Structural basis of human 5,10-methylenetetrahydrofolate reductase (MTHFR) regulation by phosphorylation and S-adenosylmethionine inhibition

D. Sean Froese[1]*, Jola Kopec[2], Elzbieta Rembeza[2], Gustavo Arruda Bezerra[2], Anselm Erich Oberholzer[3], Terttu Suormala[1], Seraina Lutz[1], Rod Chalk[2], Oktawia Borkowska[2], Matthias R. Baumgartner[1,4], Wyatt W. Yue[2]*

[1]Division of Metabolism and Children's Research Center, University Children's Hospital, Zurich, Switzerland
[2]Structural Genomics Consortium, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom
[3]Structural Biology Community Laenggasse (sbcl), 3000 Bern, Switzerland
[4]radiz – Rare Disease Initiative Zurich, Clinical Research Priority Program for Rare Diseases, University of Zurich, Zurich, Switzerland

*To whom correspondence may be addressed:
D.S.F.
Division of Metabolism and Children's Research Center, University Children's Hospital,
Steinwiesstrasse 75, CH-8032 Zürich, Switzerland
Tel.: +41 44 266 7156
Email: sean.froese@kispi.uzh.ch

W.W.Y
Structural Genomics Consortium, University of Oxford,
Old Road Campus Research Bldg., Roosevelt Dr., Headington, Oxford OX3 7DQ, United Kingdom.
Tel.: +44 1865 617757
Fax: +44 1865 617575
E-mail: wyatt.yue@sgc.ox.ac.uk

Abstract

The folate and methionine cycles are crucial to the biosynthesis of lipids, nucleotides and proteins, and production of the global methyl donor S-adenosylmethionine (SAM). 5,10-methylenetetrahydrofolate reductase (MTHFR) represents a key regulatory connection between these cycles, generating 5-methyltetrahydrofolate for initiation of the methionine cycle, and undergoing allosteric inhibition by its end product SAM. Our 2.5 Å resolution crystal structure of human MTHFR reveals a unique architecture, appending the well-conserved catalytic TIM-barrel to a eukaryote-only SAM-binding domain. The latter domain of novel fold provides the predominant interface for MTHFR homo-dimerization, positioning the N-terminal serine-rich phosphorylation region into proximity with the C-terminal SAM-binding domain. This explains how MTHFR phosphorylation, identified on 11 N-terminal residues (16-total), increases sensitivity to SAM binding and inhibition. Finally, we demonstrate the 25-amino-acid inter-domain linker enables conformational plasticity and propose it to be a key mediator of SAM regulation.

Introduction

In humans, the folate and methionine cycles both generate products essential to cellular survival. Folate, the major cellular carrier of single carbon units, is required for the synthesis of purines and thymidine monophosphate. Within the methionine cycle, the methylation of homocysteine to methionine by methionine synthase (EC 2.1.1.13) produces an essential amino acid which may be used for protein synthesis or, crucially, be further converted to S-adenosylmethionine (SAM), a vitally important donor for the methylation of DNA, RNA and proteins as well as the creation of numerous methylated compounds. These two cycles intersect at the enzyme 5,10-methylenetetrahydrofolate reductase (MTHFR; E.C. 1.5.1.20). MTHFR catalyzes the physiologically irreversible reduction of 5,10-methylene-tetrahydrofolate (CH_2-THF) to 5-methyl-tetrahydrofolate (CH_3-THF), a reaction requiring FAD as cofactor and NADPH as electron donor. Since the product CH_3-THF is exclusively used by methionine synthase, and only the demethylated form (THF) may be recycled back to the folate cycle, MTHFR commits THF-bound one-carbon units to the methionine cycle.

In accordance with this essential role, major and minor deficiencies of human MTHFR are the direct or indirect causes of human disease. Severe MTHFR deficiency (MIM #607093) is inherited in an autosomal recessive manner and is the most common inborn error of folate deficiency[1] with ~200 patients known[2]. To date, over 100 different clinically relevant mutations in *MTHFR* have been described, the majority of which are of the missense type (n=70, >60%) and "private"[2]. Milder enzyme deficiencies, due to single nucleotide polymorphisms (SNPs) of the *MTHFR* gene, have been associated with various common disorders. The most studied of these is p.Ala222Val (c.665C>T in NM_001330358, commonly annotated as c.677C>T), identified as a risk factor for an overwhelming number of multifactorial disorders, including: vascular diseases, neurological diseases, various cancers, diabetes and pregnancy loss (see e.g. review by [3]).

Human MTHFR is a 656 amino acid multi-domain protein (Fig. 1). The catalytic domain is conserved across evolution, and crystal structures of MTHFR from *Escherichia* (*E.*) *coli*[4-7] and *Thermus thermophilus*[8], in which the catalytic domain constitutes the entire sequence (Fig. 1), have been solved. These structures reveal the catalytic domain to form a $\beta_8\alpha_8$ (TIM) barrel and have uncovered residues critical for binding the cofactor FAD[4], the electron donor NADPH (NADH in bacteria[7]) and the product CH_3-THF[5,6]. The bacterial structures, together with activity assay of trypsin cleaved porcine MTHFR[9], indicate that the catalytic domain is sufficient for the entire catalytic cycle. Eukaryotic MTHFR orthologs additionally possess a C-terminal regulatory domain that is connected to the catalytic domain by a linker sequence (Fig. 1). This C-terminal domain is able to bind SAM, resulting in allosteric inhibition of enzymatic activity[10], an effect which is very slow[11] and can be reversed by binding to S-adenosylhomocysteine (SAH)[12,13], the demethylated form of SAM.

Human MTHFR further contains a 35 amino acid serine-rich region at the very N-terminus which is not found in MTHFR orthologs of bacteria, yeast or even lower animals (Fig. 1). This region has been identified to be multiply phosphorylated following heterologous expression in insect cells[14] and yeast[15], or following immunoprecipitation from human cancer cell lines[16]. Phosphorylation has been associated with moderately decreased catalytic activity[14-16] and increased total inhibition mediated by SAM[14]. Although phosphorylation mapping of this region has been thus far unsuccessful, scanning mutagenesis has revealed substitution of alanine for threonine at position 34 (p.Thr34Ala) to almost completely block phosphorylation[14,15], suggesting Thr34 is the priming position. The cellular relevance of this modification remains unclear, although one group has suggested that phosphorylation at Thr34 can be accomplished by CDK1/cyclin B1[16] and at Thr549 by polo-like kinase 1[17] whereby they posit a role in histone methylation and replication.

The repertoire of bacterial MTHFR structures to date does not provide any mechanistic insight into the enzymatic regulation by phosphorylation and SAM binding, because both features are absent in prokaryotes. To this end, we have combined structural, biophysical and biochemical data of human

MTHFR to provide a molecular view of MTHFR function and regulation in higher eukaryotes. We have identified specific phosphorylation sites and demonstrate a distinct relationship between phosphorylation, conformational change and SAM inhibition. Further, using our 2.5 Å resolution crystal structure of the almost full-length human protein, we reveal that the regulatory domain utilizes a novel topology to bind SAH/SAM and transmit a catalytic inhibition signal by long range conformational change, most likely through the linker region. These novel insights highlight conformational plasticity as an important mediator of MTHFR regulation.

Results

Identification of phosphorylated residues by mass spectrometry

To examine the phosphorylation status of human (*Hs*)MTHFR, we generated full-length recombinant human MTHFR (*Hs*MTHFR$_{1-656}$) by baculovirus expression in Sf9 cells. Mass spectrometry-based phosphorylation mapping (with 92% coverage) identified 16 separate phosphorylation sites in *Hs*MTHFR$_{1-656}$ following purification from Sf9 cells (called here "*as purified*") (Fig. 2A). All phosphorylation sites were considered to have partial occupancy, since no residues were phosphorylated in every tryptic peptide analysed (data not shown). Of these, 11 phosphorylated amino acids (Ser9, Ser10, Ser18, Ser20, Ser21, Ser23, Ser25, Ser26, Ser29, Ser30, Thr34) were within the N-terminal serine-rich region, including the putative phosphorylation determining residue Thr34 (Fig. 2A). Additionally, we found phosphorylation of three further amino acids in the catalytic domain (Tyr90, Thr94, Ser103) and two in the regulatory domain (Ser394, Thr451). Up to ten phosphorylation sites were identified to be occupied simultaneously, whereby treatment with calf intestine alkaline phosphatase (CIP) resulted in removal of 9 (Fig. 2B) or 10 (Fig. 2C) phosphate groups, as identified by denaturing and native mass spectrometry, respectively. To examine the importance of the N-terminal serine-rich region to global protein phosphorylation, we produced recombinant *Hs*MTHFR$_{38-644}$, which removes the N-terminal 37 amino acids, including the entire serine-rich region (Fig. 1) as well as the poorly conserved C-terminal 12 amino acids predicted to be of high disorder (Suppl. Fig. 1). *As purified Hs*MTHFR$_{38-644}$ was not found to be phosphorylated by phosphorylation mapping (Suppl. Fig. 2A), or native mass spectrometry (Suppl. Fig. 2B), and treatment with CIP did not alter the protein molecular mass (Fig. 2D). Therefore, the primary determinant of *Hs*MTHFR phosphorylation resides within the N-terminus.

Phosphorylation does not alter MTHFR kinetic parameters

Phosphorylation has been described to alter MTHFR kinetics, resulting in moderately decreased catalytic activity as measured by the NADPH-menadione oxidoreductase assay[14,16]. To investigate this more thoroughly, we used a very sensitive HPLC-based activity assay which monitors the full enzymatic reaction in the physiological direction and allows determination of kinetic values[18]. Overall, we found similar kinetic values for *Hs*MTHFR$_{1-656}$ and *Hs*MTHFR$_{38-644}$ (Table 1). Compared to non-phosphorylated *Hs*MTHFR$_{38-644}$, phosphorylated *Hs*MTHFR$_{1-656}$ had only slightly decreased specific activity (34.0 ± 1.3 versus 30.8 ± 1.5 µmol • min^{-1} • mg^{-1}) and K_{cat} values (51.4 ± 4.2 versus 40.7 ± 3.2 sec^{-1}), suggesting turn-over was not meaningfully reduced. These values are comparable but higher than previous determinations from recombinant *Hs*MTHFR (12.4 µmol/min/mg[13]) and purified porcine MTHFR (19.4 µmol/min/mg[19]). Importantly, CIP treated *Hs*MTHFR$_{1-656}$ (dephosphorylated), retained nearly identical activity values to mock treated *Hs*MTHFR$_{1-656}$ (i.e. assay without addition of CIP; phosphorylated) (Table 1), confirming that phosphorylation does not decrease protein turn-over. Likewise, Michaelis-Menten constants (K_m) for the substrate CH_2-THF (range: 21.3 - 25.5 µM) and electron donor NADPH (range: 23.5 - 35.5 µM) were very similar for all four proteins (Table 1), and comparable to that of MTHFR from human fibroblast lysates (CH_2-THF: 26 µM; NADPH: 30 µM[18]). Thus, we conclude that our assay is sensitive and specific, and that phosphorylation does not significantly alter MTHFR kinetic properties.

