REAL-LIFE EXAMPLES
The FTO gene codes for a novel member of the non-heme dioxygenase superfamily.

Luis Sánchez Pulido & Miguel Angel Andrade
Background

- Genetic variants in the FTO (fat mass and obesity associated) gene have been associated with an increased risk of obesity. 
  *Science, Nat. Genet., PLOS Genet.....*

- However, the function of its protein product has not been experimentally studied and previously reported sequence similarity analyses suggested the absence of homologs in existing protein databases.

![Description from UniProt for UniProt ID 'Q9C0B1']

- Here, we present the first detailed computational analysis of the sequence and predicted structure of the protein encoded by FTO.
### HHpred

<table>
<thead>
<tr>
<th>No</th>
<th>Hit</th>
<th>Prob</th>
<th>E-value</th>
<th>P-value</th>
<th>Score</th>
<th>SS</th>
<th>Cols</th>
<th>Query</th>
<th>HMM</th>
<th>Template HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2iuw_A Alkylated repair protein</td>
<td>97.0</td>
<td>3.2E-21</td>
<td>2.2E-25</td>
<td>85.3</td>
<td>0.0</td>
<td>1682</td>
<td>1-212</td>
<td>49-228</td>
<td>1-237</td>
</tr>
<tr>
<td>2</td>
<td>2fdi_A Alkylated DNA repair protein</td>
<td>96.6</td>
<td>3.1E-12</td>
<td>2.2E-16</td>
<td>54.2</td>
<td>0.0</td>
<td>1740</td>
<td>1-212</td>
<td>15-202</td>
<td>1-281</td>
</tr>
<tr>
<td>3</td>
<td>2f4i_A Hypothetical protein TM</td>
<td>45.1</td>
<td>0.019</td>
<td>1.3E-06</td>
<td>20.5</td>
<td>0.0</td>
<td>282</td>
<td>1-212</td>
<td>1-281</td>
<td>1-238</td>
</tr>
<tr>
<td>4</td>
<td>1li2t_A HYD protein; 1.04A [Homo sapiens]</td>
<td>38.9</td>
<td>0.038</td>
<td>2.7E-06</td>
<td>19.4</td>
<td>0.0</td>
<td>201</td>
<td>1-120</td>
<td>42-61</td>
<td>61</td>
</tr>
<tr>
<td>5</td>
<td>1zgk_A Kelch-like ECH-associat</td>
<td>35.6</td>
<td>0.055</td>
<td>3.9E-06</td>
<td>18.8</td>
<td>0.0</td>
<td>241</td>
<td>1-212</td>
<td>1-24</td>
<td>1-360</td>
</tr>
<tr>
<td>6</td>
<td>1xme_A Cytochrome C oxidase po</td>
<td>53.8</td>
<td>0.06</td>
<td>4.3E-06</td>
<td>18.7</td>
<td>0.0</td>
<td>241</td>
<td>1-212</td>
<td>1-24</td>
<td>1-568</td>
</tr>
<tr>
<td>7</td>
<td>1x13_A NAD(P) transhydrogenase</td>
<td>28.8</td>
<td>0.12</td>
<td>8.6E-06</td>
<td>17.6</td>
<td>0.0</td>
<td>241</td>
<td>1-212</td>
<td>1-24</td>
<td>1-401</td>
</tr>
<tr>
<td>8</td>
<td>1xx8_A Divalent cation tolerant</td>
<td>27.1</td>
<td>0.15</td>
<td>1.1E-05</td>
<td>17.3</td>
<td>0.0</td>
<td>271</td>
<td>1-212</td>
<td>1-27</td>
<td>1-147</td>
</tr>
<tr>
<td>9</td>
<td>2fbm_A Y chromosome chromodomain</td>
<td>25.1</td>
<td>0.19</td>
<td>1.4E-05</td>
<td>17.0</td>
<td>0.0</td>
<td>251</td>
<td>1-212</td>
<td>1-25</td>
<td>1-291</td>
</tr>
<tr>
<td>10</td>
<td>1lknP_Transcription factor SK</td>
<td>23.7</td>
<td>0.23</td>
<td>1.6E-05</td>
<td>16.7</td>
<td>0.0</td>
<td>241</td>
<td>1-212</td>
<td>1-24</td>
<td>1-92</td>
</tr>
<tr>
<td>11</td>
<td>1lae_P53; 1.5A [Homo sapiens]</td>
<td>22.2</td>
<td>0.28</td>
<td>2.0E-05</td>
<td>16.4</td>
<td>0.0</td>
<td>311</td>
<td>56-89</td>
<td>1-31</td>
<td>1-31</td>
</tr>
<tr>
<td>12</td>
<td>2oza_A Probable 6-pyruvyl tet</td>
<td>20.2</td>
<td>0.37</td>
<td>2.6E-05</td>
<td>16.0</td>
<td>0.0</td>
<td>251</td>
<td>188-212</td>
<td>1-36</td>
<td>1-138</td>
</tr>
<tr>
<td>13</td>
<td>1vk0_A Inositol-3-phosphate sy</td>
<td>20.0</td>
<td>0.38</td>
<td>2.7E-05</td>
<td>15.9</td>
<td>0.0</td>
<td>281</td>
<td>185-212</td>
<td>1-28</td>
<td>1-537</td>
</tr>
</tbody>
</table>

