Journal of Computer-Aided Molecular Design, **12:** 241–257, 1998.
KLUWER/ESCOM

Dihydrofolate reductase: A potential drug target in trypanosomes and leishmania

Fabio Zuccotto[a], Andrew C.R Martin[b], Roman A. Laskowski[b,c], Janet M. Thornton[b,c] and Ian H. Gilbert[a,*]
[a]*Welsh School of Pharmacy, University of Wales Cardiff, King Edward VII Avenue, Cardiff CF1 3XF, U.K.*
[b]*Department of Biochemistry and Molecular Biology, Biomolecular Structure and Modelling Unit, University College London, Gower Street, London WC1E 6BT, U.K.*
[c]*Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, U.K.*

Received 26 July 1997; Accepted 9 October 1997

Key words: comparative molecular modelling, dihydrofolate reductase, drug design, protein structure

Summary

Dihydrofolate reductase has successfully been used as a drug target in the area of anti-cancer, anti-bacterial and anti-malarial chemotherapy. Little has been done to evaluate it as a drug target for treatment of the trypanosomiases and leishmaniasis. A crystal structure of *Leishmania major* dihydrofolate reductase has been published. In this paper, we describe the modelling of *Trypanosoma cruzi* and *Trypanosoma brucei* dihydrofolate reductases based on this crystal structure. These structures and models have been used in the comparison of protozoan, bacterial and human enzymes in order to highlight the different features that can be used in the design of selective anti-protozoan agents. Comparison has been made between residues present in the active site, the accessibility of these residues, charge distribution in the active site, and the shape and size of the active sites. Whilst there is a high degree of similarity between protozoan, human and bacterial dihydrofolate reductase active sites, there are differences that provide potential for selective drug design. In particular, we have identified a set of residues which may be important for selective drug design and identified a larger binding pocket in the protozoan than the human and bacterial enzymes.

Introduction

Dihydrofolate reductase (DHFR) is a ubiquitous enzyme that has wide-scale application as a drug target. DHFR is responsible for the reduction of dihydrofolate (Figure 1a) to tetrahydrofolate, an important co-factor in the biosynthesis of thymine. The reaction occurs at the N5-C6 bond and requires the transfer of two hydrogen nuclei to the substrate molecule. It is thought that one hydrogen is transferred to C6 from the reduced nicotinamide ring of the NADPH cofactor (net hydride ion transfer) while the other, which attaches to N5, is donated as a proton from a water molecule [1,2]. Inhibition of DHFR leads to cell death through lack of thymine as the cell has no alternative biosynthetic pathway [3]. The structure of DHFR is conserved poorly enough between species to make it possible to design selective inhibitors with therapeutic use. Currently, non-selective DHFR inhibitors, such as methotrexate (MTX) (Figure 1b), are used in the treatment of cancer, whilst selective inhibitors are used in the treatment of bacterial infections and malaria.

Trypanosomiasis and leishmaniasis are major health problems, especially in the Third World, and the current drugs available to treat these diseases suffer from poor clinical efficacy, side effects, increasing problems due to drug resistance and most require hospitalisation for administration. Therefore, there is an urgent need to develop new drugs for their treatment. We are interested in the design and synthesis of selective inhibitors for DHFR from protozoan parasites

*To whom correspondence should be addressed.

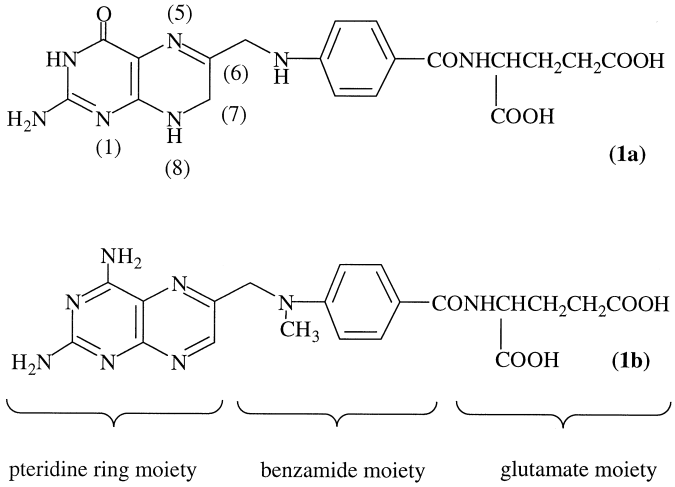

Figure 1. (a) The natural DHFR substrate dihydrofolate and (b) the universal DHFR inhibitor MTX.

responsible for trypanosomiasis and leishmaniasis as potential drugs to treat these diseases.

