The methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase (MeTr) enzyme was originally isolated from the acetogenic bacteria Moorella (formerly Clostridium) thermoacetica [612]. This bacterium, as some other anaerobic organisms, can obtain its entire carbon and energy source from CO or H₂/CO₂ by the Wood-Ljungdahl or acetyl-CoA pathway [613-615]. Initially, CO₂ is reduced to formate and converted to N⁵-methyltetrahydrofolate (CH₃THF) by a formate dehydrogenase and a series of four different tetrahydrofolate (THF)-dependent enzymes. Then, MeTr catalyses the transfer of the methyl group of CH₃THF to the cobalt centre of a corrinoid/iron-sulfur protein (C/Fe-SP). Afterwards, the methyl group is transferred from cob(III)amide to CO dehydrogenase/acetyl-CoA synthase (CODH/ACS) to form a methylnickel intermediate [616], which finally combines with a Fe-CO species [617,618] and CoA to form acetyl-CoA.

The physiological acceptor of the methyl group of CH₃THF is C/Fe-SP, an 88 kDa heterodimeric protein with 55 and 33 kDa subunits [619,620]. The small subunit binds 5’-methoxybenzimidazolylcobamide, and the large subunit contains a [4Fe-4S]²⁺/¹⁺ cluster [621] that is involved in reductive activation of the protein to maintain the cobalt centre in the cob(I)amide state [622].

MeTr is different from other enzymes of the Wood-Ljungdahl pathway as it lacks metals or other prosthetic groups, and is one of the few oxygen-stable proteins [623]. The gene encoding MeTr has been cloned, sequenced and actively expressed in E. coli [624]. The active enzyme, a homodimeric structure with a subunit molecular mass of 28 kDa [623], has been crystallized both as the apo protein [625] and as a MeTr•CH₃THF complex (Doukov, Ragsdale et al., unpublished results 2003).
Figure 4-1 Crystal structure of MeTr.
Cartoon representation of the crystal structure of MeTr. Helixes are shown in red, loops in green, and β sheets in yellow. The enzyme crystallizes as a homodimeric complex (A) (1FY6), where the TIM-barrel structures adopt an antiparallel orientation. The active site is located at the top of the barrel, as observed in (B), where one of the units of the crystal structure of MeTr is shown complexed with N5-methyltetrahydrofolate (CH$_3$THF) (Doukov, Ragsdale et al., Unpublished results 2003).
4.1.1 Structure

The crystal structure of the *apo* MeTr [625] is publicly available (PDB code 1F6Y), and a MeTr•CH$_3$THF complex has been recently determined (Doukov, Ragsdale *et al*., Unpublished results 2003) (Figure 4-1B). Dr Doukov and Prof Ragsdale (University of Nebraska-Lincoln) provided to Prof Gready in confidence the coordinates of two possible models of the MeTr•CH$_3$THF complex, obtained from the refinement of the same X-ray diffraction data by two different programs (CNS [626] and SHELX [627]). When the crystal structures of the *apo* and the substrate-bound enzyme are compared, it is clear that no major rearrangements are required to accommodate CH$_3$THF within the active site.

MeTr crystallizes as a homodimer, and each monomer presents a TIM α-β barrel structure (Figure 4-1A). The main structure of the barrel is made up of eight β strands surrounded by eight major α helices. The CH$_3$THF binding site is located at the top of the TIM barrel (Figure 4-1B), in a polar cup-like region that is relatively solvent exposed and with several acidic residues. The pterin ring is partially buried within the TIM barrel, with most of the residues involved in pterin binding conserved among other methyl transferases and dihydropteroate synthases (DHPS). These include Asp and Asn residues (Asp43, Asp75, Asn96 and Asp160), which form H-bond interactions with the nitrogen and oxygen atoms of the pterin ring (Figure 4-2).

The conserved residue Asp160 constitutes what has been called the ‘pterin hook’ [625], commonly present in dihydrofolate reductases, and which usually establishes H-bond interactions with the N2, N3 and O4 atoms of the pterin ring, either directly or through a water molecule. The H-bond network at the bottom of the pterin ring, formed by D43, N45, D75 and N96, is highly conserved among methyltransferases and DHPS families [625]. This array of polar residues can be expected to be involved not only in CH$_3$THF binding, but also in the polarization of the pterin ring, modifying its reactivity and facilitating its protonation during the methyl transfer reaction. The crystal structure of the complex MeTr•CH$_3$THF shows two crystal water molecules within H-bond distance from the pterin ring. My observations suggest that both waters appear to
be interacting with the O4 atom of the pterin ring, with one of them in the same position as in the apo structure, within H-bond distance to the carboxylic group of D160 and the N atom of the backbone of Ile163 and Gly196. While one face of the pterin ring is exposed to the solvent environment, from where the cob(I)alamin cofactor is expected to approach, the other face is oriented towards the inside of the protein, which is lined by a series of non-polar residues (I120, V194, G196 and I227), most of them conserved among methyltransferases and DHPSs. The glutamic tail of CH₃THF extends towards the opening of the barrel, and appears to be interacting with R207 and N202.

Figure 4-2 Disposition of the ligand methyftetrahydrofolate (CH₃THF) within the active site of MeTr.

Selected active-site residues (Asp 43, Asn 45, Asp 75, Asn 96, Asp 160, Ile 163, Gly 196, Asn 199, Gln 202 and Arg 207) that appear to be forming hydrogen-bond interactions (shown as dotted lines) with the ligand, and the ligand itself are represented in tubular form. Two crystallographic water molecules interacting with the pterin ring are shown as red spheres. Only the side-chains of the residues are shown, except for Ile 163 and Gly 196, which appear interacting with water molecules through their backbone atoms. Putative H-bond interactions are shown as dotted lines. Most of the residues involved in binding of the pterin ring (Asp43, Asp75, Asn96 and Asp160) are conserved among other methyl transferases and dihydropteroate synthases.
There is a strong structural homology between MeTr and dihydropteroate synthases (DHPS) from *Staphylococcus aureus* (1AD4) and *E. coli* (1AJZ), and a remarkable conservation of the active-site residues that interact with the pterin ring [625]. However, while in the DHPS enzymes there is a potential proton donor for the activation of the pterin ring (Lys203 in *S. aureus* DHPS and Lys221 in *E. coli*), in MeTr there is no obvious H-bond donor to the N5 atom of CH$_3$THF as the position equivalent to Lys203 in DHPS is occupied by Asn199 in MeTr. This lack of a clear proton donor has important implications for the reaction mechanism (see below). Another structural difference between MeTr and DHPS are the loop structures closing the active site. While in DHPS two loops partially block the access to the active site, in MeTr much shorter loops leave this region solvent exposed. This is likely a necessary feature to allow access of the bulky cob(I)alamin cofactor to the methyl group of CH$_3$THF.

A model for the interaction between cobalamin and the binary complex MeTr•CH$_3$THF [625] suggests that the Co-cofactor (cobalamin or analogue) closes the cup formed by the $\beta$ strands on the TIM barrel (Figure 4-1). The formation of this cage around the methyl group has been suggested to reduce the entropy of the reaction by constraining the mobility of the bound reactants and limiting solvent accessibility to the active site. Interestingly, this binding arrangement does not seem likely to support the random bi-bi binding mechanism suggested for the system [628], as an initial binding of the cobalamin group is expected to block the active site, preventing further binding of the substrate CH$_3$THF.

### 4.1.2 Reaction

MeTr catalyses the reversible transfer of the N5-methyl group of CH$_3$THF to the cob(I)alamin centre of a corrinoid iron-sulphur protein (C/Fe-SP) [629], in a reaction similar to the first half-reaction of methionine synthase [630]. It has been proposed that both CH$_3$THF ($K_m = 2.0\pm0.3$ µM) and C/Fe-SP ($K_m = 12\pm4$ µM) bind randomly and independently to form the reactive ternary complex [628]. The Co(I) centre of the C/Fe-SP protein attacks the N5-methyl group of CH$_3$THF in a rate-limiting $S_N2$ reaction, forming an intermediate
methylcob(III)amide species. Although the cobalt centre undergoes a formal two-electron oxidation during the process, it is not considered a redox reaction as the methyl group is transferred as a cation.

The same mechanism appears to take place when vitamin B12, a free cobalamin, is the cofactor [629]. Although the $K_m$ for CH$_3$THF is similar irrespective of whether B12 or cob(I)amide is used, the $k_{cat}/K_m$ for C/Fe-SP for methyl transfer is $\sim$100 fold higher than for B12. However, as the cellular concentration of B12 in *C. thermoaceticum* is 10 times higher compared with that of C/Fe-SP, the difference between the activation rates of the two cofactors *in vivo* is only $\sim$7 to 17-fold [629].

As the removal of a methyl group from a tertiary amine is not a facile reaction, even with a strong nucleophile like cob(I)amide, it has been suggested that the N5 group of CH$_3$THF must be activated by protonation, electrophilic coordination or oxidation before the reaction can occur [631]. Moreover, the stereochemical characteristics of the SN2 displacement mechanism require the proton at the N5 position to be on the opposite side of the pterin-ring plane from which the nucleophilic cob(I)amide approaches the methyl group (i.e. towards the protein and not the solvent environment). The enzyme could achieve this specific arrangement by binding the unprotonated form of CH$_3$THF, followed by protonation at the correct orientation.