Interestingly, we found no increase in the specific activity of MTHFR proteins following addition of exogenous FAD to the assay buffer (Table 1). Since FAD is required as cofactor for the MTHFR reaction, this suggests the cofactor was already bound to the *as purified* protein, presumably acquired during cellular expression. This is consistent with native mass spectrometry, which identified monomeric and dimeric forms of *as purified Hs*MTHFR$_{1-656}$ which, in addition to phosphorylation, contained equivalent units of FAD (Fig. 2C). *Hs*MTHFR$_{38-644}$ also presented as monomeric and dimeric forms bound to equivalent units of FAD (Suppl. Fig. 2B), suggesting phosphorylation has no effect in this regard. Supplementation with FAD, however, helped rescue activity either during, or to a lesser

extent following, incubation of these proteins at 46°C for 5 minutes (Table 1). Therefore, this cofactor, which is important for protein stability, may be lost under heat treatment. In our experiments, *Hs*MTHFR$_{1-656}$ was markedly more sensitive to heat inactivation than *Hs*MTHFR$_{38-644}$, but this heat sensitivity was not affected by the phosphorylation state of the protein, and therefore likely rather reflects overall protein stability.

MTHFR phosphorylation increases protection of and sensitivity to SAM

In addition to phosphorylation and FAD, native mass spectrometry identified the dimeric form of *as purified Hs*MTHFR$_{1-656}$ to contain 0, 1 or 2 units of SAM (Fig. 2E). Like FAD, SAM was likely acquired during cellular expression. However, following CIP treatment, the SAM bound to MTHFR was found to degrade to SAH, a chemical transition which did not occur during mock treatment of the protein (Fig. 2E). Correspondingly, *as purified* dimeric *Hs*MTHFR$_{38-644}$, which is not phosphorylated, was found to be bound to 0, 1 or 2 units of SAH, but not SAM (Suppl. Fig. 2C). Thus, phosphorylated MTHFR appeared to protect thermally unstable SAM from degradation to SAH, while the non-phosphorylated protein was unable to perform this function.

Phosphorylation has been identified to affect the maximum degree of inhibition of MTHFR by SAM, whereby phosphorylated protein was found to be maximally ~80% inhibited, while phosphatase treated protein was maximally ~60% inhibited[14]. At high concentrations of SAM (> 200 µM), we were able to inhibit all recombinant *Hs*MTHFR proteins by over 95%, regardless of the phosphorylation state (Fig. 3). However, at low SAM concentrations we found phosphorylated *Hs*MTHFR$_{1-656}$ to be more sensitive to SAM inhibition than *Hs*MTHFR$_{38-644}$ and dephosphorylated *Hs*MTHFR$_{1-656}$ (Fig. 3). Further analysis revealed *as purified* and mock treated *Hs*MTHFR$_{1-656}$ to have inhibition constants (K_i) of ~ 3 µM, while CIP treated *Hs*MTHFR$_{1-656}$ was approximately 2-fold less sensitive to SAM inhibition, and *Hs*MTHFR$_{38-644}$ 7-fold less sensitive (Fig. 3 - inset; Table 1). Thus, although phosphorylation does not directly affect MTHFR enzymatic activity, it increases the protein's sensitivity to SAM inhibition.

Human MTHFR has evolved an extensive linker to connect and interact with its two domains

We determined the 2.5 Å resolution structure of *Hs*MTHFR$_{38-644}$ in complex with FAD and SAH by multiple-wavelength anomalous dispersion using the selenomethionine (SeMet) derivatized protein (Table 2). The identity of both ligands is guided by well-defined electron density (Suppl. Fig. 3), and in line with native mass spectrometry analysis for this construct (Suppl Fig. 2B). The *Hs*MTHFR$_{38-644}$ protomer folds into two globular domains (Fig. 4A) to form an overall elongated molecule. As predicted from bacterial structures, the N-terminal catalytic domain (aa 40-337) consists of an 8α/8β TIM barrel, adorned with 3 extra α-helices (α8, α9 and α11) (Fig. 4A, Suppl. Fig. 4). The C-terminal regulatory domain (aa 363-644) makes up a novel fold of two five-stranded β-sheets arranged side-by-side in the core, flanked by a number of α-helices (Fig. 4A, Suppl. Fig. 4). The two domains do not contact each other directly, but instead are connected by an extended linker region encompassing aa 338-362 (Fig. 4A), with its amino acid sequence rich in Arg (n=4), Pro (4), Glu (3) and Leu (3). Mediated by three β-turns, this linker makes multiple intricate contacts with both domains, and changes direction twice in traversing between the catalytic and regulatory domains.

The *Hs*MTHFR$_{38-644}$ structure allows the mapping of the 70 inherited missense mutations known to cause severe MTHFR deficiency, which lie on 64 different residues of the polypeptide (Suppl. Fig. 5). Twice as many of the mutation sites are found in the catalytic domain (n=38) as the regulatory domain (20), with the remainder (6) found in the linker region. By proportion, however, the linker region has a higher density (24% of the sequence) of mutation sites than the catalytic (11%) and regulatory (7%) domains. Additionally, a number of sites in the catalytic and regulatory domains are in direct contact with the linker region. Further, the most severe mutations, those found either homozygously or in conjunction with a truncating mutation to result in enzymatic activity below 1.5% of control activity in patient fibroblasts[20], cluster in the catalytic domain and the first two aa of the linker region, most of

which are located where the linker meets the catalytic domain (Suppl. Fig. 5). Together, this analysis underscores the importance of the linker region to proper protein function.

The *Hs*MTHFR$_{38-644}$ structure displays an asymmetric dimer with inter-domain flexibility

The *Hs*MTHFR$_{38-644}$ structure reveals a homodimer (Fig. 4B), consistent with native mass spectrometry (Suppl. Fig 2B) and previous investigation of mammalian MTHFR by size exclusion chromatography and scanning transmission electron microscopy[9]. It was previously thought that MTHFR homodimerizes in a head-to-tail manner, where the regulatory domain of one subunit interacts with the catalytic domain of the other subunit[13]. Unexpectedly, in our structure dimerization is mediated almost entirely by the regulatory domain (Fig. 4B), although the first ordered residue in chain A (Glu40) is located around 5-6 Å from the regulatory domain of chain B (e.g. Glu553, Arg567). The N-terminal sequence that is either not present (Ser-rich phosphorylation region, aa 1-37), or present but disordered (aa 38-39) in the *Hs*MTHFR$_{38-644}$ structure will likely project towards the interface of the two regulatory domain (Fig. 4C), and may contribute further to the dimer contacts.

The essential interfacial residues from the regulatory domain are contributed predominantly from the two central β-sheets, including a β-turn (β11-β12), strand β16, and the loop encompassing Asn386-Asn391 (Suppl. Fig. 6), which buries in total ~1330 Å^2 of accessible surface. Half of the sites of missense mutations in the regulatory domain causing MTHFR deficiency (n=10, Suppl. Fig. 5) either participate in, or are within in two residues of, the dimerization site.

Within the homodimer, each of the two catalytic domains is presented away from the dimeric interface and their active sites are at opposite ends of the overall shape and face away from each other (Fig. 4B). In this arrangement, the catalytic domain is not involved in oligomerization, unlike bacterial and archaeal MTHFR proteins (Suppl. Fig. 7). This said, the N-terminus of the *Hs*MTHFR$_{38-644}$ construct is projecting towards the dimer interface. A direct consequence of the dimeric architecture is that the *Hs*MTHFR catalytic domain displays a large degree of flexibility in relative orientation with the regulatory domain. In fact, this is reflected in our structure whereby the catalytic domain of one dimer subunit (chain A) is ordered, while that of the other dimer subunit (chain B) is highly disordered, to the extent that only main chain atoms of the amino acid 40-58, 129-134 and 155-342 in chain B could be modeled.

Dynamics of MTHFR observed by solution scattering

Our *Hs*MTHFR$_{38-644}$ crystal structure has captured the snapshot of an asymmetric dimer whereby the two catalytic domains have different orientations with respect to their own regulatory domains (Suppl. Fig. 8). We applied small angle x-ray scattering (SAXS) to understand better the different conformational variations assumed by the protein in solution. Superimposition of the theoretical scattering curve back-calculated from the crystal structure dimer against experimental data obtained from *Hs*MTHFR$_{38-644}$ in solution revealed a poor fit (Chi^2 14.8; Fig. 5A and B), suggesting this is not the predominant conformation in solution. However, by employing CORAL[21] to simulate relaxation of the relative orientations of the catalytic and regulatory domains (by allowing flexibility in residues 338-345 of the linker), and thus also permitting rigid body movement of these subunits in relative orientation to each other, we obtained a significantly improved fit (Chi^2 5.5; Fig. 5A). Here, the best model was also represented by an asymmetric dimer, but in this case with catalytic domains extended and rotated in comparison to the regulatory domains (Fig. 5B). Thus, consistent with our finding from the crystal structure, *Hs*MTHFR retains a significant degree of intra- and inter-domain conformational flexibility in solution.

To further investigate the influence of phosphorylation on protein conformation, we next collected SAXS data for full-length *Hs*MTHFR$_{1-656}$ *as purified* (i.e. phosphorylated and bound with SAM) and treated with CIP (i.e. dephosphorylated and bound with SAH). Superimposition of the experimental scattering curves between these two conditions indicates that *as purified Hs*MTHFR$_{1-656}$ has slightly

larger dimensions than CIP-treated *Hs*MTHFR$_{1-656}$ as indicated by its larger LogI(0) (Fig. 5C), although both protein forms are consistent with a dimeric configuration. To clarify the difference in overall shape, we fitted the theoretical scattering curves of the *Hs*MTHFR$_{38-644}$ rigid body CORAL models to the SAXS experimental data of phosphorylated (Chi^2 of 22.2 +/-0.4) and dephosphorylated (Chi^2 of 30.7 +/-1.8) MTHFR. Since the hits were clearly different, the protein conformations represented by the experimental data are also different. Likewise, the theoretical scattering curve of *Hs*MTHFR$_{38-644}$ observed in the crystal presents a good fit to the experimental data of dephosphorylated MTHFR (Chi^2 of 5.9), but not to that of phosphorylated MTHFR (Chi^2 11.6). This data was further corroborated by charge radius analysis of native phosphorylated and dephosphorylated *Hs*MTHFR$_{1-656}$ by electrospray ionization mass spectrometry, which found a significantly different charge-distribution of protein ions between the two protein forms, indicating altered flexibility (Suppl. Fig 9). Altogether, we interpret that the phosphorylated SAM-bound form of the protein presents a different conformation to the dephosphorylated SAH-bound form.

Subtle features of the eukaryotic catalytic domain provide for NADPH specificity

The MTHFR catalytic domain adopts a TIM-barrel structure evolutionarily conserved across all kingdoms. In addition to *Hs*MTHFR$_{38-644}$, we further determined the catalytic domain structure of the yeast homolog MET12 (*Sc*MET12$_{1-301}$) to 1.56 Å resolution (Table 2). This enables for the first time a structural comparison across mammalian (*Hs*MTHFR), low-eukaryotic (*Sc*MET12) and bacterial (*E. coli*, *H. influenzae*, *T. thermophilus*) orthologues. Consistent with their sequence conservation (Suppl. Fig. 10), the catalytic domains have highly superimposable folds (main chain RMSD: 1.85 Å), although distinct local differences are found in low homology loop regions (Fig. 6A, a-b) and helices (Fig. 6A, c-d).

