ARTICLE

OBC
www.rsc.org/obc

A proposed model of *Mycobacterium avium* complex dihydrofolate reductase and its utility for drug design †

Prashant S. Kharkar and Vithal M. Kulkarni*
Pharmaceutical Division, Institute of Chemical Technology, University of Mumbai, Matunga, Mumbai, 400 019, India

Received 12th December 2002, Accepted 3rd February 2003
First published as an Advance Article on the web 13th March 2003

A homology model of *Mycobacterium avium* complex dihydrofolate reductase (MAC DHFR) was constructed on the basis of the X-ray crystal structure of *Mycobacterium tuberculosis* (Mtb) DHFR. The homology searching of the MAC DHFR resulted in the identification of the Mtb DHFR structure (PDB 1DF7) as the template for the model building. The MAC enzyme sequence was aligned to that of the Mtb counterpart using a modified Needleman and Wunsch methodology. The initial geometry to be modeled was copied from the template, either fully or partially depending on whether the residues were conserved or not, respectively. Using a randomized modeling procedure, 10 independent models of the target protein were built. The cartesian average of all the model structures was then refined using molecular mechanics. The resulting model was assessed for stereochemical quality using a Ramachandran plot and by analyzing the consistency of the model with the experimental data. The structurally and functionally important residues were identified from the model. Further, 5-deazapteridines recently reported as inhibitors of MAC DHFR were docked into the active site of the developed model. All the seven inhibitors used in the docking study have a similar docking mode at the active site. The network of hydrogen bonds around the 2,4-diamino-5-deazapteridine ring was found to be crucial for the binding of the inhibitors with the active site residues. The 5-methyl group of the inhibitors was located in a narrow hydrophobic pocket at the bottom of the active site. The relative values of the three torsion angles of the inhibitors were found to be important for the proper orientation of the inhibitor functional groups into the active site.

Introduction

Mycobacterium avium complex (MAC), a group of microorganisms, causes one of the most significant systemic bacterial infections in patients with acquired immunodeficiency syndrome (AIDS). About 60–70% of the patients with advanced AIDS are infected with MAC.[1–4] The CD4+ cell counts in patients with advanced AIDS may become reduced to below 100, and a correlation has been reported between the CD4+ cell count and MAC disease.[5] These microbes can adapt to the extremes of their surroundings, *e.g.*, pH and temperature, and are resistant to most antibiotics and available antimycobacterial agents.[3] There is an urgent need to develop antimycobacterial agents targeted against these microorganisms.

Dihydrofolate reductase (DHFR) is a key enzyme involved in cellular metabolism and catalyzes the NADPH-dependent reduction of dihydrofolate (FH_2) to produce tetrahydrofolate (FH_4). Tetrahydrofolate and its derivatives serve as one carbon donors in the synthesis of purines, pyrimidines and several amino acids. Thus the inhibition of DHFR leads to a cellular deficiency of tetrahydrofolate cofactors, disrupting the biosynthesis of purines and pyrimidines and resulting in cell death.[6,7] DHFR has been identified as a target for the action of several classes of drugs important in the treatment of malaria, bacterial infections, parasitic infestations and cancer. Examples of such drugs include trimethoprim, pyrimethamine and methotrexate. Availability of the crystal structures of these inhibitors with DHFRs from various sources resulted in the identification of potential binding interactions with the enzyme and this information was used for the design of better analogs.[8–10]

In the late 1990s significant progress was made on the design of MAC DHFR inhibitors.[11–15] Recently Suling *et al.* reported a class of 2,4-diamino-5-deazapteridine inhibitors of MAC DHFR as potential antimycobacterial agents.[14] They have successfully identified a potent inhibitor with a selectivity index of 2300 [IC_{50} (hDHFR)/IC_{50} (MAC DHFR)]. Pharmacophore hypotheses were developed for this series of MAC DHFR and human DHFR inhibitors.[16] The pharmacophore, in the absence of the three-dimensional (3D) structure of MAC DHFR, may provide important information regarding the 3D arrangement of chemical features in a molecule that is essential for major binding interactions with MAC DHFR.

Although the amino acid sequences of DHFRs from bacteria, fungi and mammals have been characterized and substrate specificity and inhibitor selectivity of DHFR have been investigated through structure-based approaches,[17,18] a detailed structure–function analysis has not been rigorously approached in mycobacteria. The availability of the crystal structure of *Mycobacterium tuberculosis* (Mtb) DHFR recently has prompted the structure-based design of mycobacterial DHFR inhibitors. In the absence of the crystal structure of MAC DHFR, it is of great interest to understand the nature and structural requirements of its inhibitor-binding site. These studies will be facilitated by information about the 3D structure, which would allow a detailed analysis of the enzyme–inhibitor interactions and help the rational design of new MAC DHFR inhibitors as antimycobacterial agents. Here we elucidate the 3D structure of MAC DHFR by so-called comparative protein modeling or homology modeling to gain major structural insights into the binding mode of the 5-deazapteridine inhibitors of MAC DHFR and to assist the design of novel inhibitors using the information obtained from these studies.

The mere knowledge of a protein's sequence, or primary structure, does not allow a detailed understanding of its function. The unique, well-defined 3D structure of a protein allows scientists to gain insight into the active site of the protein or the way it interacts with small molecules and other proteins. The 3D structure also dictates the way in which the protein performs its biological function. With the recent genome sequencing revolution, the determination of complete genome sequences of various organisms has already become routine. However, the

† Electronic supplementary information (ESI) available: details of the calculations. See http://www.rsc.org/suppdata/ob/b2/b212211a/

DOI: 10.1039/b212211a

experimental determination of the 3D structure of the proteins encoded in these genomes is currently a very laborious process that is often hampered by difficulties in obtaining sufficient protein, diffracting crystals and many other technical aspects. In some cases it can take years before the structure of a protein is determined, and in other cases, such as membrane proteins, current methods are not always applicable. The number of the solved 3D structures increases only slowly compared to the rate of sequencing novel cDNAs, and no structural information is available for the vast majority of protein sequences registered in the SWISS-PROT database. Thus, predictive methods have gained much interest.

