

# A Novel N-Terminal Domain May Dictate the Glucose Response of Mondo Proteins

Lisa G. McFerrin<sup>1\*</sup>, William R. Atchley<sup>1,2</sup>

**1** Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, United States of America, **2** Department of Genetics, North Carolina State University, Raleigh, North Carolina, United States of America

## Abstract

Glucose is a fundamental energy source for both prokaryotes and eukaryotes. The balance between glucose utilization and storage is integral for proper energy homeostasis, and defects are associated with several diseases, e.g. type II diabetes. In vertebrates, the transcription factor ChREBP is a major component in glucose metabolism, while its ortholog MondoA is involved in glucose uptake. Both MondoA and ChREBP contain five Mondo conserved regions (*MCR1-V*) that affect their cellular localization and transactivation ability. While phosphorylation has been shown to affect ChREBP function, the mechanisms controlling glucose response of both ChREBP and MondoA remain elusive. By incorporating sequence analysis techniques, structure predictions, and functional annotations, we synthesized data surrounding Mondo family proteins into a cohesive, accurate, and general model involving the *MCRs* and two additional domains that determine ChREBP and MondoA glucose response. Paramount, we identified a conserved motif within the transactivation region of Mondo family proteins and propose that this motif interacts with the phosphorylated form of glucose. In addition, we discovered a putative nuclear receptor box in non-vertebrate Mondo and vertebrate ChREBP sequences that reveals a potentially novel interaction with nuclear receptors. These interactions are likely involved in altering ChREBP and MondoA conformation to form an active complex and induce transcription of genes involved in glucose metabolism and lipogenesis.

**Citation:** McFerrin LG, Atchley WR (2012) A Novel N-Terminal Domain May Dictate the Glucose Response of Mondo Proteins. PLoS ONE 7(4): e34803. doi:10.1371/journal.pone.0034803

**Editor:** Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

**Received:** August 16, 2011; **Accepted:** March 8, 2012; **Published:** April 10, 2012

**Copyright:** © 2012 McFerrin, Atchley. 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:** The authors have no support or funding to report.

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

\* E-mail: Lisa.McFerrin@gmail.com

## Introduction

Glucose is a carbohydrate in the form of a simple sugar that is an important source of energy for both eukaryotes and prokaryotes. However, glucose regulation is complex and not well understood. Extensive work has been devoted to the function of individual components within known metabolic pathways, e.g. [1–3], yet our understanding of their coordinated roles in response to different metabolic and cancerous conditions is just beginning to take shape. The discovery of additional regulatory factors such as ChREBP and MondoA broach this issue, but still need to be incorporated in current models of glucose sensing and regulation.

### SREBF1 and ChREBP promote glucose storage in mammals

In mammals, the liver is the primary organ that controls energy homeostasis by processing glucose for energy or storage. In fasting conditions, the liver produces glucose via *de novo* synthesis (gluconeogenesis) or decomposition of glycogen (glycogenolysis). Glucose can then be converted to pyruvate through glycolysis and subsequently enter the citric acid (TCA) cycle within mitochondria to produce energy. In contrast, when excess carbohydrates are consumed, glucose can be stored according to two major pathways. Insulin induced enzymes trigger the glycogen synthase pathway to store glucose as glycogen. Alternatively, glucose can be converted to triglycerides through the *de novo* lipogenesis pathway for a more compact form of storage. Triglycerides within the liver

can be further packaged into lipoproteins (i.e. VLDL, LDL, HDL) and transported into the blood stream and other tissues.

Initially, sterol regulatory element binding transcription factor 1 (SREBF1) was identified as the major factor involved in glucose metabolism and insulin response [4]. However, knockout experiments revealed an additional factor was necessary for the full glucose-dependent transactivation of certain lipogenic genes, e.g. acetyl-CoA carboxylase (ACC) and fatty acid synthase (FAS) [5–7]. The discovery of a conserved carbohydrate response element (ChORE) consisting of two E-boxes separated by exactly 5 residues (CACGTG<sub>5</sub>CACGTG) within the promoters of such genes facilitated the identification of this glucose responsive element [8]; ChORE binding protein ChREBP has subsequently been implicated in transactivation of several genes that regulate the *de novo* lipogenesis pathway, e.g. liver pyruvate kinase (L-PK), malic enzyme (ME), glucose phosphoisomerase (GPI), ACC, and FAS [9].

ChREBP protein, also named WBSR14, MondoB and MLXIPL, has a paralog in vertebrates named MondoA or MLXIP. Interestingly, MondoA and ChREBP have overlapping yet distinct expression profiles, which underly their downstream effects and separate roles in regulating genes involved in glucose metabolism. MondoA can restrict glucose uptake and influences energy utilization, while ChREBP signals energy storage through *de novo* lipogenesis [10,11]. Only a single Mondo gene has been identified in invertebrate animals [12], including *Drosophila melanogaster* (*dmondo/mio*) and *Caenorhabditis elegans* (*mml-1*)

*T20B12.6*) [13]. We refer to this single ortholog in invertebrates as non-vertebrate Mondo. In addition, while domain names are not generally italicized, we adopt this naming convention to avoid confusion with protein references.

### ChREBP and MondoA are glucose responsive

Current evidence shows both ChREBP and MondoA are glucose responsive, whereby they are mainly located in the cytoplasm under low glucose conditions and have increased nuclear accumulation and transactivation of target genes in high glucose medium [10,14,15]. This nuclear translocation and DNA binding is dependent upon the dimerization to obligate partner Mlx, a Max-like transcription factor, which is ubiquitously expressed. Mlx and Mondo proteins contain a C-terminal basic Helix-Loop-Helix-Leucine Zipper (*bHLH*ζ) domain responsible for DNA binding and dimerization as well as a dimerization and cytoplasmic localization (*DCD*) domain that must be masked prior to nuclear entry [16,17]. As shown for MondoA, dimerization through either the *bHLH*ζ or *DCD* region is sufficient to block this cytoplasmic retention signal (CRS), but not sufficient for nuclear translocation [16–18].

Since MondoA and ChREBP are mainly cytoplasmic proteins, it was surprising to find that trapping them within the nucleus in low glucose conditions was not sufficient to replicate the transactivation potential [19,20]. Consistent with this, both MondoA and ChREBP are known to shuttle between the cytoplasm and nucleus in both low and high glucose conditions, yet have increased transactivation only under high glucose. In contrast, proteins lacking the N-terminus are able to constitutively transactivate genes in both glucose mediums [16,21–23], indicating additional N-terminal domains within MondoA and ChREBP contribute to their nuclear accumulation and transactivation in response to glucose [16,21].

### N-terminal conserved regions regulate ChREBP and MondoA activity

MondoA and ChREBP proteins have five Mondo Conserved Regions (*MCR I–V*) in their N-terminus. These have previously been reported as *PADRE1*, *PADRE2*, and *MADRE* [24] as well as a low glucose inhibitory domain (*LID*) which spans *MCR I–IV* and glucose responsive activation conserved element (*GRACE*) which contains *MCR V* [21]. The distances between *MCR II*, *MCR III*, and *MCR IV* are also conserved, implying they act as a functional module, while the regions linking *MCR I* and *MCR V* vary between MondoA and ChREBP [18]. *MCR II* contains a strong CRM1 dependent nuclear export signal (NES), almost identical to the high affinity LxxLFxxLSV motif. In contrast, *MCR IV* in ChREBP contains a bipartite nuclear localization signal (NLS) that mediates its nuclear entry [15,25]. Between these two regions *MCR III* contains a binding motif recognized by the 14-3-3 protein that is involved in ChREBP and MondoA cytoplasmic retention, transactivation, and nuclear export [16,22,26]. The functions of *MCR I* and *MCR V* are not as clear, although *MCR I* is necessary for glucose dependent transactivation in ChREBP [27] and *MCR V* is within the *GRACE* region responsible for transactivation [21].

The N-terminal *LID*, containing *MCR I–IV*, possesses a robust repressive mechanism that regulates the strong transactivation region within the *GRACE*. Contrary to prediction, individually deleting or mutating *MCR I*, *II*, *III*, or *IV* also abolishes MondoA or ChREBP transactivation in response to glucose [17,22,23]. Hence the *LID* participates in repression in low glucose and activation in high glucose, where no individual *MCR* can sufficiently replicate the glucose response. Moreover, reversing the order of *LID* and *GRACE* regions results in a constitutively active ChREBP protein,

indicating its structure and intramolecular contacts are major factors in regulating its function [21].

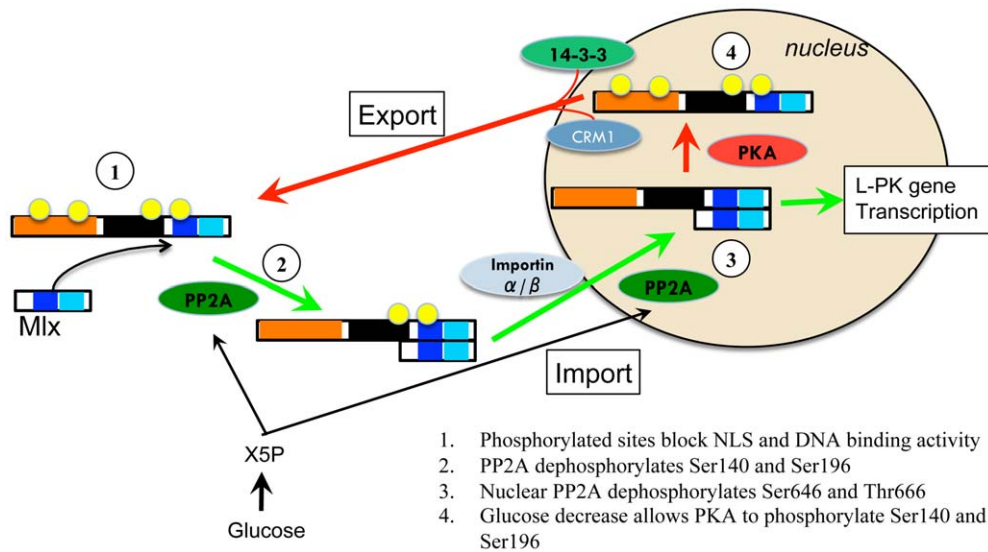
Deletion and mutation constructs further show each *MCR* seems to have multiple and often opposing function. *MCR I* is necessary for glucose response, since alterations to *MCR I* (ChREBP: Δ1–71, Δ1–58; MondoA: Δ1–100, H78A/H81A/H88A) block transactivation in high glucose, yet mimicking phosphorylation (ChREBP: S56D) enhances it [16,19,22,27]. Likewise, altering the NES in *MCR II* (ChREBP: L89A, F90A; MondoA: F130A, M133A, Δ125–137) mildly enhances transactivation, while other mutations in *MCR II* (ChREBP: L86A/L93A, T85A, L95A, Δ72–99; MondoA: L129A) completely block it [17,28,29]. In *MCR III*, abrogating 14-3-3 protein binding sites (ChREBP: R128A, W130A; MondoA: I166A/W167A/R168A) inhibit transactivation, but so do mutations (ChREBP: N123A, I126A, Δ100–115) that are still capable of interacting with 14-3-3 [17,19,22].

Intriguingly, changes within *MCR IV* have even more diverse effects. Some changes (ChREBP: Δ141–197, Δ158–181) likely block the NLS and thus prevent transactivation [23,25], one change (ChREBP: Δ144–196) reduces transactivation function yet also removes glucose dependent inhibition [22], while another change (MondoA: Y210D/W211D/K212) increases nuclear accumulation and transactivation [17]. While *MCR V* shows no repressive effects in the absence of *MCR I–IV*, changes to it (ChREBP: Y275A/V276A/G277A, L289A/Q290A/P291A; MondoA: Δ282–324) within the full-length sequence lead to an increase in nuclear accumulation and transactivation [17,19]. Although the cellular conditions, site mutations, and reporter assays in these studies greatly vary, they individually and in combination suggest that the *MCRs* cooperatively repress and activate MondoA and ChREBP function in response to glucose.

### Current models of ChREBP and MondoA glucose response are incomplete

To properly balance glucose storage and usage, extracellular signals instigate the expression and phosphorylation of proteins involved in the lipogenic pathway. ChREBP contains several such phosphorylation sites [21]. A ChREBP based phosphorylation model postulates that during starvation glucagon increases the concentration of cAMP in hepatocytes, which triggers the phosphorylation of ChREBP by cAMP dependent protein kinase A (PKA) [15]. Phosphorylation of ChREBP site Ser196 causes an adjacent bipartite nuclear localization signal (NLS) in *MCR IV* to be blocked and ChREBP to be sequestered in the cytosol [25]. Conversely, dephosphorylation events mediate a conversion to energy storage rather than usage after a high carbohydrate meal. Increased glucose and thus accelerated glycolytic flux increases the concentration of intermediate metabolite Xylulose-5-phosphate (X5P) within the pentose phosphate shunt, which stimulates protein phosphatase 2A (PP2A) [30]. Cytosolic PP2A mediated dephosphorylation of S196 in ChREBP results in its nuclear localization, while ChREBP DNA binding and transactivation is enhanced by further dephosphorylation of sites S626 and T666 via X5P activated PP2A in the nucleus (Figure 1) [29].