It was found that the rate constant of the methyl-transfer reaction increases as the pH is lowered ($pK_a \sim 5.0$) [629], following a pH profile that closely matches that for the N5 group of free CH$_3$THF ($pK_a \sim 5$). However, it was observed that both the forward (methylation of C/Fe-SP) and the reverse (methylation of THF) reactions showed a similar pH dependency. That is, the rate increased as the pH was lowered, despite deprotonation of THF being expected during the reverse reaction. Moreover, the $k_{cat}/K_m$ for C/Fe-SP was not a function of the concentration of protonated CH$_3$THF, and no significant deuterium isotope effect was observed. Therefore, it was proposed that the pH dependence of the reaction was more likely to be connected to the protonation of a residue within MeTr rather than to the direct protonation of the substrate [629].
Spectroscopic UV-visible analysis suggested that a hydrophobic region of the MeTr protein containing a buried tryptophan residue becomes more solvent accessible at pH values close to the kinetic $pK_a$ [632]. The pH-dependent conformational change of the enzyme did not appear to affect the binding of the ligands, as the binding of both substrates for the forward reaction is pH independent [628]. Nor was it found to play a role in the protonation and activation of CH$_3$THF, as the extent of proton uptake upon CH$_3$THF binding appeared to depend only on the ionisation state of N5 and not on the ionisation state of any other site either in the ligand or in the protein [633]. It was suggested, however, that the conformational change was involved in the stabilization of the transition state of the reaction [633], as an explanation of the pH dependency. Furthermore, the binding of CH$_3$THF to the unprotonated form of MeTr was said to slightly elevate the $pK_a$ for the conformational change of the protein [628].

Kinetic studies indicate that CH$_3$THF binds to MeTr in the unprotonated form and then undergoes rapid protonation [633]. The proton transfer must occur much faster than the transmethylation step, which is the rate-limiting step of the overall reaction. The ionisation state of bound CH$_3$THF was further studied using NMR spectroscopy [633]. It was found that the $pK_a$ of $^{13}$CH$_3$THF increases by 1 $pK_a$ unit upon binding to MeTr both in H$_2$O and D$_2$O. The interaction of the ligand with MeTr resulted in a change of the NMR spectra that was associated with a stronger binding of the protonated than the unprotonated form of CH$_3$THF, despite the pH independency of the dissociation constant for CH$_3$THF [633].

The X-ray crystal structure does not show any obvious proton donor within the active site interacting directly with N5. Therefore, although specific H-bond interactions with the protein environment might increase the $pK_a$ at the N5 position of CH$_3$THF, proton transfer must involve an indirect pathway. The proton taken up by MeTr-bound CH$_3$THF is expected to originate from solvent, either directly or indirectly, and not from the enzyme itself, as otherwise no proton uptake would be measured when MeTr is titrated with CH$_3$THF in the presence of a pH indicator. The binding of C/Fe-SP to the binary complex
MeTr•CH₃THF introduces an electronegative nucleophile adjacent to the pterin ring, and is expected to further increasing the $pK_a$ of CH₃THF and facilitate its activation by protonation.

While it is most likely that protonation occurs directly at the N5 position, due to the lack of a clear proton donor, scenarios in which protonation at other sites, e.g. C8a, activate the ligand for the methyl-transfer reaction have been suggested [625]. On one hand, protonation of N5 would probably require participation of a solvent molecule. Residue Asn199, located over N5 and O4, could position a water molecule to serve as the proton donor. Alternatively, Asp160 and the conserved bridging-water molecule could stabilise a tautomeric form of the pterin ring and facilitate proton transfer from a water molecule, similar to what has been proposed for dihydrofolate reductase (DHFR) [634]. On the other hand, a proton transfer to C8a of the pterin ring [625] would shift CH₃THF to a quinonoid form, generating a positive charge over N5 and increasing its electrophilicity. In this case, Asp75 appears as a possible proton donor, provided its $pK_a$ is slightly elevated by interactions with Asp43 and Asn96.

4.2 Cobalamin-dependent methionine synthase (MetH)

Cobalamin-dependent methionine synthase (MetH) presents not only structural [625,635] and functional similarities with MeTr, but it also lacks the proton donor to the N5 position found in DHPS. Although MetH and MeTr are expected to use an indirect strategy to protonate and activate CH₃THF, significantly different protonation mechanisms have been suggested from independent experimental studies.

Cobalamin-dependent methionine synthase (MetH) catalyses the methylation of homocysteine by CH₃THF via a methyl-Cob(III)amide intermediate. During the catalytic cycle, the methylcobalamin form of the enzyme is demethylated by homocysteine, producing cob(I)alamin and methionine. CH₃THF is used to remethylate the cob(I)alamin cofactor, forming methylcobalamin and THF. Each ligand binds to one of four separate domains of this 1227-residue protein [636]. The first 350 residues constitute the homocysteine-binding domain with an active-site zinc ion [637], which activates homocysteine for nucleophilic attack on methylcobalamin [638]. The next 300 residues form the CH₃THF binding
Chapter 4  

MeTr

domain, which shows significant homology with MeTr [635]. The third module, 250 amino acids long, constitutes the cobalamin-binding domain [639]. Finally, the C-terminal module is involved in the reactivation of the enzyme when the cob(I)alamin prosthetic group is oxidized to cob(II)alamin during turnover [640].

The two methyl-transfer reactions involved in the turnover cycle of MetH have generally been assumed to proceed by $S_N2$ mechanisms, as the overall transfer of the methyl group proceeds with retention of the stereochemistry [641]. Similarly to the methyl-transfer reaction catalysed by MeTr, the $\text{CH}_3\text{THF}$ molecule is expected to be activated by protonation at N5. This protonation step could take place upon $\text{CH}_3\text{THF}$ binding, as suggested for MeTr, or at a later stage, either prior or during the methyl-transfer reaction; this is where experiments seem to differ for the two enzymes, as discussed next.

### 4.2.1 Protonation and activation of $\text{CH}_3\text{THF}$ within MetH

When the MetH enzyme (either the inactive cob(II)alamin or the active methylcobalamin form) was titrated with $\text{CH}_3\text{THF}$ in the presence of the pH indicator phenol red at pH 7.8, no proton uptake or release was observed [642]. It was suggested that the proton destined for N5 must be either already bound to the free MetH enzyme or taken up at a later stage. The affinity of $\text{CH}_3\text{THF}$ for MetH was seen to be pH independent with a $K_d$ of 5 µM [642]. UV-Visible spectral changes associated with the binding of $\text{CH}_3\text{THF}$ to MetH$^{(2-649)}$ (homocysteine and $\text{CH}_3\text{THF}$ binding domains of MetH) over a pH range from 5.5 to 8.5 seem to indicate that the ligand binds in its unprotonated form [643]. Furthermore, the spectral changes associated with the binding process were interpreted as indicative of the ligand being transferred into a more hydrophobic environment, and formation of the binary complex was said to decrease the $pK_a$ at N5 [643]. Moreover, binding at low pH was said to be associated with proton release rather than proton uptake.

The second-order rate constant for the reaction between MetH$^{(2-649)}$$\cdot$$\text{CH}_3\text{THF}$ and exogenous cob(I)alamin was seen to increase as the pH was lowered according to an apparent $pK_a$ of 5.9±0.1 [644]. In the reverse reaction, the transfer of the methyl group from methylcobalamin to THF, the second-order
rate constant decreased as the pH was lowered. The pH dependence of the MetH-catalysed reaction is consistent with either a protonation of a general-acid/base catalyst on the enzyme, which needs to be protonated for the reaction of CH₃THF with cob(I)alamin, and deprotonated for the reaction of THF with methylcobalamin, or the direct protonation of CH₃THF following the formation of the ternary complex. The latter mechanism was considered more likely due to the lack of a suitable residue at H-bond distance from N5 that could act as direct proton donor [644].

4.2.2 Role of aspartic residues within the MetH active site

The contribution of the aspartic residues D399, D434 and D522 (corresponding to residues D43, D75 and D160 in MeTr, see Figure 4-2) to ligand binding and activity of MetH was explored by single-point Asp→Asn mutations in E. coli MetH(2-649) [643]. No catalytic activity was detected for the D434N and D522N mutants, and a 40-fold reduction was seen for D399N. No binding of CH₃THF was detected for D522N and an 8-fold decrease of affinity was observed for D399N, while D434N showed almost no change. These results demonstrate the importance of these well-conserved aspartic residues in the binding of CH₃THF and in the catalysis of the methyl-transfer reaction.

4.3 Comparisons of MeTr and MetH

Not only do both enzymes, MeTr and MetH, catalyse the methylation of a cob(I)alamin group using CH₃THF as a substrate, but their structures are strikingly similar. The superposition of MeTr and the CH₃THF-binding domain of MetH (using the Dali Server [561]) results in a Cα-RMSD of 1.8 Å (Figure 4-3A) and a level of sequence identity of 30% for the superimposable residues. If the conserved residues are mapped over the surface of the MeTr protein (Figure 4-3B), it can be seen that most active-site amino acids are absolutely conserved between the two proteins.

A comparison of the active sites (Figure 4-4) shows that most residues that appear hydrogen bonded to the ligand CH₃THF, and in particular to the pterin ring, are almost the same in both enzymes.
Figure 4-3 Structural comparison of MeTr and the CH$_3$THF-binding domain of MetH.