In *Hs*MTHFR$_{38-644}$, clear electron density for FAD was observed in the TIM barrel of chain A (Suppl. Fig. 3A). However, there is high disorder in the TIM-barrel of chain B particularly around the FAD binding site, implying a low ligand occupancy of the ligand, although native mass spectrometry of the crystallized construct indicated two FADs bound per homodimer (Suppl. Fig. 2B). Analysis of the FAD binding residues in *Hs*MTHFR chain A (Fig. 6B) and in *Sc*MET12 (not shown) reveals perfect overlap with those predicted from the *Ec*MTHFR structure[4]. These include Thr129, Arg157, Ala175 and Ala195 of *Hs*MTHFR, which were associated with *in vitro* FAD-responsiveness when mutated in severe MTHFR deficiency[20,22] (Fig. 6B; Suppl. Fig. 5).

The bi bi kinetic mechanism of MTHFR necessitates the electron donor NAD(P)H and substrate CH_2-THF to interact in turn with FAD for transfer of the reducing equivalent, and hence to share the same binding site. In our structures, the FAD ligand adopts a conformation poised to expose the *si* face of the isoalloxazine ring for the incoming NAPDH and CH_2-THF. However, instead of trapping the electron donor or substrate (despite multiple attempts at co-crystallization), the binding site in *Sc*MET12$_{1-301}$ and subunit A of *Hs*MTHFR$_{38-644}$ is blocked by a crystal packing interaction from a nearby symmetry mate, making π-π stacking interactions with the FAD ligand (Suppl. Fig. 11). By contrast, no crystal packing interaction is found in the chain B binding site of *Hs*MTHFR$_{38-644}$, explaining the overall mobility and disorder of its catalytic domain.

Superimposing the *Hs*MTHFR$_{38-644}$ structure with that of *Ec*MTHFR bound with NADH (Fig. 6C) and CH_3-THF (Fig. 6D) demonstrates that the human enzyme has largely preserved the same shared binding site found in prokaryotes, with Gln228, Gln267, Lys270, Leu271 and Leu323 likely to be important for interacting with both NAD(P)H and CH_3-THF. *Ec*MTHFR preferentially utilizes NADH[23], and its NADH-bound structure reveals a highly uncommon bent conformation[24] for the electron donor, where the nicotinamide ring stacked over the adenine base to mediate π-π interactions[7]. Our activity assay of *Hs*MTHFR$_{38-644}$ and *Hs*MTHFR$_{1-656}$ clearly demonstrates an ~100-fold preference for NADPH compared to NADH as electron donor (Table 1), in agreement with previous enzyme studies from pig[11,25] and rat[11] MTHFRs.

Within the *Hs*MTHFR active site, we did not identify any obvious differentiating features surrounding the modeled NADH, which could indicate how the extra 2'-monophosphate group on the NADPH ribose is accommodated (Suppl. Fig. 12). It is also unclear if *Hs*MTHFR actually binds NADPH in a similar manner as NADH for *Ec*MTHFR, considering there is only one report in literature documenting a compact stacked conformation for NADPH[26]. Modelling an NADPH ligand with such stacked conformation onto the *Hs*MTHFR$_{38-644}$ structure reveals severe steric clashes with helix α8 (not shown), which creates the floor of the NAD(P)H binding site (e.g. via Gln267, Lys270 and Leu271). Helix α8 is poorly aligned with bacterial and low eukaryotic orthologues in both amino acid sequence (Suppl. Fig 10) and structural topology (Fig. 6A). The equivalent helix in *Ec*MTHFR harbours the residue Phe223, which is crucial to NADH binding[7] and moves to accommodate substrate release[5]. Notably, this residue is not conserved in *Hs*MTHFR and *Sc*MET12, replaced by Leu268 and Ala230, respectively. (Suppl. Fig 10). Therefore, given its position and mobility, we propose that residue(s) on helix α8 in *Hs*MTHFR may play a role in the specificity for NADPH and likely also substrate binding/release.

A novel fold for the SAM-binding regulatory domain

The *Hs*MTHFR$_{38-644}$ structure provides the first view of the 3D arrangement of the regulatory domain unique to eukaryotic MTHFR. The core of this fold comprises two mixed β-sheets of 5 strands each (β9↑-β17↑-β16↓-β12↑-β11↓ and β10↓-β13↑-β18↓-β14↑-β15↓) (Suppl. Fig. 4). Strand β10 from one sheet forms a continuous segment with β11 from the other sheet, and similarly β12 from one β-sheet continues onto β13 of the other sheet. The threading of the two central β-sheets are interspersed with three loop extensions containing different numbers of α-helices (α12-α15, α16, and α17-α18). To the best of our knowledge, the MTHFR regulatory domain represents a unique SAM binding architecture distinct from the 18 known classes of SAM-dependent methyltransferases and non-methyltransferases (Gana et al. 2013)(Suppl. Fig. 13). Further, a DALI search of this domain (Holm and Laakso 2016) did not yield any structural homolog, and we found no existing annotation in PFAM/CATH/SCOP databases and no sequence for this domain beyond eukaryotic MTHFR homologs. Therefore, this appears to be a novel fold utilized only by MTHFR for SAM binding/inhibition.

In our structure, SAH is bound in an extended conformation within the part of the regulatory domain (Fig. 7A) that faces the catalytic domain. Indeed, part of the binding site is constituted by the linker region itself. The ligand is sandwiched between the loop segment preceding α15 (N_{456}DEPLAAET$_{464}$) and the first strand β10 (T_{481}INSQ$_{485}$) of the central β-sheets, where a number of conserved residues are found. For example, Thr481 (conserved in 96% of 150 orthologues; Consurf[27]) and Ser484 (98%) hydrogen-bond to the SAH adenine moiety, while Glu463 (99%) and Thr464 (62%) fixate the ribose hydroxyl groups. The strongest sequence conservation in the SAH binding site is found around the homocysteine moiety, including Pro348 (invariant) and Trp349 (99%) from the linker region, as well as Thr560 and Thr573 (both invariant) at the start and end of the β15-β16 turn. The SAH homocysteine sulfur atom is loosely contacted by Glu463 (3.8 Å) and Ala368 (3.7 Å). SAM is expected to bind to the same site in the regulatory domain, in a similar extended configuration as SAH and requiring the same set of binding residues. However, the additional methyl group in the sulphonium centre of SAM would create steric clash to the Ala368 position of the structure (inter-residue distance ~2.0 Å between heteroatoms, and <1.5 Å between hydrogen atoms) (Suppl. Fig. 14). Although not strictly conserved (45% of 150 orthologues), conservation of Ala368 follows a similar evolutionary pattern as the MTHFR domain organisation (Fig. 1): in higher animals alanine is invariant; lower animals may accommodate a serine; while lower eukaryotes often incorporate a bulky residue (e.g. lysine) (see Suppl. Fig. 15). Therefore, in higher organisms such humans, SAM binding likely results in conformational rearrangement of the loop region containing Ala368 to accommodate its methyl moiety.

SAM-dependent conformational change is mediated by the linker region

Since there is no direct interface between the active-site of the catalytic domain and any part of the regulatory domain (Fig. 4), SAM binding must elicit enzymatic inhibition via a conformational change propagated from the regulatory to catalytic domain. The most likely effector of this conformational change is the extended linker region (defined as aa 338-362), since it makes multiple contacts to both the regulatory and catalytic domains (Fig. 4) and forms part of the SAM/SAH binding site (Fig. 7A). To investigate the potential of this region to elicit conformational change following SAM binding, we generated recombinant *Hs*MTHFR proteins consisting of the regulatory domain alone attached to progressively shorter linker regions, where the N-terminus of these constructs would become Pro348 (*Hs*MTHFR$_{348-656}$), Arg357 (*Hs*MTHFR$_{357-656}$) and Arg377 (*Hs*MTHFR$_{377-656}$) (Suppl. Fig. 1; Fig. 7B). All three constructs are sufficient to bind SAM and SAH, as demonstrated by dose-dependent increases in thermostability by differential scanning fluorimetry when exposed to increasing concentrations of each ligand (Suppl. Fig. 16A). This again reinforces the catalytic and regulatory domains as separate binding modules for their cognate ligands (FAD/NADPH/CH_3-THF vs SAM/SAH respectively).

We employed analytical size exclusion chromatography (aSEC) as a means to study solution behaviour of the MTHFR regulatory domain in response to SAM/SAH binding. Exposure of MTHFR$_{348-656}$ to either SAH or SAM resulted in shifts of elution volume (V_e) compared to *as purified* (apo-) protein, which we interpreted as changes in the overall protein conformation, rather than changes in oligomeric states. Our assumption is based on the native mass spectrometry data (Fig. 2C; Suppl. Fig 2B) showing that SAM and SAH do not alter the oligomeric states observed for the *Hs*MTHFR$_{1-656}$ and *Hs*MTHFR$_{38-644}$ proteins. Importantly, for MTHFR$_{348-656}$, SAM resulted in a leftward V_e shift (suggestive of a larger hydrodynamic volume), and SAH a rightward shift (suggestive of a smaller hydrodynamic volume) (Fig. 7B). By contrast, MTHFR$_{357-656}$ showed conformational change only when exposed to SAM, and MTHFR$_{377-656}$ did not change conformation when exposed to either ligand (Fig. 7B). A similar pattern of results were observed when using purified recombinant mouse MTHFR of the same protein boundaries (Suppl. Fig. 16A and B). Therefore, we conclude that residues within 357-377 must contribute to conformational change upon SAM binding.

Next we carried out site-directed mutagenesis to define residues involved in SAM binding, and/or SAM-mediated conformational change as observed in the aSEC experiment. We reasoned that mutation of Glu463 (which hydrogen-bonds a ribose oxygen) could lead to loss of SAH/SAM binding, and thus conformational change. Indeed, conservative mutation of Glu463 to either aspartate (p.E463D) or glutamine (p.E463Q) on MTHFR$_{348-656}$ resulted in protein that could no longer bind SAM (Suppl. Fig. 16C), nor change conformation in its presence (Fig. 7C). We further hypothesized that mutation of Ala368 (in close proximity to the SAM/SAH sulphonium centre) to a smaller residue (glycine: p.A368G) may not have an effect on binding or conformational change, while mutation to a larger residue (leucine: p.A368L), might reduce the ability of the linker region to sense SAM binding. Correspondingly, p.A368L resulted in protein which retained the ability to bind SAM, but was less sensitive to change in its presence, while p.A368G did not change either of these properties. (Fig. 7C, Suppl. Fig.16C). These experiments conclusively pinpoint Glu463 as crucial to SAM binding, and Ala368 to SAM sensing, representing a mechanism that could transmit a ligand-bound signal from regulatory to catalytic domain of the protein.

Discussion

Catalytic regulation by phosphorylation and SAM binding distinguishes human MTHFR from its bacterial (which do not have phosphorylation or SAM binding regions) and lower eukaryotic (which do not have a phosphorylation region) counterparts. Until now, the molecular basis of how these two allosteric events modulate the catalytic machinery was entirely unknown, due to the absence of a structural context. Now, our structure-guided study has provided 2 major discoveries in this area: (1) identification of an extensive linker region involved in both SAM-binding and purveying the binding signal to inhibit catalysis by conformational change; and (2) demonstration of the concerted effects of phosphorylation and SAM binding, individually mediated by regions more than 300 amino acids apart.