While this simple model is attractive, it is not complete and several issues remain unresolved. Foremost, mimicking the phosphorylation status in ChREBP is not sufficient to activate transcriptional machinery in low glucose [21]. Moreover, MondoA is glucose responsive although it does not contain many of the phosphorylation sites found in ChREBP. In light of recent work, new evidence indicates phosphorylation of glucose by hexokinase to form G6P has a direct impact on the activation of MondoA and ChREBP, although the mechanism is still not known [17,31]. How G6P is able to promote transactivation within the



**Figure 1. Phosphorylation model depicting ChREBP response to glucose.** Image adapted from [29]. 1) In low glucose conditions, sites S140/S196/S626/T666 are phosphorylated and block the NLS and DNA binding activity. 2) Upon glucose stimulation, X5P activates PP2A to dephosphorylate S140/S196 in the cytosol, unblocking the NLS, and allowing ChREBP to enter the nucleus. 3) Nuclear PP2A dephosphorylation of S626/T666 increases DNA binding. 4) Decreased glucose levels increase PKA activity to phosphorylate S140/S196 and shuttle ChREBP back to the cytoplasm.

doi:10.1371/journal.pone.0034803.g001

*GRACE* and override the N-terminal repression imposed by the *LID* region is an important, yet unanswered question. In addition, low glucose repression seems to be independent of a cofactor and is likely a result of protein conformation [23]. Determining the function and interactions of *MCRs* within the N-terminus is of great importance to understanding MondoA and ChREBP glucose response and transactivation of genes involved in glucose metabolism. Since a significant fraction of tumor cells exhibit an increase in glucose metabolism and direct glucose into *de novo* lipogenesis [32,33], understanding the specific roles of MondoA and ChREBP in glucose regulation can directly affect the treatment of such diseases.

Herein, we synthesize the current knowledge of Mondo family proteins and domains into a cohesive, accurate, and generalized model to address Mondo activation in response to glucose. First, we hypothesize that MondoA and ChREBP domains function analogously and defend that their overall conservation implies similar structure and function among Mondo proteins. Second, we identify a novel domain and propose it is involved in sensing changing glucose levels and altering Mondo transactivation potential. Finally, we form a unified model based on current data that explains MondoA and ChREBP subcellular localization and transactivation in response to glucose. Together, this information forms a more complete picture for how Mondo proteins, in general, respond to elevated glucose levels and creates a series of testable hypothesis, which can be experimentally validated to refine our understanding of glucose metabolism.

## Results

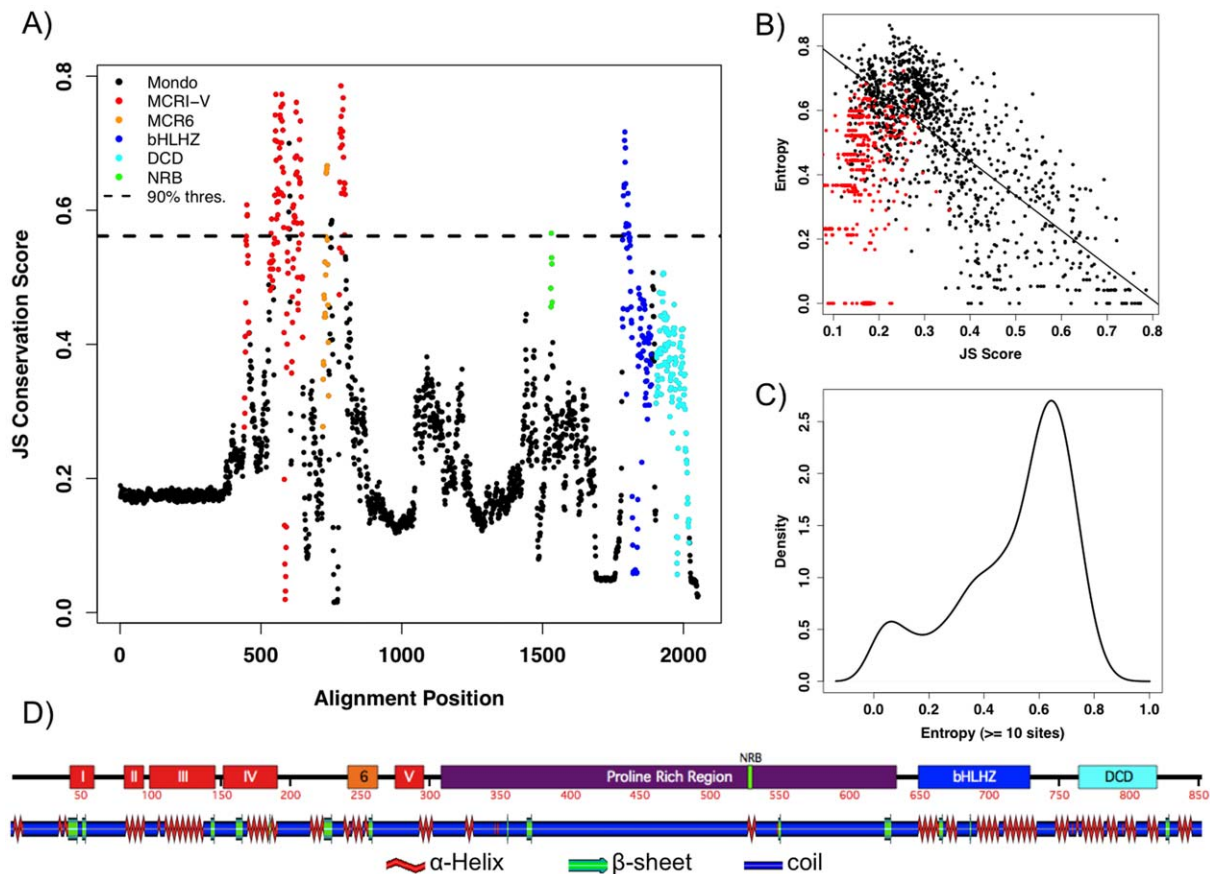
### *MCRI-V*, *bHLHZ*, and *DCD* domains are conserved among Mondo protein sequences

According to previous reports [16,18], the similarity within Mondo protein sequences is largely contained within the *MCRI-V*, *bHLHZ*, and *DCD* domains. However, the BLOCKS and MEME approaches in these papers were used to simply present delimited regions of increased conservation without commenting on the

constraints or functional contribution of each residue. Here we include orthologous Mondo proteins from several ancient and intermediate lineages, such as the Placozoa *Trichoplax adhaerens* and Cnidaria *Nematostella vectensis* to help explicate the evolution of Mondo conserved domains as well as the imposed functional constraints.

To more precisely identify and quantify the conservation within Mondo family proteins among diverse organisms, we created a multiple sequence alignment consisting of numerous species sampled across the animal kingdom (see Methods). This allowed us to directly observe the conservation of each alignment column through the Jenson-Shannon Divergence (JS) score (Figure 2), which rates each site by an autocorrelated conservation value [34]. Since conservation is a powerful predictor for detecting functional sites, sites within more conserved regions have higher JS values and are thus more likely to affect protein function (Figure 2a). Similarly, entropy (H) measures the amount of information or variability within an alignment column where conserved sites have low entropy values. As expected, sites within the *MCR*, *bHLHZ*, or *DCD* regions are highly conserved and have correspondingly high JS and low H values.

However, the relationship between JS and H is nonlinear due to several autapomorphies within the full sequence alignment (Figure 2b). In these cases, sequence specific insertions or poor prediction of exon boundaries for unannotated sequences create alignment columns with just a single or few residues. By removing alignment positions with less than ten residues, we were able to recover the correlation between entropy and JS scores ( $r^2 = 0.55$ ), as well as reveal two peaks in entropy values (Figure 2c). From this reduced dataset, 127 (11.6%) sites are considered highly conserved with  $H < 2.0$ , while most other sites are variable. Since JS values are scored using an adjacency window, the JS distribution is smoothed to form a single peak and there is no clear delineation of conserved and variable sites (data not shown). In accordance with entropy values, setting an arbitrary 90% threshold ( $JS > 0.5597$ ) shows the most conserved sites are within the *MCR* and *bHLHZ* regions (Figure 2a).



**Figure 2. Mondo sequence and structure conservation.** **A) JS Conservation Score.** All Mondo sequences were used to construct an alignment of homologous sites. Black dots represent alignment columns, while sites within domains are colored: red: *MCR1-V*, orange: Myc box II-like (*MCR6*), green: nuclear receptor box, blue: basic helix-loop-helix-zipper, cyan: *DCD*. The dashed line sets the 90% threshold for JS scores **B) JS and Entropy Comparison.** red: sites with less than 10 residues, black at least 10 residues, where linear regression was performed on the latter with intercept = 0.8745, slope = -1.0803,  $r^2 = 0.55467$ . **C) Entropy Distribution.** Distribution of entropy values for sites with at least 10 residues **D) Domains and Secondary Structure.** Consensus secondary structure for ChREBP shown alongside its sequence domains. doi:10.1371/journal.pone.0034803.g002

High JS scores were also observed for two new and potentially important regions. The first region, which we name Mondo Conserved Region 6 (*MCR6*), was previously reported as a MBII-like region located between *MCRIV* and *MCRV* [18]. However, the MBII-like region designated by the previous alignment showed little similarity in amino acid composition. From our dataset, we were able to improve the alignment and identify a highly conserved [ST]DTLF[ST] motif, where [ST] indicates either a serine or threonine. The conservation of *MCR6* residues, as well as *MCR1-V*, is depicted by the weblogs in Figure 3, where larger letters indicate more conserved sites. Based on the distribution of amino acids, we propose *MCR6* be defined by the 12 residue sequence signature [MLD][SNED][EDML] [FIM][ST]DTLF [ST][STM][LTI].

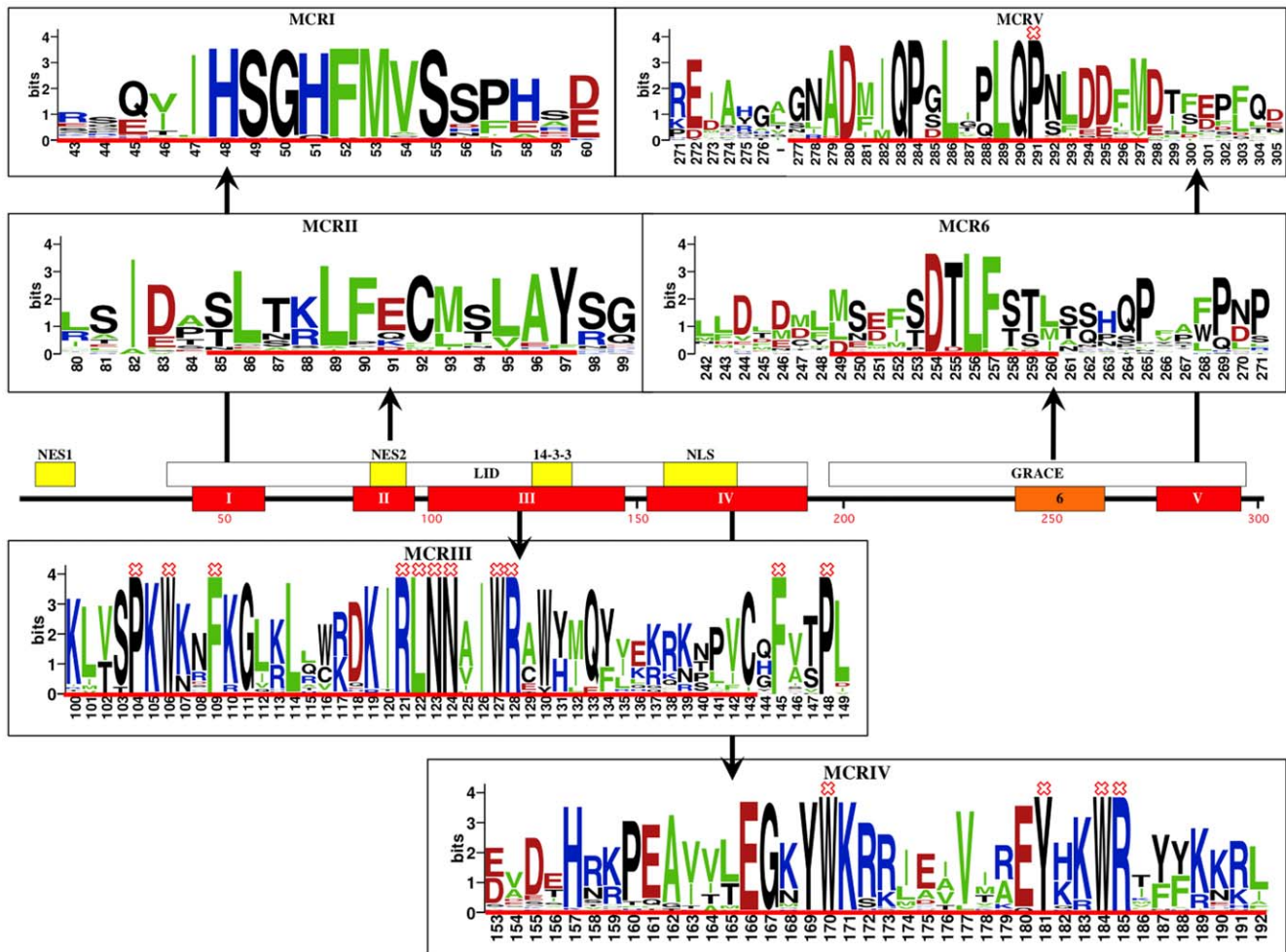
#### Mondo proteins exhibit divergent domains

JS scores also revealed a novel LxQLLT motif located within the central region of ChREBP and non-vertebrate Mondo protein sequences, but not MondoA (Figure 4). This sequence conforms to the LxxLL nuclear receptor box (*NRB*) signature that participates in the ligand dependent activation of nuclear receptors. *NRBs* are found within nuclear receptor coactivators such as the SRC-1 family of proteins (pfam ID: PF08832), which typically have multiple repeats of this motif, each sufficient for ligand interaction with several nuclear receptors [35]. Non-vertebrate Mondo and

ChREBP proteins only contain one putative *NRB*. Interestingly, ChREBP and nuclear receptor HNF4 $\alpha$  have adjacent recognition sequences in the promoter sequence of liver pyruvate kinase (L-PK) [9,36–38]. Full activation of the L-PK gene requires both ChREBP and HNF4 $\alpha$  [37], and ChREBP:HNF4 $\alpha$ :CBP is recruited as a complex to the L-PK promoter region in a glucose dependent manner [39]. Taking this into consideration, it is reasonable to assume that the ChREBP *NRB* is capable of activating HNF4 $\alpha$ .