(A) Superimposed structures of MeTr (Doukov, Ragsdale et al. unpublished results 2003) in blue and MetH [635] in green. It is clear that not only is the general structure of the two proteins almost the same (Cα RMSD 1.8 Å), but also the binding conformation of the ligand CH$_3$THF is strikingly similar. (B) Conserved residues between MeTr and MetH (30% sequence identity, shaded in grey) are mapped onto a surface representation of MeTr (blue). Most residues lining the active-site region are the same in both enzymes, the blue patch near N3 of the pterin ring corresponds to conservative substitutions: Ile in MeTr and Val in MetH.
The main difference between the two complexes is the positioning of the glutamic group of CH$_3$THF, which appears rotated 180° in one structure compared with the other (Figure 4-3A). This difference, however, is expected to have little, if any influence, on the reactive properties of the pterin ring.

![Figure 4-4 Schematic representation of the MeTr active site.](image)

Figure 4-4 Schematic representation of the MeTr active site.

Only selected residues (Asp 43, Asn 45, Asp 75, Asn 96, Asp 160, Ile 163, Asn 199, Gln 202 and Arg 207) are shown. The only two residues that differ in the active site of MetH (shown in green) are Ile163, which is a valine, and Gln202, which is a phenylalanine. While Ile 163 (or Val) interacts through its backbone group, the positioning of the Glu tail of CH$_3$THF appears flipped 180° in the active site of MetH compared with the conformation found within MeTr and, therefore, does not interact with the Phe residue in place of Gln 202. This difference, however, is not expected to affect significantly the reactivity of the pterin ring.

The structural similarity between MeTr and the CH$_3$THF-binding domain of MetH suggests that they should catalyse the transfer of the methyl group of CH$_3$THF to cob(I)alamin using a similar mechanism of reaction. However, in addition to some similarities, many experimental analyses of these two enzymatic systems have led to contradictory conclusions (Table 4-1). It is unclear if the experimentally observed properties correspond indeed to different protonation mechanisms within MeTr and MetH, or if the discrepancies are due to the different adopted experiments or in the interpretation of the results.
### 4.4 Motivations and Goals

Given the structural similarities between MeTr and MetH, two enzymes that catalyse the same reaction, the conflicting experimental conclusions regarding the protonation and activation of the ligand CH₃THF are puzzling. The lack of a direct proton donor over the N5 position of CH₃THF have lead to different hypotheses regarding the mechanism of protonation/activation of the substrate, which probably follows an indirect pathway.

To contribute to the understanding of these systems, a computational approach was used to study the protonation and activation of the ligand CH₃THF within the MeTr active site. The ionisation states of both the ligand and active-site aspartic acid residues were investigated in order to understand the influence of the protein environment on the reactivity of CH₃THF, and to try to find some clue to the apparently different experimental behaviour of the two enzymes.

---

#### Table 4-1. Experimental observations regarding the protonation and activation of CH₃THF within the active site of MeTr and MetH.

<table>
<thead>
<tr>
<th></th>
<th>MeTr</th>
<th>MetH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agreements</strong></td>
<td>The forward reaction rate (methylation of cob(I)alamin) increases as the pH is lowered in association with an apparent $pK_a \sim 5.8$ (MeTr) [629]/5.9±0.1 (MetH) [644].</td>
<td>The $K_d$ for CH₃THF is pH independent [628,642].</td>
</tr>
<tr>
<td></td>
<td>The $K_d$ for CH₃THF is pH independent [628,642].</td>
<td>Binding of the cob(I)alamin species increases the $pK_a$ of bound CH₃THF [628,644].</td>
</tr>
<tr>
<td></td>
<td>The reverse reaction rate, methylation of THF using methylcobalamin as cofactor, decreases as the pH is lowered in association with an apparent $pK_a \sim 5.9$ [629,644]</td>
<td></td>
</tr>
<tr>
<td><strong>Disagreements</strong></td>
<td>Titration of MeTr with CH₃THF (or THF) in the presence of phenol red at pH 6.8-5.2, produced changes associated with a $pK_a \sim 5.5-6.0$ [633].</td>
<td>Titration of MetH with CH₃THF (or THF) in the presence of phenol red at pH 7.8 does not indicate proton uptake [642].</td>
</tr>
<tr>
<td></td>
<td>$^{13}$C NMR studies suggest an increase of 1 $pK_a$ unit at the N5 position of CH₃THF upon binding to the enzyme [633].</td>
<td>UV-Visible spectral changes of CH₃THF binding to MetH over pH 8.5-5.5 indicate a decrease in the $pK_a$ at N5 (below 5) [643].</td>
</tr>
<tr>
<td></td>
<td>MeTr binds the protonated form of the CH₃THF 10-fold stronger than the unprotonated form [633].</td>
<td>Binding of protonated CH₃THF to MetH is accompanied by proton release, suggesting that the unprotonated form is the preferred ligand [643].</td>
</tr>
</tbody>
</table>
Given the size and the complexity of the system, the analysis of possible protonation profiles was done in a progressive and systematic manner. Initially, the ability of computational methods to predict the protonation pattern and $pK_a$ of several folate-like molecules in an aqueous environment was studied. Then, the size and complexity of the modelled system was increased progressively, by gradually incorporating active-site residues and the influence of the protein environment. Different computational methods were employed, depending on the size of the system and the level of accuracy required, including QM calculations, ONIOM calculations, and molecular dynamics (MD) thermodynamic integrations (TI) for the determination of free energies.

4.5 Methods

4.5.1 Theoretical calculations of protonation energies

In order to determine the accuracy of simple QM calculations to predict both protonation positions and $pK_a$ values, different protonated and unprotonated forms of pterin, folate, and their derivatives (Table 4-3) were analysed.

The molecules were manually constructed using GaussView [645], and the final structures optimised in Gaussian [646] at the HF/6-31+G** level using the PCM implicit water solvent model, or at the HF/6-31G* PCM level followed by single point calculations at HF/6-31+G** PCM. This was necessary due to convergence problems with some of the systems.

The final energies were used to predict both relative protonation patterns, in order to assess the reliability of the method to predict the most likely position for protonation, and to calculate protonation energies, which were correlated with experimentally determined $pK_a$s in a similar approach to that originally published by Gready [647].

For the qualitative predictions of protonation sites, only the oxidized and reduced forms of 6-methyl-pterin and folate were analysed. The protonation energies of four different sites in the pterin rings were compared, including N1, N5, O4 and N8. When present, both the ionised and the neutral forms of the pABA-Glu tail were considered.
For the calculation of protonation energies and correlations with experimental $pK_a$s, several other species were considered in order to cover a $pK_a$ range from 2.3 to 10.5 (Table 4-3). This involved the inclusion of some heterocyclic anions. As both protonated and unprotonated pABA-Glu tails showed comparable results during qualitative predictions, only selected protonated forms of the glutamic moiety of folate ($pK_a = 4.7-4.6$) and its derivatives were analysed for the correlation studies. The Glu group was considered to be protonated in the case of folate (N1 $pK_a$ 2.35 [648]) and dihydrofolate (N5 $pK_a$ 2.59 [649]) and ionised for tetrahydrofolate (N5 $pK_a$ 4.82 [650]) and CH$_3$THF (N5 $pK_a$ 5.05 [629]). The final protonation energy values calculated at HF/6-31+G** PCM implicit solvent were correlated with experimentally determined $pK_a$s to develop a calibration curve.

Given that large basis sets that include polarization and diffuse functions, such as 6-31+G**, may result in an overestimation of the polarization of molecules and affect, therefore, the correlation between protonation energy and $pK_a$, similar analysis were carried out the HF/3-21G PCM level (not shown). This second set of calculations resulted in a less accurate correlation, supporting, therefore, the use of the 6-31+G** basis set for this kind of calculations.

A similar analysis was done using in vacuo HF/6-31+G** and HF/3-21G protonation energies (not shown). Both basis sets provided very similar results, where molecules of different total charge were found to cluster together, giving separate correlation patterns for the neutral and the positively-charged species. This is in agreement with the original work of Gready [647], where two different protonation plots for neutral pterin and pterin anions were constructed based on HF calculations at the STO-3G and 3-21G levels. The inclusion of an implicit solvent representation provides, therefore, more realistic protonation energies, as the effect of the water environment on the stabilization of the unprotonated and protonated species is taken into account, while at the same time allowing to include molecules with different total charge within the same correlation plot.

The energy differences between the QM calculations for different protonated forms of CH$_3$THF with implicit solvent representation (HF/6-31+G** PCM) and single-point in vacuo HF/6-31+G** calculations, were taken as indicative of the
solute-solvent interaction energies. These values proved to give an energy pattern very similar to the relative free energies of solvation obtained by the thermodynamic integration approach.

4.5.2 QM fragment calculations

A simplified representation of the ligand CH₃THF and the protein active site was defined to perform QM calculations (Figure 4-5).

![Figure 4-5](image_url)

**Figure 4-5 Representation of the active site and ligand used during fragment QM calculations.**

A total of 114 atoms were considered including: the pterin ring of CH₃THF; side chains of residues D43, N45, D75, N96, D160 and N199; the backbone of residue I163; and two water molecules. In order to preserve the general conformation of the active site, atoms marked with an asterisk were kept fixed during optimisation. Five different protonation positions in the pterin ring (N1, O4, N5, N8 and C8a) and different protonated forms of the residues D75 and D160 were analysed. Insert table: A total of seven single-protonated and 10 double-protonated complexes were studied.