Little has been done to investigate DHFR as a potential therapeutic target in these organisms. However recently the crystal structure of DHFR from *Leishmania major* has been published [4], and sequence data have been obtained for DHFR from *Trypanosoma cruzi* [5], and *Trypanosoma brucei* [6]. Interestingly, in these protozoa, DHFR and thymidylate synthase form a bifunctional enzyme which is dimeric, whilst in mammals and bacteria DHFR is found as a monofunctional enzyme [7].

Homology models have been successfully used in a number of cases of structure-based drug design [8]. Therefore, as a first step towards the design of potential drugs to treat leishmaniasis and African and South American trypanosomiases, we built models of DHFR from *T. cruzi* and *T. brucei* based on the *L. major* structure and used these models together with the *L. major* structure in a comparison with the known structures of human and bacterial DHFRs to identify differences in the DHFR active site which may be exploited for the design of selective inhibitors of the protozoan enzymes.

Materials and methods

Coordinates of DHFR for human (PDB code: 1drf, binary complex with folate), *Lactobacillus casei* (3dfr, ternary complex with MTX and NADP) and *Escherichia coli* (4dfr, binary complex with MTX; 5dfr, uncomplexed; 6dfr, binary complex with NADP; 7dfr, ternary complex with folate and NADP) were obtained from the Brookhaven Protein Databank (PDB) [9]. The coordinates of the enzyme from *L. major* (ternary complex with MTX and NADPH) were kindly supplied by David A. Matthews.

Sequence alignment was performed using the Needleman and Wunsch algorithm [10] as implemented in the programme NW [11]. Phylogenies were calculated using the phylogeny inference package PHYLIP [12].

Homology modelling was carried out using MODELLER [13] to generate models of the *T. cruzi* and *T. brucei* DHFR structures. The A and B chains from the *L. major* DHFR were used as multiple templates. (Chain A is the chain from one monomer of the enzyme and chain B is from the other monomer: the two monomers are identical and the dimeric protein is symmetrical. However, in the crystal structure the A chain is better resolved than the B chain.) An alternative model of the *T. cruzi* enzyme was also generated using *E. coli* (4dfr), *L. casei* (3dfr), chicken liver (8dfr), human (1drf) and the *L. major* structures as multiple templates. Both models were assessed to be of similar quality as judged by PROCHECK [14]. It was decided to use the model generated from the *L. major* A and B chains in the following study as the *T. cruzi* enzyme has a much higher homology to the *L. major* enzyme than to other enzymes for which the structure is known. Multiple sequence alignment was performed using MULTAL [15] as implemented in CAMELEON [16] and then manually modified to move gaps from regions of helices and strands to those of loops. For the alignment procedure,

the *L. major* DHFR published sequence was used as there were slight discrepancies between the published sequence and that derived from the crystal structure (residues 126–127, Asp-Val in the sequence, were found to have been refined as Ala-Ala in the crystal structure; presumably there was poor electron density for these residues; residues 135–137 have been refined as Leu-Ala-Glu instead of Ala-Glu-Ala reported in the published sequence; residues 1–3, 115–120 and 198–203 are not defined in the crystal structure). Least-squares fitting of the structures and root mean square deviation (rmsd) calculations were performed using the McLachlan algorithm [17] as implemented in ProFit [18]. The protein structures were visualised using MACROMODEL [19], QUANTA [20], GRASP [21] and RASMOL [22]. Solvent accessibility surface area was calculated using the Lee and Richards algorithm [23] as implemented in the programme ACCESS [24].

In order to establish whether conformational change occurs during binding of ligands and to identify particularly mobile residues, rmsd values over active site residues were calculated for the *E. coli* DHFR crystal structures in different bound states.

The differences in shape between the human DHFR active site and the other enzyme active sites were studied by generating the surfaces of the active site cavities using the programme SURFNET [25]. The active sites were also investigated in terms of different electrostatic potential distributions. The Poisson–Boltzmann equation for the binding site region was qualitatively solved using the programme GRASP. In order to simplify the discussion of different residues, a standardised numbering scheme was introduced such that all structures were numbered relative to the human enzyme.

Results

DHFR sequence alignment

A study of the sequence alignment results suggests that there is a high degree of conservation amongst protozoan DHFRs; they display a higher degree of homology and identity to each other than to the enzymes from other species (Table 1). Their percentage identity is 46–58% compared with the identity between the bacterial DHFRs of 28%.

Interestingly, the identity between the human and protozoan and the human and bacterial enzymes is quite similar. In the first case there is an average identity of 26.52% with a maximum of 27.23% for *T. cruzi* DHFR, whilst for the bacterial enzymes there is a slightly lower average value of 21.88% with a maximum of 23.33% for *E. coli* DHFR; it might have been expected that the human and protozoan structures would have greater homology than the human and bacterial structures as both are eukaryotic organisms whilst bacteria are prokaryotic.