Conversely, MondoA, but not ChREBP, localizes specifically to the outer mitochondrial membrane (OMM) when in the cytosol [10]. Mitochondria import stimulating factor (MSF) was identified as a mitochondrial chaperone and is a member of the 14-3-3 protein family [40]. Chaperone proteins transport cargo proteins to the mitochondria that contain a presequence located in the distal N-terminus. Generally, mitochondrial surface proteins cleave this preprotein sequence, which allows the mature protein to enter through the mitochondrial membrane. However, some OMM proteins have a distal N-terminal, preprotein sequence that is not cleaved. In these few cases, this sequence is used for mitochondrial targeting, but not cleavage or import [41].

We find that MondoA, but not ChREBP or non-vertebrate Mondo proteins, are predicted to contain mitochondrial targeting peptides within the first 42 residues, as specified by the program TargetP [42]. MondoA is not known to enter the mitochondria



**Figure 3. Mondo conserved regions.** MondoA and ChREBP have five previously defined and uniquely conserved regions, i.e. *MCR*I-V. These have been grouped into the *LID* and *GRACE* regions in ChREBP, and annotated for nuclear export signals (NES1, NES2),  $\alpha$ -helix necessary for 14-3-3 binding, and a bipartite nuclear localization signal. These domains, along with newly identified *MCR*6, are highly conserved among Mondo sequences, with Mondo invariant positions marked with a red 'X'. Weblogs depicting the particularly conserved sites and regions were created using the full Mondo alignment, with the previously defined MCR regions designated by a red line. We use the red line in *MCR*6 to accentuate the 12 residues with increased conservation in this region. Amino acids are colored so basic (HKR) residues are blue, acidic (DE) are red, and hydrophobic (AVLIFM) are green. Numbering is according to human ChREBP sequence. doi:10.1371/journal.pone.0034803.g003

[10] or predicted to contain a transmembrane region that inserts into the OMM. Hence we propose the N-terminus sequence of MondoA induces mitochondrial transport via 14-3-3, where it interacts with receptors located on the OMM. This novel function may further contribute to glucose sensing and regulation in skeletal muscle, where MondoA is preferentially expressed.

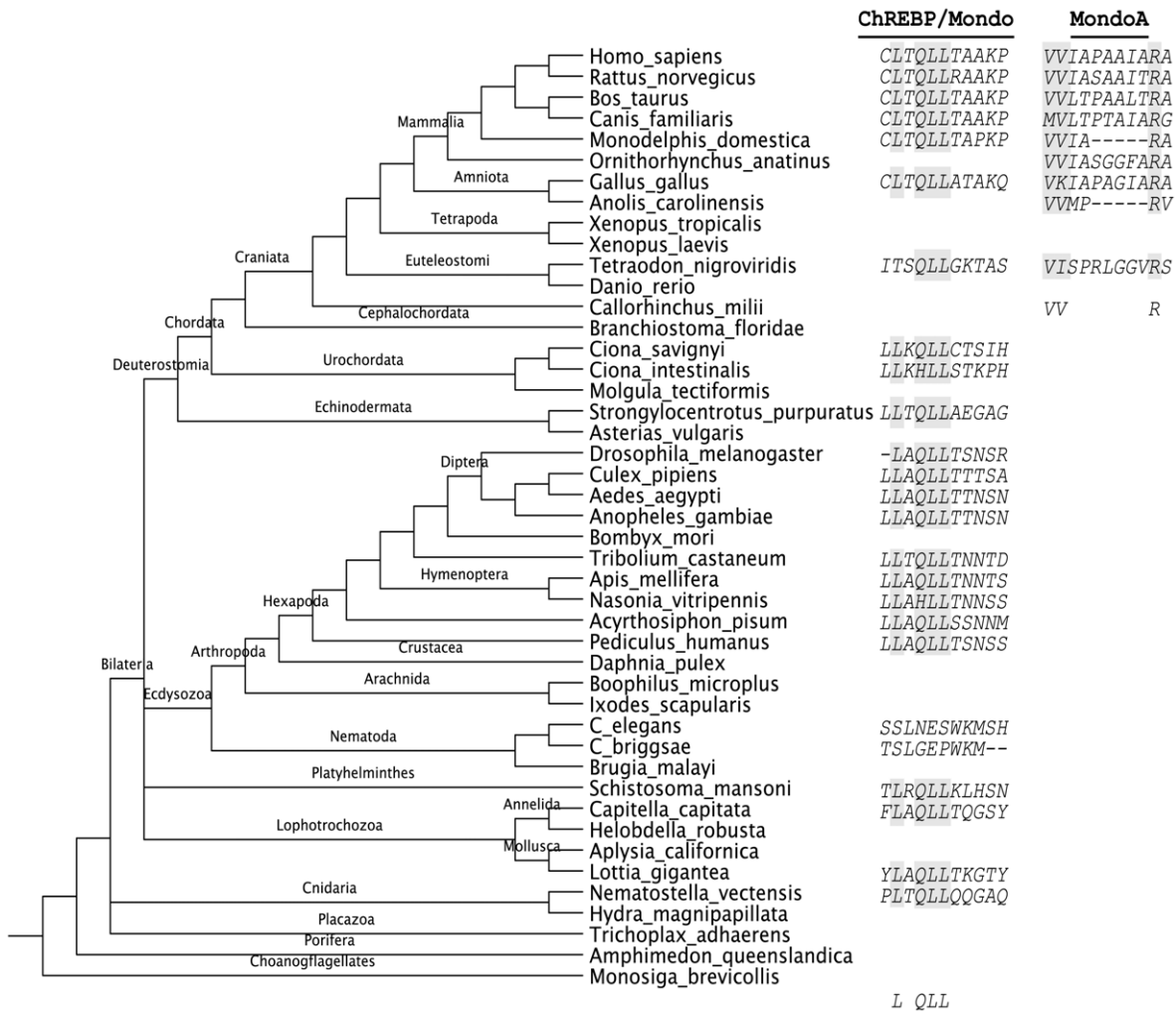
**The importance of *MCR* and *DCD* invariant positions**

By isolating columns with zero entropy and hence no variation, we identify 24 invariant sites within the Mondo sequence alignment, all of which are contained within the *MCR* and *DCD* regions (see Figures 3 and 5). We hypothesize that these sites are crucial for proper function of Mondo family proteins and find that many have been reported as essential for MondoA or ChREBP interactions or transactivation.

*MCR*III contains two groupings of invariant residues P104/W106/F109 and R121/L122/N123/N124/W127/R128 (human ChREBP numbering used throughout, except when directly referencing MondoA). Accordingly, the  $\alpha$ -helix spanning ChREBP

sites 116–135 is essential for 14-3-3 binding as is R128A [22], suggesting the invariant RLNN and WR residues are involved in 14-3-3 interactions. However an N123A mutation demonstrates it not necessary for 14-3-3 binding, but is essential for transactivation [19]. In comparison, mutation to MondoA sites **P144A/K145A/W146A** (human MondoA numbering, invariant sites in bold) did not affect 14-3-3 binding and no other phenotypic variations were reported [16]. However, we found that a serine or threonine immediately precedes P104 in all sequences, indicating this may be an important phosphorylation site for Mondo proteins.

Sites F145 and P148 are also invariant, yet have not been previously included in a specific *MCR* sequence. These residues (bold) are within a conserved [KR]<sub>x</sub>[KRN][NSTP][PLIV][VFI][CIV]<sub>x</sub>F[AVI][STV]P[LIV] motif that is located directly downstream of *MCR*III (underlined). With the exception of upstream insertions within tunicate *Molgula tectiformis* (KILRRYGY), and nematodes *C. elegans* (KKQP) and *Brugia malayi* (RPDKD), this conserved region is contiguous with the remainder of *MCR*III and thus we include these additional sites within *MCR*III (Figure 3). As



**Figure 4. Nuclear receptor box conservation.** A LxQLLT motif is largely conserved among animals. Since we could not obtain the full sequence of all sampled species (shown in the species tree), many display alignment gaps, which do not necessarily indicate they lack the putative *NRB*. However, MondoA in vertebrates exhibits a divergent sequence and lacks the *NRB*. doi:10.1371/journal.pone.0034803.g004

before, the prevalence of serine and threonine residues before P148 suggests a putative phosphorylation site in Mondo family proteins, with the exception of orthologous MML-1 proteins in nematodes, which have a valine instead.

*MCRIV* sites W170/Y181/W184/R185 are also invariant, along with P291 of *MCRV*. Analogous to ChREBP sites Y169/W170/K171, alanine mutations of MondoA sites Y211/W212/K213 resulted in nuclear accumulation in low and high glucose as well as three-fold induction of TXNIP reporter gene in L6 myoblasts [17]. Similar results were observed for L289A/Q290A/P291A mutation in ChREBP with two-fold ACC gene reporter expression in 832/13 cells [19]. Hence these sites are likely involved in repression of Mondo family proteins.

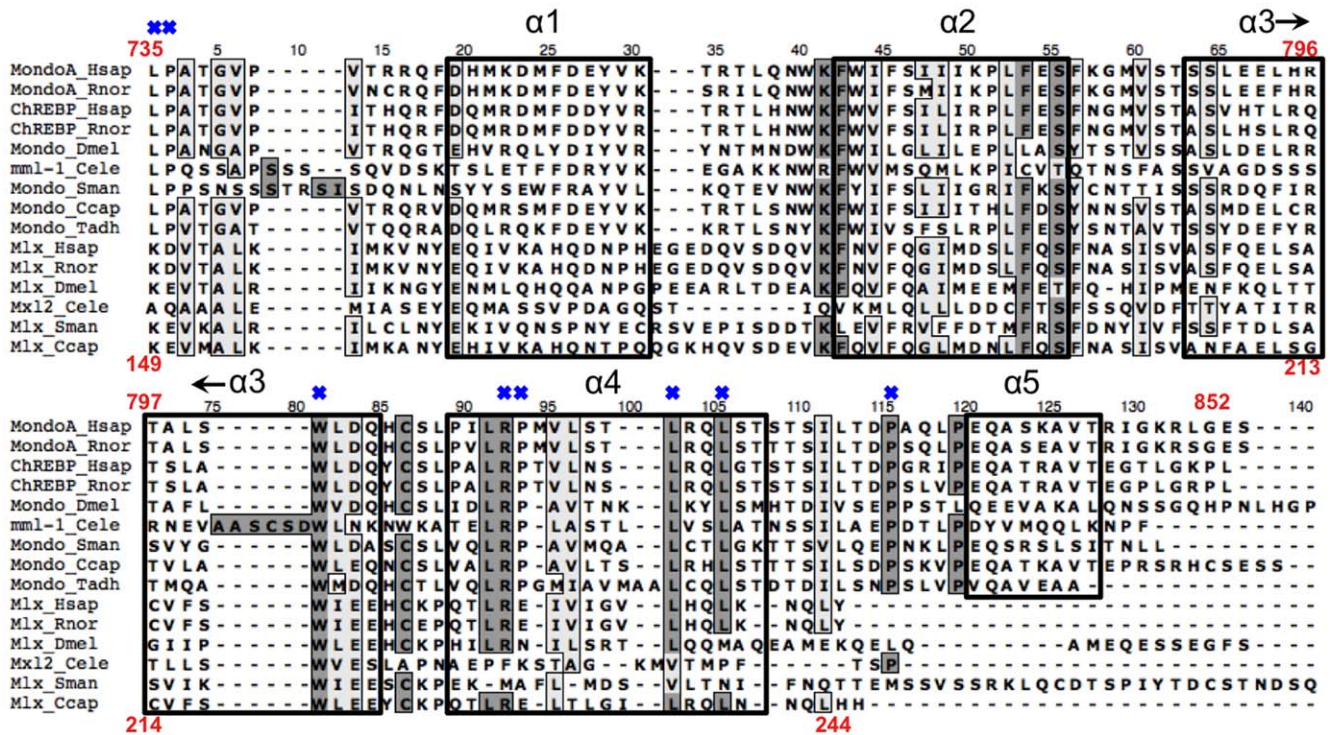
The remaining eight invariant positions are within the *DCD* region, represented by ChREBP sites L735, P736, W801, R812, P813, L819, L822, and P832. While their function is unknown, sites L735/P736 are located directly after the *bHLH $\zeta$*  and may be important for correctly orienting the *DCD* domain. The conservation of this region is addressed in later sections.