With the aid of advanced computational resources and quantum and statistical mechanics, protein models and simulations of their function could be obtained. The basic assumption is that the information the protein needs in order to fold into its unique 3D structure lies entirely in its amino acid sequence.[19] It is widely accepted that the native 3D structure of a protein has the lowest free energy possible for its combination of amino acids. Thus, in principle, finding the unique 3D structure of a protein given its amino acid sequence alone is a computable problem. However, protein structure prediction *in silico* has proven to be a very difficult task. It is not yet fully understood how a protein folds *in vivo*, or what are the precise energetic determinants of the protein folding.

Proteins from different sources and sometimes with diverse biological functions can have similar sequences, and it is generally accepted that the high sequence similarity is reflected by distinct structure similarity. Indeed, the relative mean square deviation (RMSD) of the C_α co-ordinates for protein cores sharing 50% residue identity is expected to be around 1 Å. This fact served as the premise for the development of comparative protein modeling[20–22] (often called modeling by homology or knowledge-based modeling). Comparative model building consists of the extrapolation of the structure for a new (target) sequence from the known 3D structure of the related family members (template structures).

While the high-precision structures required for detailed studies of protein–ligand interactions can only be obtained experimentally, theoretical protein modeling provides biologists with "low-resolution" models which hold enough essential information about the spatial arrangement of important residues to guide the design of experiments. The rational design of many site-directed mutagenesis experiments could, therefore, be improved if more of these "low-resolution" theoretical model structures were available.

In the present investigation, we have constructed a homology model of MAC DHFR on the basis of the known Mtb DHFR crystal structure. Such a model should include structural information from all known enzymes and lead to a more accurate model structure. The model can be used to explain major interactions of the inhibitors with MAC DHFR. For this purpose, the docking of some MAC DHFR inhibitors into the active site of the model was explored. The structurally and functionally important residues identified allowed for a better understanding of the structure–function relationships of the enzyme. The mode of the enzyme–inhibitor interactions would be useful in developing more potent antimycobacterial agents.

Results and discussion

Modeling of MAC DHFR and general features of the model

In the present investigation, a model of MAC DHFR in a complex with methotrexate and NADPH was constructed on the basis of the Mtb DHFR crystal structure.

The amino acid sequence of MAC DHFR was used as the query sequence for searching the MOE-Homology databank. The homology searching resulted in the identification of the Mtb DHFR structure (~ 70% sequence identity with the MAC DHFR) as the template structure. Both sequences were aligned using modified Needleman–Wunsch methodology.[23] The sequence alignment is shown in Fig. 1. There were two major insertions in the MAC DHFR model from residues 2–5 (region R1) and from residues 91–94 (region R2). The residues 168–181 (region R3) have no counterpart in the Mtb DHFR and formed an extended turn structure on the surface of the model. These insertions, exposed on the molecular surface, did not cause significant perturbations in the protein backbone folding or structural core relative to those in Mtb DHFR. The residues 168–181 in MAC DHFR were modeled with little confidence, as the structural information for these residues was absent. This segment contained mostly polar or charged residues. The RMSD between the MAC DHFR model and the Mtb DHFR X-ray structure was found to be 1.053 Å for the C_α atoms, 1.137 Å for the main chain atoms and 1.5445 Å for all atoms.

```
                    10                  20                  30
MAC   MTRAEVGLVW   AQSTSGVIGR   GGDIPWSVPE
Mtb   M....VGLI W   AQATSGVIGR   GGDIPWRLPE
      R1
                    40                  50                  60
MAC   DL TRFKEVTM   GHTVI MGRRT   WESLPAKVRP
Mtb   DQAHFRE ITM   GHTIVMGRRT   WDSLPAKVRP

                    70                  80                  90
MAC   LPGRRNVVVS   R RPDFVAEGA   RVAGSLEAAL
Mtb   LPGRRNVVLS   RQADFMASGA   EVVGSLEEAL

                   100                 110                 120
MAC   AYAGSDPAPW   VIGGAQI Y LL   ALPHATRCEV
Mtb   ....TSPETW   VIGGGQVYAL   ALPYATRCEV
      R2
                   130                 140                 150
MAC   TEI E IDLRRD   DDDALAPALD   DSWVGETGEW
Mtb   TEVDIGLPRE   AGDALAPVDL   ETWRGETGEW

                   160                 170                 180
MAC   LASRSGLRYR   FHSYRRDPRS   SVRGCSPSRP S
Mtb   RFSRSGLRYR   LYSYHRS...   ..........
                                          R3
```

Fig. 1 Sequence alignment of MAC DHFR and Mtb DHFR. Conserved residues are shown in red. Regions R1 and R2 are insertions in MAC DHFR as compared to Mtb DHFR. Region R3 has no counterpart in Mtb enzyme.

To evaluate the quality of the modeled structure, various tools like a Ramachandran plot, from the MOE-Stereochemical quality evaluation functionality were used. After refinement, 75.69% of the residues were in the core, 19.88% were in the allowed region, 3.8% were in the generously allowed region and 0.55% were in the outside or disallowed region. Thus, a total of 95.7% of the residues of the modeled structure after minimization were in the allowed region, which indicated that the backbone dihedral angles ϕ and ψ in the model were reasonably accurate.

The MAC DHFR model exhibits the same general fold (present in DHFRs from other species) that is dominated by a central β-sheet with flanking α-helices.[24] The central β-sheet is made up of seven parallel strands and a C-terminal antiparallel strand. The secondary structure elements, loops and turns of MAC DHFR are shown in Fig. 2.

Regarding the protonation state of the key residues in the MAC DHFR structure, Asp31 (Asp27 in Mtb DHFR) was treated as negatively charged aspartate. Asp 31 is important for catalysis and is conserved in most of the DHFRs. No attempts were made to change the protonation states of the other key residues in MAC DHFR. The co-ordinates of most of the conserved residues were copied directly from the Mtb DHFR crystal structure and all of them were treated as such. The cofactor was treated as NADPH. All the inhibitors were treated as neutral molecules.