Long-distance cross-talk between phosphorylation and SAM-binding

This work presents the first mapping of the entire phosphorylation landscape of *Hs*MTHFR, revealing phosphorylated Ser/Thr not only at the far N-terminus (n=11) as predicted from the sequence, but also within the catalytic (3) and regulatory (2) domains. Many of the N-terminal phosphorylation sites identified are consistent with previous mutation analysis[14], including Thr34[14-16]. The phosphorylated residues detected in the catalytic and regulatory domains were not reported before. Interestingly, two phosphorylated Ser are located within the FAD binding site, although their physiological significance is currently unclear. Contrary to the recent observation of Li et al.[17] we did not identify phosphorylation of Thr549.

An important finding with regards to MTHFR phosphorylation is that it does not directly alter the catalytic parameters of the enzyme, as determined by a sensitive HPLC-based activity assay. Perhaps this is not too surprising, since the first ordered residue of the structure, Glu40 (i.e. immediately following the phosphorylation region aa 1-37), is far removed from the catalytic site. Instead, MTHFR phosphorylation exerts a long-range influence on the SAM binding status at the regulatory domain some 300 amino acids away, by causing an increased sensitivity of the enzyme to SAM inhibition, but with no overall changes on total SAM inhibition. Phosphorylation likely enhances SAM sensitivity in two interdependent ways. Firstly, it enables protection of bound SAM from spontaneous degradation to SAH, a phenomenon widely observed for SAM-bound enzymes *in vitro*[28] (unpublished observations) and *in crystallo*[29], to avoid dis-inhibition by SAH. Secondly, phosphorylation could induce a conformational change to the protein that primes an inhibition ready state. The SAM Ki differences between phosphorylated and dephosphorylated protein, while relatively small (2-3 µM versus 6-7 µM), are likely to be physiologically relevant. Intracellular SAM concentrations are reported to be 1 - 3 µM in human cells[30,31], and the mTORC1 linked starvation sensor SAMTOR, which recognizes SAM for nutrient sensing, has a SAM dissociation constant of 7 µM[32].

It is not immediately clear if global phosphorylation or phosphorylation of only specific residues contributed to the results we found. Evolutionary conservation of the identified phosphorylation sites varies from absolute invariance through to yeast (e.g.Ser394), to poor conservation even among animals (e.g. Ser9 and Ser10) (Suppl. Fig 15). Truncated recombinant *Hs*MTHFR$_{38-644}$ was not identified to be phosphorylated by mass spectrometry or crystallography, suggesting that phosphorylation at the far N-terminus primes the other phosphorylation events within the catalytic and regulatory domains. This is consistent with previous observations[14-16] that removal of Thr34 results in non-phosphorylated protein *in vitro*. It remains to be seen whether phosphor-Thr34 alone, or other sites at the Ser-rich region, primes other phosphorylation events *in vivo*.

Inter-domain conformational change is integral to MTHFR regulatory properties

We identified the MTHFR regulatory domain to constitute a novel SAM binding fold whose appendage to the well conserved catalytic TIM-barrel is a relatively recent and contained evolutionary event. A similar phenomenon of domain organisation is found in several eukaryotic enzymes, for example those involved in amino acid metabolism (e.g. cystathionine β-synthase, CBS[33]; phenylalanine hydroxylase,

PAH[34], whereby the additional metabolite-binding modules, not found in their bacterial counterparts, serve to fine-tune catalysis in response to the more intricate higher eukaryotic metabolic and signalling cues. We propose that MTHFR belongs to this class of allosteric enzymes that share a common mechanism – to regulate catalysis through steric sequestration of the catalytic site, in a ligand-dependent manner (SAM for MTHFR and CBS; phenylalanine for PAH).

In all of these enzymes, inter-domain conformational change is central to the allosteric mechanism, bringing about a rearrangement of the relative orientation between the regulatory and catalytic domains. Often this requires the flexibility of an inter-domain linker that adopts different conformations to mediate the domain-domain rearrangement. Our data are consistent with the MTHFR linker region being an indispensable component of this mechanism, as supported by the concentration of deleterious disease mutations found in this region. The MTHFR linker has a function beyond merely joining the two domains physically, but actively partakes in the allosteric mechanism by (i) acting as a SAM sensor that contributes to binding the effector ligand, and (ii) purveying the SAM-bound inhibition signal to the catalytic domain. The MTHFR linker is aptly suited for these dual roles, as it makes extensive contacts with the regulatory domain (e.g. SAM binding site) and catalytic domain (e.g. helices α3 and α4).

Our domain truncation and mutagenesis experiments coupled with size exclusion chromatography have dissected the regions responsible for SAM/SAH binding and binding-mediated conformational change. Notably, aa 348-376 contribute to, but are not essential for the SAM/SAH binding site; while aa 357-376 are essential for SAM-mediated conformational change. Within aa 357-376, Ala368 is in direct vicinity of SAM/SAH sulphonium centre, and its mutation to a bulkier residue blocked conformational change without affecting SAM/SAH binding. We therefore posit that SAM binding causes a change in the linker conformation (e.g. via Ala368), which in turn translates to a change in the catalytic domain, resulting in decreased enzyme activity.

Although our crystal structure represents a static snapshot of the enzyme state (likely a dis-inhibited state due to SAH binding), evidence for inter-domain conformational changes is provided by the following data. Firstly SAXS analysis between SAM-bound phosphorylated protein and SAH-bound dephosphorylated protein reveals inter-domain flexibility, consistent with subtle, but distinguishable changes to the protein dimensions. Secondly, the two chains in the crystal asymmetric unit show varying intrinsic order of the catalytic domain with respect to the regulatory domain. Additional genetic data from our lab are also in accord[20], as patient fibroblasts homozygous for p.His354Tyr, a linker residue which contacts helix α3 in catalytic domain, exhibited a 5-fold decrease in Ki for SAM.

So in what aspects could the SAM-bound signal influence the catalytic domain, seeing that its kinetic parameters remain largely unaltered? One possibility is an effect on the stability or integrity of FAD, the essential cofactor. We observed that supplementation with FAD enabled rescue of activity to our recombinant MTHFR. This is indicative of cofactor loss, in agreement with previous findings[13], and suggestive of FAD being only loosely bound, as exemplified in chain B of our structure. Furthermore a number of MTHFR mutations[20,22] and polymorphisms[13] are shown to affect FAD responsiveness. It is therefore possible that the inter-domain flexibility we observed, communicated by the SAM-bound signal, would alter the orientation of catalytic domains with respect to the rest of the protein, in a similar manner as the multi-domain enzymes CBS and PAH[35]. Such structural conformations are supported by overlays between chains A and B in our structure, and between apo- and holo-subunits in *Tt*MTHFR[8]. In the case of MTHFR, the active site could be more sequestered (leading to FAD bound) or more exposed and mobile (leading to FAD loss) as a consequence.

SAM-binding and phosphorylation act in concert as on/off and dimmer switches, respectively

The SAM/SAH ratio is regarded as an indicator of a cell's methylation potential and is a crucial indicator of the cells' capacity to perform DNA methylation or create compounds which require methyl groups for assimilation. In the face of a low SAM/SAH ratio, meaning methyl donor deficiency, MTHFR

is dis-inhibited, increasing the production of CH_3-THF to improve throughput of the methionine cycle and replenish SAM levels. Conversely, a high SAM/SAH ratio means abundant methylation capacity, in which case SAM mediated allosteric inhibition of MTHFR turns off CH_3-THF production, thereby lowering methionine cycle activity and concomitantly generation of SAM. This “on/off” switch is especially powerful at high SAM levels, as illustrated by almost complete inhibition of recombinant *Hs*MTHFR at > 200 µM SAM. Although these types of concentrations are unlikely to be seen inside the cell, *Hs*MTHFR has been further outfitted with a “dimmer” switch, whereby protein phosphorylation increases sensitivity to SAM-mediated inhibition at normal (1-10 µM) cellular SAM levels. In this regard, phosphorylation allows linkage of the methionine cycle to other cellular pathways (e.g. cell cycle) through specific kinase activities (as suggested by[16,17]).

The clear correlation we observed between phosphorylated MTHFR with SAM binding in solution (*vs* dephosphorylated MTHFR and SAH binding) leads us to interpret that the two regulatory properties act in concert. In fact, the architecture of the *Hs*MTHFR homodimer is smartly tailor-made to facilitate this correlation. (1) The dimeric interface is entirely constituted by the regulatory domain to form a scaffold, while leaving each SAM binding site on a different face for its sensing and signal transmission functions; (2) The lack of contacts between inter-monomeric catalytic domains allows for the intrinsic intra-monomeric mobility with respect to the regulatory domains for signal propagation; (3) Importantly, this dimeric configuration brings the N- and C-termini of the polypeptide in proximity, projecting the phosphorylation region close to the regulatory domain dimer interface.

Concluding Remarks

In summary, we provide the first structural view of a eukaryotic MTHFR, pointing to the linker region playing a direct role in allosteric inhibition following SAM binding, and phosphorylation as a means to modulate SAM inhibition sensitivity. Modulating such finite control towards the level of a key metabolite may be of pharmacological interest, including in cancer metabolism[36,37]. Our work here constitutes a strong starting point for future, more precise investigation by structural, biochemical, and cellular studies, for example towards: identification of the kinase(s) responsible for MTHFR phosphorylation *in vivo*; combining different structural methods to delineate conformational changes of the entire protein, and revealing the molecular basis of its specificity over NADPH.

Materials and Methods

Recombinant production of MTHFR

Numbering of the nucleotide changes follows the nomenclature of NM_005957.4, which places the A of the ATG initiation codon as the +1 nucleotide. The protein is numbered according to NP_005948.3. For *E. coli* (BL21(DE3)R3) expression, DNA fragments encoding human MTHFR (IMAGE: 6374885), mouse MTHFR (IMAGE: 6834886) and yeast MET12 (clone: ScCD00096551 from Harvard Medical School) harbouring different N- and C-terminal boundaries were amplified and subcloned in pNIC28-Bsa4 vector (accession number: EF198106) in-frame with a tobacco etch virus protease cleavable N-terminal His_6-tag. For baculovirus expression, DNA fragments encoding human MTHFR (IMAGE: 6374885) harbouring different N- and C-terminal boundaries were cloned into the pFB-CT10HF-LIC vector (Addgene plasmid: 39191) in-frame with a tobacco etch virus protease cleavable C-terminal flag/His_{10}-tag. Site-directed mutations were constructed using the QuikChange mutagenesis kit (Stratagene) and confirmed by sequencing. All primers are available upon request.
Proteins expressed in *E. coli* were purified by affinity (Ni-Sepharose; GE Healthcare) and size-exclusion (Superdex 200; GE Healthcare) chromatography. Proteins expressed in insect cells in Sf9 media (ThermoFisher) were purified by affinity (Ni-NTA, Qiagen) and size-exclusion (Superdex 200) chromatography, followed by cleavage of the C-terminal tag by His-tagged tobacco etch virus protease (1:20 mass ratio) overnight at 4°C and re-passage over Ni-NTA resin. Selenomethionine (SeMet)-derivatized proteins were expressed using SelenoMethionine Medium Complete (Molecular Dimensions) and purified as above.