Surprisingly, *MCRI*, *MCRII*, and the *bHLH $\zeta$*  region lack invariant residues. However, high JS scores indicate these regions as well as others within *MCRIII* and *MCRIV* are still functionally

conserved among species. For example, divergence of the predicted protein sequence in beetle *Tribolium castaneum* (XP\_973749.2) prevents the identification of otherwise invariant residues **HSGxFMxS** within *MCR1*, where bold letters are conserved and x represents a variable site. *MCRII* in *Tribolium* is also not conserved, suggesting its N-terminal region is divergent or incorrectly identified. Regardless, most *MCRII* site variability arises from divergence in nematodes and other more distantly related species, which may indicate changes in selective pressure in Arthropoda and Deuterostoma lineages. In contrast, no single sequence is responsible for *bHLH $\zeta$*  variability, although it appears that nematode, ghost shark *Callorhynchus milii*, and sea squirt *Ciona intestinalis* often differ at otherwise conserved sites. Conservation of the *bHLH $\zeta$*  is addressed in detail in [12].

**N- and C-terminal regions of Mondo family proteins have conserved secondary structure**

Considering the extent of sequence conservation among species, we further hypothesize Mondo proteins exhibit similarity in higher order structures. As expected, we found secondary structure predictions of ChREBP, MondoA and non-vertebrate Mondo



**Figure 5. Mondo and Mlx WMC/DCD alignment.** DCD region of Mondo and Mlx sequences from *Homo sapiens* (Hsap), *Rattus norvegicus* (Rnor), *Drosophila melanogaster* (Dmel), *Caenorhabditis elegans* (Cele), *Capitella capitata* (Ccap), and *Trichoplax adhaerens* (Tadh). Red numbering on top corresponds to human ChREBP position, while the bottom represents the Mlx numbering. Sites with >75% identity or chemical similarity are shaded dark and light gray respectively, while the five (four) predicted alpha helices for MondoA and ChREBP (Mlx) are boxed. Mondo invariant positions are marked with a blue 'X'.

doi:10.1371/journal.pone.0034803.g005

proteins are comparable, and the majority of their protein sequences are random coil with several  $\alpha$ -helices and intermittent  $\beta$ -sheets (Figure 2d, Fig. S1). Predictably, the  $\alpha$ -helices and  $\beta$ -sheets overlap the *MCR*, *bHLHZ*, and *DCD* conserved regions described above, as well as *MCR6* and the *NRB* in non-vertebrate Mondo and ChREBP sequences. This implies the conserved residues are similarly orientated within the domains and Mondo family proteins are composed of the same structural elements.

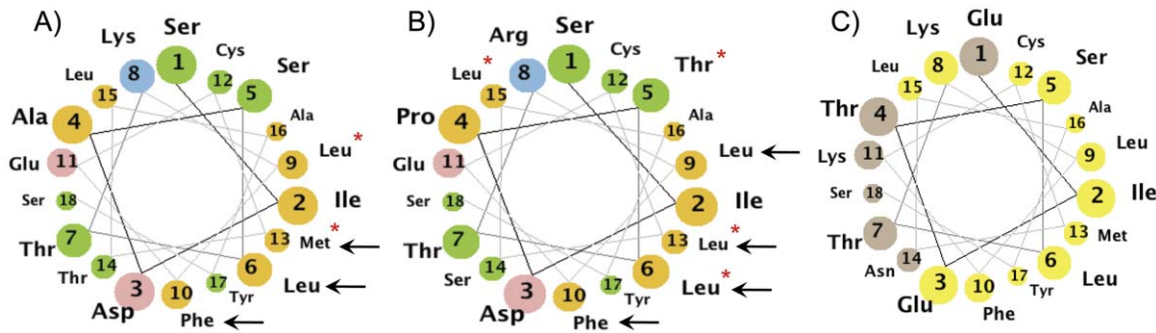
We also predict secondary structure plays a role in maintaining the function of these conserved Mondo domains. In support of this premise, experiments show the  $\alpha$ -helices comprising the *bHLHZ* and *DCD* domains are necessary for basic Mondo protein function, e.g. DNA binding, dimerization and subcellular localization [16–18]. Likewise, the three  $\alpha$ -helices within *MCRII*, *MCRIII*, and *MCRIV* correspond to a NES, 14-3-3 binding region, and NLS respectively and are critical for proper function [29]. In particular, *MCRII* residues have been found to be independently essential for transactivation in addition to CRM1 dependent nuclear export [19]. Projecting the residues of *MCRII* onto a helical wheel, we find the residues necessary for these functions are more highly conserved and located on the same side of the  $\alpha$ -helix (Figure 6). Hence the relative orientation of these residues possibly creates a surface for competitive interaction mediating a transition in functions.

### Evidence for a CRS in MCRIV

MondoA:Mlx and ChREBP:Mlx heterodimers actively shuttle between the nucleus and cytoplasm, indicating that increased nuclear accumulation in response to glucose is not simply the result of nuclear targeting (Table S1). In fact, all the *MCRs* affect

the subcellular localization of ChREBP and MondoA. Blocking the *MCRII* NES in either MondoA (M133A, F130A, MondoA $\Delta$ 125–137) or ChREBP (ChREBP $\Delta$ 86–95, ChREBP $\Delta$ 72–99, L86A/L93A, L89A, F90A) results in nuclear accumulation in either low or high glucose conditions [17,19,22,28,29]. Likewise, altering the *MCRIV* NLS in ChREBP (ChREBP $\Delta$ 158–173, ChREBP $\Delta$ 158–173, ChREBP $\Delta$ 168–190) results in cytoplasmic retention [25,28,29]. However, MondoA triple mutant Y211A/W212A/K213A, which overlaps the latter portion of the bipartite NLS in ChREBP, results in MondoA nuclear localization in low and high glucose in L6 myoblasts [17]. In addition, C-terminal sequences, optionally including *MCRV* and *MCR6*, result in nuclear accumulation for both MondoA [16] and ChREBP [21,29]. However, the inclusion of residues 224–273 in MondoA resulted in a cytoplasmic shift with most cells having equal nuclear and cytoplasmic amounts, while a MondoA mutant containing the full *MCRIV* region (MondoA:182–919) slightly reversed this effect with most cells being nuclear [16]. This suggests that MondoA *MCRIV* has opposing roles in nuclear localization.

It has been suggested that MondoA *MCRIV* contains a CRS [18] and truncation mutants indicate it is located within the latter half of *MCRIV*. The bipartite NLS in ChREBP *MCRIV* is only partially conserved in some MondoA sequences, due to a single arginine to serine mutation (MondoA:R213S) arising prior to the divergence of canines. Interestingly, the basic residues within the first portion of the NLS are conserved in MondoA, but variable in non-vertebrates, suggesting that the NLS may be weak, dispensable, or nonexistent in these proteins. As such, fusing *MCRIV* of MondoA to a heterologous NLS resulted in complete cytoplasmic localization [18]. This is independent of 14-3-3, which binds to



**Figure 6. *MCRII* helical wheel. A)** MondoA sites 121–138, **B)** ChREBP sites 81–98. Helical numbering is according to position within *MCRII* and represented by decreasing circle sizes. Black arrows point to sites indicated as essential for NES and red asterisks mark those necessary for glucose responsive transactivation. Color scheme: blue-basic, pink-acidic, orange-nonpolar, green-polar, uncharged. **C)** *Drosophila* sequence. Yellow circles have at least 75% chemical identity among all Mondo sequences. doi:10.1371/journal.pone.0034803.g006

*MCRIII* and was previously implicated in cytoplasmic localization. Together this data implies MondoA:224–273 contains a strong CRS. We found this region is similarly conserved among Mondo family proteins, with sequence signature VxxEY[KH]KWRx[-FY][FY][KR], where x represents a variable site and bold letters are invariable among Mondo sequences. Due to this conservation, we hypothesize that ChREBP and non-vertebrate Mondo proteins also contain a CRS within *MCRIV*.

Directly downstream of *MCRIV*, site S196 dephosphorylation results in the nuclear accumulation of ChREBP in low and high glucose [25]. Since MondoA and non-vertebrate Mondo proteins lack this phosphorylation site but have glucose-responsive subcellular localization, we anticipate the putative phosphorylation site 147-[TSP]-148 between *MCRIII* and *MCRIV* (ChREBP numbering) may be involved, as it is found in almost all Mondo family proteins and phosphorylated in high glucose for ChREBP triple mutant S196A/S626A/T666A [27].

#### *DCD/WMC* is conserved among Mlx and Mondo family proteins

For MondoA, and presumably ChREBP, to enter the nucleus, dimerization with Mlx must first occur. This is due to a cytoplasmic retention signal (CRS) located within the *DCD*, which is directly downstream of the *bHLH $\zeta$*  domain [16,17]. The *DCD* region provides an additional and independent interaction interface between Mondo family and Mlx proteins, which masks the CRS and allows for nuclear entry. While most of our understanding regarding this region is based on MondoA mutations, observations concerning the homologous and extended sequence WBSCR14-Mlx C-tail (*WMC*) region of ChREBP provide similar results [26,43]. Still, little is known about how the *DCD/WMC* region acts as a CRS, dimerizes, or differs between Mondo and Mlx proteins.

To determine which residues within the *DCD/WMC* potentially contribute to its structure or function, we compared Mondo and Mlx protein sequences using multiple entropy measures (see Methods). From the *DCD/WMC* alignment columns (Figure 5) containing more than three residues, sites K41, F42, W81, L91, and L102 are nearly invariant across all Mlx and Mondo sequences with entropy less than 0.1 ( $H < 0.1$ ), while columns 5, 6, 13, 21, 41, 42, 44, 55, 56, 60, 81, 82, 83, 86, 91, 96, and 102 (*DCD/WMC* alignment numbering) display conservation with functional entropy less than 0.1 ( $H_{FG} < 0.1$ ) (Figure S2). As expected, sites with  $H < 0.1$  also have  $H_{FG} < 0.1$ . This is consistent with experimental evidence, which show residues K41, F42, S54,

and F56 of MondoA and Mlx are important determinants of heterodimerization [16]. Compared to the Mondo invariant sites described previously, only W81 is invariant in both Mondo and Mlx, although L91 is conserved in all but the nematode sequences.

Based on the *DCD/WMC* conservation, our results disagree with the claim that *C. elegans* MML-1 lacks a *DCD* region [44]. We find that *C. elegans* MML-1 is conserved at 10 (58.8%) of the 17 functionally constrained sites as well as the eight invariant Mondo residues. Moreover, the *DCD/WMC* region of MML-1 is 46.7% similar and 21.3% identical to mosquito *Culex pipiens*, while nematode Mlx homolog Mxl-2 is 40% similar and 16.2% identical to the Mlx *DCD/WMC* sequence in beetle *Tribolium castaneum*. Hence, we assert that the *DCD/WMC* region is intact in *C. elegans* MML-1 and Mxl-2 proteins. Since these nematode sequences contain *MCR* and *DCD* domains that define Mondo and Mlx proteins, we further defend that MML-1 (myc- and mondo-like 1) is within the Mondo family and Mxl-2 is an ortholog of Mlx. This corroborates with the phylogenetic classification of their bHLH $\zeta$  sequences [12].

#### *DCD/WMC* structure forms an $\alpha$ -helix bundle

To determine the importance and potential interactions among conserved sites within the *DCD/WMC*, we predicted the higher order structures of this region. Secondary structure predictions of the *DCD/WMC* for MondoA, ChREBP, and non-vertebrate Mondo proteins identifies five  $\alpha$ -helices, while only four were found for Mlx sequences (Figure 5). Previously, just the *DCD* region was considered in structure prediction of ChREBP and a zipper like tertiary structure was assumed [45]. However, by including the entire *WMC* region, the powerful 3-D structure software Rosetta predicts the ChREBP *DCD/WMC* model assumes a cyclin-like confirmation with five grouped  $\alpha$ -helices, Figure S3a [46]. This predicted configuration forms a groove flanked by hydrophobic residues in alpha helices 1, 2, 3, and 4 designated by alignment sites **21**, 25, and 29 of  $\alpha 1$ , **44**, 47, 48, 49, 52 of  $\alpha 2$ , 65, 68, 73, and **82** of  $\alpha 3$ , and 88, **91**, 95, **96**, **102** and 105 of  $\alpha 4$ , where functionally conserved residues ( $H_{FG} < 0.1$ ) are in bold.

This interior region also displays increased conservation according to both entropy and ConSurf estimates (Fig. S3b). The program ConSurf estimates the evolutionary rate of each site by comparing homologous sequences and similar protein structures [47]. ConSurf predicts ChREBP residues V6, K41, F42, S55, W81, L88, and L102 (*DCD/WMC* alignment numbering) have high conservation scores and are likely functionally important.



Besides L88, these positions have low functional entropy for all Mondo and Mlx sequences, suggesting a common function.

The *DCD/WMC* of Mlx and Mondo family proteins show clear similarity, although we anticipate protein distinctions likely affect their tertiary conformation. First, our alignments show the *DCD/WMC* region of Mlx abuts the 21-residue zipper region, while the zipper and a linker region of Mondo sequences together extend for 35 residues before the *DCD/WMC* begins. In addition, Mondo invariant sites L735/P736 are alternatively conserved for charged residues (lysine and either aspartate or glutamate) in Mlx, which may affect the *DCD/WMC* orientation. Moreover, helix 5 shows considerable variability among the Mondo sequences, and may not be directly involved in protein-protein interactions, as it is completely lost in most Mlx sequences. These differences may restrict interaction between *DCD/WMC* regions and factor in the prevention of MondoA and Mlx homodimerization [16].