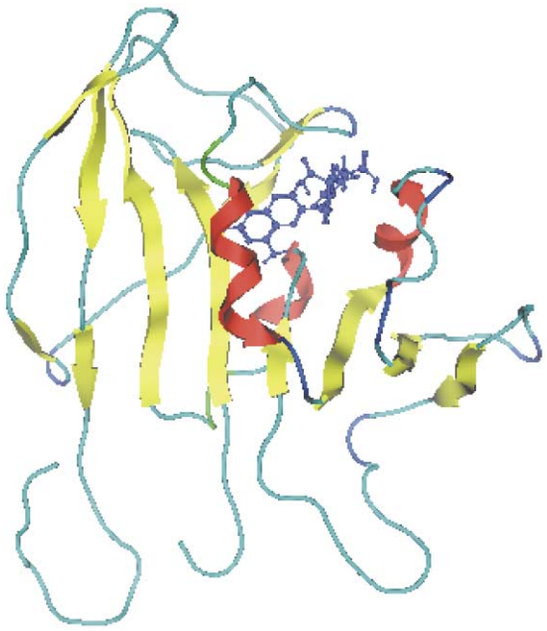


Fig. 2 Ribbon representation of MAC DHFR. Methotrexate is a ball-and-stick model.

NADPH binding to MAC DHFR

As in the Mtb DHFR crystal structure,[25] the cofactor is bound to MAC DHFR in an extended conformation. The adenine ring contacts the protein through hydrophobic interactions involving Val69 (Leu65 in Mtb DHFR), Leu110 and Ile107 on one side and through stacking interactions with the side chain of Arg71 on the other side. In addition, the adenine ring contacts residues Ser70, Gly84 and Gln106. The O2′-phosphate of the adenosyl ribose interacts with MAC DHFR through five H-bonds involving side chains of Ser70, Arg48 and Arg72 (Gln68 in Mtb DHFR) and the main-chain of Arg71. The pyrophosphate moiety forms a salt-bridge with Arg49 while the amide group of the nicotinamide ring forms three H-bonds with main-chain atoms of Ala11 and Ile18. The interactions of NADPH with the MAC DHFR are indicated in Fig. 3.

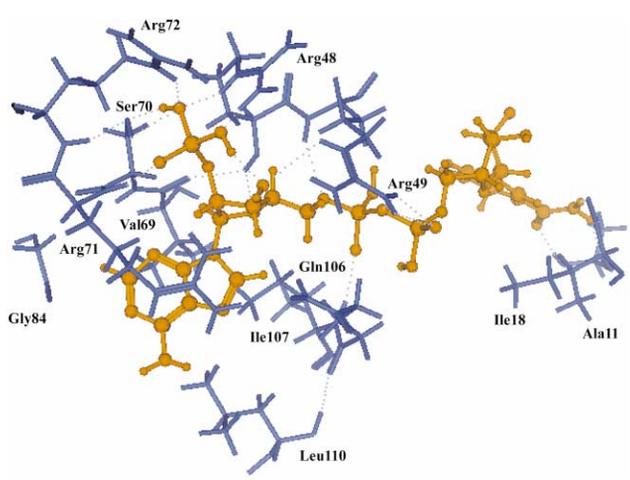


Fig. 3 The cofactor-binding site of MAC DHFR with NADPH (cofactor) shown as a ball-and-stick model. Hydrogen bonds are indicated with dotted lines.

Methotrexate binding to MAC DHFR

The 2,4-diaminopteridine ring of MTX is situated in a deep cleft and interacts strongly with MAC DHFR through five hydrogen bonds. Asp31 forms two H-bonds with N1 and the 2-NH_2 group. The negatively charged Asp31 is highly conserved and critical for catalysis.[26] Val9 (Ile5 in Mtb DHFR) and Ile102 form one and two H-bonds respectively between their main chain carbonyl oxygen atoms and the 4-NH_2 group of the MTX. This arrangement of H-bonds around the aminopteridine ring is highly conserved among the known DHFR structures from various species.[25] The aminopteridine ring also interacts with the protein through many hydrophobic interactions. On the side facing the nicotinamide ring of the cofactor, the aminopteridine ring is in contact with the Trp10, Ala11 and Ile24.

The other side of the aminopteridine ring interacts with Phe35 (π–π stacking), Leu32, Val9 and Ile102. The *p*-aminobenzoyl ring of MTX is in contact with Leu61, Pro55, Leu32 (Gln28 in Mtb DHFR), Phe35 and Leu54, which provide a hydrophobic environment for the phenyl group. Fig. 4 shows the arrangement of the MAC DFHR active site residues and their interactions with methotrexate.

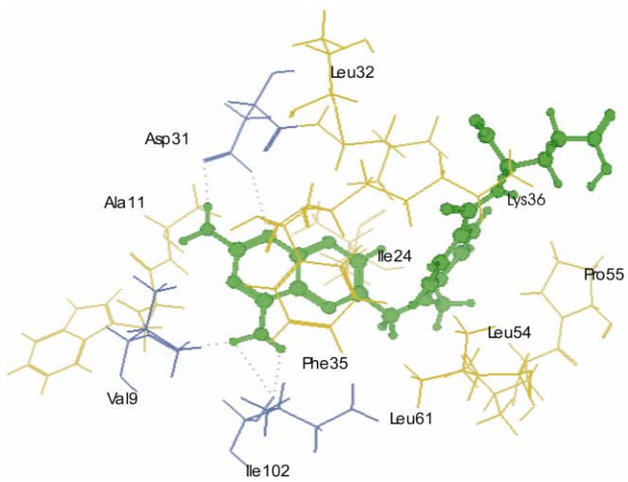


Fig. 4 Methotrexate docked into the active site of MAC DHFR. The residues shown in blue are involved in H-bonding (indicated with dotted lines) and the residues shown in gray are involved in hydrophobic interactions with MTX (green, ball-and-stick model).