### HHpred

Command: `~/usr/people/sanchez/miSOF/hhsearch -glob -ssm 0 -i ftoSelng20.hmm -d ~/usr/people/sanchez/databases/pdb70.hmm -o ftoSelng20.HHresPdb70`

<table>
<thead>
<tr>
<th>Source</th>
<th>Domain</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfamA</td>
<td>2OG-Fe(l)Oxy</td>
<td>98</td>
<td>194</td>
</tr>
</tbody>
</table>

Description from UniProt for 02BPN4:

- **hypothesised protein**

[Diagram showing UniProt description]

- [195 residues]
Pfam entry 20G-Fell_Oxy

Accession number: PF03171

20G-Fe(II) oxygenase superfamily

This family contains members of the 2-oxoglutarate (20G) and Fe(III)-dependent oxygenase superfamily [1]. This family includes the C-terminal of prolyl 4-hydroxylase alpha subunit. The holoenzyme has the activity EC:1.14.11.2 catalysing the reaction: Procollagen I-proline + 2-oxoglutarate + O2 ↔ procollagen II-trans-4-hydroxyproline + succinate + CO2. The full enzyme consists of a alpha2 beta2 complex with the alpha subunit contributing most of the parts of the active site [3]. The family also includes lysyl hydrolases, isopenicillin syntheses and AikB.

INTERPRO description (entry IPR005123)

This family contains members of the 2-oxoglutarate (20G) and Fe(III)-dependent oxygenase superfamily PUBMED:11275424. This family includes the C-terminal of prolyl 4-hydroxylase alpha subunit. The holoenzyme has the activity [1] catalysing the reaction:

The full enzyme consists of a alpha2 beta2 complex with the alpha subunit contributing most of the parts of the active site PUBMED:7753822. The family also includes lysyl hydrolases, isopenicillin syntheses and AikB.

For additional annotation, see the PROSITE document PRO001028

Alignment

- Seed (.41)  - Full (2565)

Domain organisation

- View 45 representative architectures
FTO Family

Description from UniProt for Q9C0B1
kiaa1752 protein (fragment)

<table>
<thead>
<tr>
<th>Source</th>
<th>Domain</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>seg</td>
<td>Low Complexity</td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td>seg</td>
<td>Low Complexity</td>
<td>155</td>
<td>167</td>
</tr>
<tr>
<td>seg</td>
<td>Low Complexity</td>
<td>346</td>
<td>356</td>
</tr>
</tbody>
</table>
Conclusions

- Our analysis suggests that human FTO is a member of the non-heme dioxygenase (Fe(II)- and 2-oxoglutarate-dependent dioxygenases) superfamily.

- Amino acid conservation patterns support this hypothesis and indicate that both 2-oxoglutarate and iron should be important for FTO function.

- This computational prediction of the function of FTO should suggest further steps for its experimental characterization and help to formulate hypothesis about the mechanisms by which it relates to obesity in humans.
The faster experimental validation that have ever seen!!!!!!!!!!!!!!!!!!
Crystal structure of the FTO protein reveals basis for its substrate specificity

Zhifu Han, Tianhui Niu, Junbiao Chang, Xiaoguang Lei, Mingyan Zhao, Qiang Wang, Wei Cheng, Jinjing Wang, Yi Feng & Jijie Chai
POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins.

CNB – CSIC & University of Namur

Initial Protein: OMP85/D15 ~ 800 Aa

Function: Outer membrane protein transport and integration
POTRA: A new domain and, sometimes, repeated
We proposed that POTRA domains could work.

In Analogy to: SicP (transport system III)

Each SptP protein (substrate to be transported) is wrapped around by three SicP chaperones.
The proposed model was accepted and validated by different experimental groups.
“The N-terminal 60 residues of the periplasmic part of FtsQ superimpose on the second POTRA domain of YaeT with an Rmsd of 1.8 Å. The regular appearance of the hydrophobic core residues was the reason for Sánchez-Pulido et al. to predict this region would adopt a POTRA domain-like fold, which the crystal structure confirms.”
**MAL protein Family**

MARVEL: a conserved domain involved in membrane apposition events.