Table 1. Analysis of the individual sequence alignment results

	L. major	*T. cruzi*	*T. brucei*	*E. coli*	*L. casei*
Human	76.23[a]	77.26	76.66	77.10	76.65
	26.18[b]	27.23	26.14	26.88	25.27
L. major	–	83.84	83.78	78.84	75.45
	–	50.00	46.25	24.14	21.65
T. cruzi	–	–	89.53	78.05	75.33
	–	–	58.33	25.96	21.28
T. brucei	–	–	–	77.39	75.86
	–	–	–	23.33	20.42
E. coli	–	–	–	–	74.73
	–	–	–	–	27.88

[a] Percentage homology using the Dayhoff mutation matrix.
[b] Percentage identity over the alignment length.

Phylogeny

The rooted phylogenic tree of the studied DHFR structures (Figure 3) was obtained using the sequence alignment in Figure 2. It can be seen that the *T. cruzi* and *T. brucei* DHFRs are closely related to one another. They are somewhat less related to *L. major* DHFR. However, the protozoan structures are more closely related to each other than to either human or bacterial ones. Surprisingly, the bacterial DHFRs are closely related to each other according to the phylogenic tree although they have a low sequence homology.

Homology modelling

In order to generate the model of *T. cruzi* DHFR, manual modification of the alignment was carried out to move gaps from regions of helices and strands to those of loops. The alignment used is shown in Figure 4. The stereochemical qualities of the *T. cruzi* and *T. brucei* DHFR models generated by MODELLER were assessed using PROCHECK [14]. For *T. cruzi*, 88.9% of residues were in the most favoured regions [26] of the

```
T.brucei        MLSLTRILRKKIPVHELAGKISRPPLRPFSVVVASDEKGGIGDGGTIPWE-IPEDMQYFR
T.cruzi         -MSLF-----KIRMPETVAEGTRLALAAFSLVVAVDEHGGIGDGRSIPWN-VPEDMKFFR
L.major         -MSRAAA-RFKIPMPETKADFAFPSLRAFSIVVALDMQHGIGDGESIPWR-VPEDMTFFK
E.coli          -------------------------MISL--IAALAVDRVIGMENAMPWN-LPADLAWFK
L.casei         -------------------------TAF--LWAQNRNGLIGKDGHLPWH-LPDDLHYFR
Human           -------------------------VGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQ

T.brucei        RVTTNLRGKNVKPSPSKRNAVVMGRKTWDSLPP-KFRPLSNRLNVVLSRSATKEQLLAGI
T.cruzi         DLTTKLRGKNVKPSPAKRNAVVMGRKTWDSIPP-KFRPLPGRLNVVLSSTLTTQHLLDGL
L.major         NQTTLLRNKKP-PTEKKRNAVVMGRKTWESVPV-KFRPLKGRLNIVLSSKATVEELLAPL
E.coli          RNTL------DKPV-------IMGRHTWESI---G-RPLPGRKNIILSSQPGTD------
L.casei         AQTV------GKIM-------VVGRRTYESF---PKRPLPERTNVVLTHQEDYQ------
Human           RMTT------TSSVEGKQNLVIMGKKTWFSIPEKN-RPLKGRINLVLSRELKEP------

T.brucei        PDPIKRAEAANDVVAVNGGLEDALRMLVSKEHTSSIE---TVFCIGGGTIYKQALCAPCV
T.cruzi         PDEEKRNLHADSIVAVNGGLEQALRLLASPNYTPSIE---TVYCIGGGSVYAEALRPPCV
L.major         PEG-QRAAAAQDVVVVNGGLAEALRLLARPLYCSSIE---TAYCVGGAQVYADAMLSPCI
E.coli          ---------DRVT--WVKSVDEAIAACGD------VP---EIMVIGGGRVYEQFL--PKA
L.casei         ---------AQGA--VVVHDVAAVFAYAK------QHLDQELVIAGGAQIFTAFK--DDV
Human           ---------PQGAHFLSRSLDDALKLTEQPELANKVD---MVWIVGGSSVYKEAMNHPGH

T.brucei        NVLQAIHRTVVRPASNSCSVFFDIPAAGTKTPEGLELVRESITDERVSTGAGGKKYQFEK
T.cruzi         HLLQAIYRTTIRASESSCSVFFRVPESGTEAAAGIEWQRETISEELTSANGNETKYYFEK
L.major         EKLQEVYLTRIYATAPACTRFFPFPPENAATAWDL-----ASSQGRRKSEAEGLEFEICK
E.coli          QKL---YLTHIDAEVEGDTHFPDYEPDDWESV--F------SE-FHDADAQNSHSYCFKI
L.casei         DTL---LVTRLAGSFEGDTKMIPLNWDDFTKV--S------SR-TVE-DTNPALTHTYEV
Human           LKL---FVTRIMQDFESDTFFPEIDLEKYKLLPEY------PGVLSDVQEEKGIKYKFEV

T.brucei        LVPRN
T.cruzi         LIPRN
L.major         YVPRN
E.coli          LERR-
L.casei         WQKKA
Human           YEKND
```