<table>
<thead>
<tr>
<th>Protonated forms studied</th>
<th>MeTr</th>
<th>MeTr-D75H</th>
<th>MeTr-D160H</th>
</tr>
</thead>
<tbody>
<tr>
<td>unp</td>
<td>SP</td>
<td>SP</td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>SP</td>
<td>DP</td>
<td>DP</td>
</tr>
<tr>
<td>N5</td>
<td>SP</td>
<td>DP</td>
<td>DP</td>
</tr>
<tr>
<td>N8</td>
<td>SP</td>
<td>DP</td>
<td>DP</td>
</tr>
<tr>
<td>O4</td>
<td>SP</td>
<td>DP</td>
<td>DP</td>
</tr>
<tr>
<td>C8a</td>
<td>SP</td>
<td>DP</td>
<td></td>
</tr>
</tbody>
</table>

* unp unprotonated, SP singly protonated, DP doubly protonated.

The initial coordinates of the ligand, protein side chains, and water molecules were taken from a crystal structure of the MeTr•CH₃THF complex (coordinates
provided by Dr Doukov and Prof Ragsdale). Optimisations were carried out at the HF/6-31G* level \textit{in vacuo} (Gaussian03 [646]), and single-point \textit{in vacuo} B3LYP/6-31G** calculations were carried out on the final structures. In order to preserve the overall structure of the fragment close to that of the crystal structure, and to prevent unrealistic displacements of active-site groups, the C\alpha atoms of the side chains, as well as the terminal carbon atoms of the Ile163 backbone, were fixed during optimisation (see Figure 4-5). All other atoms were free to move. Protonation of both the CH\textsubscript{3}THF-pterin ring (N1H\textsuperscript{+}, O4H\textsuperscript{+}, N5H\textsuperscript{+}, N8H\textsuperscript{+}, and C8aH\textsuperscript{+}) and aspartic residues (D75 and D160) were studied, including a total of 7 single-protonated and 10 double-protonated complexes (Figure 4-5, table insert).

For the prediction of p\textsubscript{K\textsubscript{d}} values, single-point calculations at the HF/6-31+G** level in PCM implicit water solvent were conducted with the optimised structures (this is the same computational model used to prepare the calibration curve of protonation energy vs p\textsubscript{K\textsubscript{d}}, see section 4.5.1).

4.5.3 ONIOM calculations

In order to include a more realistic representation of the protein environment during the protonation analysis, ONIOM calculations were carried out. The system studied included the complete MeTr protein monomer, crystal water molecules, and the ligand CH\textsubscript{3}THF.

Four different conformations of the protein were analysed. These conformations were obtained from two different sets of X-ray structure coordinates generated during refinement of the same diffraction data with two different programs (SHELX [627] and CNS [626]) (Doukov, Ragsdale \textit{et al.}, unpublished results 2003). Each individual monomer of the dimeric crystal structures was analysed separately, providing four different representations of the MeTr•CH\textsubscript{3}THF complex. Only those crystallographic water molecules interacting with the ligand, and present in all four structures, were included in the calculations.
4.5.3.1 Preparation of initial structures

The initial crystal complexes were processed as described below in order to add hydrogen atoms, relax the structures, and remove bad interactions/clashes in preparation for the ONIOM calculations.

The four different protein conformations (CNS_A, CNS_B, SHELX_A, and SHELX_B), together with six selected water molecules were processed with the *xleap* module of AMBER 8 [651]. Hydrogen atoms were added automatically; except D43 was considered to be protonated, and the other two active-site aspartic residues, D75 and D160, were set as ionised. The Amber ff03 [265] force field was used to assign atom types and charges to the protein structure, while waters were described using the TIP3P representation [283].

After superposition of the few crystal-structure models, the ligand CH$_3$THF was seen to adopt almost the same conformation in all four complexes. It was extracted from one of the models and processed with GaussView [645]. Hydrogen atoms were added to the unprotonated structure, and five different protonated forms were generated (N1H$^+$, N5H$^+$, N8H$^+$, O4H$^+$ and C8aH$^+$). As *in vacuo* QM optimisations led to folded conformations of the molecule, very different from that observed crystallographically, optimisations were performed at the HF/6-31G** level with PCM implicit water solvent representation (Gaussian03 [646]). Single-point *in vacuo* HF/6-31G* calculations were performed on the final optimised structures, and RESP partial atomic charges were fitted using the *antechamber* module of AMBER 8. GAFF (General Amber Force Field) atom types, compatible with the ff03 force field used to describe the protein, were automatically assigned to the ligand using *antechamber*, and manually modified when necessary. Missing parameters were added manually by comparison with known structures.

Next, the structures of the four protein models were combined with the unprotonated form of the ligand, and the final complexes energy minimized using *Sander* (AMBER 8). The minimization was done in two stages: first, 500 steps for hydrogen atoms only, followed by 200 steps where all atoms were released, except the oxygen atoms of the six water molecules. The final
optimised structures were used as a template to generate different single- and double-protonated forms of the system (see table within Figure 4-5).

4.5.3.2 ONIOM Energy Calculations

After models for all protonated forms of the MeTr•CH₃THF complex were generated, Gaussian03 [646] was used to carry out the ONIOM calculations. The models were divided into two regions, a QM and a MM layer. The QM region was treated at the HF/6-31G* level, while the MM region was treated with Amber, using the ff96 force field [652] (only one implemented in Gaussian03). All calculations were done using mechanical embedding conditions based on previous problems observed with electronic embedding (Rostov, Cummins and Gready, unpublished results).

In a first approach, the complete ligand CH₃THF only was included within the QM region. Five different single-protonated forms of CH₃THF (N1H⁺, N5H⁺, N8H⁺, O4H⁺ and C8aH⁺) were studied within three different protonated states of the active site (MeTr, D75H-MeTr and D160H-MeTr), in four different protein conformations (CNS_A, CNS_B, SHELX_A, and SHELX_B). These sixty different systems were optimised under ONIOM mechanical embedding conditions. In order to preserve the same protein environment for all protonated forms, nearly all MM atoms were kept frozen during the optimisation. Only the complete ligand, water molecules, and side chains of residues Asp43, Asn45, Asp75, Asn96, Asp160 and Asn199 were allowed to move (see Figure 4-6 as an example). ONIOM optimisations were carried out at the HF/6-31G**:Amber level, followed by single-point B3LYP/6-31G**:Amber calculations. The energies of each single- and double-protonated forms of each of the 4 models were averaged. The most stable protonated species were selected for further studies.

In a second set of calculations, the most stable structures of the first set of ONIOM studies were analysed using a different ONIOM partition. The QM region was expanded to include several active-site residues and two water molecules directly interacting with the pterin ring (Figure 4-6). The inclusion of the side chains of Asp75 and Asp160 within the QM region allowed us to include two new single-protonated (D75H-MeTr•CH₃THF, D160H-
MeTr•CH$_3$THF) and one double-protonated (D75H-D160H-MeTr•CH$_3$THF) forms of the system, in which protons were added to the aspartic residues while the substrate CH$_3$THF was considered to be unprotonated. Optimisations were done at the HF/6-31G*:Amber ONIOM level under mechanical embedding conditions, with all but QM atoms frozen, followed by single-point B3LYP/6-31G*:Amber calculations. The energies over the four structure models were averaged.

**Figure 4-6 QM region of the MeTr•CH$_3$THF system used during ONIOM calculations.**
The complete ligand, CH$_3$THF; two water molecules; and the side chains of residues D43, N45, D75, N96, D160, and N199 were included within the QM region. The rest of the protein was simulated at the MM level.

### 4.5.4 Free energy calculations

In order to further evaluate the effect of the protein environment on the protonation pattern of CH$_3$THF, a series of free energy calculations were carried out using the thermodynamic integration approach. A similar protocol to those published by Cummins and Gready [436,440] was implemented.
AMBER 8 [651] was used to calculate the free energy associated with several alchemic transformations of the ligand in vacuo, water, and within the protein environment. The mutations between different protonated states of CH$_3$THF were done via $\lambda$-coupling (see Chapter 2, section 2.5.1.1), taking the entire ligand molecule as the perturbed group. The coupling parameter $\lambda$ was divided into discrete values, $\lambda_i$, to yield a series of windows. For each window an equilibration simulation was performed, followed by a simulation period in which data for the calculation of the perturbation free energy was collected. To estimate the total energy change, the $dV/d\lambda$ values obtained over the $\lambda_i$ windows were integrated using the trapezoid rule.

All transformations were done starting from a single common structure of the ligand CH$_3$THF, with dummy atoms attached to N1, N5, N8 and O4 (Figure 4-7). The dummy atoms were assigned the same van der Waals parameters as hydrogen atoms and zero atomic charge.