CNB – CSIC & CBM - CSIC

---

**Initial Protein:** MAL

**Function:** Membrane Traffic

Apical Zone of polarized Epithelial Cells - TransGolgi - Endosomes - Membrane -
Diverse... but have COMMON Functional Elements

**Conclusions:**
MARVEL domain could be used as machinery for raft organization in membrane apposition events, such as those occurring during biogenesis of transport vesicles (e.g. MAL, physins and gyrins) or the formation of specialized close contacts (kisses) in tight junctions (e.g. occludin).
Vascular permeability in ocular disease and the role of tight junctions

Kathryn K. Erickson¹, Jeffrey M. Sundstrom¹ and David A. Antonetti¹

¹ Department of Cellular and Molecular Physiology, Penn State College of Medicine, Hershey, PA 17033, USA

Structure

Structure of synaptophysin: A hexameric MARVEL domain channel protein

Christopher P. Arthur¹ and Michael H. B. Stowell¹

¹ MCD Biology, University of Colorado, Boulder, CO 80309

Regulation of EGF receptor signaling by the MARVEL domain-containing protein CKLFSF8

Edited by Veli-Pekka Lehto

Cailing Jin, Peiguo Ding, Ying Wang, Dalong Ma
HELP in experimental Characterization
Essential Genes in fusion of myoblasts into multi-nucleated fibers (myotubes)

**The MARVEL domain protein, Singles Bar, is required for progression past the pre-fusion complex stage of myoblast fusion.**

*Estrada B, Maeland AD, Gisselbrecht SS, Bloor JW, Brown NH, Michelson AM.*

Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA.

Multinucleated myotubes develop by the sequential fusion of individual myoblasts. Using a convergence of genomic and classical genetic approaches, we have discovered a novel gene, singles bar (sing), that is essential for myoblast fusion. sing encodes a small multipass transmembrane protein containing a MARVEL domain, which is found in vertebrate proteins involved in processes such as tight junction formation and vesicle trafficking where--as in myoblast fusion--membrane apposition occurs. sing is expressed in both founder cells and fusion competent myoblasts preceding and during myoblast fusion.
Protein Domains
identification in the twilight zone of protein sequence analysis.

A common ancestry for BAP1 & Uch37 Regulators

De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome

Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB

Bohring-Opitz syndrome is characterized by severe intellectual disability, distinctive facial features and multiple congenital malformations. We sequenced the exomes of three individuals with Bohring-Opitz syndrome and in each identified heterozygous de novo nonsense mutations in ASXL1, which is required for maintenance of both activation and silencing of Hox genes. In total, 7 out of 13 subjects with a Bohring-Opitz phenotype had de novo ASXL1 mutations, suggesting that the syndrome is genetically heterogeneous.
ASXL1 & BAP1

Beisel & Paro
Silencing chromatin: comparing modes and mechanisms.
What we know about ASXL1?

ASXL1 and BAP1

---

ASXL1 and BAP1

~200 papers

But... since the point of view of computational protein sequence analysis

ASXL1 is a
UNKNOWN Protein
ASXL1_HUMAN

Fh  DEU  PHD

DEU
DEUBAD domain  PHD  PHD domain

Fh
Forkhead domain
Stereo views of the recognition helices and DNA interactions.

FORKHEAD Domain

Tsai K et al. J. Biol. Chem. 2006;281:17400-17409
Non-specific DNA backbone binding

Sequence-specific DNA binding

Human HNF3G numbering
<table>
<thead>
<tr>
<th>Protein</th>
<th>Accession</th>
<th>Start</th>
<th>End</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASXL1</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
<tr>
<td>ASXL3</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
<tr>
<td>ASXL2</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
<tr>
<td>G3Y8H</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
<tr>
<td>E3PM2</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
<tr>
<td>R0VU7</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
<tr>
<td>D4N02</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
<tr>
<td>E2B6P</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
<tr>
<td>ASX</td>
<td>XP_545117</td>
<td></td>
<td></td>
<td>Human</td>
<td>PPIL1X49RSALN</td>
</tr>
</tbody>
</table>
DEUBAD domain
Multiple domain architectures

- ASXL1_HUMAN
- ASX_DROME
- Q4WDB3_ASMFY
- GAT27_ARATH
- INO80G/NFRKB_HUMAN
- Q9LZU7_ARATH
- hRpn13/ADRM1_HUMAN
Functional Hypothesis

Biochemical Knowledge

PubMed
Scheuermann et al., 2010; Bott et al., 2011.

Known

Ubiquitin C-terminal Hydrolase domain

Uch37-like domain

Conserved region

Pru domain

Forkhead domain

DEUBAD domain

PHD domain
Structure of Proteasome Ubiquitin Receptor hRpn13 and Its Activation by the Scaffolding Protein hRpn2

Xiang Chen, Byung-Hoon Lee, Daniel Finley, Kylie J. Walters.
BAP1 and Uch37 are homologous proteins!!!!