Crystallization, Structural Determination and Analysis

Purified native *Sc*$MET12_{1-302}$ as well as SeMet-derivatized and native *Hs*$MTHFR_{38-644}$ were concentrated to 15-20 mg/ml, and crystals were grown by siting drop vapour diffusion at 20°C. The mother liquor conditions are summarized in Table 1. Crystals were cryo-protected in mother liquor containing ethylene glycol (25% v/v) and flash-cooled in liquid nitrogen. X-ray diffraction data were collected at the Diamond Light Source and processed using XIA2[38]. The *Hs*$MTHFR_{38-644}$ structure was solved by selenium multi-wavelength anomalous diffraction phasing using autoSHARP[39], and subjected to automated building with BUCCANEER[40]. The SeMet model was used to solve the native structure of *Hs*$MTHFR_{38-644}$ by molecular replacement using PHASER[41]. This structure was refined using PHENIX[42], followed by manual rebuilding in COOT[43]. Phases for *Sc*$MET12_{1-302}$ were calculated by molecular replacement using 3APY as model. Atomic coordinates and structure factors for both *Sc*$MET12_{1-302}$ (accession code: To Be Deposited) and *Hs*$MTHFR_{38-644}$ (accession code: 6CFX) have been deposited in the Protein Data Bank. Data collection and refinement statistics are summarized in Table 1.

MTHFR Assay

All enzymatic assays, including SAM inhibition and thermolability, were performed using the physiological forward assay described by Suormala et al.[18] with modifications as described by Rummel et al.[44] and Burda et al.[20,22]. Only minor adaptations were made for use with pure protein, including using the substrate CH_2-THF at a concentration of 75 µM, reducing the assay time to 7 minutes and the addition of BSA to keep purified proteins stable. Prior to assay, purified proteins were diluted from 15-20 mg/ml to 1 mg/ml in 10 mM HEPES-buffer pH 7.4, 5% glycerol and 500 mM NaCl followed by successive dilutions of 1:100 and 1:32 in 10 mM potassium-phosphate, pH 6.6 plus 5 mg/ml BSA, to a final MTHFR concentration of 312.5 ng/ml. All *K*m values were derived using a non-linear fit of Michaelis-Menten kinetics by GraphPad Prism (v6.07). For SAM inhibition, purified SAM[45] was used. The Ki was estimated following a plot of log(inhibitor) vs. response and a four parameter curve fit as performed by GraphPad Prism (v6.07).

Solution Analysis

Analytical gel filtration to assess changes in conformation was performed in the presence or absence of 250 µM SAH or SAM (both Sigma-Aldrich) as described previously[46]. Differential scanning fluorimetry in the presence of 0 - 250 µM SAM and SAH was performed as described previously[47,48].

Small angle X-ray scattering (SAXS)

SAXS experiments for the *Hs*MTHFR$_{38-644}$ and *Hs*MTHFR$_{1-656}$ (phosphorylated and dephosphorylated) were performed at 0.99 Å wavelength Diamond Light Source at beamline B21 coupled to the Shodex KW403-4F size exclusion column (Harwell, UK) and equipped with Pilatus 2M two-dimensional detector at 4.014 m distance from the sample, $0.005 < q < 0.4$ Å^{-1} ($q = 4\pi \sin\theta/\lambda$, 2θ is the scattering angle). The samples were in a buffer containing 20 mM Hepes-NaOH pH 7.5, 0.5 mM TCEP, 150 mM NaCl, 5 mM $MgCl_2$ and the measurements were performed at 20 □C. The data were processed and analyzed using the ATSAS program package[21]. The radius of gyration Rg and forward scattering I(0) were calculated by Guinier approximation. The maximum particle dimension Dmax and P(r) function were evaluated using the program GNOM[49]. To demonstrate the absence of concentration dependent aggregation and interparticle interference in the both SAXS experiments, we inspected Rg over the elution peaks and performed our analysis only on a selection of frames in which Rg was most stable. Overall, such stability of Rg over the range of concentrations observed in the SEC elution indicates that there were no concentration-dependent effects or interparticle interference. The *ab initio* model was derived using DAMMIF[50]. 10 individual models were created, then overlaid and averaged using DAMAVER[51]. CORAL rigid body modeling was performed by defining residues 338-345 as a flexible linker and allowing the catalytic subunits to move while keeping the regulatory subunits fixed.

Mass spectrometry

Native (intact) mass spectrometry was performed as outlined[52]. Denaturing mass spectrometry and phosphorylation mapping were performed as described[53]. For further details see Supplementary Methods.

References

1. Watkins, D. & Rosenblatt, D.S. Inherited Disorders of Folate and Cobalamin Transport and Metabolism. in *The Online Metabolic and Molecular Bases of Inherited Disease* (eds. Valle, D. et al.) (McGraw-Hill Medical, 2017).
2. Froese, D.S. et al. Mutation Update and Review of Severe Methylenetetrahydrofolate Reductase Deficiency. *Hum Mutat* **37**, 427-38 (2016).
3. Liew, S.C. & Gupta, E.D. Methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism: epidemiology, metabolism and the associated diseases. *Eur J Med Genet* **58**, 1-10 (2015).
4. Guenther, B.D. et al. The structure and properties of methylenetetrahydrofolate reductase from Escherichia coli suggest how folate ameliorates human hyperhomocysteinemia. *Nat Struct Biol* **6**, 359-65 (1999).
5. Lee, M.N. et al. Functional role for the conformationally mobile phenylalanine 223 in the reaction of methylenetetrahydrofolate reductase from Escherichia coli. *Biochemistry* **48**, 7673-85 (2009).
6. Pejchal, R. et al. Structural perturbations in the Ala --> Val polymorphism of methylenetetrahydrofolate reductase: how binding of folates may protect against inactivation. *Biochemistry* **45**, 4808-18 (2006).
7. Pejchal, R., Sargeant, R. & Ludwig, M.L. Structures of NADH and CH3-H4folate complexes of Escherichia coli methylenetetrahydrofolate reductase reveal a spartan strategy for a ping-pong reaction. *Biochemistry* **44**, 11447-57 (2005).
8. Igari, S. et al. Properties and crystal structure of methylenetetrahydrofolate reductase from Thermus thermophilus HB8. *PLoS One* **6**, e23716 (2011).
9. Matthews, R.G., Vanoni, M.A., Hainfeld, J.F. & Wall, J. Methylenetetrahydrofolate reductase. Evidence for spatially distinct subunit domains obtained by scanning transmission electron microscopy and limited proteolysis. *J Biol Chem* **259**, 11647-50 (1984).
10. Sumner, J., Jencks, D.A., Khani, S. & Matthews, R.G. Photoaffinity labeling of methylenetetrahydrofolate reductase with 8-azido-S-adenosylmethionine. *J Biol Chem* **261**, 7697-700 (1986).
11. Kutzbach, C. & Stokstad, E.L. Mammalian methylenetetrahydrofolate reductase. Partial purification, properties, and inhibition by S-adenosylmethionine. *Biochim Biophys Acta* **250**, 459-77 (1971).
12. Daubner, S.C. & Matthews, R.G. Purification and properties of methylenetetrahydrofolate reductase from pig liver. *J Biol Chem* **257**, 140-5 (1982).
13. Yamada, K., Chen, Z., Rozen, R. & Matthews, R.G. Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. *Proc Natl Acad Sci U S A* **98**, 14853-8 (2001).
14. Yamada, K., Strahler, J.R., Andrews, P.C. & Matthews, R.G. Regulation of human methylenetetrahydrofolate reductase by phosphorylation. *Proc Natl Acad Sci U S A* **102**, 10454-9 (2005).
15. Marini, N.J. et al. The prevalence of folate-remedial MTHFR enzyme variants in humans. *Proc Natl Acad Sci U S A* **105**, 8055-60 (2008).
16. Zhu, B., Xiahou, Z., Zhao, H., Peng, B. & Xu, X. MTHFR promotes heterochromatin maintenance. *Biochem Biophys Res Commun* **447**, 702-6 (2014).
17. Li, X. et al. Polo-like kinase 1 (PLK1)-dependent phosphorylation of methylenetetrahydrofolate reductase (MTHFR) regulates replication via histone methylation. *Cell Cycle* **16**, 1933-1942 (2017).
18. Suormala, T., Gamse, G. & Fowler, B. 5,10-Methylenetetrahydrofolate reductase (MTHFR) assay in the forward direction: residual activity in MTHFR deficiency. *Clin Chem* **48**, 835-43 (2002).
19. Matthews, R.G. Methylenetetrahydrofolate reductase from pig liver. *Methods Enzymol* **122**, 372-81 (1986).
20. Burda, P. et al. Insights into severe 5,10-methylenetetrahydrofolate reductase deficiency: molecular genetic and enzymatic characterization of 76 patients. *Hum Mutat* **36**, 611-21 (2015).
21. Petoukhov, M.V. et al. New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* **45**, 342-350 (2012).
22. Burda, P. et al. Functional characterization of missense mutations in severe methylenetetrahydrofolate reductase deficiency using a human expression system. *J Inherit Metab Dis* **40**, 297-306 (2017).