The free energy change between two protonated forms of the ligand CH$_3$THF could be calculated following different pathways, each involving a different sequence of mutations. In this case, transformations were done according to the diagram shown in Figure 4-8. The mutation of one protonated state $A$ to another $B$ was carried out in two stages, via an intermediate state $A'$ [440]. The intermediate state $A'$ had the electrostatic parameters of the final state $B$, and the same intramolecular and vdW properties as the initial state $A$. Therefore, the first stage of the transformation $A \rightarrow A'$ provides the electrostatic contribution of the free energy ($\Delta G^{Ele}$); and the second step, $A' \rightarrow B$, provides the vdW component ($\Delta G^{vdW}$).
Figure 4-7 N5-protonated form of CH$_3$THF used in alchemic transformations during free energy calculations.

Dummy atoms attached to N1, O4, N5 and N8. While atoms N5 and N8 were considered tetrahedral, N1 was treated as trigonal planar. Dummy atoms were assigned the same van der Waals parameters as the corresponding hydrogen atoms and zero atomic charge.

Figure 4-8 Mutational sequence followed during the calculation of relative free energies.

Each transformation was divided into an electrostatic and a van der Waals component. Initially, only the charges of the atoms were modified, followed by a second transformation in which atom types were mutated.

The transformations were carried out in vacuo, water, and within the protein environment. In each case both the forward ($\lambda=0\rightarrow1$) and the reverse transformations ($\lambda=1\rightarrow0$) were performed, with the final coordinates for the forward mutation used as the starting point of the reverse simulation. This was used to evaluate the hysteresis of the transformations along the chosen transformation pathway. The standard deviation of five different transformations in the forward direction (different initial velocities) was taken as the error associated with the calculations.
4.5.4.1 Preparation of initial structures

The ligand CH$_3$THF was extracted from one of the sets of MeTr•CH$_3$THF crystal structure coordinates (CNS) and processed using GaussView [645]. Dummy atoms were attached to N1, N5, N8 and O4. RESP partial atomic charges for the neutral and protonated forms of the ligand were adjusted from HF/6-31G* in vacuo calculations (Gaussian03 [646]) using the *antechamber* program of AMBER 8 [651]. GAFF atom types were assigned for the MM simulations. Dummy atoms were assigned H-like vdW parameters.

The protein structure was processed using the *antechamber* program of AMBER 8 and the ff03 force field. Hydrogen atoms were added, generating three different protonated forms of the active site: MeTr, D75H-MeTr, and D160H-MeTr. In all these cases D43 was considered to be protonated. The 3 models were combined with the unprotonated form of the ligand into MeTr•CH$_3$THF complexes and solvated in a TIP3P water box, leaving a 10 Å thick solvent layer. The systems were neutralized with the addition of 10 Na$^+$ ions (MeTr•CH$_3$THF structure) or 9 Na$^+$ ions (D75H-MeTr•CH$_3$THF and D160H-MeTr•CH$_3$THF structures). To remove bad contacts and relax the systems, 200 steps of energy minimization were carried out, followed by 50 ps (25000 steps) MD simulation of the water molecules only, and finally 50 ps simulation of the entire system. The final relaxed structures were used as starting points for all four different transformations within the three different protein environments.

4.5.4.2 Transformations in vacuo

*In vacuo* transformations were done at 300K, with a 10 Å cutoff for van der Waals and electrostatic interactions. A time step of 0.02 ps was used, and hydrogen bonds were constrained using the SHAKE algorithm. In order to maintain the solution-extended conformation of CH$_3$THF, a distance constraint was imposed between atoms N3 and C$_\alpha$ of the glutamic tail (13.5 Å minimum distance, force of 50 kcal/mol Å$^2$) during *in vacuo* simulations. Several different simulation schemes were tried in order to find conditions which produced converging energy results with minimal dispersion (not shown). Details of the finally chosen transformation protocol are given in Table 4-2.
4.5.4.3 Transformations in water

For these transformations, the CH\textsubscript{3}THF molecules were solvated in a TIP3P box of water, leaving 10 Å between the substrate and the edge of the box. The simulations were carried out under periodic boundary conditions (PBC), for the closed, isothermal, isobaric (NTP) ensemble at 300 K and 1 atm. A time step of 0.002 ps was employed, and the SHAKE algorithm was used to constrain hydrogen bonds. A cutoff of 10 Å was set for all nonbonded interactions, and PME (See Chapter 2) was used to treat electrostatic interactions beyond the 10 Å cutoff. The same distance constraint between atoms N3 and the C\textalpha of the glutamic moiety as used for in vacuo calculations was applied (calculations without the distance constraint were shown to give similar results). Several different simulation conditions for the TI calculations were tested, with the finally selected parameters summarized in Table 4-2.

<table>
<thead>
<tr>
<th></th>
<th>Vacuo</th>
<th>Water</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial relaxation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimization</td>
<td>200 steps</td>
<td>200 steps</td>
<td>200 steps</td>
</tr>
<tr>
<td>Water MD</td>
<td>Not applicable</td>
<td>20ps</td>
<td>50ps</td>
</tr>
<tr>
<td>System MD</td>
<td>20ps</td>
<td>20ps</td>
<td>50ps</td>
</tr>
<tr>
<td><strong>Electrostatic transformation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda)</td>
<td>6 windows</td>
<td>6 windows</td>
<td>21 windows</td>
</tr>
<tr>
<td>(\Delta \lambda = 0.2)</td>
<td>(\Delta \lambda = 0.2)</td>
<td>(\Delta \lambda = 0.05)</td>
<td></td>
</tr>
<tr>
<td>Equilibration</td>
<td>80ps</td>
<td>20ps</td>
<td>10ps</td>
</tr>
<tr>
<td>Collection</td>
<td>20ps</td>
<td>20ps</td>
<td>10ps</td>
</tr>
<tr>
<td>Total time</td>
<td>600ps</td>
<td>240ps</td>
<td>440ps</td>
</tr>
<tr>
<td><strong>van der Waals transformation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda)</td>
<td>11 windows</td>
<td>17 windows</td>
<td>21 windows</td>
</tr>
<tr>
<td>(\Delta \lambda = 0.1)</td>
<td>(\Delta \lambda = 0.025)</td>
<td>(\Delta \lambda = 0.05)</td>
<td></td>
</tr>
<tr>
<td>(from (\lambda = 0) to 0.2)</td>
<td>(from (\lambda = 0.2) to 1.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equilibration</td>
<td>20ps</td>
<td>20ps</td>
<td>10ps</td>
</tr>
<tr>
<td>Collection</td>
<td>20ps</td>
<td>20ps</td>
<td>10ps</td>
</tr>
<tr>
<td>Total time</td>
<td>440ps</td>
<td>680ps</td>
<td>440ps</td>
</tr>
<tr>
<td><strong>Total transformation time</strong></td>
<td>1040ps</td>
<td>920ps</td>
<td>880ps</td>
</tr>
</tbody>
</table>

4.5.4.4 Transformations inside the protein

Three different protein environments were analysed: unprotonated aspartates (MeTr), protonated Asp75 (D75H-MeTr), and protonated Asp160 (D160H-
MeTr). In all cases D43 was considered to be protonated. Simulation conditions were identical to those used during transformations in water (see above). Specific details on the simulation conditions, adjusted after systematic examination of several protocols, can be found in Table 4-2.

Final energy values for the transformations \textit{in vacuo}, water, and within the protein environment were used to calculate relative solvation energies, protein-solvation energies, and binding energies.

4.6 Results and Discussion

4.6.1 Protonation Patterns and \( pK_a \) of folates and derivatives

In order to test the ability of computational methods to predict the preferred protonation position and associated \( pK_a \) of different molecules in solution, QM HF/6-31+G** calculations with implicit solvent representation were carried out. A similar protocol to that originally published by Gready [647] was followed. A series of folate-like molecules with experimentally determined \( pK_a \) values were chosen as a test set (Table 4-3).

Table 4-3. Compilation of pterin-derived compounds used for the QM prediction of \( pK_a \).

<table>
<thead>
<tr>
<th>Name</th>
<th>Site</th>
<th>Exptl. ( pK_a )</th>
<th>Prot. Energy HF-631+G** PCM (kcal/mol)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-methyl-pterin</td>
<td>N1</td>
<td>2.31</td>
<td>-286.9662</td>
<td>[653]</td>
</tr>
<tr>
<td>6-methyl-7,8-dihydropterin</td>
<td>N5</td>
<td>4.17</td>
<td>-288.9013</td>
<td>[654]</td>
</tr>
<tr>
<td>6-methyl-5,6,7,8-tetrahydropterin</td>
<td>N5</td>
<td>5.4</td>
<td>-293.2455</td>
<td>[655]</td>
</tr>
<tr>
<td>folate (FOL)</td>
<td>N1</td>
<td>2.35</td>
<td>-286.3901</td>
<td>[648]</td>
</tr>
<tr>
<td>7,8-dihydrofolate (DHF)</td>
<td>N5</td>
<td>2.59</td>
<td>-286.2125</td>
<td>[649]</td>
</tr>
<tr>
<td>5,6,7,8-tetrahydrofolate (THF)</td>
<td>N5</td>
<td>4.82</td>
<td>-292.4979</td>
<td>[650]</td>
</tr>
<tr>
<td>5-methyl-5,6,7,8-tetrahydrofolate (CH3,THF)</td>
<td>N5</td>
<td>5.05</td>
<td>-292.1589</td>
<td>[629]</td>
</tr>
<tr>
<td>6,7-dimethyl-7,8-dihydropterin</td>
<td>N5</td>
<td>4.16</td>
<td>-289.1894</td>
<td>[656]</td>
</tr>
<tr>
<td>6,7,7-trimethyl-7,8-dihydropterin</td>
<td>N5</td>
<td>4.24</td>
<td>-288.6825</td>
<td>[656]</td>
</tr>
<tr>
<td><strong>Anions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-methyl-pterin</td>
<td>N3</td>
<td>8.10</td>
<td>-302.1152</td>
<td>[653]</td>
</tr>
<tr>
<td>6-methyl-5,6,7,8-tetrahydropterin</td>
<td>N3</td>
<td>10.5</td>
<td>-307.8472</td>
<td>[655]</td>
</tr>
<tr>
<td>6-methyl-pterin 8-oxide</td>
<td>N3</td>
<td>7.07</td>
<td>-298.7673</td>
<td>[657]</td>
</tr>
<tr>
<td>7-methyl-pterin 8-oxide</td>
<td>N3</td>
<td>6.94</td>
<td>-298.9162</td>
<td>[657]</td>
</tr>
</tbody>
</table>
Figure 4-9 Protonation patterns of pterin and derivatives.

