8. doi:10.1371/journal.pone.0069538.g005



**Figure 6. The dimer interface of PEG3-SCAN (II).** A second cluster of hydrogen bonding and salt bridge interactions at the subunit-subunit interface. doi:10.1371/journal.pone.0069538.g006

structures where phenylalanine replaces the tyrosine. Furthermore, in PEG3-SCAN there is an inter-subunit salt bridge between Glu56 and Lys101 (data not shown). Glutamate replaces the lysine in most other SCAN domains. Such sequence variations may confer a preference of different SCAN domains to form distinctive homo- and heterodimers.

Whilst most of the dimer interface excludes water molecules, with the notable exception described above, at the periphery of the SCAN dimer there are five that are involved in mediating subunit-subunit interactions (data not shown). This does not appear to be a major factor in stabilizing the dimer given the high percentage of the surface area involved in direct association as described above.

The dimer interface includes important stabilizing contributions from hydrophobic residues. A hydrophobic patch comprising Phe49, Trp74, Leu75, Ile85, Ile86 and Leu89, from  $\alpha 2$  and  $\alpha 3$ , make van der Waals interactions to equivalent residues in the partner subunit (Fig. 8). In another area, a conserved Tyr54 and Pro60 from one subunit interact with a hydrophobic core formed by Tyr54, Ile97, Ile98, Pro99, Leu102 and Leu121 from  $\alpha 3$  and  $\alpha 5$  of the partner subunit (Fig. 7). Other hydrophobic residues extending across the dimer interface include Leu52, Leu64, Leu67, Val90 and Val118 (not shown).

The strictly conserved pair of residues Pro60 and Pro99, already mentioned as contributing to the inter-subunit hydrophobic associations, are positioned at the start of  $\alpha 2$  and  $\alpha 4$  respectively (Fig. 7), and likely important to initiate formation of the two helices.

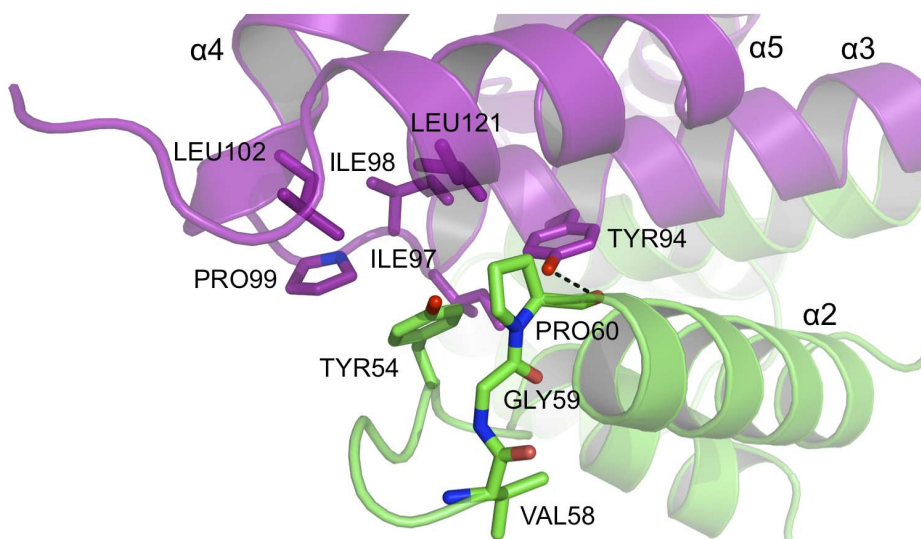
### Function of the SCAN Domain

Transcription factors form complexes that regulate the expression of target genes. For example, the proto-oncogene Jun, belonging to the basic region-leucine zipper (bZIP) family, regulates gene transcription in a specific manner by forming homo- and heterodimers. It is the combination of dimerization state and binding partners that determine activity and DNA-binding site preferences [42]. Similar control of transcription has been observed in some Cys<sub>2</sub>-His<sub>2</sub> zinc finger containing transcription factors that contain a protein-protein binding BTB/POZ domain. Self-oligomerization of this domain in the