23. Sheppard, C.A., Trimmer, E.E. & Matthews, R.G. Purification and properties of NADH-dependent 5, 10-methylenetetrahydrofolate reductase (MetF) from Escherichia coli. *J Bacteriol* **181**, 718-25 (1999).
24. van den Heuvel, R.H. et al. Structural studies on flavin reductase PheA2 reveal binding of NAD in an unusual folded conformation and support novel mechanism of action. *J Biol Chem* **279**, 12860-7 (2004).
25. Matthews, R.G. & Kaufman, S. Characterization of the dihydropterin reductase activity of pig liver methylenetetrahydrofolate reductase. *J Biol Chem* **255**, 6014-7 (1980).
26. Bhattacharyya, S., Dutta, A., Dutta, D., Ghosh, A.K. & Das, A.K. Structural elucidation of the NADP(H) phosphatase activity of staphylococcal dual-specific IMPase/NADP(H) phosphatase. *Acta Crystallogr D Struct Biol* **72**, 281-90 (2016).
27. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529-33 (2010).
28. Wang, J.X., Lee, E.R., Morales, D.R., Lim, J. & Breaker, R.R. Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol Cell* **29**, 691-702 (2008).
29. Zheng, W. et al. Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. *J Am Chem Soc* **134**, 18004-14 (2012).
30. Wise, C.K., Cooney, C.A., Ali, S.F. & Poirier, L.A. Measuring S-adenosylmethionine in whole blood, red blood cells and cultured cells using a fast preparation method and high-performance liquid chromatography. *J Chromatogr B Biomed Sci Appl* **696**, 145-52 (1997).
31. Loehrer, F.M. et al. Disturbed ratio of erythrocyte and plasma S-adenosylmethionine/S-adenosylhomocysteine in peripheral arterial occlusive disease. *Atherosclerosis* **154**, 147-54 (2001).
32. Gu, X. et al. SAMTOR is an S-adenosylmethionine sensor for the mTORC1 pathway. *Science* **358**, 813-818 (2017).
33. McCorvie, T.J. et al. Inter-domain communication of human cystathionine beta-synthase: structural basis of S-adenosyl-L-methionine activation. *J Biol Chem* **289**, 36018-30 (2014).
34. Patel, D., Kopec, J., Fitzpatrick, F., McCorvie, T.J. & Yue, W.W. Structural basis for ligand-dependent dimerization of phenylalanine hydroxylase regulatory domain. *Sci Rep* **6**, 23748 (2016).
35. Jaffe, E.K. Impact of quaternary structure dynamics on allosteric drug discovery. *Curr Top Med Chem* **13**, 55-63 (2013).
36. Sun, D.F. et al. Knock-down of methylenetetrahydrofolate reductase reduces gastric cancer cell survival: an in vitro study. *Cell Biol Int* **32**, 879-87 (2008).
37. Stankova, J., Lawrance, A.K. & Rozen, R. Methylenetetrahydrofolate reductase (MTHFR): a novel target for cancer therapy. *Curr Pharm Des* **14**, 1143-50 (2008).
38. Winter, G., Lobley, C.M. & Prince, S.M. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* **69**, 1260-73 (2013).
39. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol Biol* **364**, 215-30 (2007).
40. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-11 (2006).
41. McCoy, A.J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**, 32-41 (2007).
42. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
43. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
44. Rummel, T. et al. Intermediate hyperhomocysteinaemia and compound heterozygosity for the common variant c.677C>T and a MTHFR gene mutation. *J Inherit Metab Dis* **30**, 401 (2007).
45. Shapiro, S.K. & Ehninger, D.J. Methods for the analysis and preparation of adenosylmethionine and adenosylhomocysteine. *Anal Biochem* **15**, 323-33 (1966).
46. Froese, D.S. et al. Structure of MMACHC reveals an arginine-rich pocket and a domain-swapped dimer for its B12 processing function. *Biochemistry* **51**, 5083-90 (2012).
47. Niesen, F.H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-21 (2007).
48. Froese, D.S. et al. Thermolability of mutant MMACHC protein in the vitamin B12-responsive cblC disorder. *Mol Genet Metab* **100**, 29-36 (2010).

49. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J App Cryst* **25**, 495-503 (1992).
50. Franke, D. & Svergun, D.I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342-346 (2009).
51. Volkov, V.V. & Svergun, D.I. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860-864 (2003).
52. Chalk, R. et al. High-Throughput Mass Spectrometry Applied to Structural Genomics. *Chromatography* **1**, 159 (2014).
53. Chalk, R. Mass Spectrometric Analysis of Proteins. in *Heterologous Gene Expression in E. coli: Methods and Protocols* (ed. Burgess-Brown, N.A.) 373-396 (Humana Press, New York, 2017).
54. Svergun, D., Barberato, C. & Koch, M.H.J. CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **28**, 768-773 (1995).

Acknowledgements

This work was supported by the Olga-Mayenfisch Stiftung (to D.S.F.), the Rare Disease Initiative Zurich, a Clinical Research Priority Program from the University of Zurich (to D. S. F. and M. R. B.) and the Swiss National Science Foundation (SNSF 31003A_156907) (to M.R.B.). The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, MSD, Merck KGaA, Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda and Wellcome Trust [106169/ZZ14/Z]. We thank the Diamond Light Source for access to the beamlines and Merima Forny for sequencing.

Author Contributions

D.S.F., M.R.B. and W.W.Y. conceived of the study. D.S.F., J.K. and E.R. cloned, purified and crystallized MTHFR constructs. They also performed SAM binding and de-phosphorylation assays. J.K. and A.O. performed x-ray structure determination. G.A.B. performed and analysed the SAXS measurement. T.S. and S.L. performed and analyzed the activity assays. R.C. and O.B. performed and analyzed the mass spectrometry assays. D.S.F. and W.W.Y. wrote the manuscript with editing and proofreading from the other co-authors.

Table 1. Kinetic characterization of *Hs*MTHFR.

	Specific activity (µmol • min^{-1} • mg^{-1})			Heat-stable activity (%)			Apparent K_M values (µM)			Inhibition (µM)
Protein	ØFAD	+FAD	K_{cat} (sec^{-1})	ØFAD	assay +FAD	pre-FAD	CH2-THF	NADPH	NADH	SAM K_i (95% CI)
*Hs*MTHFR$_{1-656}$	30.5 ± 0.9	30.8 ± 1.5	40.7 ± 3.2	33.9 ± 3.1	39.8 ± 1.4	81.0 ± 3.8	22.4 ± 1.3	35.5 ± 2.4	3760 ± 410	2.7 (2.2 – 3.5)
*Hs*MTHFR$_{38-644}$	34.8 ± 1.2	34.0 ± 1.3	51.4 ± 4.2	40.7 ± 3.9	83.7 ± 2.8	101.8 ± 3.2	25.5 ± 1.7	23.5 ± 1.7	2160 ± 270	21 (19 – 23)
*Hs*MTHFR$_{1-656}$ (mock)	33.6 ± 3.1	32.9 ± 3.7	40.4 ± 4.0	32.6 ± 2.0	40.0 ± 0.8	82.2 ± 2.8	22.5 ± 3.3	30.3 ± 4.0	N.D.	3.8 (3.4 – 4.3)
*Hs*MTHFR$_{1-656}$ (CIP)	33.2 ± 1.7	34.4 ± 2.5	39.9 ± 4.2	36.7 ± 1.6	43.2 ± 1.6	81.6 ± 2.7	21.3 ± 1.8	28.7 ± 2.3	N.D.	6.4 (6.1 – 6.7)

ØFAD: FAD was not supplemented; +FAD: FAD was added to the assay buffer; pre-FAD: FAD was added to the protein before heating

N.D. = not determined

All values represent the results of at least 3 separate experiments

Table 2. Data collection and refinement statistics.

Dataset	*Hs*MTHFR$_{38-644}$ Native	*Sc*MET12$_{1-301}$
Crystallization condition	0.1M Na citrate tribasic, 22.5% PEG4K, 5% 2-propanol	0.2M Na/K tartrate, 20% PEG3350
Data collection and processing		
Beamline	Diamond I04-1	Diamond I04-1
Wavelength (Å)	0.9281	0.9282
Unit cell parameters (Å)	97.3 127.9 147.0	110.6 54.5 61.9
(°)	90.00 90.00 90.00	90.00 90.00 90.00
Space group	$P2_12_12_1$	$P2_12_12$
Resolution range (Å)	68.53-2.50 (2.64-2.50)	22.75-1.56 (1.60-1.56)
Observed/Unique reflections	64992/9272	448825/53602
Rsym (%)	8.6 (136.2)	9.2 (149.8)
Rpim (%)	3.4 (55.6)	3.3 (53.4)
CC1/2 (%)	99.8 (58.5)	99.9 (52.3)
I/sig(I)	12.8 (1.2)	14.3 (1.3)
Completeness (%)	100 (100)	99.5 (99.1)
Multiplicity	7.5 (7.0)	8.4 (8.6)
Refinement		
R_{cryst} (%)	19.74	15.97
R_{free} (%)	24.40	19.93
Wilson *B* factor (Å^2)	64.73	
Average total *B* factor (Å^2)	88.57	
Protein	88.60	
Ligand/Ion	115.18	
Water	73.35	
R.m.s.d. bond length (Å)	0.008	0.012
R.m.s.d. bond angle (°)	0.946	1.438
Model		
Built residues	1182	292
PDB code	6CFX	To be deposited

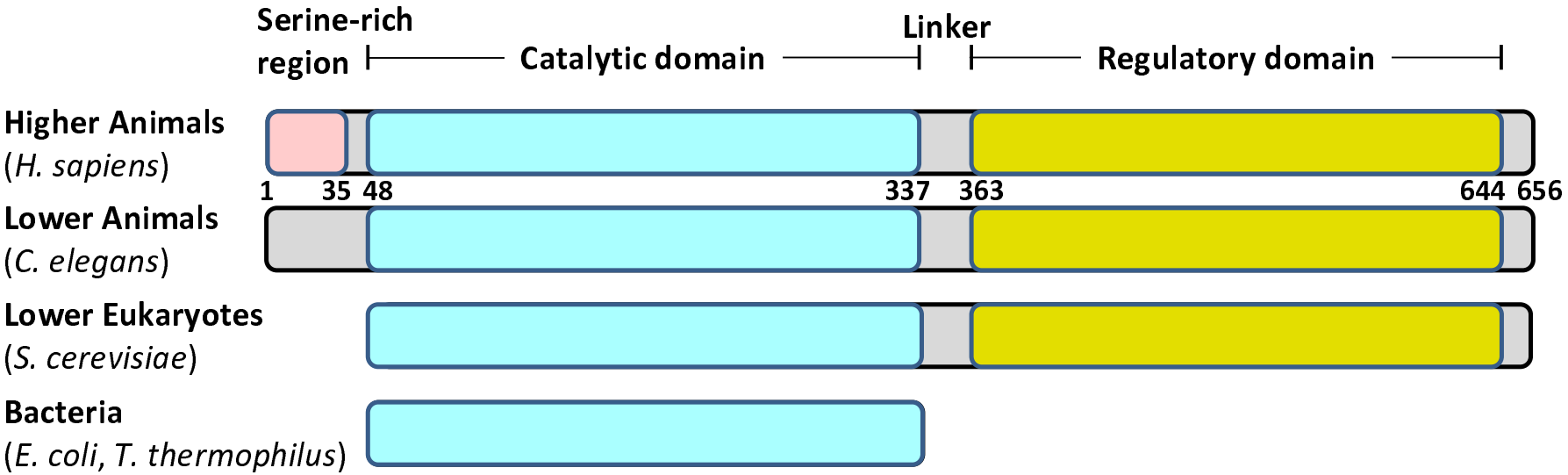

Figure 1. Schematic representation of MTHFR.

Domain organization of MTHFR orthologs across evolution. Numbers given represent approximate amino acid boundaries in human MTHFR corresponding to NP_005948. In brackets is shown representative species within each category.