The glutamate moiety of the MTX lies near the surface of the protein and interacts with Lys36, Leu32 and Thr33. The salt bridge interaction of the α-carboxyl group of MTX (with Arg32 and Arg60 of Mtb DHFR) is absent in MAC DHFR. The α-carboxyl group is oriented in a different direction, which is clearly evident from Fig. 5. This is due to the presence of hydrophobic Leu32 (Gln28 in Mtb DHFR) and Lys36 (Arg32 in Mtb DHFR) in the MAC enzyme.

In the Mtb DHFR (1DF7) structure, the B-factor of methotrexate was reported to be 20.9 $Å^2$. It was not taken into consideration. For the inhibitor methotrexate, the structure extracted from the crystal structure of Mtb DHFR was used as the starting point for docking into the MAC enzyme active site.

Modeling of 5-deazapteridines into the active site of MAC DHFR

The docked conformations of the 5-deazapteridine inhibitors (Table 1) were superimposed using the heavy atoms of the 5-deazapteridine ring and are shown in Fig. 6. The values of the energy difference between the docked conformation and the global minimum energy conformation are given in Table 2.

All the 7 inhibitors used in the docking study exhibited a similar docking mode in the active site of MAC DHFR. Most of the active site residues interacting with the deazapteridines were the same as those interacting with the methotrexate. The major interactions of the inhibitors (H-bonding and hydrophobic) with the MAC DHFR are indicated in Table 3. The docking mode of compound D28 (most active) is shown in Fig. 7.

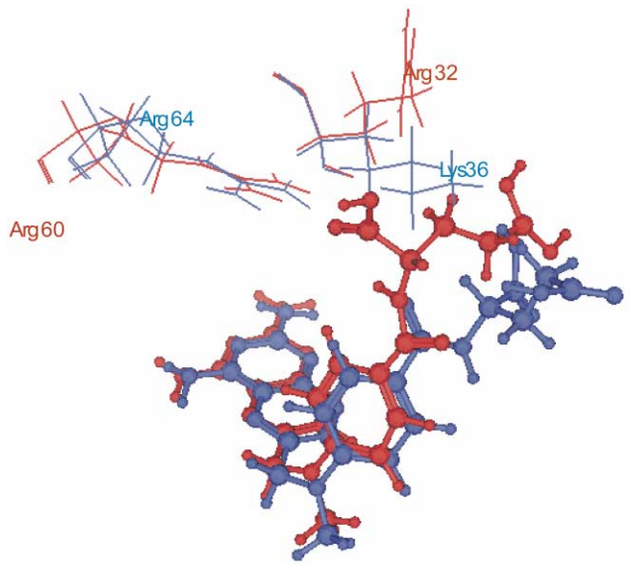


Fig. 5 Stereorepresentation of the conformations of methotrexate. The docked conformation into the active site of MAC DHFR (shown in blue) is superposed onto the Mtb DHFR X-ray structure conformation (shown in red). MAC DHFR residues (Lys36 and Arg64) are indicated in blue and the corresponding Mtb DHFR residues (Arg32 and Arg60) are shown in red. Leu32 (Gln28 in Mtb DHFR) is not shown.

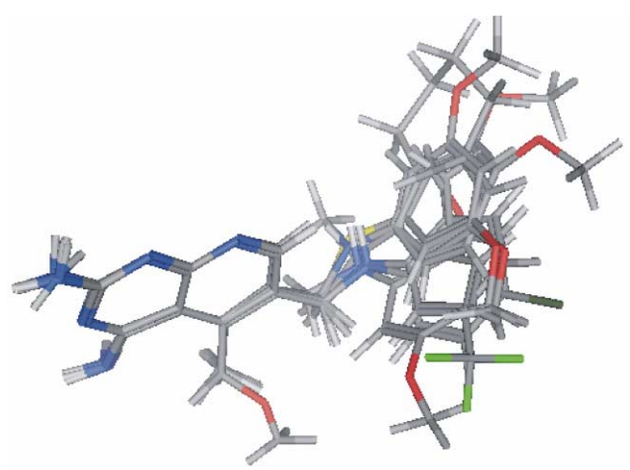


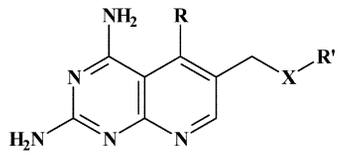
Fig. 6 Superposition of the docked conformations of the 5-deazapteridine inhibitors of MAC DHFR onto the corresponding heavy atoms of the deazapteridine ring using compound D28 as the template.

The 5-methyl group of the inhibitors was located in the hydrophobic pocket formed by the four residues Thr50, Ile102, Leu54 and Phe35. Since the pocket is narrow, the space adjacent to the 5-position of the aminopteridine ring is limited. Bulky substituents larger than a methyl group would probably produce significant steric clashes with the residues lining the hydrophobic pocket, altering the docked ligand conformation substantially and lowering the binding affinity, as in compound D29 in Table 4. The absence of a 5-methyl group in D26 indicated the loss of hydrophobic interactions with the residues of the hydrophobic pocket, resulting in lower binding affinity (MAC DHFR IC_{50} = 240 nM). The binding energies ($E_{binding}$) of each inhibitor in Table 4 were calculated as

$$E_{binding} = E_{complex} - (E_{ligand} + E_{receptor})$$

where E_{ligand} is the energy of the ligand corresponding to the overall minimum energy for the conformation search and $E_{receptor}$ is the energy of the receptor.