Relative energies of different protonated forms of the oxidized, dihydro- and tetrahydro-reduced forms of 6-methyl-pterin (A) and folate (B). QM energies were determined at the HF/6-31+G** level with an implicit solvent model (PCM), and relative energy values calculated using the O4-protonated form as reference. The results shown for folate and its derivatives (B) correspond to the species with the ionised glutamic tail, similar results were obtained for the protonated form (not shown). In every case, the species with the lowest relative energy correspond to the experimentally found one (marked with an asterisk).
To analyse protonation patterns, four different protonation sites (N1, N5, N8 and O4) on the pterin ring of 6-methyl-pterin and folate, including their dihydro- and tetrahydro- reduced forms, were analysed. The final QM relative energies, calculated at HF/6-31+G** level with an implicit solvent model (PCM), are shown in Figure 4-9.

In all cases, the protonated species with the lowest relative energy is the one found experimentally. These results suggest that relative energies obtained from HF/6-31+G** PCM calculations on different protonated forms of folate-like molecules can be used to accurately predict the preferred protonation position.

The ability to predict not only the position but also the pKa of the atom being protonated was tested. Several pterin-derived compounds with experimentally determined pKas were analysed, covering a pKa range from 2.3 to 10.5 (Table 4-3). Protonation energies determined at the QM HF/6-31+G** PCM level were plotted against experimentally determined pKas (Figure 4-10).

\[
pKa = -0.3518 \times \text{EProt} - 97.936 \\
R^2 = 0.975
\]

Figure 4-10 Correlation of theoretical protonation energies (HF/6-31+G** PCM) and experimental pKas values.

Plotted results are shown in Table 4-3. A linear fit of all points was performed, with the final equation that correlates QM protonation energies with pKas values shown within the plot. The slope correspond to 2.84 kcal/mol per pKa unit.
A very good correlation between the experimental $pK_a$ values and the theoretical protonation energies was found, with a $R^2$ value of 0.98 and a slope of 2.84 kcal/mol per $pK_a$ unit.

Calculations at the HF/3-21G PCM level resulted in less accurate correlations (not shown), suggesting that the overestimation of polarization associated with large basis sets (6-31+G**) does not have a negative effect on this particular study. Similar analysis were also carried out using in vacuo HF/6-31+G** and HF/3-21G protonation energies (not shown). Larger dispersions were found, and molecules with different total charge grouped into separate clusters with different correlation patterns. Therefore, the use of an implicit solvent model and a large basis set was found to provide the most accurate correlations between protonation energies and $pK_a$ values, while at the same time making it possible to compare molecules with different total charges with a single calibration curve.

These results have shown that QM calculations at the HF/6-31+G** PCM level can be used not only to predict protonation sites of pterin and derivatives, but also to estimate the associated $pK_a$ with a good degree of accuracy.

### 4.6.2 Influence of the protein environment on the protonation pattern and energy of CH$_3$THF

Following the predictions of protonation patterns and the correlation of protonation energies and experimental $pK_a$ values for a series of pterin-derived molecules, the effects of the protein environment on the properties of CH$_3$THF were analysed. This was done in three different ways: fragment QM calculations, ONIOM calculations and free energy calculations.

#### 4.6.2.1 Fragment QM calculations

In order to study the influence of the protein environment on the protonation pattern and $pK_a$ of CH$_3$THF, a limited model of the system was created (Figure 4-5). In this model, the ligand was represented using the pterin ring of CH$_3$THF, while the protein environment was modelled including only those residues directly H bonded to the ring. These included the side chains of D43, D45, D75, D96, D160 and D199, and the backbone of I163. Two water molecules
conserved in the crystal structures and forming H-bond bridges between O4 and the protein were also included.

During fragment calculations different protonated species of the pterin ring of CH₃THF as well as the ionisation state of active-site aspartic acid residues were considered. Asp43 and Asp75 appear within H-bond distance from each other in the X-ray structure. In order for this interaction to be stable, one of the residues is expected to be protonated. Given that D43 is the least solvent exposed of the two, it was considered protonated in all our analyses. The protonation states of the other two active-site Asp residues, D75 and D160, were systematically explored during the calculations, together with 4 different protonated forms of the ligand (N1, N5, N8 and O4).

All structures were optimised at the HF/6-31G* level, and single-point calculations were also carried out using B3LYP/6-31G*. The final relative energies of seven single-protonated and eleven double-protonated forms of the fragment are shown in Figure 4-11.

When a single proton is added to the fragment, it preferentially goes to the N5 atom of the pterin ring. N5 appear to be the preferred protonation position of the CH₃THF ligand within the protein environment (Figure 4-1A), regardless of the protonation state of D75 or D160 (Figure 4-1B).

In the case of double-protonated species, the results of HF and B3LYP calculations differ. While HF energies predict that the double protonated complex D75H-D160H is ~9 kcal more stable than N5H-D160H, B3LYP calculations give almost identical energies for both species, ~0.1 kcal difference in favour of N5H-D160H. However, the B3LYP energy differences among the three doubled-protonated species, N5H-D160H, N5H-D75H, and D75H-D160H, are all relatively small. These results might suggest that these three energetically comparable protonated forms of the system could coexist and interconvert easily.
Figure 4-11 Relative energies of single- (A) and double-protonated (B) forms of a simplified representation of a MeTr•CH₃THF complex.

The structures (see Figure 4-5) were optimised at the HF/6-31G* level, and single-point calculations were carried out at B3LYP/6-31G*. Relative energies were calculated using the N1H⁺ single-protonated species, or the N1H⁺-D75H double-protonated species as reference. When a single proton is added to the system, it preferentially goes to N5, as observed for the free ligand in water. However, when a second proton is added to the system the most stable species according to HF energies is D75H-D160H, where both protons are positioned on the aspartic residues of the active site, while the lowest B3LYP energies are for N5H-D160H and D75H⁻-D160H.
These results suggest that the protein environment does not modify the preferred protonation position for the ligand, which appears to be the same found in solution, N5. However, the interaction with several active-site residues is expected to influence the effective pK_a at which the protonation reaction will take place. Using the previously established correlation between HF/6-31+G** PCM protonation energies and experimental pK_a values (Figure 4-10), the pK_a values of several protonation reactions of this simplified fragment-model representation of the protein system were estimated (Figure 4-12).

\[
\begin{align*}
\text{CH}_3\text{THF} \cdot \text{E} - \text{D75H} & \quad \Delta E = -362.19 \quad pK_a = 6.01 \\
\text{CH}_3\text{THF} \cdot \text{E} \quad \Delta E = -364.29 \quad pK_a = 5.97 \\
\text{CH}_3\text{THF} \cdot \text{E} - \text{D160H} & \quad \Delta E = -317.29 \quad pK_a = 6.62 \\
\text{CH}_3\text{THF} \cdot \text{E} - \text{N5H} \cdot \text{E} - \text{D75H} & \quad \Delta E = -319.04 \quad pK_a = 4.61 \\
\text{CH}_3\text{THF} \cdot \text{E} - \text{N5H} \cdot \text{E} & \quad \Delta E = -369.13 \quad pK_a = 7.95 \\
\text{CH}_3\text{THF} \cdot \text{E} - \text{N5H} \cdot \text{E} - \text{D160H} & \quad \Delta E = -317.29 \quad pK_a = 6.62 \\
\end{align*}
\]

Figure 4-12 Estimated pK_a values within the protein environment.
Protonation energies (kcal/mol) and estimated pK_a values for a series of possible protonation pathways within a simplified fragment-model representation of the MeTr•CH_3HTF complex (CH_3THF•E) (see Figure 4-5). Protonation energies correspond to in vacuo calculations at the HF/6-31G* level. Single point HF/6-31+G** PCM calculations on the optimised complexes were used to estimate pK_a values using the expression: pK_a = -0.3518 E Prot - 97.936 (see Figure 4-10). Considering that this correlation was constructed for the protonation of pterin and derivatives (see section 4.6.1), the predicted pK_a values for the protonation of aspartic acid residues D75 and D160 are of arguable accuracy and appear in grey.