GAGA transcription factor promotes DNA binding affinity and specificity [43]. The SCAN domain may be playing a similar regulatory role in mammalian gene expression by mediating both homo- and selective hetero-dimerization of Cys<sub>2</sub>-His<sub>2</sub> zinc finger proteins. To date, there are no data to suggest that SCAN domains participate directly in transcriptional activation or repression. The possibility that PEG3 could homodimerize through the SCAN domain suggests however that PEG3 binds to at least some target genes as a homodimer. In particular, this might be a prerequisite for PEG3 binding palindromic DNA motifs. Consistent with this is the finding that the Zfp206 transcription factor can form a homodimer using its SCAN domain and binds to a palindromic sequence [26,44]. A 15 base pair non-palindromic DNA binding motif for PEG3 was predicted and subsequently validated using electromobility shift and promoter assays [8]. Although the full-length motif was required for maximal binding, PEG3 was observed to bind to the partial sequence as well as to other regions with degenerate sequence.

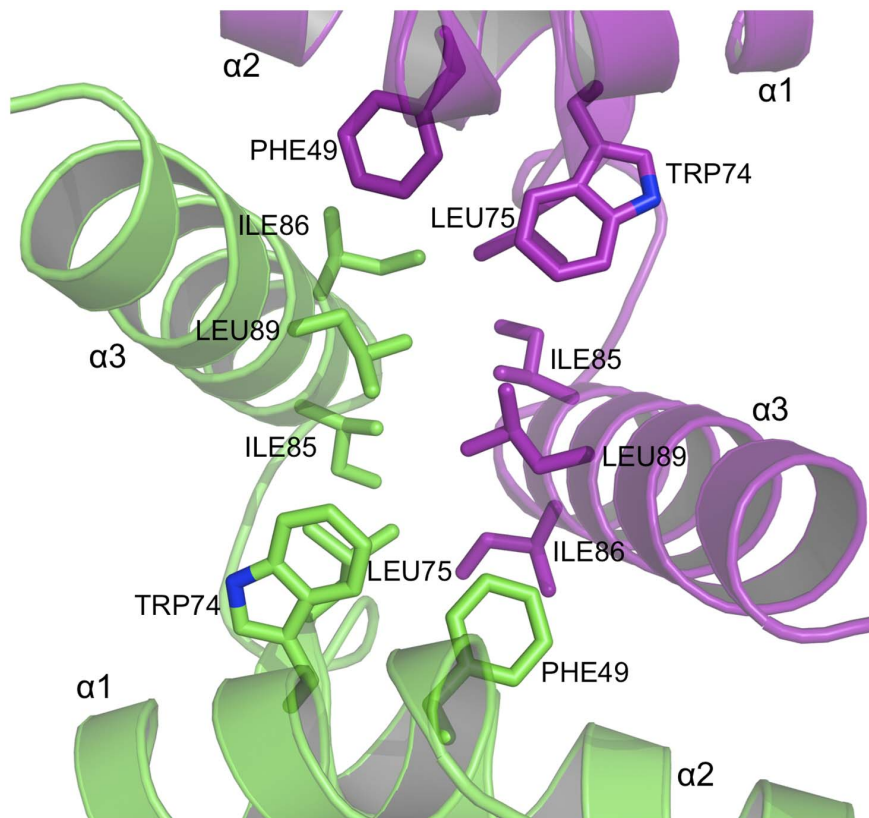
The DNA binding selectivity and affinity of PEG3 might depend on its specific isoform. To date, there are four known PEG3 isoforms. One isoform contains only two Cys<sub>2</sub>-His<sub>2</sub> zinc finger motifs, while in two other isoforms the SCAN domain is absent. Thus, the presence or absence of the SCAN domain and varying number of zinc finger motifs provides a potential additional control mechanism. A number of SCAN domain members have been reported to selectively interact with other SCAN members [23–26]. However, it has not been shown that PEG3-SCAN can form heterodimers. Superposition of a PEG3-SCAN subunit on the Zfp206 dimer structure shows a close fit and spatial conservation of key residues forming the intermolecular contacts (Fig. 2, Fig. 4, data not shown), suggesting the possibility that the PEG3-SCAN domain might form a heterodimer with Zfp206 as well as with other SCAN domain containing proteins. Future studies, focusing on the binding partners of PEG3-SCAN are required to inform on the role of PEG3 in gene regulation.