Table 1 Deazapteridine inhibitors of MAC DHFR used in the docking study



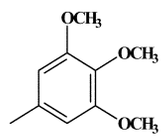
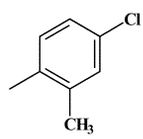
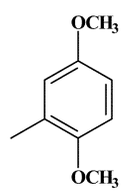
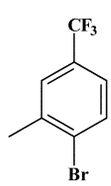
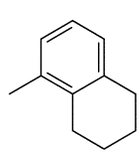
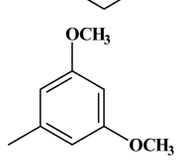
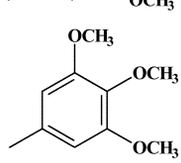
Compound	X	R	R′	IC_{50}/nM
D26	NH	H		240
D28	NH	CH_3		0.19
D29	NH	CH_2OCH_3		2400
D49	NH	CH_3		0.82
D52	NCH_3	CH_3		1.9
D68	CH_2	CH_3		1.5
D70	S	CH_3		4.5

Table 2 Energy differences (ΔE) (kcal mol^{-1}) between the docked conformer energy (E_1) and the global minimum energy (E_2)

Compound	E_1	E_2 [a]	ΔE
D26	10.628	3.324	7.304
D28	−9.738	−15.110	5.273
D29	229.408	4.763	224.645
D49	1.093	−4.274	5.364
D52	41.234	29.150	12.084
D68	14.218	5.572	8.646
D70	26.626	14.081	12.545

[a] The conformational search was performed using a systematic search protocol. The rotatable bonds in all molecules were searched from 0–359° in 10° increments. The minimum energy conformation thus obtained was minimized using an MMFF94 force field with distance-dependent dielectric and 12 Å nonbonded cutoff.

Table 3 Major interactions of the 5-deazapteridines with MAC DHFR

	Hydrogen bonding		Hydrophobic
Compound	Total no.	Residues involved	Residues involved
D26	2	Asp31, Ile102	Val9, Ala11, Phe35, Ile24, Leu32, Val58, Pro55
D28	3	Asp31, Val9, Ile102	Val9, Ile102, Ala11, Phe35, Ile24, Thr50, Leu54, Leu32, Val58, Pro55
D29	2	Asp31, Val9	Val9, Ile102, Ala11, Phe35, Ile24, Leu54, Leu62, Leu32, Val58, Pro55
D49	3	Asp31, Val9, Ile102	Val9, Ile102, Ala11, Phe35, Ile24, Thr50, Leu54, Leu32, Val58, Pro55
D52	1	Asp31	Val9, Ile102, Ala11, Phe35, Ile24, Thr50, Leu54, Leu32
D68	3	Asp31, Val9, Ile102	Val9, Ile102, Ala11, Phe35, Ile24, Thr50, Leu54, Leu32, Val58, Pro55
D70	3	Asp31, Val9, Ile102	Val9, Ile102, Ala11, Phe35, Ile24, Thr50, Leu54, Leu32, Val58, Pro55

Table 4 Binding energies of the 5-deazapteridines interacting with the active site of MAC DHFR

Compound	E_{steric}/kcal mol^{-1}	$E_{electrostatic}$/kcal mol^{-1}	E_{total}/kcal mol^{-1}
D26	−17.4099	−10.3971	−27.807
D28	−15.1128	0.838	−14.2748
D29	17.6497	−5.6511	11.9986
D49	−15.976	−2.1701	−18.077
D52	−12.2989	−1.1139	−13.4128
D68	−15.9527	−0.2278	−16.1805
D70	−14.9688	−0.2916	−15.2604

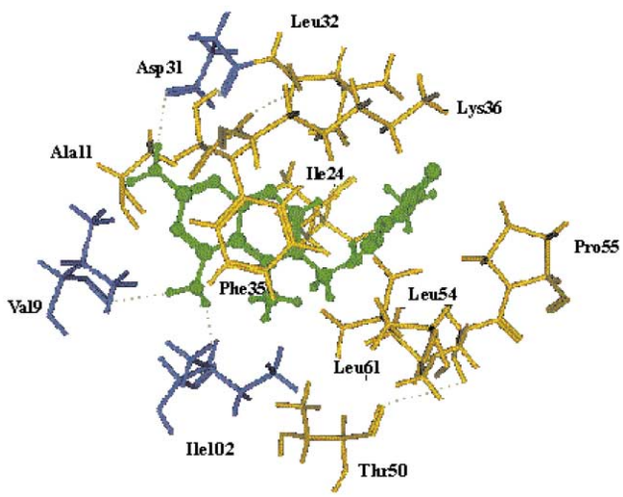
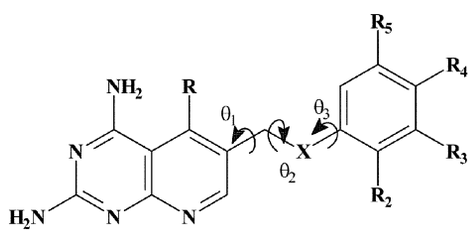


Fig. 7 Compound D28 docked into the active site of MAC DHFR. The residues shown in blue are involved in H-bonding (indicated with dotted lines) and the residues shown in gray are involved in hydrophobic interactions with D28 (green, ball-and-stick model).

The inhibitors have three principal degrees of freedom (θ_1, θ_2 and θ_3). The values of θ_1, θ_2 and θ_3 are indicated in Table 5. When X = N/NCH$_3$ (Table 1), the values of θ_3 range from 7.4 to 19.8°, θ_2 from 81.9 to 92.9° with D29 showing significant deviation (θ_2 = 61.7°) and of θ_1 from 0 to −10.3° with D26 (θ_1 = −50.8°), D29 (θ_1 = 14.8°) and D52 (θ_1 = −40.8°) showing significant deviations. The torsion angle θ_2 was particularly important as its value near 90° resulted in a conformation that positioned the two aromatic rings nearly perpendicular to each other which was necessary for the proper orientation of the interacting groups. In the case of compound D52 (θ_1 = −40.8°), the tetrahydronaphthalene ring was found to be located at a position slightly away from the entrance of the active site of MAC DHFR with subsequent loss of hydrophobic interactions with the corresponding hydrophobic residues. This is evident from the slightly higher value of the steric energy (−12.299 kcal mol^{-1}) (Table 4) as compared to other inhibitors except compound D29. In compounds with X = CH$_2$ (D68) and X = S (D70), the values of θ_2 lie near 90° while the values of θ_1 and θ_3 deviate considerably from the others. This results in the significant disposition of the aromatic ring (R′) from the corresponding position observed in the active compounds (D28, D49).