### Mondo proteins have disparate Proline and Glutamine Rich Regions

In contrast to the structured N- and C-terminus, the central region of Mondo proteins is mainly composed of random coil. Both MondoA and ChREBP proteins contain a proline rich region (*PRR*) within their proximal region that is retained among most vertebrates. However, we were unable to find any identifiable stretch of homology between MondoA and ChREBP *PRRs* and the *PRR* is not found within any non-vertebrate species. Instead, most non-vertebrates contain a glutamine rich region (*GRR*) (Table 1). The prevalence and length of these low complexity regions suggests the central region contains an imprecise function, such as indiscriminate scaffolding regions as seen in other *PRR* and *GRR* containing proteins [48,49] and may contribute to Mondo transactivation of target genes.

### *MCR6* involvement in Glucose Dependent Activation

Recent evidence shows that MondoA and ChREBP activation is dependent upon glucose phosphorylation by hexokinase, which metabolizes glucose to form glucose-6-phosphate (G6P) [17,20,31]. Induction of 2-deoxyglucose (2-DG), which is a glucose analog that can be phosphorylated but not further metabolized, promotes MondoA nuclear accumulation, increases promoter occupancy and recruits histone H3 acetyltransferase thereby activating gene transcription [17]. Similarly, 2-DG dose dependently increased the transactivation ability of Gal4-ChREBP, while hexokinase inhibitor d-mannoheptulose and glycolytic enzymes PFK1 and PFK2 decreased ChREBP activity [31]. This suggests that MondoA and ChREBP activation is directly invoked by glucose phosphorylation. Moreover, the N-

terminus of Drosophila ortholog dMio activates a luciferase reporter comparable to Gal4-ChREBP levels in a glucose responsive manner [21]. Domain swapping of the LID region of ChREBP with that of either MondoA or dMio resulted in a strong glucose response, suggesting that the LID and GRACE regions are interchangeable among homologs and Mondo proteins, in general, are glucose responsive. As such, we hypothesize regulation of Mondo family proteins is expected to occur through a G6P mediated signaling cascade, direct binding of G6P to an allosteric mechanism, or both.

To investigate the presence of an allosteric G6P binding region within Mondo proteins, we first examined the binding region of known G6P interactors (Figure 7), i.e. glucokinase (GK), hexokinase (HKI-III), G6P phosphatase (G6Pase), phosphoglucose mutase (PGM), glucose phosphate isomerase (GPI), G6P dehydrogenase (G6PDH), and glutamine:fructose-6-phosphate amidotransferase (human: Gfat1, *E.coli*: Glms). Since glucose is essential among prokaryotes and eukaryotes, the enzymes and binding regions involved in glucose metabolism are highly conserved. Interestingly, we find the G6P binding region is similar among GK, GPI, and Gfat1, with serine and threonine residues forming hydrogen bonds with the 6-phosphate molecule (Figure 7b). Moreover, the phosphate recognizing residues of GPI and Gfat1 are in close proximity in the linear sequence, forming an Sx[ST]xxT motif, where x indicates a residue not involved in 6-phosphate recognition. This is distinct from G6PDH and PGM, which have HYxxK and SKN motifs, respectively.

We propose G6P binds to Mondo proteins within the highly conserved *MCR6* region, which contains an Sx[ST]xx[ST] motif similar to that found in GPI and Gfat1. Our alignments show MondoA consists of residues 281-**SDTLFS**-287, while ChREBP contains a 253-**SDTLFT**-258 motif. This putative G6P recognition motif is also preserved in non-vertebrate Mondo sequences, where serine and threonine are likely to interchangeably form hydrogen bonds with the 6-phosphate molecule. We predict this motif is associated with recognizing the phosphate group of G6P, which is consistent with the correlation between MondoA and ChREBP activation and glucose phosphorylation.

While the strict conservation of Sx[ST]xx[ST] within *MCR6* among animals is evidence for its functional importance among Mondo proteins, this short motif has low specificity and is predicted to occur in several sequence locations. By plotting the location of each Sx[ST]xx[ST] motif for each Mondo sequence (Fig. S4), we find that this motif is not distinctly conserved elsewhere in the alignment, suggesting these residues in *MCR6* are functionally constrained.

In addition, *MCR6* is located within the GRACE region, which is sufficient for ChREBP transactivation [21]. Interestingly, mutations to the only other conserved domain within the ChREBP GRACE region, *MCRV*, show an increase in transactivation [19]. ChREBP:299–645, which is downstream of the GRACE region and encompasses the Proline Rich Region, is also sufficient for transactivation. Meanwhile, ChREBP:197–479, which overlaps the GRACE and PRR, shows a significant increase in fold activation of a luciferase reporter, suggesting a synergy between these domains [21]. This is compatible with the TAD domain found in MondoA 322–445 [11], which overlaps its PRR.

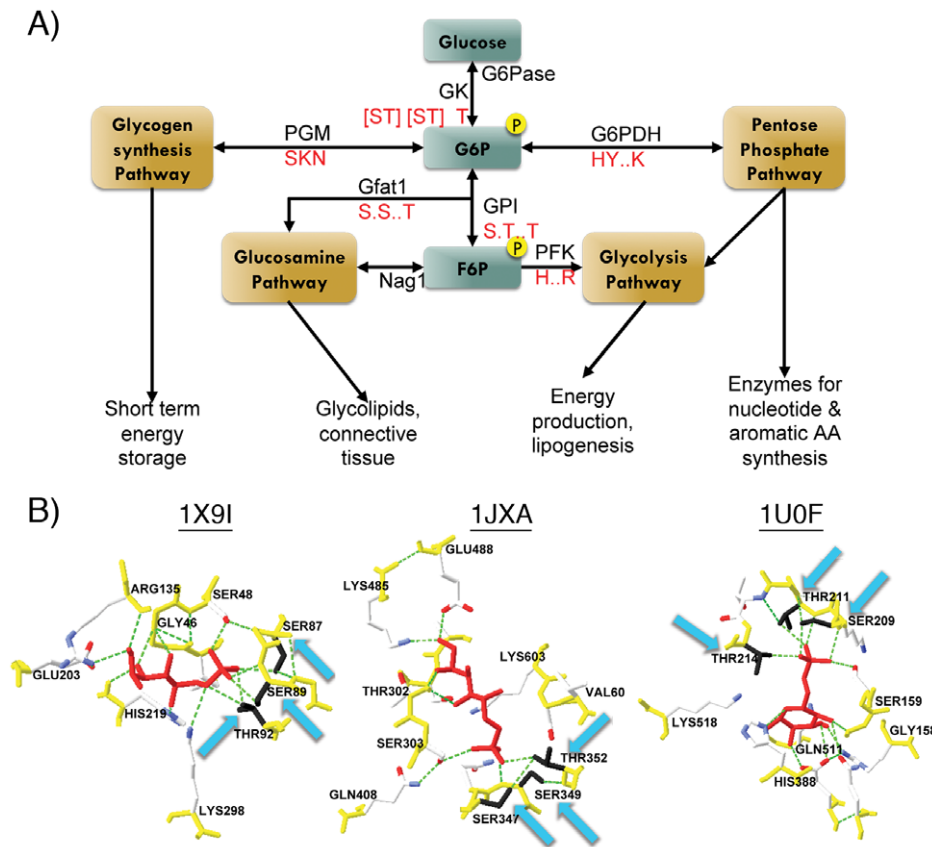
We hypothesize *MCR6* of the GRACE region harbors a TAD that contributes to the recruitment of coactivators such as CBP/p300, which are known to interact with ChREBP [39]. To test this, we searched the entire sequence of each Mondo protein for the nine amino acid transactivation domain (9aa TAD) signature that is recognized by coactivators TAF9, MED15, CBP, and p300 [50]. Although individual sequences displayed multiple hits using

**Table 1.** Proline and Glutamine Rich Region.

	Proline	Glutamine	Neither	Missing
Vertebrates	16	0	1	3
Non-vertebrates	0	10	6	10
Length	<i>mean</i>	355.75	543	462.14
	<i>sd</i>	48.9	173.4	104.3

Existence of Proline Rich and Glutamine Rich Regions in the proximal domain of Mondo sequences as predicted by ScanProsite. Neither indicates the central region is intact, yet ScanProsite did not identify a *PRR* or *GRR* region. Missing denotes sequences where the central region was only partially or not recovered. Length is calculated by the number of amino acids between the *MCRV* and *bHLHZ* of full length sequences.

doi:10.1371/journal.pone.0034803.t001



**Figure 7. G6P binding region.** **A)** Glucose metabolism pathways. Glucose is phosphorylated in the liver by GK to form G6P. G6P can then enter the pentose phosphate pathway by interacting with G6PDH, the glycogen synthesis pathway by binding to PGM, or form F6P by GPI isomerization. Residues involved in these interactions are shown in red, with dots indicating nonbinding sites within a linear sequence and spaces denoting larger linear distances. **B)** G6P interacting protein structures. The structures for GPI in Rat (1U0F), phospho-glucose/phospho-mannose protein in archaea (1X9I), and GlmS in *E. coli* (1JXA) have been crystallized. The backbone of residues within 5 Angstrom of G6P (red) are yellow and hydrogen bonds are shown by a green dashed line. We indicate the residues conforming to the G6P recognition motif with blue arrows and color the side chains black. doi:10.1371/journal.pone.0034803.g007

the 9aa TAD regular expression (see Methods), the only common occurrence was in *MCR6* where we observed two overlapping 9aa TAD motifs. ChREBP was restricted to motif 1 (ChREBP:250-SDISDTLFT-258), while MondoA and Mondo sequences also matched motif 2 (MondoA:283-DTLFSTLSS-291); conserved sites within the overlapping regions are in bold and underlined. Of the 34 sequences in our dataset containing *MCR6*, nineteen contained both motif 1 and 2, five only had motif 2, eight only had motif 1, trematode *Schistosoma mansoni* matched an intermediate sequence, and sea anemone *Nematostella vectensis* matched neither. Although there was no clear preference for either motif, we propose the TAD is located in *MCR6*, and consider that the presence of multiple TAD motifs within this region may provide variable specificity for binding cofactors.

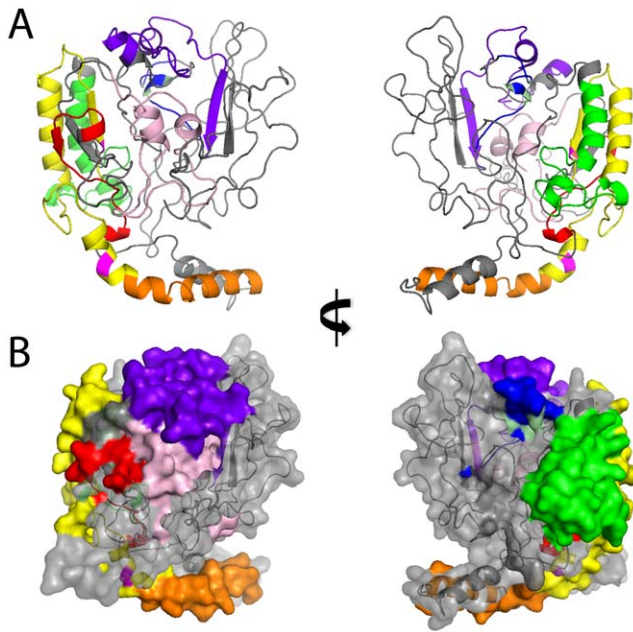
#### LID and GRACE regions have intramolecular contacts in N-terminal Predicted Structure

The LID region, containing *MCR1-IV*, is necessary to repress transactivation in low glucose conditions and promote transactivation in high glucose conditions [21]. However, how the *MCR1-IV* domains individually and cooperatively operate is not clear. To better understand how *MCR1-IV* switches between repressive and activating functions, we predicted the protein structure for MondoA and ChREBP N-terminal sequences.

From the sequence and secondary structure predictions of 3D-Jury, the N-terminus of MondoA was most similar to Estrone Sulfatase (ES, PDB ID: 1p49) (Figure 8) and also showed a likeness to similar sulfatase structures (PDB ID: 1auk, 1fsu). As expected, the N-terminus of ChREBP also shows structural similarity to 1p49 and resembles the MondoA conformation (Figure 9a).

The putative MondoA and ChREBP protein structures are compatible with the accessibility of their known domains. The protruding  $\alpha$ -helices in MondoA and ChREBP correspond to *MCR1* and its CRM1 dependent NES in the predicted structure (Figure 8, orange). This is concordant with the CRM1-SNUPN structure, where the NES of SNUPN forms an extended amphipathic  $\alpha$ -helix that protrudes away from the rest of the molecule and binds a hydrophobic groove in CRM1 [51]. The exposure of *MCR1* (Figure 8, yellow) also allows for its  $\alpha$ -helix to interact with known binding partner 14-3-3. The orientation of *MCR1* and *MCR4* (Figure 8, green)  $\alpha$ -helices closely position S140 and S196 in ChREBP, so they are both situated near *MCR4* (Figure 8, purple; Fig. S5). This conformation agrees with evidence implicating S196 and S140 phosphorylation affects nuclear accumulation and 14-3-3 interaction [29] as well as the interaction model hypothesized by Davies et al. [23].