It has been suggested that some SCAN domains act as specific cellular localization sequences. For example, the SCAN domain of Zfp42 is reported to be essential for targeting the protein to promyelocytic leukemia nuclear bodies in the cell nucleus [45].



**Figure 7. The dimer interface of PEG3-SCAN (III).** A pronounced hydrophobic patch occurs at each end of the assembly to stabilize the dimer. The conserved Tyr94 extends across the dimer interface, contributes to hydrophobic interactions and donates a hydrogen bond to the carbonyl of Pro60.

doi:10.1371/journal.pone.0069538.g007



**Figure 8. The dimer interface of PEG3-SCAN (IV).** A group of conserved, aliphatic and aromatic residues form a hydrophobic core to stabilize the dimer.

doi:10.1371/journal.pone.0069538.g008

However, since localization studies using different constructs of PEG3 have indicated that the SCAN domain was not required for nuclear localization [12] then such a general function is unlikely.

It is not known whether SCAN domains interact with any other protein motifs in addition to self-association with other SCAN members. The function of PEG3 in regulation of TNF and Wnt signal transduction pathways has been implied by an ability to bind TRAF2, the E3 ubiquitin ligase Siah1 and  $\beta$ -catenin [10–12]. A yeast two-hybrid screen showed residues outside the SCAN domain (residues 268–502) to be required for interaction with full-length TRAF2 as well as a Siah2 fragment missing the RING (really interesting gene) domain (residues 106–326) [10]. The same study reported the interaction of a mouse PEG3 with Siah proteins by immunoprecipitation, while the later experiments using deletion generated constructs of human PEG3 revealed that the N-terminus including the SCAN domain (residues 1–268) were required for binding the full-length Siah1. Many proteins that interact with Siah1 contain the consensus Siah1 binding motif Val-x-Pro [46,47]. This amino acid sequence is also present within the SCAN domain of PEG3, Val58-Gly59-Pro60, on a short segment of extended structure leading to  $\alpha 2$  (Fig. 7). We sought to investigate whether the human PEG3-SCAN domain (residues 40–130) interacted with human Siah1. A construct of Siah1 without the RING domain (residues 90–282) was cloned, expressed and purified separately and combined with PEG3-SCAN in one-to-one stoichiometry. The sample was analyzed on a size exclusion column, but there was no evidence of complex formation (data not shown). An NMR study revealed no differences in chemical shifts in the two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC (heteronuclear single quantum coherence spectroscopy)

spectra of  $^{15}\text{N}$ -labeled Siah1 upon addition of unlabeled PEG3-SCAN (data not shown) indicative of a lack of interactions between the polypeptides. The lack of an association suggests that residues outside the SCAN domain might be required for interaction with Siah1 but there is no other obvious Siah1 binding motif. The PEG3-SCAN structure reveals that the Val58-Gly59-Pro60 motif is an extended conformation immediately prior to  $\alpha 2$ . The proline is buried and involved in inter-subunit interactions, and the valine side chain tucked down towards the side of  $\alpha 2$ , away from the surface of the protein. We speculate that if this is indeed a recognition site for Siah1 then conformational changes might be required to allow for complex formation. The conditions under which we investigated the potential interaction may not have been correct to allow such changes or alternatively the presence of the Val-x-Pro sequence is coincidence. Either way, our observations suggest that the data invoking Siah1 interactions with PEG3 should be re-evaluated.

### Concluding Remarks

In summary, we have determined the structure of SCAN domain from PEG3, a predicted transcription factor, and compared it with available SCAN domain structures. Our results show this domain forms a stable homodimer and we provide an analysis of the residues forming the dimer interface. The sequence alignment and an overlay of PEG3-SCAN with available SCAN domain structures shows overall structural conservation and serves to identify key residues important to the creation of the PEG3-SCAN dimer interface.