Table 5 Values of torsion angles (θ_1, θ_2, θ_3) of 5-deazapteridine inhibitors of MAC DHFR



Compound	θ_1/deg	θ_2/deg	θ_3/deg
D26	−50.8	89.3	15.2
D28	0.0	81.9	19.8
D29	14.8	61.7	12.2
D49	−10.3	84.5	11.3
D52	−40.8	92.9	7.4
D68	−33.8	74.7	33.3
D70	−50.1	95.2	3.5

Table 6 RMSD of the active site residues of the MAC DHFR

	RMSD/Å		
Compound	C_α	Main chain	All
D26	0.1846	0.2378	0.3189
D28	0.1143	0.1634	0.2553
D29	0.1465	0.1948	0.3276
D49	0.1318	0.1972	0.2769
D52	0.1642	0.2224	0.2345
D68	0.1327	0.2270	0.3415
D70	0.1238	0.1939	0.2946

The RMSD (Å) of the active site residues of the MAC DHFR are indicated in Table 6. For compounds with R = H (D26), and R = CH$_2$OCH$_3$ (D29), the RMSD values for C_α, main chain atoms and all atoms were higher than the corresponding values for compounds with R = CH$_3$. The higher RMSD could be attributed to the nature of the R group. In D29, the R group could not be accommodated properly due to the limited size of the hydrophobic cavity while in D26 (R = H), the molecule could not be anchored suitably in the active site without the CH$_3$ group. This was evident from the loss of important hydrogen bonding interactions with MAC DHFR

residues. Each of the H-bonds of the 4-NH_2 group in D26 and D29 with Val9 and Ile102 respectively was absent.

A pharmacophore model of the 5-deazapteridines has been proposed recently.[16] It constitutes the two ring nitrogen atoms (H-bond acceptor), the 5-methyl group (hydrophobic) and the aromatic ring (R_2) (ring aromatic) (Table 1). The 5-methyl group, a hydrophobic feature, occupies the hydrophobic cavity in the active site as described previously. We could not find any of the MAC DHFR residues directly interacting with the ring nitrogens featured as H-bond acceptors in the pharmacophore hypothesis. The ring aromatic feature correctly matched with the aromatic ring attached to X (R_2).

Experimental

Computational approaches

All molecular modeling studies described herein were performed on a 750 MHz Intel Pentium III processor running Windows 98 using Molecular Operating Environment (MOE) 2001.01 and 2002.03 molecular modeling software.[27]

Homology searching

The first step in building a protein model by homology is to identify proteins whose 3D structures are known and which are homologous to the protein of interest. MOE-Search PDB module searches for protein structures that are homologous to a query amino acid sequence. The search tool uses MOE's protein alignment capability to test the query sequence against each entry in the MOE homology databank. This databank is a library of alignments, which were generated by clustering the protein database, augmented by the sequence-only data available from public domain databases. Aligning a query sequence against a relatively small set of reliable and diverse pre-built alignments, rather than against all possible candidates individually, greatly increases the likelihood of uncovering remote homologous structures. The amino acid sequence of MAC DHFR[28] obtained from the EMBL database (Accession No. 030463) was used as the query sequence. The alignment parameters such as gap start, gap extend, amino acid substitution matrix, gonnet[29] and thresholds used for screening candidates homologues i) percent identity and ii) Z-score were used with their default values.

Percent identity score reflects both the degree of residue similarity between the query and the candidate, as well as the amount by which the query sequence is stretched in the alignment. The higher the threshold, the lower the number of matches and faster the search. The default value of 20 was used in the present study. The Z-score is an estimate of the statistical significance of the alignment score.

The threading options used in the homology searching were hydrophobic fitness score (Hfc)[30] and structure prediction.

Multiple sequence and structure alignment

The homology searching is followed by multiple sequence and structure alignment. The aim of this step is to match each residue in the target sequence to its corresponding residue in the template structure, allowing for insertions and deletions.[31] An underlying assumption is that the chains to be aligned are all related. The alignment procedure can use sequence-only (*i.e.*, residue identities) and sequence-derived information (*i.e.*, predicted secondary structure) as well as structure-based information when computing a multiple-sequence alignment.

MOE-Align implements a modified version of the alignment methodology originally introduced by Needleman and Wunsch.[23] In this approach, alignments are computed by optimizing a function based on residue similarity scores (obtained from applying an amino acid substitution matrix to pairs of aligned residues) and gap penalties. Penalties are imposed for introducing and extending gaps in the sequence with respect to another. The final optimized value is referred to as the alignment score. When aligning multiple sequences, the method attempts to optimize the so-called sum of pairs score, *i.e.*, the sum of all separate pair wise (one chain against one chain) scores.

The four stages involved in the multiple sequence alignment include:

a) Initial pair wise build-up
b) Round-robin realignment
c) Randomized iterative refinement
d) Structure-based realignment

All the default settings in the MOE-Align panel were used for the multiple sequence and structure alignment.

3D Model building

MOE-Homology uses a database driven methodology to build full-atom models for protein sequences based on template structure(s). The modeling procedure comprises three steps – 1) an initial partial geometry to be modeled is specified. The initial geometry is copied either from a specified primary template chain, or from selected regions of various different template chains. Where residue identity is conserved between the template and the model, all coordinates are copied; otherwise, only backbone coordinates are used.

2) Using a Boltzmann-weighted randomized modeling procedure[32] combined with specialized logic for the proper handling of insertions and deletions,[33] a user specified number of independent models of the target protein structure are built and written to a molecular database. Each of these intermediate models is evaluated by a residue packing quality function, which is sensitive to the degree to which the hydrogen bonding opportunities are satisfied.

3) A final model is determined and submitted to a user-specified degree of energy minimization. The coordinates of the final model are generated in one of two ways: either of the average of the atom coordinates of all intermediate models, or using the coordinates of the intermediate model that scored best according to the packing quality function.