The placement of *MCR4* near the ends of *MCR1* (Figure 8, red), *MCR1*, and *MCR4* allows for interaction among these domains

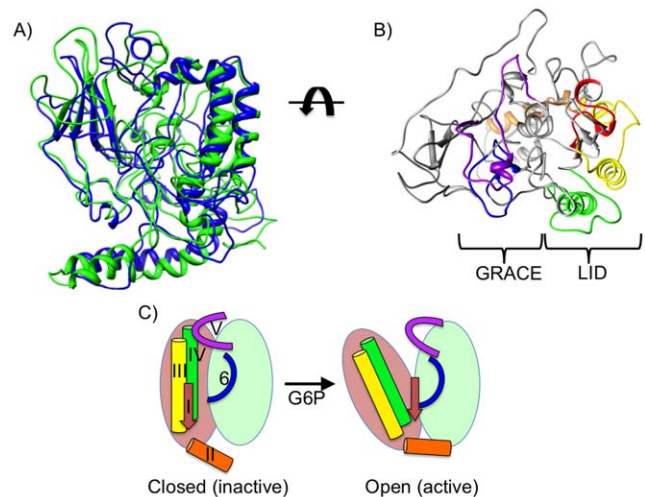


**Figure 8. MondoA N-terminus structure.** Predicted structure for MondoA:1–490. **A)** Ribbon structure. **B)** Filled structure. *MCRVI* is red, *MCRVII* is orange, *MCRVIII* is yellow, *MCRIV* is green, *MCRVI* is blue, and *MCRV* is purple. In addition, the first 42 residues potentially targeting MondoA to the OMM are light pink, and putative phosphorylation sites S143 and T187 are magenta, and the serine and threonine residues of *MCRVI* are pale green. Left and right images are rotated 180 degrees. doi:10.1371/journal.pone.0034803.g008

and corresponds to the proposed linkage between *LID* and *GRACE* regions mediated by multiple contacts with *MCRV* [21,23]. *MCRVI* (Figure 8, blue) is adjacent to *MCRIV* and may also have a binding interface. Considering the potential role of *MCRVI* in G6P binding and transactivation, this interaction may affect the glucose response, as seen for proteins with *MCRIV* deletions that lack glucose dependent regulation [22]. Viewing the predicted structure from the top (Figure 9), it is easy to see how the *LID* can contact and possibly release from the *GRACE* region to conditionally block the binding of coactivators and regulate the transactivation of target genes.

## Discussion

Conservation in sequence, domains, and glucose response for MondoA and ChREBP proteins suggest they are mechanistically similar. Based on the elevated JS conservation scores and persistence of secondary structures across sequences, the distal regions of Mondo proteins are likely to exhibit similar structure and function. The presence of *MCRI-V*, *MCRVI*, *bHLHZ*, and *WMC/DCD* regions in diverse organisms dates the origin of these regions to as early as the divergence of cnidarians around 600 million years ago [52]. Moreover, conservation of Mondo proteins and domains throughout animal evolution suggests the glucose responsive transactivation observed in MondoA and ChREBP has been preserved as well. Similar to the explanation for the emergence of energy homeostasis in bilaterians [21], cnidarians also possess muscular, nerve, and gastroderm or “stomach” cells, which contribute to the formation of an internal environment and rise of signaling factors important for homeostatic regulation, e.g. Mondo proteins and nuclear receptors.



**Figure 9. LID and GRACE interaction.** **A)** MondoA (green) and ChREBP (blue) overlay of N-terminal predicted structure. **B)** Topical view of MondoA:1–490 ribbon structure. *MCRVI* and *MCRVI* are part of the *GRACE* region, while the *LID* includes *MCRIV*. *MCR* domains are colored as in Figure 8. **C)** Predicted allosteric affect of G6P binding to *MCRVI*. *MCRVII* and *MCRVIII* release from *MCRVI*, while *MCRVI* and *MCRVII* lock the “open” conformation to separate the *LID* and *GRACE* regions and support transactivation. doi:10.1371/journal.pone.0034803.g009

## *MCRVI* involvement in G6P recognition and transactivation

Initial models of Mondo and Mlx function were solely dependent upon the subcellular localization of these proteins. Since ChREBP, MondoA, and Mlx are largely cytoplasmic, it was predicted that nuclear transport would be sufficient for the transactivation of their gene targets. However, multiple experiments have shown that trapping ChREBP:Mlx or MondoA:Mlx in the nucleus, mutating the NES, or altering the phosphorylation of particular residues does not result in constitutive activation of reporter constructs [17,19,29].

Recently, MondoA nuclear accumulation has been attributed to both increased nuclear import, increased promoter occupancy, and decreased nuclear export in response to glucose derivative 2-DG [17]. ChREBP transactivational ability is also correlated to G6P abundance [31], suggesting that MondoA and ChREBP glucose response is directly mediated by G6P. Similarities in *MCRVI* sequence with known G6P binding sites, and particularly the 6-phosphate molecule, strongly suggest that *MCRVI* is an allosteric G6P binding region.

We defend that the putative function of *MCRVI* in G6P allosteric activation and recruitment of coactivators is not mutually exclusive. Since MondoA and ChREBP have increased transactivation in response to G6P, its binding may trigger a conformational change that further exposes *MCRVI* and facilitates cofactor interaction. The structure of GPI and Gfat1 proteins suggest that G6P binds within a largely hydrophilic pocket, while the 9aa TAD structure is variable and often disordered prior to forming an  $\alpha$ -helix conformation upon cofactor binding [50]. The predicted structure of *MCRVI* in MondoA and ChREBP displays an exposed pocket suitable for G6P binding as well as a flexible, coil region capable of making protein interactions (Figure S5).

## Model of G6P mediated Mondo Glucose Response

Based on our structure predictions and published sequence annotations, we propose the following model for Mondo glucose

responsive transactivation. First, Mlx and Mondo family proteins readily form heterodimers within the cytoplasm, allowing Mlx:Mondo complexes to actively shuttle between the cytosol and nucleus. Second, *MCRV* interacts with the *LID* region, possibly through specific contacts with *MCRI*, *MCRIII*, and/or *MCRIV*, to block the transactivation region. Third, increased glucose and consequently G6P concentrations trigger signaling mechanisms that block the putative CRS in *MCRIV*. Fourth, G6P binding to *MCR6* causes an allosteric conformational change that “unlocks” *LID* and *MCRV* contacts, “pivots” *MCRII* so that it is buried, and “pins” *MCRI* in between the *LID* and *GRACE* so that Mondo remains in an open conformation. Finally, once in this open conformation, G6P may be released and cofactors such as CBP/p300 may bind to *MCR6* thereby activating Mondo proteins. In addition, non-vertebrate Mondo and ChREBP proteins interact with nuclear receptors, such as HNF4 $\alpha$ , through the *NRB*, which activate these cofactors and increase transactivational potential. This model is in accordance with previous models based on protein manipulations as explained below.

First, MondoA and ChREBP monomers are confined to the cytosol and MondoA requires Mlx dimerization prior to nuclear localization [16,17]. MondoA and ChREBP dimers have also been observed to actively shuttle between the nucleus and cytosol in numerous cell types (Table S1) and can be sequestered in the nucleus by NES inhibitor leptomycin B (LMB), whereas blocking MondoA and Mlx dimerization results in purely cytoplasmic monomers. Phosphorylation sites have been observed by mass spectrometry throughout ChREBP, except the *DCD/WMC* region, indicating Mlx dimerization is independent of phosphoregulation [27]. Conservation of *DCD/WMC* residues and similarity in both secondary and tertiary structure predictions implies monomer cytoplasmic retention and Mlx dimerization is consistent among Mondo family proteins. Thus it is likely that ChREBP and non-vertebrate Mondo proteins actively bind to available Mlx and are capable of shuttling to the nucleus as has been shown for MondoA.

Second, the *LID* region is responsible for regulating the otherwise constitutively active *GRACE* region in ChREBP. Inverting the *LID* and *GRACE* regions results in constitutive activation, showing the structural organization of these regions is important for ChREBP regulation [21]. Combinatorial deletions in ChREBP show *MCRII* has minimal repressive effects, while *MCRI*, *MCRIII* and *MCRIV* decrease transactivation in the presence of *MCRV* [22]. *MCRV* does not repress transactivation in the absence of *MCRIV*, yet mutations to *MCRV* increase transactivation when the *LID* is present [19]. Individual deletions of *MCRIV* were unable to alleviate low glucose repression [23], suggesting *MCRV* represses transcription conditionally upon multiple contacts within the *LID* region. From our structural prediction, it is likely the *MCRV* contacts *MCRIII* and *MCRIV* near residues S140 and S196, respectively (Fig. S5). These sites are known to affect the cytoplasmic localization of ChREBP as well as 14-3-3 binding, which is required for transactivation [19,22]. Although MondoA and non-vertebrate Mondo proteins do not have these phosphorylation sites, our results based on sequence, domain, and structure similarity still support the notion that *LID* repression acts through multiple intramolecular contacts and is common among all Mondo proteins.

Third, it has been suggested that MondoA *MCRIV* contains a CRS [18] and truncation mutants indicate it is located within the latter half of *MCRIV*. We find this region is highly conserved and likely to have the same interaction properties among Mondo proteins. Since increasing G6P abundance accelerates the rate of nuclear import for MondoA [17] and PP2A mediated dephosphorylation of S196 in ChREBP just downstream of *MCRIV* also

results in increased nuclear abundance [25], we predict Mondo nuclear accumulation is, at least in part, governed by a common mechanism, specifically G6P mediated relief of a CRS in *MCRIV*.

Fourth, it has been proposed that G6P allosterically affects the transactivation of MondoA and ChREBP [17,23,31]. *MCR6* provides an appropriate interface for G6P binding and also contacts the *LID* domain, particularly with *MCRIV* in our predicted structure. *MCRIV* is involved in general repression, where all mutants lacking this region show increased expression of reporters in a luciferase assay [21]. Additional deletion mutants show that *MCRI*, *MCRII*, and *MCRIII* are all necessary to overcome *MCRIV* repression and form an active complex. Thus G6P binding may break hydrogen bonds of *MCRIV* with these domains, thereby unlocking the repression of *GRACE* by *LID* and allowing these regions to separate.

Since glucose activated MondoA and ChREBP results in increased nuclear accumulation, we also expect the NES to be overpowered in high glucose medium. 14-3-3 binding has previously been attributed to blocking the NES, although *MCRII* is also necessary for recruiting a histone H3 acetyltransferase (HAT) cofactor. Since the *LID* region is not independently sufficient for MondoA or ChREBP transactivation [21], *MCRII* recruitment of a HAT cofactor must be a secondary effect. Based on the predicted N-terminus structure, it is plausible that *MCRII* pivots to make necessary contacts outside of the *LID* domain to help fix the separation between *LID* and *GRACE*.

*MCRI* is also required for glucose transactivation, but is not sufficient for full transactivation [22]. Hence *MCRI* may also form intrastructural contacts necessary for alleviating *LID* repression or interacting with activating cofactors. The position of *MCRI* near the interior of the predicted protein suggests it may act as a pin to wedge the *LID* and *GRACE* regions apart. Phosphorylation of S56 adjacent to *MCRI* increases ChREBP transactivational potential [27], possibly by facilitating this conformational change (Fig. S5).

*MCRIII* contains two essential regions. 14-3-3 and its binding region in *MCRIII* are required for ChREBP transactivation as is ChREBP:100–115 that is not necessary for 14-3-3 interaction. 14-3-3 has been shown to bind ChREBP constitutively [22], promote cytoplasmic retention, nuclear export, and transactivation. While the necessity of S140 phosphorylation for 14-3-3:ChREBP interaction is under contention [22,29], it may affect the binding orientation as non-phosphorylated motifs can bind 14-3-3 in the opposite direction [53]. While S140 and S196 have been analyzed in ChREBP, we propose phosphorylation of the highly conserved T147/P148 site has a broader impact on Mondo family protein interactions and possibly affects 14-3-3 binding.

Moreover, the conserved *MCRIII* sequence corresponding to ChREBP:100–115 may affect Mondo phosphorylation status. According to the functional site prediction server ELM [54], this region matches a MAPK kinase-docking motif. Kinase docking domains are typically located 50–100 residues upstream of the phosphorylation site and characterized by a cluster of positively charged residues preceding a  $\Phi \times \Phi$  hydrophobic sequence [55–57]. Conserved sequence 105-**KWKxFKG**[LIV][KR]**L**-114 conforms to this motif, where positively charged residues are underlined and hydrophobic residues are in bold. Interestingly, W106 and F109 are invariant, and may contribute to interaction interface specificity. Moreover, 103-[ST]P-104 (human ChREBP numbering) residues directly precedes this motif in all Mondo sequences, but has not been identified as a phosphorylation site. Recent evidence also suggests that ChREBP activity in high glucose is dependent upon *O*-linked glycosylation, which targets sequences similar to phosphorylation motifs [58]. We anticipate the conditional status (e.g. phosphorylation, glycosylation, orientation, or intramolecular

contacts) of these sites and domains are important for the activation of Mondo in response to changes in glucose levels.

Finally, MondoA and ChREBP recruit cofactors to promote transcriptional activation. Since mutants lacking the N-terminus have exceptionally high transactivational ability, G6P may only be necessary for relieving *LID* repression from *GRACE*. Hence G6P may be released from *MCR6* in the active/open conformation, thereby permitting *MCR6* access to cofactors. MondoA was shown to recruit a histone H3 acetyltransferase [17], while ChREBP is known to interact with CBP/p300 [39], which has histone acetyltransferase (HAT) function. *MCR6* matches the 9aa TAD motif depicting the CBP/p300 interaction region. Since *MCR6* is within the *GRACE* region, which is sufficient for transactivation [21], and mutating *MCRV* increases the transactivation potential [19], we deduce that *MCR6* acts as a TAD for Mondo proteins.