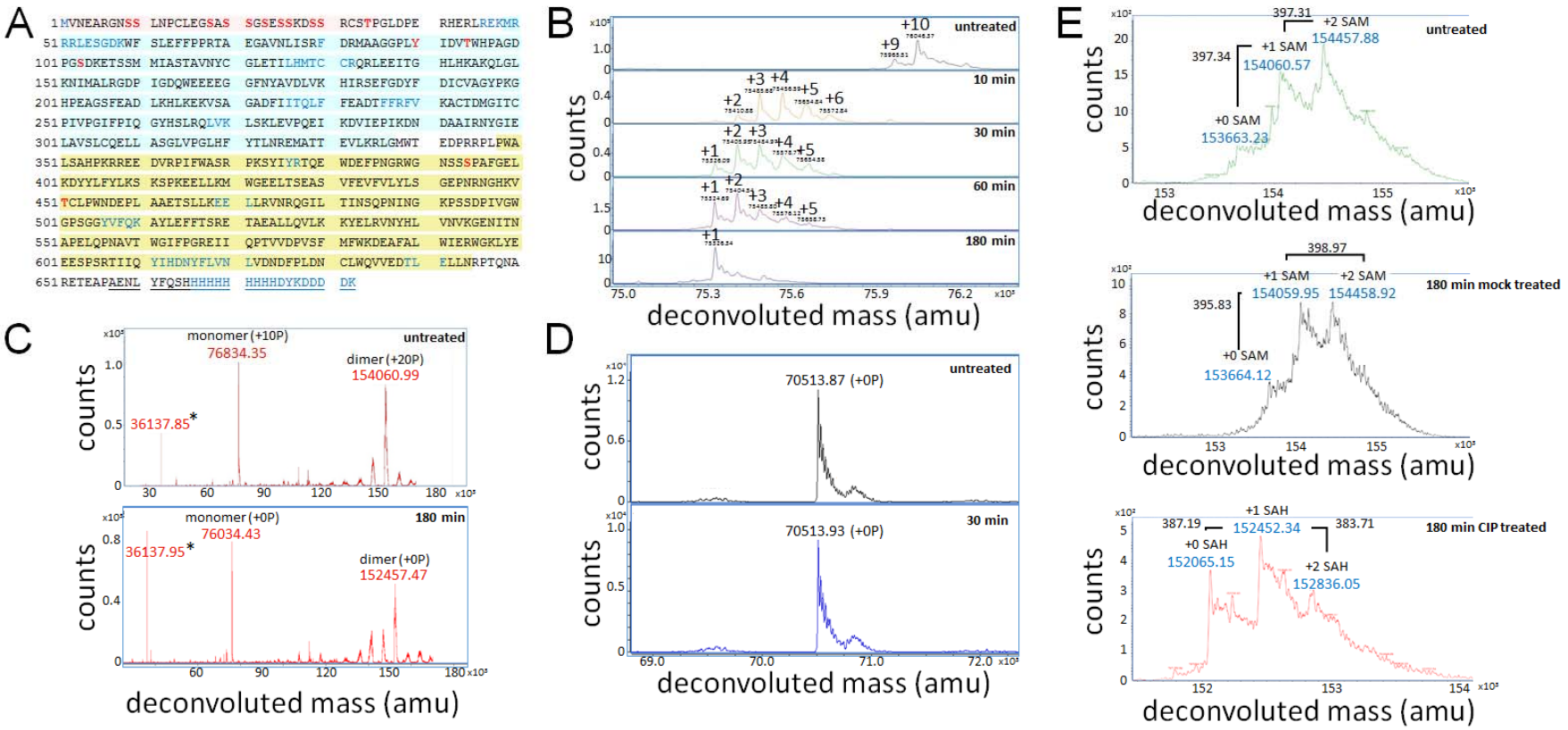

Figure 2. Phosphorylation status of *Hs*MTHFR$_{1-656}$ and *Hs*MTHFR$_{38-644}$.

A. Phosphorylation mapping of *Hs*MTHFR$_{1-656}$. The protein sequence is given as amino acids in single letter code, including the C-terminal His/flag-tag (underlined). Black font represents amino acids identified by the mass spectrometer (covered), blue font represents amino acids not identified (non-covered), red font represents phosphorylated amino acids. Domains are coloured as in Figure 1.

B. Dephosphorylation of *Hs*MTHFR$_{1-656}$ following treatment with CIP. Treatment time at 37°C is given. Large number above peaks represents number of phosphate groups attached, small number represents atomic mass. Proteins were analyzed by denaturing mass spectrometry. amu: atomic mass units.

C. Native mass spectrometry analysis of *Hs*MTHFR$_{1-656}$ before and after treatment with CIP. Upper panel: as purified (untreated) protein. Monomer represents protein bound to 1 FAD plus 10 phosphate groups (expected mass: 76831.16 amu); dimer represents protein bound to 2 FADs and 1 SAM

plus 20 phosphate groups (expected mass: 154060.74 amu). Lower panel: protein following 180 min treatment with CIP. Monomer represents protein bound to 1 FAD (expected mass: 76031.16 amu); dimer represents protein bound to 2 FADs and 1 SAH (expected mass: 152446.74 amu). Expected sizes: protein without first methionine, 75245.6 amu; FAD, 785.56 amu; SAM, 398.44 amu; SAH, 384.42 amu, phosphate, 80.00 amu. * indicates a truncated protein representing amino acids 353-663 (expected mass: 36136.6 amu).

D. *Hs*MTHFR$_{38\text{-}644}$ before and after treatment with CIP. Treatment time at 37°C is given. Protein was analyzed by denaturing mass spectrometry.

E. Native mass spectrometry of *Hs*MTHFR$_{1\text{-}656}$ identifying sequential binding of SAM or SAH. Graphs represent areas zoomed in on dimeric protein. Upper panel: *As purified* (untreated) protein. Middle panel: control protein (heated in assay buffer without CIP). Bottom panel: treated protein (heated in assay buffer with CIP). Expected size of protein with 2 FAD-bound and 20 phosphates: 153662.3. Expected size of protein with 2 FAD-bound and 0 phosphates: 152062.32. Expected size of SAM: 398.44, SAH: 384.41.

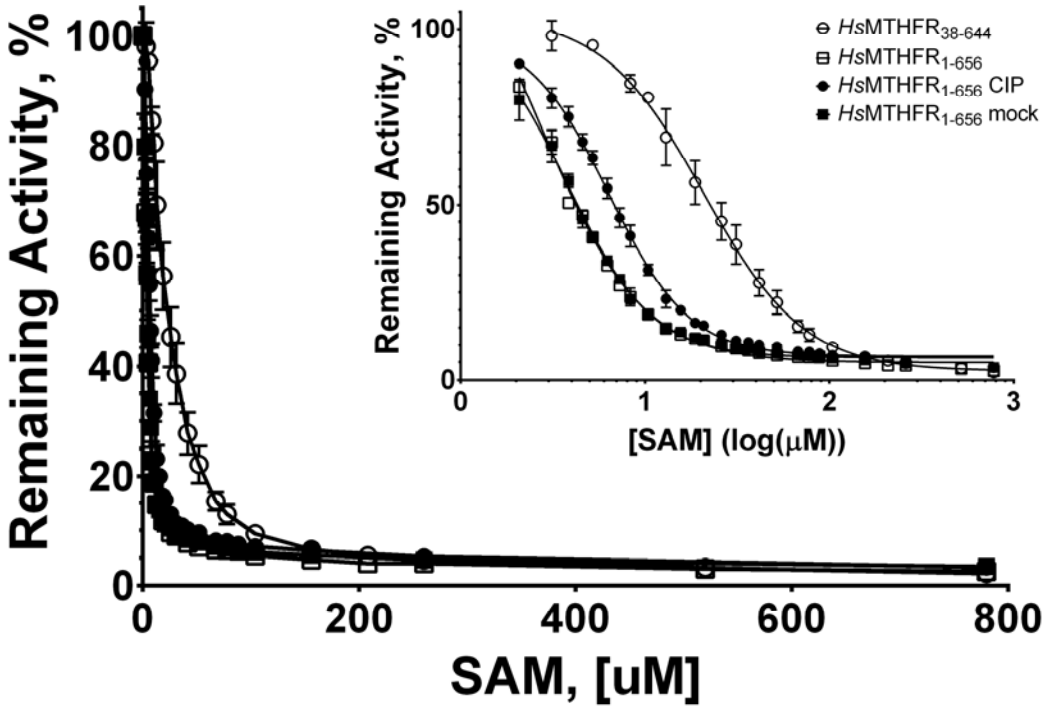

Figure 3. SAM inhibition of *Hs*MTHFR$_{1-656}$ and *Hs*MTHFR$_{38-644}$.

Inhibition of MTHFR catalytic activity following pre-incubation with various concentrations of SAM. Remaining activity represents percentage of activity compared to MTHFR incubated without SAM. **Inset:** Replot of percent activity remaining against SAM concentration transformed by log10 to reveal differences between truncated (*Hs*MTHFR$_{38-644}$) and dephosphorylated full-length (*Hs*MTHFR$_{1-656}$ CIP) protein with phosphorylated full-length protein (*Hs*MTHFR$_{1-656}$; *Hs*MTHFR$_{1-656}$ mock). Inhibitory constants (K_i's) for SAM calculated from this graph were calculated as described in the Materials and Methods and are provided in Table 1.

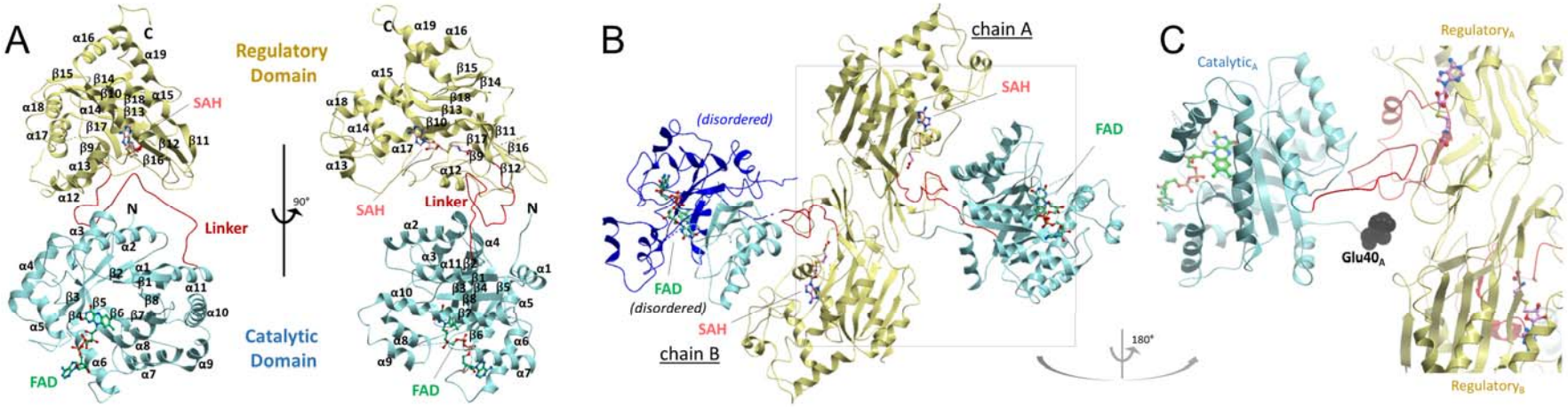

Figure 4. Structural overview of *Hs*MTHFR.

A. Orthogonal views of *Hs*MTHFR showing the catalytic domain (cyan), the linker (red) and the regulatory domain (yellow). Bound FAD (green) and SAM (pink) are shown in sticks. Dotted lines indicate disordered regions that are not modeled in the structure. α-helices and β-sheets are labeled, and correspond to the multiple sequence alignment in Suppl. Fig. 3 and the topology in Suppl. Fig. 4.

B. Homodimer of *Hs*MTHFR as seen in the crystal. Chain A represents the more ordered protomer. In chain B, the part of the catalytic domain that was poorly ordered is represented in dark blue and only the main chain was modeled. The FAD in this subunit was also partially disordered.

C. Juxtaposition of the N-terminus of *Hs*MTHFR protomer towards both regulatory domains of the homodimer. The first residue observed in the structure, Glu40 of chain A, is shown in black spheres. Other coloured features are as described for panel A.