Mycobacterium tuberculosis (Mtb) DHFR (PDB 1DF7) structure[25] was used as the primary template for building the homology model of MAC DHFR. Ten intermediate models were generated and the final model was taken as the cartesian average of all the intermediate models. The cofactor (NADPH) extracted from the Mtb DHFR X-ray structure was added to the MAC DHFR model. The binary complex was used for docking of the inhibitor (methotrexate, MTX) into the active site.

Docking

MOE-Dock was used to search for favorable binding configurations between a small, flexible ligand and a rigid macromolecular target, *i.e.*, protein. Searching was conducted within a specified 3D docking box, using simulated annealing as the search protocol and an MMFF94 forcefield.[34] The search protocol tries to optimize both purely spatial contacts and electrostatic interactions. MOE-Dock performs a series of independent docking runs and writes the resulting conformations and their energies to a molecular database file. During the calculations, the ligand molecule takes on conformations from the search trajectory.

A simulated annealing *run* consists of a sequence of Monte Carlo *cycles*, each cycle consisting of a number of *moves*, or steps. The temperature is held constant during each cycle, and is systematically reduced from one cycle to the next.

The energy of a configuration is the sum of the electrostatic and dispersive interaction energy between the ligand and the target as well as the intra-molecular energy of the ligand due to its conformation. To calculate the interaction energies between

the ligand and the target, MOE-Dock can use either the built-in potential function or *grid-based potential fields*.

For the grid-based docking of methotrexate, simulated annealing was used as the search protocol with a total of 25 runs, 6 cycles per run and 8000 steps per cycle and an initial temperature of 1000 K. The best docking orientation was selected on the basis of H-bonding interactions, binding energy, hydrophobic interactions and conformational energy difference. Once the final model structure has been built, its refinement is typically desirable. The model built so far may include steric clashes between atoms because of the different composition of residues in the new protein and the template protein. The resulting ternary complex was subjected to energy minimization using the following parameters: a distance-dependent relative permittivity ($\varepsilon = 4R$), nonbonded cutoff of 12 Å, an MMFF94 force field, and 200 steps of steepest descent followed by 200 steps of conjugate gradient and truncated Newton minimization until an energy gradient tolerance of 0.1 kcal mol^{-1} Å^{-1} was satisfied. Energy minimization was started with the insertions (residues 2–5, 91–94 and 168–181) sequentially with the rest of the structure fixed. The cofactor binding site, active site (8 Å in space from the cofactor and the inhibitor respectively) and the enzyme backbone were fixed and the structure minimized as described previously. In the next step only the active site and the cofactor-binding site were fixed and the structure minimized. The active site and the cofactor binding site backbones along with the atoms involved in the H-bonding interactions with the inhibitor and the cofactor were fixed in the preceding step. Finally, a conjugate gradient of the full protein was performed until the rms gradient energy was lower than 0.1 kcal mol^{-1} Å^{-1}. During the optimization procedure, the structure was checked periodically with the MOE-Stereochemistry Evaluation tool.

Docking of the inhibitors into the active site

The 5-deazapteridine inhibitors in Table 1 were built from the crystal structure of methotrexate (1DF7) using MOE-Molecule Builder. The molecule to be docked was placed into the active site of the developed model containing methotrexate and NADPH. The heavy atoms of the 5-deazapteridine ring were superposed onto the respective atoms of the methotrexate diaminopteridine ring system. Methotrexate was deleted from the active site and the molecule was docked into the active site of MAC DHFR as described previously using an MMFF94 force field and non-bonded cutoff 12 Å. The best orientation of the inhibitor into the active site was selected according to the criteria described previously. To optimize the enzyme–inhibitor interactions, the inhibitor was fixed and the residues in contact with the inhibitor (within 8 Å) were minimized using the steepest descents method until the gradient was less than 1 kcal mol^{-1} Å^{-1}. The system was then minimized using conjugate gradients to a maximum gradient of 0.1 kcal mol^{-1} Å^{-1}. The minimized complex was then subjected to 4 picosecond (ps) equilibration at 300 K and 20 ps hybrid Monte Carlo (HMC) conformational search at the same temperature and using a time step of 0.001 ps. The frames were saved every 100 iterations.

The HMC method accelerates the conventional molecular dynamics (MD) search by periodically performing Monte Carlo (MC) steps across torsional energy barriers, thereby addressing the barrier-crossing problem that is encountered in the standard MD algorithm. As a result, an HMC run often generates a more diverse set of conformers than a standard MD of equal length.

The minimum energy conformation was further minimized to an RMS gradient of 0.1 kcal mol^{-1} Å^{-1} as described previously. The minimized complex was used for subsequent analysis.

Conclusions

A homologous 3D model of MAC DHFR was built on the basis of the crystal coordinates of the known Mtb DHFR. The reliability of the model was assessed using Ramachandran plots and by analyzing the developed model with the experimental data on Mtb DHFR. The overall structure of the resulting MAC DHFR model is similar to the known crystal structure of the Mtb enzyme. The model retains the core structure characteristics of the DHFRs. All of the three insertions expose to the molecular surface. The structurally and functionally important residues such as cofactor and inhibitor binding residues, *e.g.*, hydrophobic Leu32 (hydrophilic Gln28 in Mtb DHFR) in the active site of the MAC enzyme, *etc.*, were identified from the model. These residues may significantly affect the binding of the inhibitors in the active site of the enzyme. These analyses may provide a basis for probing the structure–function relationship in the MAC DHFR. Site-directed mutagenesis of the specific residues should establish their importance for the substrate/inhibitor specificity. The model holds enough essential information about the spatial arrangement of important residues to guide the design of experiments. Identification of these residues will allow the rational design of specific inhibitors.

The careful building process and the subsequent comparison with the experimental results indicated that our model is reliable enough to be used for structure and function studies. Further investigation is still needed in order to gain insight into the molecular mechanism of the interaction between MAC DHFR and its inhibitors. Future work in this direction will be design of lead inhibitors based on the active site and the extensive structure–activity relationships.