The SCAN domain of PEG3 appears to function as a convenient dimerization domain. Gel filtration chromatography and NMR studies revealed no interaction between the SCAN domain and the potential PEG3-binding protein Siah1. Future studies will be needed to determine if indeed the SCAN domain of PEG3 interacts with other SCAN family members as well as other protein motifs. The validation of binding partners would represent a crucial step towards unraveling the biological role of PEG3 itself.

## References

1. Relaix F, Weng X, Marazzi G, Yang E, Copeland N, et al. (1996) Pw1, a novel zinc finger gene implicated in the myogenic and neuronal lineages. *Dev Biol* 177: 383–396.
2. Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor III A from *Xenopus* oocytes. *EMBO J* 4: 1609–1614.
3. Berg JM (1988) Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc Natl Acad Sci U S A* 85: 99–102.
4. Lee MS, Gippert GP, Soman KV, Case DA, Wright PE (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 245: 635–637.
5. Dang DT, Pevsner J, Yang VW (2000) The biology of the mammalian Kruppel-like family of transcription factors. *Int J Biochem Cell Biol* 32: 1103–1121.
6. Pavletich NP, Pabo CO (1991) Zinc-finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252: 809–817.
7. Iuchi S (2001) Three classes of C<sub>2</sub>H<sub>2</sub> zinc finger proteins. *Cell Mol Life Sci* 58: 625–635.
8. Thiaville MM, Huang JM, Kim H, Ekram MB, Roh TY, et al. (2013) DNA-binding motif and target genes of the imprinted transcription factor PEG3. *Gene* 512: 314–320.
9. Deng Y, Wu X (2000) Peg3/Pw1 promotes p53-mediated apoptosis by inducing Bax translocation from cytosol to mitochondria. *Proc Natl Acad Sci U S A* 97: 12050–12055.
10. Relaix F, Wei XJ, Wu X, Sassoon DA (1998) Peg3/Pw1 is an imprinted gene involved in the TNF-NFκB signal transduction pathway. *Nat Genet* 18: 287–291.
11. Relaix F, Wei X, Li W, Pan J, Lin Y, et al. (2000) Pw1/Peg3 is a potential cell death mediator and cooperates with Siah1a in p53-mediated apoptosis. *Proc Natl Acad Sci U S A* 97: 2105–2110.
12. Jiang X, Yu Y, Yang HW, Agar NYR, Frado L, et al. (2010) The imprinted gene PEG3 inhibits Wnt signalling and regulates glioma growth. *J Biol Chem* 285: 8472–8480.
13. Li LL, Keverne EB, Aparicio SA, Ishino F, Barton SC, et al. (1999) Regulation of maternal behaviour and offspring growth by paternally expressed Peg3. *Science* 284: 330–333.
14. Swaney WT, Curley JP, Champagne FA, Keverne EB (2007) Genomic imprinting mediates sexual experience-dependent olfactory learning in male mice. *Proc Natl Acad Sci U S A* 104: 6084–6089.
15. Swaney WT, Curley JP, Champagne FA, Keverne EB (2008) The paternally expressed gene Peg3 regulates sexual experience-dependent preferences for estrous odors. *Behav Neurosci* 122: 963–973.
16. Champagne FA, Curley JP, Swaney WT, Hasen NS, Keverne EB (2009) Paternal influence on female behaviour: the role of Peg3 in exploration, olfaction, and neuroendocrine regulation of maternal behaviour of female mice. *Behav Neurosci* 123: 469–480.
17. Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA (1991) The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc Natl Acad Sci U S A* 88: 3608–3612.
18. Bardwell VJ, Treisman R (1994) The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* 8: 1664–1677.
19. Albagli O, Dhordain P, Deweindt C, Lecocq G, Leprince D (1995) The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth Differ* 6: 1193–1198.
20. Williams AJ, Khachigian LM, Shows T, Collins T (1995) Isolation and characterization of a novel zinc-finger protein with transcriptional receptor activity. *J Biol Chem* 270: 22143–22152.
21. Pengue G, Calabrò V, Bartoli PC, Pagliuca A, Lania L (1994) Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Res* 22: 2908–2914.