ChREBP and non-vertebrate Mondo transactivation may additionally rely on the interaction with nuclear receptors. Interestingly, nuclear receptors are specific to metazoans, and not found in sponges although present in cnidarians [59]. This agrees with our identification of Mondo proteins and the *NRB* motif.

Excluding MondoA, an LxQLLT sequence matching the *NRB* motif was conserved within the central region among non-vertebrate Mondo and ChREBP proteins. Tellingly, ChREBP, HNF4 $\alpha$ , and CBP/p300 form a complex necessary for full activation of lipogenic enzyme L-PK. The HNF4 $\alpha$  and ChREBP binding domains are directly adjacent within the promoter of this gene, indicating they are also juxtaposed within the complex. Since most nuclear receptors depend upon interaction with a *NRB* for activation, ChREBP may be fulfilling this role. This interaction may also help explain the relationship of activation between ChREBP and other nuclear receptors such as FXR and COUPTF-II [60].

In conclusion, MondoA and ChREBP are important glucose responsive genes involved in energy homeostasis. While ChREBP has evolved to have unique phosphoacceptor sites, the conservation of *MCRI-V*, *MCR6*, *bHLHZ*, and *DCD/WMC* domains indicates all Mondo family proteins are regulated by common mechanisms. Although their formal structure is not known, we predict their regulation is largely governed by intramolecular contacts. We further postulate that binding of G6P causes an allosteric conformational change, which forms an open, active complex where the *LID* repression is released from *GRACE* and permits interaction with coactivators such as CBP/p300.

## Methods

Full-length Mondo family protein sequences were obtained by surveying multiple genome databases as described in [12]. ClustalW, Dialign, and MAFFT were used to align the sequences and merged according to consensus regions and manual adjustment to construct a single, optimal alignment. Mondo Conserved Regions were specified as in [18] and depicted by weblogs [61].

## Sequence Conservation

Both the Jenson-Shannon Divergence (JS) score and entropy values were used to determine sequence conservation. For a multiple sequence alignment, the JS heuristic employs window-based extension that considers the conservation of sequentially neighboring sites and quantifies each score based on a weighted distribution of amino acids [34]. Hence the mutual information based JS score rates the conservation of each site by incorporating the autocorrelation of adjacent sites, where highly conserved sites have JS scores close to one and variable positions close to zero.

Entropy values were computed by the FastaEntropy program written by Andrew Fernandez. Entropy is a statistical measure of the amount of information or variation and, when applied to sequence alignments, can depict the conservation of sites, with lower entropy values signifying increased conservation [62]. Traditionally protein entropy is calculated by the Shannon Entropy equation based on the proportion of the 20 possible amino acids at each site. However, this method does not account for shared physicochemical properties among amino acids. To account for this, we also used a functional group entropy measure developed by [63] that is based on eight distinct categories of amino acids grouped according to physicochemical similarities. This method accentuates sites that are functionally constrained yet variable, e.g. conservation of I, V, L, M hydrophobic residues.

Site conservation is also highly correlated with structural and functional importance. To estimate and project the contribution of conserved sites on protein structures, we used the Consurf program available at <http://consurf.tau.ac.il/> [64]. Consurf predicts functionally important regions in a given protein structure by estimating the phylogenetic relationship of homologs with similar known tertiary structure and ranking the evolutionary rate at each site [47]. Within this scheme, nine indicates site conservation and zero site variability.

## Identification of Functional Domains and Motifs

The presence of functional domains or motifs was determined by individually analyzing each sequence using multiple online tools. The presence of proline rich and glutamine rich regions was predicted by the ExPasy program ScanProsite [65]. Additional motifs, such as the MAPK kinase docking domain, were predicted using regular expression patterns by the Eukaryotic Linear Motif resource (ELM) [54], while the 9aa TAD server was used to specifically evaluate putative CBP/p300 binding regions [50]. The MAPK docking motif in ELM is characterized by the regular expression  $\{[KR]\{0,2\}[KR].\{0,2\}[KR].\{2,4\}[ILVM].[ILVF]\}$ , while the 9aa TAD regular expression is  $\{[GSTDENQWYM]\{KRHC GP\}[FLIVMW]\{KRHC GP\}\{CGP\}\{KRHC GP\}[FLIVMW][FLIVAMW]\{KRHC GP\}\}$ ; residues within brackets '['] are permitted and residues within braces '{ }' are prohibited.

## Characterizing the G6P recognition pocket

The structure of several G6P binding proteins has been crystallized, with specific attention to the G6P binding region, and deposited in the Protein Data Bank (PDB). During glucose metabolism in mammals, glucokinase (GK) or hexokinase (HKI-III) converts glucose to G6P [66–68], which can be reversed by G6P phosphatase (G6Pase) in the liver. G6P can be further metabolized by phosphoglucose mutase (PGM) to promote glycogen storage [69,70], glucose phosphate isomerase (GPI) to produce fructose-6-phosphate (F6P) and continue in the glycolytic pathway [71], or G6P dehydrogenase (G6PDH) to enter the pentose shunt of glycolysis [72,73]. Another enzyme, glutamine:fructose-6-phosphate amidotransferase (human: Gfat1, *E.coli*: Glms), can interact with G6P and F6P to promote the production of glycolipids through the glucosamine pathway [74–76].

We compared the G6P interacting residues described in the literature for each of these proteins to identify common features for metabolite recognition.

## Structural prediction of the DCD and N-terminal region of Mondo

Correctly predicting protein structures from amino acid sequences has been a goal within computational biology for the

last several decades. The reliability of structure predictions often depends on the availability of homologous structure templates that allow for protein threading or homology modeling methods. These methods use a database of known structures to select a template with local or global similarities in secondary structure that can be used to fit the query model.

Secondary structure predictions for human, mouse, *C. elegans* and *Drosophila* Mondo sequences were formed by NPS@, which builds a consensus based on the individual secondary structure predictions of DPM, DSC, GOR1, GOR3, HNNC, MLRC, PHD, Predator, and SOPM programs [77]. Sequences exhibited similar secondary structure predictions with compatible alignments of alpha helices and beta sheets. We depict the secondary structure by the representative human ChREBP graphic (Figure 2) produced using Polyview [78].

While using structure prediction programs is straightforward, each method can form diverse structures and evaluating their accuracy is difficult. The metaserver 3D-jury addresses this concern by aggregating and comparing multiple structure predictions from several servers and ranking them based on structural similarity to create a more accurate consensus prediction [79]. Rosetta has also been accepted as a leading protein prediction software with particular application to *ab initio* design [80]. A structure prediction for ChREBP *DCD/WMC* was previously determined by The Human Proteome Folding Project using Rosetta and deposited at the yeast resource center [81,82].

For determining the N-terminal structure, we used 3D-Jury on MondoA sequence 1–490 and ChREBP sequence 1–360. The 3D-Jury metaserver compares and ranks structural predictions from sequence only (EsyPred3, FFAS03, GRDB, Pfam-basic, Pfram-metabasic) and threading methods (3D-PSSM, FUGUE, INUB, mGenThreader, SAM-T02, samt06), whereby structure predictions are evaluated by the fit of each model and ranked according to their similarity to other models [79]. MondoA most closely matched the PDB structure (1p49A) of human estrone sulfatase using the INUB Hybrid Fold Recognition method with a Jscore of 29.67. The N-terminal protein structures were modeled by the program Modeller 9.1 [83] and images were produced by Chimera [84].

## Supporting Information

**Figure S1 ChREBP and non-vertebrate Mondo Secondary Structure.** Consensus secondary structure predictions are overlaid each sequence, with *MCRs* colored red and the *bHLH $\zeta$*  and *DCD* domains colored blue. **A)** *H. sapiens* MondoA. The *PRR* is colored green **B)** *D. melanogaster* Mondo. The *NRB* is colored green. (TIF)

**Figure S2 *DCD/WMC* entropy.** *DCD/WMC* region of all Mondo and Mlx sequences. Numbering corresponds to position in the alignment, shown in Figure 6. Low entropy values indicate site conservation for either a particular amino acid (red: AA) or physiochemical trait (black: FG), e.g. hydrophobic, although low entropy may also result from gaps in the alignment. The dotted line marks an arbitrary threshold of  $H = 0.1$  to indicate highly conserved sites. (TIF)

## References

1. Luo W, Hu H, Chang R, Zhong J, Knabel M, et al. (2011) Pyruvate Kinase M2 Is a PHD3-Stimulated Coactivator for Hypoxia-Inducible Factor 1. *Cell* 145: 732–744.
2. Maddocks OD, Vousden KH. Metabolic regulation by p53. *J Mol Med* 89: 237–245.

**Figure S3 *DCD/WMC* structure.** Rosetta and Human Proteome Folding Project prediction for ChREBP *DCD/WMC* domain. **A)** A cluster of five alpha helices is predicted within the *DCD/WMC* region of ChREBP. **B)** Hydrophobic (red) residues line the interior groove of  $\alpha 2$ ,  $\alpha 3$  and  $\alpha 4$ , while hydrophilic (blue) residues coat the exterior. **C):** Filled *DCD* structure in the same (left) and reversed (right) orientation as above, using ConSurf conservation coloring (maroon: highly conserved, white: neutral, teal: variable). Highly conserved residues are labeled according to the human ChREBP sequence and the *WMC/DCD* alignment numbering. (TIF)

(TIF)

**Figure S4  $S_x[ST]xx[ST]$  motif locations.** We provide evidence that G6P may bind an  $S_x[ST]xx[ST]$  motif in Mondo proteins. This motif has low complexity and is found throughout Mondo sequences, but is only consistently conserved among species for MondoA (red), ChREBP (blue) and non-vertebrate Mondo (black) in the glucose-responsive region containing *MCR6*. Numbering corresponds to position in the alignment. (TIF)

(TIF)

**Figure S5 ChREBP open and closed protein conformation.** Predicted structure of ChREBP in the closed (**A–C**) and open (**D–F**) conformation. Images **B** and **E** are 180 degree rotations of **A** and **D**, respectively, while **C** and **F** depict the structure from an overhead view. Domains are colored as for MondoA in Figures 8 and 9: *MCR1*-red, *MCR2*-orange, *MCR3*-yellow, *MCR4*-green, *MCR5*-purple, *MCR6*-blue. In addition, we have highlighted the proposed NES1 (light pink), the serine and threonine residues in *MCR6* (pale green), and the relevant and putative phosphorylation sites (magenta). Phosphorylation site S140 is located within *MCR3* (**A**, **D**, **C**, **F**), S196 is downstream of *MCR4* (**C**, **F**), while the putative phosphorylation sites S103 (near *MCR2*) and T147 are only accessible in the open conformation (**D**). (TIF)

(TIF)

**Table S1 Cell type specific nuclear accumulation of MondoA and ChREBP in response to glucose.** Values represent the (~approximate) percentage of cells with Mondo transcripts located in the cytoplasm (C), nucleus (N), or both (B) for low and high glucose medium in rat hepatocytes, 832/13 insulinoma cells, INS-1 pancreatic cells, L6 myoblasts, COS-7 and HEK293 kidney cells, and NIH3T3 fibroblasts. (DOCX)

## Acknowledgments

We would like to thank Steve Spiker for his critical reading of this manuscript and Patrick Carroll for his insightful discussion on MondoA and ChREBP protein function.

## Author Contributions

Conceived and designed the experiments: LGM WRA. Performed the experiments: LGM. Analyzed the data: LGM. Contributed reagents/materials/analysis tools: LGM WRA. Wrote the paper: LGM WRA.