Figure 2. Sequence alignment of the studied DHFR structures automatically obtained with MULTAL.

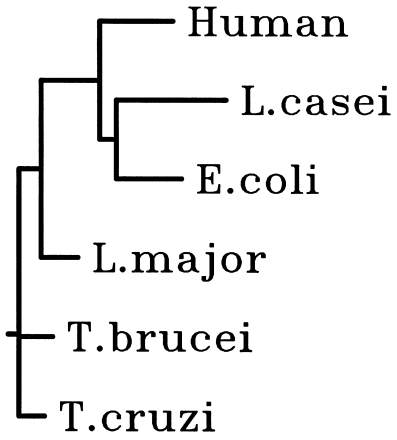

Figure 3. Rooted phylogenetic tree of the studied DHFR structures.

Ramachandran plot [27] with 9.0% in additionally allowed regions, giving a total of 97.9% of residues. For *T. brucei*, 88.6% of residues are in the most favoured regions and 9.5% in additionally allowed regions (a total of 98.1%). The template *L. major* enzyme (chain A) gave 73.2% and 24.0% of residues in the most and additionally allowed regions, respectively (97.2% in total). In our models all residues in the disallowed regions of the Ramachandran plot are loop residues; they are not near the folate binding site and are not relevant to our study of the DHFR active site. Other stereochemical parameters such as peptide planarity, bad non-bonded interactions, main-chain hydrogen bonding energy, overall G-factor and standard deviations of χ_1 angles were also examined and all showed better or similar values to those expected for a 2.9 Å resolution crystal structure (the resolution of the *L. major* template structure).

The differences of the models from chains A and B of the parent structure were also assessed by cal-

```
tcruzi          MSL----FKIRMPETVAEGTRLALAAFSLVVAVDEHGGIGDGRSIPWNVPEDMKFFRDLT
lmchA           ---AAARFKIPMPETKADFAFPSLRAFSIVVALDMQHGIGDGESIPWRVPEDMTFFKNQT
lmchB           MSRAAARFKIPMPETKADFAFPSLRAFSIVVALDMQHGI---------VPEDMTFFKNQT

tcruzi          TKLRGKNVKPSPAKRNAVVMGRKTWDSIPPKFRPLPGRLNVVLSSTLTTQHLLDGLPDEE
lmchA           TLLRNKKP-PTEKKRNAVVMGRKTWESVPVKFRPLKGRLNIVLSSKATVEELLAPL----
lmchB           TLLRNKKP-PTEKKRNAVVMGRKTWESVPVKFRPLKGRLNIVLSSAAAAAAL--------

T. cruzi        KRNLHADSIVAVNGGLEQALRLLASPNYTPSIETVYCIGGGSVYAEALRPPCVHLLQAIY
LmchA           ---AAAQAAVVVNGGLLAELRLLARPLYCSSIETAYCVGGAQVYADAMLSPCIEKLQEVY
LmchB           ------AAAVVVNGGLLAELRLLARPLYCSSIETAYCVGGAQVYADAMLSPCIEKLQEVY

T.cruzi         RTTIRASESSCSVFFRVPESGTEAAAGIEWQRETISEELTSANGNETKYYFEKLIPRN
LmchA           LTRLYATAPACTRFFPFPP----------WDLAS-SQGRRKSEAEGLEFEICKYVPRN
LmchB           LTRLY-------RFFPFPP----------WDLAS-SQGRRKSEAEGLEFEICKYVPRN
```

LmchA = *Leishmania major* chain A
LmchB = *Leishmania major* chain B

Figure 4. Alignment used for MODELLER.