It is nice to see conservation of interaction partners between paralogs!

---

**Diagram Details:**
- **ASXL1_HUMAN**
- **BAP1_HUMAN**
- **hRpn13/ADRM1_HUMAN**
- **INO80G/NFRKB_HUMAN**
- **Uch37/UCHL5_HUMAN**

**Deubiquitination Domains:**
- DEUBAD domain
- Forkhead domain
- PHD domain
- Pru domain
- Conserved region
- Uch37-like domain
- Ubiquitin C-terminal Hydrolase domain

**References:**
- Scheuermann et al., 2010; Bott et al., 2011.
- Hamazaki et al., 2006; Yao et al., 2006; Qiu et al., 2006; Chen et al., 2010.

**Known:**
ASXL1_HUMAN

BAP1_HUMAN

hRpn13/ADRM1_HUMAN

INO80G/NFRKB_HUMAN

Uch37/UCHL5_HUMAN

---

Scheuermann et al., 2010; Bott et al., 2011.

Hamazaki et al., 2006; Yao et al., 2006; Qiu et al., 2006; Chen et al., 2010.

Known
Distinct Modes of Regulation of the Uch37 Deubiquitinating Enzyme in the Proteasome and in the Ino80 Chromatin-Remodeling Complex

Tingting Yao¹, Ling Song², Jingji Jin¹, Yong Cai¹, Hidehisa Takahashi¹, Selene K. Swanson¹, Michael P. Washburn¹, Laurence Florens¹, Ronald C. Conaway¹, Robert E. Cohen³, Joan W. Conaway¹, ¹, ¹, ⁴, ⁵

What is doing a Ubiquitin protease in a Histone-Exchange Complex
A Scheuermann et al., 2010; Bott et al., 2011.
B Hamazaki et al., 2006; Yao et al., 2006; Qiu et al., 2006; Chen et al., 2010.
C Yao et al., 2008. Known

**DEUBAD domain**
**Forkhead domain**
**PHD domain**
**Pru domain**
**Conserved region**
**Uch37-like domain**
**Ubiquitin C-terminal Hydrolase domain**
Both interact with Uch37 and....
The experimentalist did not realize that both proteins share an homologous domain.
ASXL1_HUMAN

BAP1_HUMAN

hRpn13/ADRM1_HUMAN

INO80G/NFRKB_HUMAN

Uch37/UCHL5_HUMAN

---

DEUBAD domain
Forkhead domain
PHD domain
Pru domain
Conserved region
Uch37-like domain
Ubiquitin C-terminal Hydrolase domain

A Scheuermann et al., 2010; Bott et al., 2011.
B Hamazaki et al., 2006; Yao et al., 2006; Qiu et al., 2006; Chen et al., 2010.
C Yao et al., 2008. Known
Why do we analyse sequences??
because....

*Thanks to the recognition of homology between proteins, we can*

**TRANSFER INFORMATION**

- **Structural**
  *from HOMOLOGOUS proteins of known structure (X-Ray, NMR o EM)*

- **Functional**
  *from experimentally characterised HOMOLOGOUS proteins*
ASXL1_HUMAN

BAP1_HUMAN

hRpn13/ADRM1_HUMAN

INO80G/NFRKB_HUMAN

Uch37/UCHL5_HUMAN

DEUBAD domain

PHD domain

Forkhead domain

Pru domain

Conserved region

Uch37-like domain

Ubiquitin C-terminal Hydrolase domain

A Scheuermann et al., 2010; Bott et al., 2011.

B Hamazaki et al., 2006; Yao et al., 2006; Qiu et al., 2006; Chen et al., 2010.

C Yao et al., 2008. Known
Conclusions:

ASXL1 has two previously undetected domains:

I) a Forkhead DNA binding domain and,
II) a Deubiquitinase Adaptor domain shared by the known Uch37 regulators: ADRM1 (also known as hRpn13) and NFRKB.

Our analysis reveals a common ancestry for the BAP1 and Uch37 regulators in: PR-DUB (Polycomb repressive deubiquitinase), INO80, and Proteosome complexes.

In summary,

the identification of the DEUBAD and Forkhead domains in the human ASXL1 protein may help to shed light on its critical function in regulating the histone code under non-pathological and pathological conditions.

Thanks to Chris Ponting
The discovery of sequence similarities among proteins is a powerful approach to infer structural, functional and evolutionary relationships.

Computational Protein Analysis can lead to experimentally-tractable hypotheses and result in a better understanding of protein functions.

Truly convinced that:
“Nothing in Biology Makes Sense Except in the Light of Evolution”
Theodosius Dobzhansky

Thanks to Protein Sequence Analysis
IMAGINATION CAN FLY
(in the right direction)

Opening avenues for Experimental Research