Acknowledgements

The authors gratefully acknowledge support for this research from the University Grants Commission (UGC), New Delhi, under its DSA and COSIST programmes. PSK thanks UGC for the award of a senior research fellowship.

References

1 D. Schurmann, S. D. Nightingale, F. Bergmann and B. Ruf, *Infection*, 1997, **25**, 274–280.
2 M. Ghassemi, B. R. Andersen, V. M Reddy, P. R. Gangadharam, G. T. Spear and R. M. Novak, *J. Infect. Dis.*, 1995, **171**, 68–73.
3 V. M. Reddy, *Front. Biosci.*, 1998, **3**, d525–d531.
4 C. R. Horsburgh, Jr., *N. Engl. J. Med.*, 1991, **324**, 1332–1338.
5 S. D. Nightingale, L. T. Byrd, P. M. Southern, J. D. Jockush, S. X. Cal and B. A. Wynne, *J. Infect. Dis.*, 1992, **165**, 1082–1085.
6 L. F. Kuyper, in *Computer-Aided Drug Design: Methods and Applications*, eds. T. J. Perun and C. L. Propst, Marcel Dekker Inc., New York, NY and Basel, 1989, vol. 2, pp. 327–369.
7 J. McCormack, in *Comprehensive Medicinal Chemistry*, eds. C. Hansch and P. G. Sammes, 1990, vol. 2, pp. 271–298.
8 K. W. Volz, D. A. Matthews, R. A. Alden, S. T. Freer, C. Hansch, B. T. Kaufman and J. Kraut, *J. Biol. Chem.*, 1982, **257**, 2528–2536.
9 D. A. Matthews, R. A. Alden, J. T. Bolin, D. J. Filman, S. T. Freer, R. Hamlin, W. G. J. Hol, R. L. Kisluik, E. J. Pastore, L. T. Plante, N. Xuong and J. Kraut, *J. Biol. Chem.*, 1978, **253**, 6946–6954.
10 J. T. Bolin, D. J. Filman, D. A. Matthews, R. C. Hamlin and J. Kraut, *J. Biol. Chem.*, 1982, **257**, 13650–13662.
11 S. S. Meyer, S. K. Majumdar and M. H. Cynamon, *Antimicrob. Agents Chemother.*, 1995, **39**, 1862–1863.
12 W. J. Suling, R. C. Reynolds, E. W. Barrow, L. N. Wilson, J. R. Piper and W. W. Barrow, *Antimicrob. Agents Chemother.*, 1998, **42**, 811–815.
13 C. M. Shoen, O. Choromanska, R. C. Reynolds, J. R. Piper, C. A. Johnson and M. H. Cynamon, *Antimicrob. Agents Chemother.*, 1998, **42**, 3315–3316.
14 W. J. Suling, L. E. Seitz, V. Pathak, L. Westbrook, E. W. Barrow, S. Zywno-van-Ginkel, R. C. Reynolds, J. R. Piper and W. W. Barrow, *Antimicrob. Agents Chemother.*, 2000, **44**, 2784–2793.

15 A. Gangjee, E. Elzein, S. F. Queener and J. J. McGuire, *J. Med. Chem.*, 1998, **41**, 1409–1416.
16 A. K. Debnath, *J. Med. Chem.*, 2002, **45**, 41–53.
17 J. H. Chan, J. S. Hong, L. F. Kuyper, D. P. Baccanari, S. S. Joyner, R. L. Tansik, C. M. Boytos and S. K. Rudolph, *J. Med. Chem.*, 1995, **38**, 3608–3616.
18 V. M. Gokhale and V. M. Kulkarni, *J. Comput. Aided Mol. Des.*, 2000, **14**, 493–506.
19 C. Anfinsen, *Science*, 1973, **181**, 223–227.
20 T. L. Blundell, B. L. Sibanda, M. J. E. Sternberg and J. M. Thornton, *Nature*, 1987, **326**, 347–352.
21 M. S. Johnson, N. Srinivasan, R. Sowshamini and T. L. Blundell, *CRC Crit. Rev. Biochem. Mol. Biol.*, 1994, **29**, 1–68.
22 K. Srinivasan, K. Guruprasad and T. L. Blundell, in *Protein Structure Prediction: A Practical Approach*, ed. M. J. E. Sternberg, IRL Press at Oxford University Press, Oxford, 1996, pp. 111–140.
23 S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.*, 1970, **48**, 443–453.
24 D. A. Matthews, R. A. Alden, J. T. Bolin, S. T. Freer, R. Hamlin, N. Xuong, J. Kraut, M. Poe, M. Williams and K. Hoogsteen, *Science*, 1977, **197**, 452–455.
25 R. Li, R. Sirawaraporn, P. Chitnumsub, W. Sirawaraporn, J. Wooden, F. Athappilly, S. Turley and W. G. J. Hol, *J. Mol. Biol.*, 2000, **295**, 307–324.
26 E. E. Howell, J. E. Villafranka, M. S. Waren, S. J. Oatley and J. Kraut, *Science*, 1986, **231**, 1123–1128.
27 MOE 2002.03 software is available from Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Canada, H3A 2R7.
28 S. Zywno-van-Ginkel, T. P. Dooley, W. J. Suling and W. W. Barrow, *FEMS Microbiol. Lett.*, 1997, **156**, 69–78.
29 D. Fischer and D. Eisenberg, *Protein Sci.*, 1996, **5**, 947–955.
30 E. S. Huang, S. Subbiah and M. Levitt, *J. Mol. Biol.*, 1995, **252**, 709–720.
31 G. J. Barton, in *Protein Structure Prediction: A Practical Approach*, ed. M. J. E. Sternberg, IRL Press at Oxford University Press, Oxford, 1996, pp. 31–64.
32 M. Levitt, *J. Mol. Biol.*, 1992, **226**, 507–533.
33 T. Fechteler, U. Dengler and D. Schomberg, *J. Mol. Biol.*, 1995, **253**, 114–131.
34 (*a*) T. A. Halgren, *J. Am. Chem. Soc.*, 1990, **112**, 4710–4723; (*b*) T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519; (*c*) T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 730–748.