Table 2. All-atom rms deviations of conserved residues from the model *T. cruzi* DHFR structure to human (1drf) and bacterial (3dfr and 4 dfr) structures

Residue	1drf	3dfr	4dfr
35	–	0.614	–
60	0.372	–	1.327
68	–	0.560	0.585
115	0.370	–	–
179	0.748	–	0.789

culating the rmsd of the backbone atoms between the model and parent structures. The rmsd values of backbone atoms (N,C^{α},C,O) from the *L. major* DHFR chain A were 0.56 Å for *T. cruzi* and 0.55 Å for *T. brucei* DHFR structures (equivalent regions from the sequence alignment utilised for the modelling were used for this purpose). Rmsd calculations were also performed to assess the quality of the side chains in the model of *T. cruzi*. The set of active site residues that are conserved across the DHFR structures (1drf, 3dfr, 4dfr) and *T. cruzi* DHFR was used (Table 2). Most deviations found were small (of the order of 0.3–1.33 Å). Given that there are no major deviations from the parents, the models were considered satisfactory.

From an analysis of the comparative modelling submissions to the recent 'Critical Assessment of Structure Prediction' (CASP2) meeting, one would expect an rmsd of 1.25–3.0 Å from the parent structure when modelling from a structure having 50% sequence identity [28]. However, this is highly sensitive to the accuracy of the alignment and the number and size of insertions or deletions. In the region of the active site, there are no insertions or deletions and the alignment is unambiguous, and we can therefore be reasonably confident that the active site is well predicted. Side-chain orientations will be less well predicted, but have been inherited from the parent structure. Any predictions which depend critically on such information have been treated with caution.

General description of binding

Images of the enzyme were prepared using GRASP with the ligand in the active site (Figure 5). The position of the ligand was determined either from the crystal structure or by using the coordinates of MTX from the *L. major* structure in the case of the modelled structures. A study of the mode of binding in the DHFR binary and ternary complexes shows that the ligand molecule (folate for the human DHFR and MTX for *L. major*, *E. coli* and *L. casei* DHFRs) lies in a cleft in the protein. The pteridine ring is roughly perpendicular to the protein surface and is mostly buried. The remainder of the ligand (the benzamide and glutamate moieties) lies in a groove along the protein surface with considerable accessibility to solvent. The

Table 3. Equivalent active site residues

Human 1drf	*E. coli* 4dfr	*L. casei* 3 dfr	*T. brucei* model	*T. cruzi* model	*L. major* crystal structure
Ile^{7}	Ile^{5}	Leu^{4}	Val^{32}	Val^{26}	Val^{30}
Val^{8}	Ala^{6}	Trp^{5}	Val^{33}	Val^{27}	Val^{31}
Ala^{9}	Ala^{7}	Ala^{6}	Ala^{34}	Ala^{28}	Ala^{32}
Gly^{20}	Asn^{18}	Gly^{17}	Gly^{45}	Arg^{39}	Glu^{43}
Asp^{21}	Ala^{19}	His^{18}	Thr^{46}	Ser^{40}	Ser^{44}
Leu^{22}	Met^{20}	Leu^{19}	Ile^{47}	Ile^{41}	Ile^{45}
Trp^{24}	Trp^{22}	Trp^{21}	Trp^{49}	Trp^{43}	Trp^{47}
Arg^{28}	Pro^{25}	Pro^{24}	Pro^{52}	Pro^{46}	Pro^{50}
Asn^{29}	Ala^{26}	Asp^{25}	Glu^{53}	Glu^{47}	Glu^{51}
Glu^{30}	Asp^{27}	Asp^{26}	Asp^{54}	Asp^{48}	Asp^{52}
Phe^{31}	Leu^{28}	Leu^{27}	Met^{55}	Met^{49}	Met^{53}
Arg^{32}	Ala^{29}	His^{28}	Gln^{56}	Lys^{50}	Thr^{54}
Tyr^{33}	Trp^{30}	Tyr^{29}	Tyr^{57}	Phe^{51}	Phe^{55}
Phe^{34}	Phe^{31}	Phe^{30}	Phe^{58}	Phe^{52}	Phe^{56}
Gln^{35}	Lys^{32}	Arg^{31}	Arg^{59}	Arg^{53}	Lys^{57}
Thr^{38}	Thr^{35}	Thr^{34}	Thr^{62}	Thr^{56}	Thr^{60}
Thr^{39}	Leu^{36}	Val^{35}	Thr^{63}	Thr^{57}	Thr^{61}
Met^{52}	Met^{42}	Val^{41}	Met^{82}	Met^{76}	Met^{79}
Thr^{56}	Thr^{46}	Thr^{45}	Thr^{86}	Thr^{80}	Thr^{83}
Ser^{59}	Ser^{49}	Ser^{48}	Ser^{89}	Ser^{83}	Ser^{86}
Ile^{60}	Ile^{50}	Phe^{49}	Leu^{90}	Ile^{84}	Val^{87}
Pro^{61}	Gly^{51}	Pro^{50}	Pro^{91}	Pro^{85}	Pro^{88}
Asn^{64}	Arg^{52}	Lys^{51}	Phe^{94}	Phe^{88}	Phe^{91}
Pro^{66}	Pro^{53}	Pro^{53}	Pro^{96}	Pro^{90}	Pro^{93}
Leu^{67}	Leu^{54}	Leu^{54}	Leu^{97}	Leu^{91}	Leu^{94}
Lys^{68}	Pro^{55}	Pro^{55}	Ser^{98}	Pro^{92}	Lys^{95}
Arg^{70}	Arg^{57}	Arg^{57}	Arg^{100}	Arg^{94}	Arg^{97}
Ile^{114}	Val^{93}	Ile^{96}	Cys^{159}	Cys^{153}	Cys^{155}
Val^{115}	Ile^{94}	Ala^{97}	Ile^{160}	Ile^{154}	Val^{156}
Gly^{116}	Gly^{95}	Gly^{98}	Gly^{161}	Gly^{155}	Gly^{157}
Tyr^{121}	Tyr^{100}	Phe^{103}	Tyr^{166}	Tyr^{160}	Tyr^{162}
Phe^{134}	Tyr^{111}	Leu^{114}	His^{182}	Tyr^{176}	Tyr^{178}
Val^{135}	Leu^{112}	Val^{115}	Arg^{183}	Arg^{177}	Leu^{179}
Thr^{136}	Thr^{113}	Thr^{116}	Thr^{184}	Thr^{178}	Thr^{180}
Phe^{179}	Phe^{153}	Tyr^{155}	Phe^{233}	Phe^{227}	Ile^{218}