22. Sander TL, Stringer KF, Maki JL, Szauter P, Stone JR, et al. (2003) The SCAN domain defines a large family of zinc finger transcription factors. *Gene* 310: 29–38.
23. Williams AJ, Blacklow SC, Collins T (1999) The zinc finger-associated SCAN box is a conserved oligomerization domain. *Mol Cell Biol* 19: 8526–8535.
24. Sander TL, Haas AL, Peterson MJ, Morris JF (2000) Identification of a novel SCAN box-related protein that interacts with MZF1B. *J Biol Chem* 275: 12857–12867.
25. Schumacher C, Wang H, Honer C, Ding W, Koehn J, et al. (2000) The SCAN domain mediates selective oligomerization. *J Biol Chem* 275: 17173–17179.
26. Liang Y, Huimei Hong F, Ganesan P, Jiang S, Jauch R, et al. (2012) Structural analysis and dimerization profile of the SCAN domain of the pluripotency factor Zfp206. *Nucleic Acids Res* 40: 8721–8732.
27. Ivanov D, Stone JR, Maki JL, Collins T, Wagner G (2005) Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. *Mol Cell* 17: 137–143.
28. Peterson FC, Hayes PL, Waltner JK, Heisner AK, Jensen DR, et al. (2006) Structure of the SCAN domain from the tumor suppressor protein MZF1. *J Mol Biol* 363: 137–147.
29. Adamson CS, Jones IM (2004) The molecular basis of HIV capsid assembly—five years progress. *Rev Med Virol* 14: 107–121.
30. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server. Walker IJM, editor. Humana Press. 571–607.
31. Polekhina G, House CM, Traficante N, Mackay JP, Relaix F, et al. (2002) Siah ubiquitin ligase is structurally related to TRAF and modulates TNF-α signaling. *Nat Struct Biol* 9: 68–75.
32. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67: 271–281.
33. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67: 235–242.
34. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni L, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40: 658–674.
35. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67: 355–367.
36. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501.
37. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35: 375–383.
38. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372: 774–797.
39. Bond CS, Schüttelkopf AW (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments *Acta Crystallogr D Biol Crystallogr* 65: 510–512.
40. Schrödinger LLC. The PyMOL molecular graphics system, Version 1.5.0.4.
41. Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A* 32: 922–923.
42. Halazonetis TD, Georgopoulos K, Greenberg ME, Leder P (1988) c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. *Cell* 55: 917–924.
43. Espinas ML, Jimenez-Garcia E, Vaquero A, Canudas S, Bernues J, et al. (1999) The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. *J Biol Chem* 274: 16461–16469.
44. Yu HB, Kumarso G, Hong FH, Stanton LW (2009) Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. *J Biol Chem* 284: 31327–31335.
45. Noll L, Peterson FC, Hayes PL, Volkman BF, Sander T (2008) Heterodimer formation of the myeloid zinc finger 1 SCAN domain and association with promyelocytic leukemia nuclear bodies. *Leukemia Res* 32: 1582–1592.
46. House CM, Frew IJ, Huang HL, Wiche G, Traficante N, et al. (2003) A binding motif for Siah ubiquitin ligase. *Proc Natl Acad Sci U S A* 100: 3101–3106.
47. House CM, Hancock NC, Moller A, Cromer BA, Fedorov V, et al. (2006) Elucidation of the substrate-binding site of Siah ubiquitin ligase. *Structure* 14: 695–701.
48. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) ClustalW2 and ClustalX version 2. *Bioinformatics* 23: 2947–2948.
49. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60: 2256–2268.

## Acknowledgments

We acknowledge Dr. Derek Ogg for excellent support and Dr. Navratna Vajpai for running the NMR experiment.

## Author Contributions

Conceived and designed the experiments: VR WNH. Performed the experiments: VR. Analyzed the data: VR TCE WNH. Contributed reagents/materials/analysis tools: VR. Wrote the paper: VR TCE WNH.