5. Stoeckman AK, Ma L, Towle HC (2004) Mlx is the functional heteromeric partner of the carbohydrate response element-binding protein in glucose regulation of lipogenic enzyme genes. *J Biol Chem* 279: 15662–15669.
6. Tabor DE, Kim JB, Spiegelman BM, Edwards PA (1998) Transcriptional activation of the stearoyl-CoA desaturase 2 gene by sterol regulatory element-binding protein/adipocyte determination and differentiation factor 1. *J Biol Chem* 273: 22052–22058.
7. Magaña MM, Osborne TF (1996) Two tandem binding sites for sterol regulatory element binding proteins are required for sterol regulation of fatty-acid synthase promoter. *J Biol Chem* 271: 32689–32694.
8. Rufo C, Teran-Garcia M, Nakamura MT, Koo SH, Towle HC, et al. (2001) Involvement of a unique carbohydrate-responsive factor in the glucose regulation of rat liver fatty-acid synthase gene transcription. *J Biol Chem* 276: 21969–21975.
9. Ma L, Robinson LN, Towle HC (2006) ChREBP/Mlx is the principal mediator of glucose-induced gene expression in the liver. *J Biol Chem* 281: 28721–28730.
10. Sans C, Satterwhite D, Stoltzman C, Breen K, Ayer D (2006) MondoA-Mlx Heterodimers Are Candidate Sensors of Cellular Energy Status: Mitochondrial Localization and Direct Regulation of Glycolysis. *Molecular and Cellular Biology* 26: 4863–4871.
11. Billin AN, Eilers AL, Coulter KL, Logan JS, Ayer DE (2000) MondoA, a novel basic helix-loop-helix-leucine zipper transcriptional activator that constitutes a positive branch of a max-like network. *Molecular and Cellular Biology* 20: 8845–8854.
12. McFerrin LG, Atchley WR. Evolution of the Max and Mlx networks in animals. *Genome Biol Evol* 3: 915–937.
13. Yuan J, Tirabassi RS, Bush AB, Cole MD (1998) The *C. elegans* MDL-1 and MXL-1 proteins can functionally substitute for vertebrate MAD and MAX. *Oncogene* 17: 1109–1118.
14. Ma L, Tsatsos NG, Towle HC (2005) Direct role of ChREBP/Mlx in regulating hepatic glucose-responsive genes. *J Biol Chem* 280: 12019–12027.
15. Yamashita H, Takenoshita M, Sakurai M, Bruick RK, Henzel WJ, et al. (2001) A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. *Proc Natl Acad Sci USA* 98: 9116–9121.
16. Eilers AL, Sundwall E, Lin M, Sullivan AA, Ayer DE (2002) A novel heterodimerization domain, CRM1, and 14-3-3 control subcellular localization of the MondoA-Mlx heterocomplex. *Molecular and Cellular Biology* 22: 8514–8526.
17. Peterson C, Stoltzman C, Sighinolfi M, Han K, Ayer D (2010) Glucose Controls Nuclear Accumulation, Promoter Binding, and Transcriptional Activity of the MondoA-Mlx Heterodimer. *Molecular and Cellular Biology* 30: 2887–2895.
18. Billin AN, Ayer DE (2006) The Mlx network: evidence for a parallel Max-like transcriptional network that regulates energy metabolism. *Curr Top Microbiol Immunol* 302: 255–278.
19. Davies MN, O'Callaghan BL, Towle HC (2008) Glucose activates ChREBP by increasing its rate of nuclear entry and relieving repression of its transcriptional activity. *J Biol Chem* 283: 24029–24038.
20. Stoltzman CA, Peterson CW, Breen KT, Muoio DM, Billin AN, et al. (2008) Glucose sensing by MondoA/Mlx complexes: a role for hexokinases and direct regulation of thioredoxin-interacting protein expression. *Proc Natl Acad Sci USA* 105: 6912–6917.
21. Li M, Chang B, Imamura M, Pongvarin N, Chan L (2006) Glucose-dependent transcriptional regulation by an evolutionarily conserved glucose-sensing module. *Diabetes* 55: 1179–1189.
22. Li M, Chen W, Pongvarin N, Imamura M, Chan L (2008) Glucose-Mediated Transactivation of Carbohydrate Response Element-Binding Protein Requires Cooperative Actions from Mondo Conserved Regions and Essential Trans-Acting Factor 14-3-3. *Mol Endocrinol* 22: 1658–1672.
23. Davies M, O'callaghan B, Towle H (2010) Activation and Repression of Glucose-stimulated ChREBP Requires the Concerted Action of Multiple Domains within the MondoA Conserved Region. *AJP: Endocrinology and Metabolism*. 38 p.
24. Cairo S, Merla G, Urbinati F, Ballabio A, Reymond A (2001) WBSCR14, a gene mapping to the Williams-Beuren syndrome deleted region, is a new member of the Mlx transcription factor network. *Human Molecular Genetics* 10: 617–627.
25. Kawaguchi T, Takenoshita M, Kabashima T, Uyeda K (2001) Glucose and cAMP regulate the L-type pyruvate kinase gene by phosphorylation/dephosphorylation of the carbohydrate response element binding protein. *Proc Natl Acad Sci USA* 98: 13710–13715.
26. Merla G, Howard C, Antonarakis SE, Reymond A (2004) The subcellular localization of the ChoRE-binding protein, encoded by the Williams-Beuren syndrome critical region gene 14, is regulated by 14-3-3. *Human Molecular Genetics* 13: 1505–1514.
27. Tsatsos NG, Davies MN, O'Callaghan BL, Towle HC (2008) Identification and function of phosphorylation in the glucose-regulated transcription factor ChREBP. *Biochem J* 411: 261–270.
28. Fukasawa M, Ge Q, Wynn R, Ishii S, Uyeda K (2010) Coordinate regulation/localization of the carbohydrate responsive binding protein (ChREBP) by two nuclear export signal sites: Discovery of a new leucine-rich nuclear export signal site. *Biochem Biophys Res Commun* 391: 1166–1169.
29. Sakiyama H, Wynn RM, Lee WR, Fukasawa M, Mizuguchi H, et al. (2008) Regulation of nuclear import/export of carbohydrate response element-binding protein (ChREBP): interaction of an alpha-helix of ChREBP with the 14-3-3 proteins and regulation by phosphorylation. *J Biol Chem* 283: 24899–24908.
30. Kabashima T, Kawaguchi T, Wadzinski BE, Uyeda K (2003) Xylulose 5-phosphate mediates glucose-induced lipogenesis by xylulose 5-phosphate-activated protein phosphatase in rat liver. *Proc Natl Acad Sci USA* 100: 5107–5112.
31. Li M, Chen W, Harmancey R, Nuotio-Antar A, Imamura M, et al. (2010) Glucose-6-phosphate mediates activation of the carbohydrate responsive binding protein (ChREBP). *Biochem Biophys Res Commun*. 6 p.
32. Matés JM, Segura JA, Campos-Sandoval JA, Lobo C, Alonso L, et al. (2009) Glutamine homeostasis and mitochondrial dynamics. *Int J Biochem Cell Biol* 41: 2051–2061.
33. Tong X, Zhao F, Mancuso A, Gruber JJ, Thompson CB (2009) The glucose-responsive transcription factor ChREBP contributes to glucose-dependent anabolic synthesis and cell proliferation. *Proc Natl Acad Sci USA* 106: 21660–21665.
34. Capra JA, Singh M (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics* 23: 1875–1882.
35. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, et al. (1998) Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* 395: 137–143.
36. Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, et al. (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science* 303: 1378–1381.
37. Xu J, Christian B, Jump DB (2006) Regulation of rat hepatic L-pyruvate kinase promoter composition and activity by glucose, n-3 polyunsaturated fatty acids, and peroxisome proliferator-activated receptor-alpha agonist. *J Biol Chem* 281: 18351–18362.
38. Zhang P, Metukuri M, Bindom S, Prochownik E, O'doherty R, et al. (2010) c-Myc Is Required for the ChREBP-Dependent Activation of Glucose-Responsive Genes. *Mol Endocrinol*. 13 p.
39. Burke SJ, Collier JJ, Scott DK (2009) cAMP opposes the glucose-mediated induction of the L-PK gene by preventing the recruitment of a complex containing ChREBP, HNF4alpha, and CBP. *FASEB J* 23: 2855–2865.
40. Schleiff E (2000) Signals and receptors—the translocation machinery on the mitochondrial surface. *J Bioenerg Biomembr* 32: 55–66.
41. Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N (2009) Importing mitochondrial proteins: machineries and mechanisms. *Cell* 138: 628–644.
42. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2: 953–971.
43. Tsatsos NG, Towle HC (2006) Glucose activation of ChREBP in hepatocytes occurs via a two-step mechanism. *Biochem Biophys Res Commun* 340: 449–456.
44. Pickett C, Breen K, Ayer D (2007) A *C. elegans* Myc-like network cooperates with semaphorin and Wnt signaling pathways to control cell migration. *Developmental Biology* 310: 226–239.
45. de Luis O, Valero MC, Jurado LA (2000) WBSCR14, a putative transcription factor gene deleted in Williams-Beuren syndrome: complete characterisation of the human gene and the mouse ortholog. *Eur J Hum Genet* 8: 215–222.
46. Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. *Meth Enzymol* 383: 66–93.
47. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38: W529–W533.
48. Kay BK, Williamson MP, Sudol M (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* 14: 231–241.
49. Guo L, Han A, Bates DL, Cao J, Chen L (2007) Crystal structure of a conserved N-terminal domain of histone deacetylase 4 reveals functional insights into glutamine-rich domains. *Proc Natl Acad Sci USA* 104: 4297–4302.
50. Piskacek S, Gregor M, Nemethova M, Grabner M, Kovarik P, et al. (2007) Nine-amino-acid transactivation domain: establishment and prediction utilities. *Genomics* 89: 756–768.
51. Dong X, Biswas A, Sütel KE, Jackson LK, Martinez R, et al. (2009) Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* 458: 1136–1141.
52. Ryan JF, Burton PM, Mazza ME, Kwong GK, Mullikin JC, et al. (2006) The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, *Nematostella vectensis*. *Genome Biol* 7: R64.
53. Ottmann C, Yasmin L, Weyand M, Veesenmeyer JL, Diaz MH, et al. (2007) Phosphorylation-independent interaction between 14-3-3 and exoenzyme S: from structure to pathogenesis. *EMBO J* 26: 902–913.
54. Gould CM, Diella F, Via A, Puntervoll P, Gemünd C, et al. (2010) ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Research* 38: D167–180.
55. Sharrocks AD, Yang SH, Galanis A (2000) Docking domains and substrate-specificity determination for MAP kinases. *Trends in Biochemical Sciences* 25: 448–453.
56. Tanoue T, Maeda R, Adachi M, Nishida E (2001) Identification of a docking groove on ERK and p38 MAP kinases that regulates the specificity of docking interactions. *EMBO J* 20: 466–479.
57. Raman M, Chen W, Cobb MH (2007) Differential regulation and properties of MAPKs. *Oncogene* 26: 3100–3112.

58. Sakiyama H, Fujiwara N, Noguchi T, Eguchi H, Yoshihara D, et al. (2010) The role of O-linked GlcNAc modification on the glucose response of ChREBP. *Biochem Biophys Res Commun.* 6 p.
59. Escriva H, Langlois MC, Mendonça RL, Pierce R, Laudet V (1998) Evolution and diversification of the nuclear receptor superfamily. *Ann N Y Acad Sci* 839: 143–146.
60. Poupeau A, Postic C. Cross-regulation of hepatic glucose metabolism via ChREBP and Nuclear Receptors. *Biochim Biophys Acta.*
61. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Research* 14: 1188–1190.
62. Shannon CE (1951) Prediction and entropy of printed English. *The Bell System Technical Journal* 30: 50–64.
63. Atchley WR, Terhalle W, Dress A (1999) Positional dependence, cliques, and predictive motifs in the bHLH protein domain. *J Mol Evol* 48: 501–516.
64. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, et al. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research* 33: W299–W302.
65. Gattiker A, Gasteiger E, Bairoch A (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl Bioinformatics* 1: 107–108.
66. Aleshin AE, Kirby C, Liu X, Bourenkov GP, Bartunik HD, et al. (2000) Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes relevant to allosteric regulation. *J Mol Biol* 296: 1001–1015.
67. Kamata K, Mitsuya M, Nishimura T, Eiki J, Nagata Y (2004) Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* 12: 429–438.
68. Mulichak AM, Wilson JE, Padmanabhan K, Garavito RM (1998) The structure of mammalian hexokinase-1. *Nat Struct Biol* 5: 555–560.
69. Regni C, Shackelford GS, Beamer IJ (2006) Complexes of the enzyme phosphomannomutase/phosphoglucomutase with a slow substrate and an inhibitor. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62: 722–726.
70. Zhang G, Dai J, Wang L, Dunaway-Mariano D, Tremblay LW, et al. (2005) Catalytic cycling in beta-phosphoglucomutase: a kinetic and structural analysis. *Biochemistry* 44: 9404–9416.
71. Graham Solomons JT, Zimmerly EM, Burns S, Krishnamurthy N, Swan MK, et al. (2004) The crystal structure of mouse phosphoglucose isomerase at 1.6 Å resolution and its complex with glucose 6-phosphate reveals the catalytic mechanism of sugar ring opening. *Journal of Molecular Biology* 342: 847–860.
72. Cosgrove MS, Gover S, Naylor CE, Vandeputte-Rutten L, Adams MJ, et al. (2000) An examination of the role of asp-177 in the His-Asp catalytic dyad of *Leuconostoc mesenteroides* glucose 6-phosphate dehydrogenase: X-ray structure and pH dependence of kinetic parameters of the D177N mutant enzyme. *Biochemistry* 39: 15002–15011.
73. Kotaka M, Gover S, Vandeputte-Rutten L, Au SW, Lam VM, et al. (2005) Structural studies of glucose-6-phosphate and NADP<sup>+</sup> binding to human glucose-6-phosphate dehydrogenase. *Acta Crystallogr D Biol Crystallogr* 61: 495–504.
74. Teplyakov A, Obmolova G, Badet-Denisot MA, Badet B (1999) The mechanism of sugar phosphate isomerization by glucosamine 6-phosphate synthase. *Protein Sci* 8: 596–602.
75. Teplyakov A, Obmolova G, Badet B, Badet-Denisot MA (2001) Channeling of ammonia in glucosamine-6-phosphate synthase. *Journal of Molecular Biology* 313: 1093–1102.
76. Nakaishi Y, Bando M, Shimizu H, Watanabe K, Goto F, et al. (2009) Structural analysis of human glutamine:fructose-6-phosphate amidotransferase, a key regulator in type 2 diabetes. *FEBS Lett* 583: 163–167.
77. Combet C, Blanchet C, Geourjon C, Deléage G (2000) NPS@: network protein sequence analysis. *Trends Biochem Sci* 25: 147–150.
78. Porollo A, Meller J (2007) Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D. *BMC Bioinformatics* 8: 316.
79. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 19: 1015–1018.
80. DiMaio F, Terwilliger TC, Read RJ, Wlodawer A, Oberdorfer G, et al. Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* 473: 540–543.
81. Malmström L, Riffle M, Strauss CE, Chivian D, Davis TN, et al. (2007) Superfamily assignments for the yeast proteome through integration of structure prediction with the gene ontology. *PLoS Biol* 5: e76.
82. Riffle M, Malmström L, Davis TN (2005) The Yeast Resource Center Public Data Repository. *Nucleic Acids Res* 33: D378–382.
83. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics Chapter 5: Unit 5.6.*
84. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612.