The residues of different DHFR structures are numbered according to their files. The numbering from human DHFR will be used when referring to residue position.

benzamide and glutamate moieties are perpendicular to the pteridine ring, when bound to the enzyme. In the bound state, the energy of the MTX conformation is significantly above its global minimum energy of MTX. (The global minimum energy of MTX was calculated using a Monte Carlo energy minimisation conformational search using the MM3* force field and the GB/SA solvation model as implemented in batchmin, MACROMODEL.)

Dihydrofolate active site residues

In order to use the models and structures for selective drug design, it is important to be able to identify differences in structures between enzymes from the different species.

The folate active sites were defined by selecting all residues involved in the binding of the different ligands in the DHFR complexes [4, 29, 30] and these were compared with the corresponding residues from other species (Table 3). Of the 35 residues included in the table, 11 are absolutely conserved (Ala^{9}, Trp^{24},

Table 4. General properties of the variable position in DHFR active sites

Residue	Human	*E. coli*	*L. casei*	*T. brucei*	*T. cruzi*	*L. major*
7	HS	*	HS	HS	HS	HS
8	HS	HS	PL	*	*	*
20	S	PS	*	*	S	–S
21	–S	HS	PL	PS	PS	PS
22	HS	HS	*	HS	HS	HS
28	+L	HS	HS	HS	HS	HS
29	PS	HS	–S	–S	–S	–S
30	–S	–S	–S	–S	–S	–S
31	HL	HS	HS	HL	HL	HL
32	+L	HS	PS	PS	+L	PS
33	PL	PL	*	*	HL	HL
35	PS	+L	+L	+L	+L	+L
39	PS	HS	HS	*	*	*
52	HL	*	HS	*	*	*
60	HS	*	HL	HS	*	HS
61	HS	S	*	*	*	*
64	PS	+L	+L	HL	HL	HL
68	+L	HS	HS	PS	HS	*
114	HS	HS	*	PS	PS	PS
115	HS	HS	HS	HS	HS	*
121	PL	*	HL	*	*	*
134	HL	PL	HS	PL	PL	PL
135	HS	HS	*	+L	+L	HS
179	HL	*	PL	*	*	HS

H: hydrophobic residue; P: hydrophilic residue; S: small residue; L: large residue; (+): positively charged residue; (−): negatively charged residue; *: same residue as human DHFR.

Phe^{34}, Thr^{38}, Thr^{56}, Ser^{59}, Pro^{66}, Leu^{67}, Arg^{70}, Gly^{116} and Thr^{136} in the human sequence). A further 11 residues show high similarity or variation in just one species (positions 7, 8, 22, 30, 33, 52, 60, 61, 115, 121 and 179, human numbering). The remaining 13 residues are variable (positions 20, 21, 28, 29, 31, 32, 35, 39, 64, 68, 114, 134 and 135, human numbering). This suggests a significant degree of variation in the structure of the active site between species. These differences are likely to account for the species selectivity of various DHFR inhibitors and may be exploited in the design of novel inhibitors.