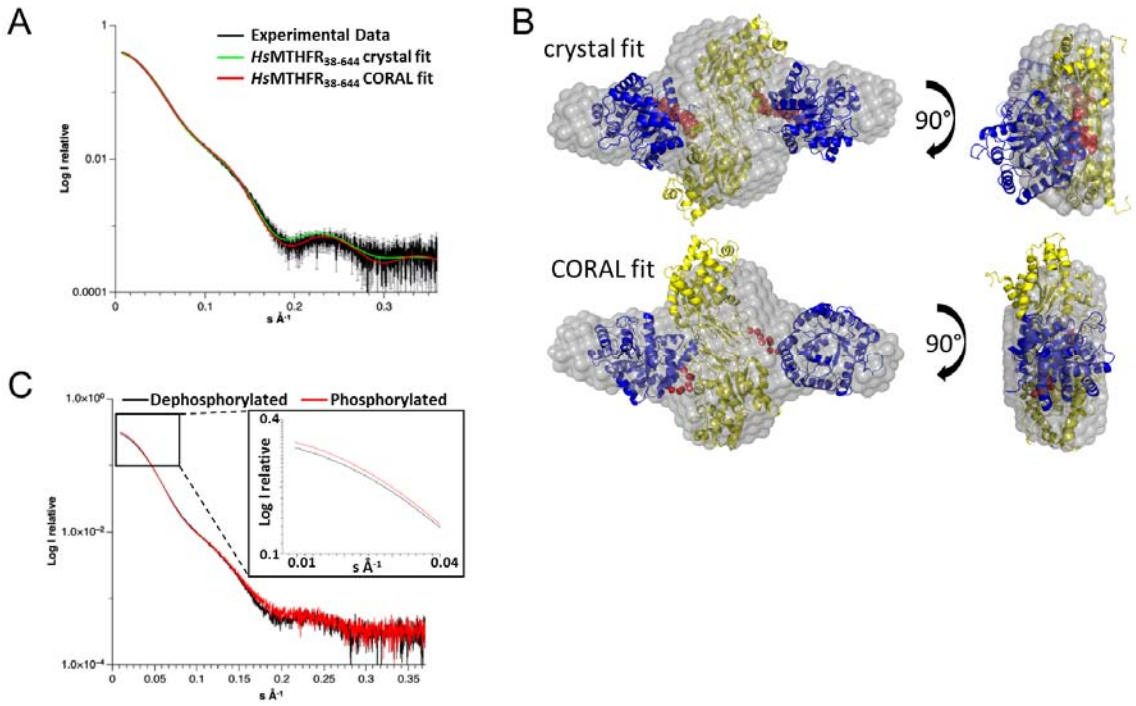

Fig 5. SAXS analysis of *Hs*MTHFR$_{38-644}$ and *Hs*MTHFR$_{1-656}$ phosphorylated and dephosphorylated.

A. SAXS analysis of *Hs*MTHFR$_{38-644}$. Experimental scattering profile is shown in black, theoretical scattering curve of the *Hs*MTHFR$_{38-644}$ dimer observed in the crystal is in green and that of the rigid body modeling by CORAL is in red. Chi^2 was determined by CRYSOL[54].

B. Superimposition of the *ab initio* model envelope (gray surface) with the *Hs*MTHFR$_{38-644}$ crystal structure (upper panels) and the rigid body model from CORAL (lower panels). For each, the catalytic subunit is shown in blue, the regulatory subunit is shown in yellow and the linker region in red.

C. Main panel: Superimposition of experimental scattering profiles from *Hs*MTHFR phosphorylated (red) and *Hs*MTHFR dephosphorylated (black). Inset: Zoom in the low s-values region.

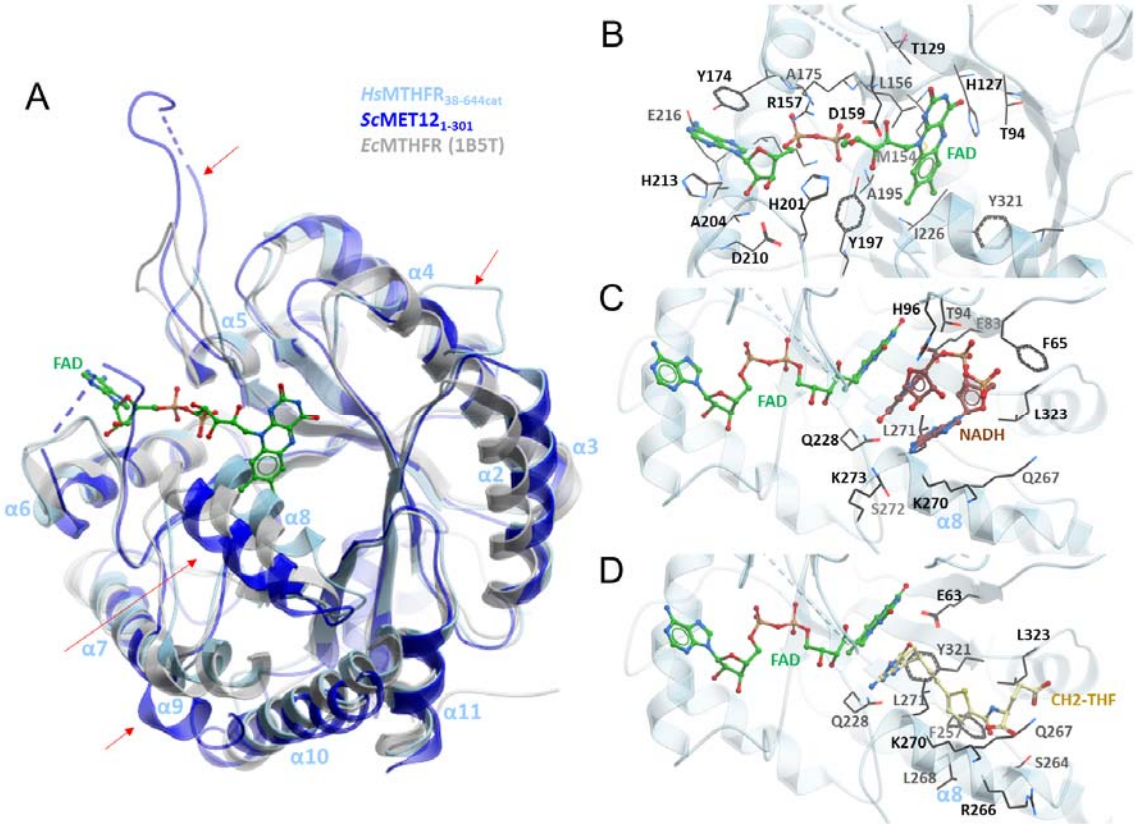

Figure 6. Structural examination of the *Hs*MTHFR$_{38-644}$ catalytic domain.

A. Structural alignment of *Hs*MTHFR$_{38-644}$ (cyan) with *Ec*MTHFR (grey) and ScMET12$_{1-301}$. Four sites of important differences are indicated by arrows (a-d). α-helices of *Hs*MTHFR$_{38-644}$ are indicated for orientation.

B. Binding pocket of FAD. FAD is shown in green sticks, residues contributing to FAD binding are labelled and shown in black sticks, grey cloud represents the FAD binding pocket.

C. Binding pocket of NAD(P)H. NADH is taken from an overlay of *Ec*MTHFR (PDB: IZRQ) with *Hs*MTHFR$_{38-644}$ but for clarity *Ec*MTHFR is not shown. FAD is shown in green sticks, NADH in brown sticks, residues expected to contribute to NADH binding are labelled and shown in black sticks, grey cloud represents the FAD/NADH binding pocket.

D. Binding pocket of CH_3-THF. CH_3-THF is taken from an overlay of *Ec*MTHFR (PDB: 2FMN) with *Hs*MTHFR$_{38-644}$ but for clarity *Ec*MTHFR is not shown. FAD is shown in green sticks, CH_3-THF in yellow sticks, residues expected to contribute to CH_3-THF binding are labelled and shown in black sticks, grey cloud represents the FAD/ CH_3-THF binding pocket.

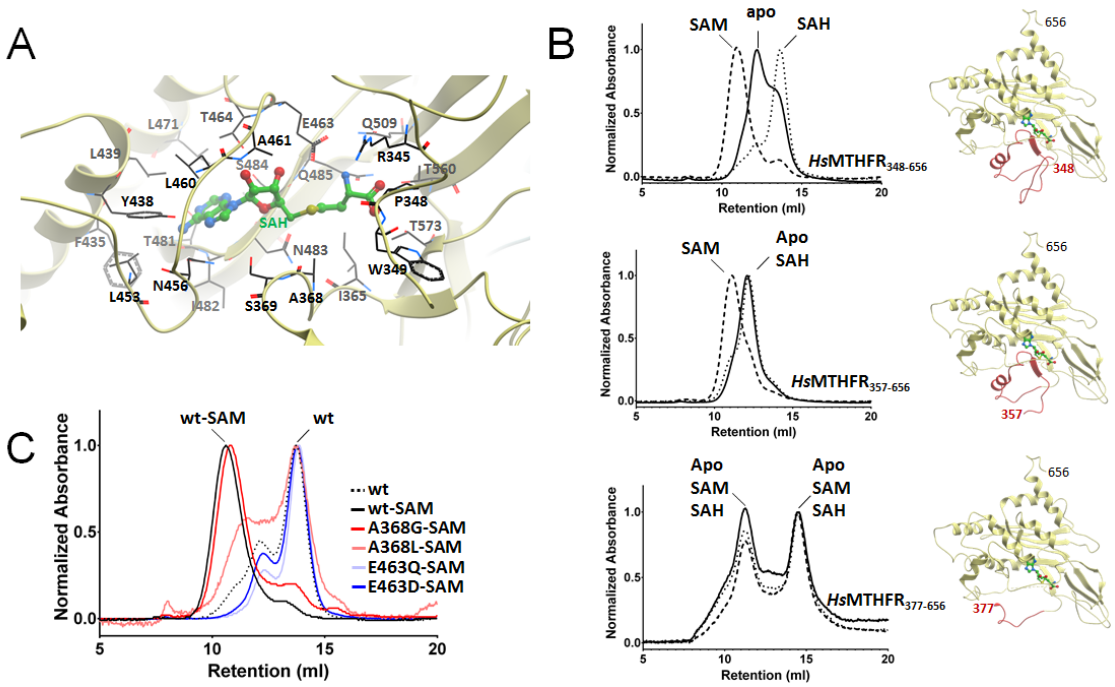

Figure 7. SAH/SAM binding and conformational change.

A. The SAH binding site. Amino acids that contribute to binding are labeled and shown in black sticks. SAH is shown in green sticks.

B. Size exclusion chromatography of *Hs*MTHFR with various N-terminal truncations following incubation with SAM (dashed lines), SAH (dotted lines), or buffer (apo; solid line). For each N-terminally truncated construct, the corresponding structure is shown.

C. Size exclusion chromatography of *Hs*MTHFR$_{348-656}$ proteins without (wt) or with (wt-SAM) pre-incubation with SAM. Mutated *Hs*MTHFR$_{348-656}$ proteins were pre-incubation with SAM.