The predicted pK_a of 5,6-dimethyl-5,6,7,8-tetrahydropterin in water is 5.02 (from HF/6-31+G** PCM calculations, using the pK_a correlation shown in section 4.6.1), this suggests an increase ~3 units to 7.91 upon binding to the active site of MeTr (ionised D75 and D160). A smaller increase, ~1.6 unit (to 6.55), is also predicted for the binding to a D160-protonated form of the protein. However, if D75 is protonated instead, a decrease of 0.4 units (to 4.51) on N5 is...
expected upon formation of the binary complex. Although not necessarily accurate, $pK_a$ predictions for the aspartic groups suggest that protonation of the ligand would inhibit their protonation or decrease their $pK_a$. In summary, these results indicate that the binding of CH$_3$THF to the active site of MeTr increases the $pK_a$ of N5, except when D75 is protonated. Therefore, binding may be accompanied by proton uptake if both aspartic residues are ionised or if only D160 is protonated. It could also be the case that the binding of CH$_3$THF to a D75- or D160-protonated protein triggers a proton re-shuffling, and a transfer occurs from an aspartic residue to the pterin ring without any proton uptake from the solvent.

4.6.2.2 ONIOM calculations

Although fragment calculations can account partially for the effect of the protein environment on the reactivity and properties of the ligand, this simple representation of the system and the constraints which need to be imposed on the mobility of side chains during optimisation may result in an inadequate model of the real system. In order to incorporate the effect of the complete enzyme environment, and allow for a greater flexibility of the structure during the optimisation process, ONIOM calculations were carried out.

During ONIOM calculations, the system to be studied is divided into several regions, each treated at a different level of theory. While the region of interest, in this case the ligand and active-site residues, is treated with expensive and more accurate QM approaches, less critical portions of the system, in this case the rest of the protein, are simulated at the MM level.

Four different structures of the protein (CNS_A, CNS_B, SHELX_A, and SHELX_B), obtained from two different refinements of a crystal structure of the dimeric form of the MeTr•CH$_3$THF complex, were analysed. Only water molecules H bonded to the ligand, and conserved among the four different structures, were included. In a first approach, only the complete CH$_3$THF was included within the QM region, while water molecules and the complete protein were simulated using MM (see section 4.5.3 for further details).
Five single-protonated and ten double-protonated species were analysed (Figure 4-13). The structures were optimised at the HF/6-31G* level, using Amber for the MM region, and the mechanical embedding implementation of ONIOM. The final energy values showed that for two of the protein conformations (CNS_A and SHELX_B) O4H⁺-CH₃THF is the preferred single-protonated species, while the other two (CNS_B and SHELX_A) indicate that N5H⁺-CH₃THF is preferred (Figure 4-13). In the case of the double-protonated structures, N5H⁺-D75H, O4H⁺-D75H and N5H⁺-D160H had the lowest energies.

**Figure 4-13 ONIOM energies of different protonated forms of the MeTr•CH₃THF complex.**

Total ONIOM energy (HF/6-31G*:Amber, mechanical embedding) of different single- and double-protonated forms of the MeTr•CH₃THF complex. Only the complete ligand was included within the QM region, and four different starting structures of the enzyme were analysed (CNS_A, CNS_B, SHELX_A, and SHELX_B). Relative energies were calculated using the N1H⁺ and the N5H⁺-D75H complexes as reference structures for the single- and double-protonated species, respectively.
As the variability observed in the energy of different protein starting structures did not allow confident identification of the most stable protonated species, a larger QM region was defined in order to provide a better representation of the active site (see section 4.5.3 for further details). The side chains of residues D43, D45, D75, N96, D160 and N199, together with two water molecules were included within the QM region (Figure 4-6). Only those protonated states that presented low energies in the previous calculations, including the single-protonated N1H\(^+\), N5H\(^+\) and O4H\(^+\) species and the double-protonated N5H\(^+\)-D75H, N5H\(^+\)-D160H, and O4H\(^+\)-D75H forms were analysed further. Also, in this model, incorporation of active-site aspartic residues within the QM region allowed study of new single (D75H, and D160H) and double (D75H-D160H) protonated structures that were not accessible before.

The results of the ONIOM calculations with the extended QM region seem to indicate that the aspartic residues of the active site have a higher pK\(_a\) than the ligand CH\(_3\)THF (Figure 4-14). When a single proton is added to the active site, it preferentially goes to D160, and when a second proton is added, the double-protonated D75H-D160H species is the most stable one. These results suggest that in order to protonate the ligand CH\(_3\)TF, all aspartic residues within the active site must be neutralized first.

The results for the ONIOM energies do not agree with those of our previous fragment calculations. In the case of the single-protonated species, fragment calculations predict that the N5 position of CH\(_3\)THF is the preferred protonation position, and not the aspartic acid residues. However, both methods suggest that the double-protonated species D75H-D160H is the most stable one. Taking both studies into consideration, it seems that three different positions within the active site of the MeTr•CH\(_3\)THF complex, N5 of CH\(_3\)THF, D75, and D160, are readily accessible for protonation and have comparable pK\(_a\)s. Further studies on the system are required to predict accurately the relative order of protonation of these positions.
Figure 4-14 ONIOM energies of MeTr•CH₃THF complexes (larger QM region).

Total ONIOM energy (HF/6-31G*:Amber, mechanical embedding) of different single- and double-protonated forms of the MeTr•CH₃THF complex. Four different starting structures of the protein were analysed (CNS_A, CNS_B, SHELX_A, and SHELX_B), with only the extended QM region (Figure 4-6) allowed to relax during the optimisations. Relative energies were calculated using the N1H⁺ and the N5H⁺-D75H complexes as reference structures for the single- and double-protonated species, respectively.

4.6.2.3 Free Energy Calculations Using Thermodynamic Integration

It is well known that the protein active site provides a special environment for a reaction to occur, by modifying the reactivity of substrates and orchestrating the reaction. Taking into account the presence of several potential protonation sites within the pterin ring of CH₃THF, free energy calculations may be useful to indicate whether the protein environment has been designed to favour one over another, in a way different from that encountered in solution.
Initially, the interaction of bulk solvent with different protonated states of CH$_3$THF was studied, followed by calculations involving the protein environment.

4.6.2.3.1 Relative solvation energy

The transfer of a molecule from the gas phase to an aqueous environment is accompanied by a change in free energy, the solvation energy. This energy change reflects the stability of the molecule within the new environment; the larger the released energy, the more stable the molecule is in water.

Different protonated species of CH$_3$THF are expected to have different solvation energies. Although the direct calculation of solvation energies is a difficult process, the relative solvation energy of two different species can be calculated simply using a perturbational approach. During this calculation, one molecule is mutated into the other, and the free energy of the process is determined according to the thermodynamic integration approach (see section 4.5.4 for further details).

In this case, the free energies of the alchemic transformations between different protonated states of CH$_3$THF were calculated (Figure 4-15). The final results for these calculations show that the protonated species with the largest solvation energy is N1H$^+$-CH$_3$THF (Figure 4-16).

$$
\Delta G_{\text{solv}}^1 = \Delta G_{\text{solv}}^2 = \Delta G_{\text{transf}}(\text{water}) - \Delta G_{\text{transf}}(\text{gas})
$$

Figure 4-15 Calculation of relative energies of solvation.

Schematic representation of the transformation process used to determine the relative solvation energies of different protonated forms of CH$_3$THF. The thermodynamic cycle used to calculate the energy difference between the N5- and the N1-protonated forms of CH$_3$THF is shown here as an example. The alchemic transformations were carried out both in vacuo and in water, and these values were used to estimate the relative solvation energies. Finally, relative solvation energies were obtained the equation shown.
Figure 4-16 Relative solvation free energies.

Free energy of solvation differences ($\Delta \Delta G_{solv}$) between pairs of protonated forms of CH$_3$THF. Free energy calculations were done using the thermodynamic integration approach as implemented in AMBER 8 [651]. Five independent transformations for each pair of protonated species were carried out both in vacuo and in water, and error bars indicate the standard deviation for the final energies (see section 4.5.4 for further details).

This indicates that the N1-protonated molecule has the most favourable interactions with the water environment, probably due to the charge distribution pattern over the pterin ring and its accessibility to solvent molecules. However, this result does not indicate that N1H$^+$-CH$_3$THF is the most stable species in solution, as the internal energy of the molecule has not been taken into account.

The pattern observed for the relative free energies of solvation closely resembles the energy difference between in vacuo and implicit solvent PCM QM calculations of different protonated forms of CH$_3$THF (Figure 4-17).
Figure 4-17 Free energy of solvation and interaction energies.

Comparison between relative free energies of solvation (blue) and QM solvent interaction energies (HF red, and B3LYP green) for a series of protonated forms of CH$_3$THF. Free energies were obtained using the MM thermodynamic integration approach, and different protonated forms of CH$_3$THF were interconverted both in vacuo and in explicit solvent. QM interaction energies were calculated as the difference between in vacuo and implicit solvent PCM calculations. The structures were first optimised at the HF/6-31+G** PCM level, and single-point calculations were carried out in vacuo and at the B3LYP/6-31+G** level. Relative energies were determined using N5H$^+$-CH$_3$THF as the reference structure.