Carrying out a similar comparison for only the protozoan structures gives 23 conserved residues, 6 similar residues and 6 variable residues, implying a much higher degree of similarity within the active site of the protozoans. Among the bacterial DHFRs only 15 residues are conserved and 7 are similar while more than one third are variable. At the 24 positions which are not absolutely conserved across the three groups, there is a much higher degree of conservation within the protozoan DHFRs. This offers the possibility of designing drugs specific for protozoa but with broad specificity.

The differences between human, protozoan and bacterial enzymes are highlighted in Table 4. Here the differences have been categorised in terms of size, charge and hydrophilicity/hydrophobicity of the residues. Of the 24 sites which are variable in sequence, some have conserved properties (e.g. position 7: Ile for human DHFR, Val for protozoan DHFR) while others are variable and likely to be of more importance in terms of the design of active selective inhibitor molecules (e.g. position 35: Gln for human DHFR, Lys or Arg for protozoan DHFR).

Solvent accessibility

The absolute solvent accessibility and the relative solvent accessibility (RSA) were calculated using a 1.4 Å probe in order to estimate the general availability of the residues to interact with a ligand molecule at different positions in the active site region (Table 5).

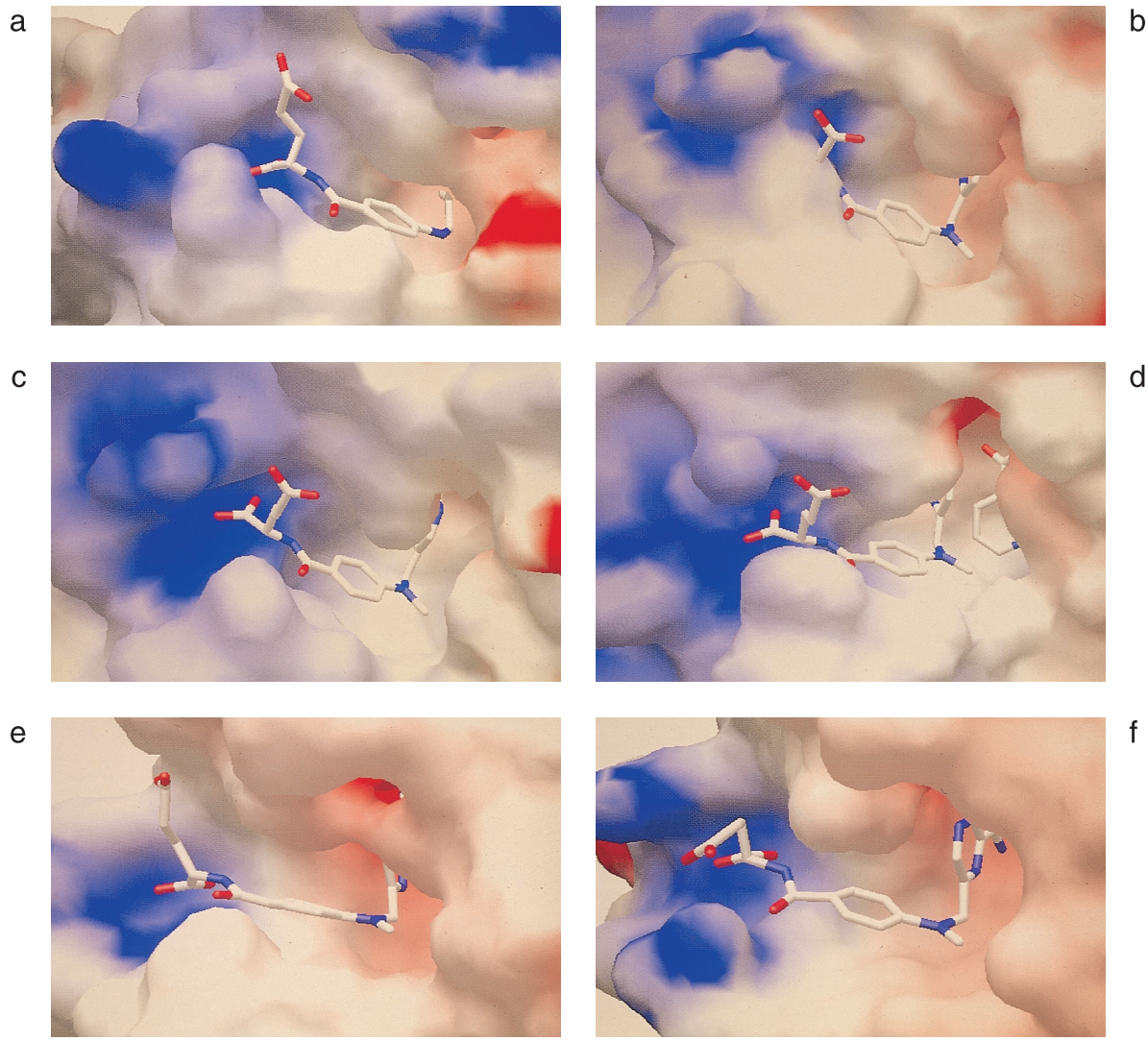

Figure 5. Electrostatic distribution in the dihydrofolate active site cavity (GRASP): (a) human; (b) *T. cruzi*; (c) *T. brucei*; (d) *L. major*; (e) *L. casei*; and (f) *E. coli* DHFRs. For the *T. cruzi* and *T. brucei* structures, the MTX molecule from the *L. major* complex is shown.

Residues with an absolute solvent accessibility of less than 2 $Å^2$ in the unliganded structure were considered buried and eliminated from the list of potentially useful residues as they cannot make direct contact with any ligand (residues 38, 52, 114 and 136). Significant differences in solvent accessibility between species were observed at residues 28 and 33. At position 28, the human residue shows a much higher solvent accessibility (124 $Å^2$) than in other species (52–63 $Å^2$). At position 33, the human and bacterial residues show negligible solvent accessibility whilst larger values (28–48 $Å^2$) are observed with the other species.