In both cases N1H$^+$-CH$_3$THF presents the most favourable interactions with the solvent and N5H$^+$-CH$_3$THF the least stabilizing ones. HF relative interaction energies are closer to the solvation free energies than the B3LYP values. The main difference between the MM and the QM energy pattern is the relative positioning of O4H$^+$-CH$_3$THF. Free energy calculations predict this species to be ~10 kcal/mol more stable than N8H$^+$-CH$_3$THF, while QM calculations suggest that they have similar solvent interaction energies.

Although N1H$^+$-CH$_3$THF might have the preferred charge distribution around the pterin ring to establish the most stable interactions with water solvent, both
theoretical calculations (see section 4.6.1) and experimental analysis [629] have shown that N5H⁺-CH₃THF is the preferred protonated form of the ligand, despite having the least favourable solvation energy. This is likely a result of the internal energy of the molecules, which favour N5H⁺-CH₃THF over N1H⁺-CH₃THF.

### 4.6.2.3.2 Free energy changes and the protein environment

After analysing the relative solvation energy of different protonated forms of CH₃THF in water, the effect of the protein environment was studied (Figure 4-18).

![Thermodynamic cycles used to determine relative protein-solvation energies (A) and relative binding energies (B).](image)

The free energies associated with the unphysical transformations between protonated forms of CH₃THF (ΔG\text{transf}) in different environments (gas, water, and protein) were calculated using the MM thermodynamic integration approach as implemented in AMBER 8 [651]. The transformations between N5H⁺-CH₃THF and N1H⁺-CH₃THF are shown as an example.

Two different free energy calculations were carried out, one similar to the solvation energy, to determine the free energy change associated with moving...
the molecule from vacuum to the protein active site (Figure 4-18A), and another set of calculations to study the passage of the ligand from a water environment to the protein active site, giving the binding energy (Figure 4-18B).

The alchemic transformations between protonated CH$_3$THF forms were done within the protein environment (see section 4.5.4.4 for details). When the relative free energy changes resulting from the transfer of the molecules from the gas phase to the active-site environment were calculated, a very similar profile to that of the solvation energies was obtained (Figure 4-19).

![Figure 4-19](image)

**Figure 4-19 Free energy of solvation in water and within the protein environment.**

Free energy differences for water-solvation ($\Delta G_{\text{water}} - \Delta G_{\text{gas}}$) (as in Figure 4-1) and proteinsolvation ($\Delta G_{\text{protein}} - \Delta G_{\text{gas}}$) of different protonated forms of CH$_3$THF. The transformations between protonated species were done using the thermodynamic integration approach in gas, water, and the protein environment. Three different protonation states of the enzyme active site were considered, including the unprotonated form (MeTr) and the single protonated Asp75-H and Asp160-H forms. Error bars indicate the standard deviation in the final energies of five independent transformation processes.
The best interactions are observed for the N1-protonated form of the ligand, either with water or the protein environment. Similarly, the N5H\(^+\)-form shows the least stable ones. These results suggest that the active-site environment, regardless of the protonation state of residues D75 and D160, is better prepared to accommodate the N1H\(^+\)-protonated form of the ligand, similar to what is found in aqueous solution. However, this finding does not imply that the absolute stabilization energy within the active site is similar to that found in an aqueous environment, only that the relative interactions energies of different protonated forms of CH\(_3\)THF are similar.

In order to compare the affinity of MeTr for different protonated forms of CH\(_3\)THF, relative binding free energies were determined. These were calculated as the energy difference between the transformations in water ($\Delta G_{\text{water}}$) and the transformations in the protein environment ($\Delta G_{\text{protein}}$) (Figure 4-20).

![Figure 4-20 Relative binding free energies.](image)

Binding free energy differences ($\Delta \Delta G = \Delta G_{\text{transf(protein)}} - \Delta G_{\text{transf(water)}}$) (Figure 4-18) for different protonated forms of CH\(_3\)THF binding to MeTr. The transformations between protonated species were done using the thermodynamic integration approach. Three different protonation states of the enzyme active site were considered, comprising the unprotonated form (MeTr) and the single protonated Asp75-H and Asp160-H forms. Error bars indicate the standard deviation in the final energies of five independent transformations processes.
Regardless of the protonation state of active-site aspartic residues, the differences in free binding energies found for the different protonated forms of CH₃THF are quite small, in the order of 2-3 kcal/mol, and the errors associated with them are approximately the same magnitude. Therefore, no clear conclusions could be drawn from the present studies. The intrinsic errors usually associated with this kind of free energy calculation suggest that simulations with for the same system different and/or better parameters are not likely to provide significantly different results. In consequence, a different approach involving more accurate QM free energy calculations is necessary to establish accurately the differences in binding free energies among several protonated forms of CH₃THF.

4.7 Conclusions

In the present chapter I have presented a series of studies intended to elucidate the protonation and activation of the ligand CH₃THF within the active site of MeTr. Despite the sequence and structural homologies between MeTr and the CH₃THF-binding domain of MetH, significantly different protonation mechanisms have been proposed. Experimental results on both systems have lead to significantly different, if not contradictory, explanations for the activation step of CH₃THF by protonation. While the binding of CH₃THF to MeTr has been said to be accompanied by a decrease in the pKₐ of N5 and a net proton uptake from the solvent, binding to MetH has been associated with a decrease in the pKₐ of N5 and no proton exchange with solution. Careful comparison of the active sites of both enzymes, and the nature of the interactions between the ligand and active site residues, clearly suggests that these two systems are likely to have a common mechanism. Therefore, a series of computational studies was carried out in order to define the most likely protonation state and activation of CH₃THF within the active site of MeTr.

The accuracy of QM calculations to predict the protonation position and pKₐ of 6-methyl-pterin derivatives, including CH₃THF, was initially assessed. It was found that HF/6-31+G** calculations with implicit solvent representation (PCM) are adequate to predict qualitatively the preferred protonation position of
several pterin-derivatives, and also provide a good correlation between final protonation energies and experimentally determined $pK_a$ values.

In order to incorporate the effect of the protein environment on the protonation properties of CH$_3$THF, QM calculations were performed on a fragment representation of the ligand and active site. This simplified representation included only the pterin ring of CH$_3$THF and a number of active-site residues H-bonded to the ring (Figure 4-5). Final energies suggested that when a single proton is added to the system, it preferentially goes to the N5 position of the N5,6-dimethyl-5,6,7,8-tetrahydropterin ring, as observed in aqueous solution. However, when two protons are added, the double-protonated species N5H$^+$-D75H, N5H$^+$-D160H, and D75H-D160H, all show very similar protonation energies, and it is unclear if there is an evidently preferred one in solution.

From QM calculations of the fragment representation of the MeTr•CH$_3$THF system it was estimated that the $pK_a$ of the N5 atom of the N5-methyl-5,6,7,8-tetrahydrofolate ring may increase $\sim$3 $pK_a$ units upon binding to the unprotonated active site, or by $\sim$1.6 $pK_a$ units if residue Asp160 is protonated. However, binding to the Asp75-protonated form may result in a decrease of the N5 $pK_a$. Therefore, binding of the ligand to the active site of the enzyme could result in proton uptake from the solution only if Asp75 is ionised. As the protonation state of the active-site aspartic residues is not known, further analysis are needed to accurately establish a preferred protonation pattern.

In a complementary approach to analyse the influence of the complete MeTr protein environment on the protonation state of CH$_3$THF, ONIOM calculations were carried out. This time the results showed that the aspartic acid residues D75 and D160 are expected to protonate before the ligand. It seems necessary, therefore, to predict the $pK_a$ of these groups before focusing attention on the activation and protonation of the ligand. The interaction between D160 and the pterin ring of CH$_3$THF resembles that found within dihydrofolate reductase, where Asp27, the ‘pterin hook’, interacts with the pterin ring of folate or DHF. Cummins and Greedy [508] used a fragment approach to analyse the most likely substrate and active-site protonation sites in dihydrofolate reductase. Their calculations showed that for both folate and DHF the carboxyl oxygen of Asp27
is expected to protonate first, followed by protonation of N8 (folate) or N5 (DHF); supporting the preliminary results on MeTr.

Finally, the MM thermodynamic integration method was used to calculate relative free energies for different protonated forms of the ligand via a thermodynamic cycle approach. Relative solvation energies showed that the N1-protonated form of CH$_3$THF presents the most favourable charge distribution over the pterin ring to interact with the water environment, while the N5 protonated form shows the least favourable interactions. Similar patterns were found when free energy changes were compared between the protein environment and the gas phase. These results seem to indicate that different protonated forms of CH$_3$THF establish basically the same sort of interactions whether in water or within the active site (e.g. the protein environment does not favour one protonation state over another in a differential way to that found in solution). Attempts to determine the relative free binding energies of the different protonated species of CH$_3$THF to the active site of MeTr were unsuccessful. Small energy differences and large associated errors prevented a definitive assessment of the results. The intrinsic errors of MM thermodynamic integration calculations suggest that a different technique will be required to accurately predict small difference in binding energies, such as those of this system.

In summary, structure and sequence comparisons between MeTr and MetH clearly suggest that both enzymes will employ a similar mechanism for the protonation and activation of the CH$_3$THF substrate. However, after several computational approaches, I have been unable to resolve whether binding of the ligand is accompanied by proton uptake from the solvent or not. The importance of the aspartic acid residues D75 and D160 and their protonation state is now evident, and it is a problem that needs to be addressed before focusing on the protonation state of the ligand.