The 24 non-conserved residues of Table 4 were then classified as 'buried' or 'exposed' positions of the active site. We considered a position to be buried if the average of the RSA over the six DHFR structures was less than 15% and considered it to be exposed if greater than that. It is expected that an exposed residue is more likely to be involved in ligand binding than a buried one. In some cases the categorisation of buried/exposed differed between species. For instance, residue 39, classified as exposed (its average RSA value over the six structures is 22.4%), shows a quite low RSA in protozoan structures (average 12.0%) but a much higher value in bacterial (average 33.4%) and

Table 5. Residues surface accessibility (absolute and relative values)

Position	Human		*L. major*		*T. cruzi*		*T. brucei*		*E. coli*		*L. casei*	
	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)
7	6	4	6	4	6	4	6	4	7	4	8	5
8	11	7	5	4	5	3	4	3	9	9	15	6
9	20	19	17	15	29	27	33	31	28	26	16	15
20	48	60	74	43	210	88	51	63	126	88	55	68
21	76	54	57	49	47	40	67	48	95	88	108	59
22	60	34	76	44	75	43	82	47	45	23	85	48
24	16	6	47	19	54	22	58	24	53	21	25	10
28	124	52	52	38	61	45	55	40	53	39	63	46
29	72	50	52	30	54	31	58	34	25	24	51	37
30	17	10	23	16	10	7	10	3	18	13	16	11
31	103	52	115	59	109	56	112	58	121	68	81	45
32	135	57	79	57	138	69	138	77	55	51	127	69
33	3	2	28	14	48	24	40	19	8	3	30	14
34	44	22	42	21	39	20	38	19	38	19	41	21
35	80	45	110	55	130	55	152	64	108	54	108	45
39	44	32	20	14	16	12	14	10	66	37	45	30
56	26	18	33	24	30	21	32	23	29	21	37	27
59	33	29	49	42	47	41	48	41	75	65	60	51
60	26	15	23	15	23	13	37	21	51	29	23	12

continued...

Table 5. cntd

Position	Human		*L. major*		*T.cruzi*		*T. brucei*		*E. coli*		*L. casei*	
	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)	Absolute (Å^2)	RSA (%)
61	83	61	90	66	85	63	68	50	51	63	131	97
64	57	40	123	62	68	34	118	59	139	58	56	28
66	54	40	34	25	21	15	21	16	30	22	42	31
67	13	7	27	15	2	1	21	12	17	9	24	13
68	135	67	112	56	72	53	61	52	79	58	70	51
70	10	4	10	4	–	–	8	4	26	11	25	10
115	15	10	14	9	15	8	20	11	14	8	15	14
116	3	4	4	5	3	4	2	3	–	–	–	–
121	8	4	7	4	5	2	6	3	11	5	–	–
134	6	3	19	9	16	8	23	13	6	3	16	9
135	–	–	–	–	26	11	20	8	–	–	–	–
179	18	9	55	32	18	9	15	8	8	4	7	4

Residues having an absolute solvent accessibility lower than 2 Å^2 have not been considered and are shown with a (–). Residues 38, 52, 114 and 136 have not been included as they have absolute solvent accessibility lower than 2 Å^2 in all species.

Table 6. RSA in the bound and unbound human DHFR (1drf)

Position	RSA		Position	RSA	
	Unbound	Bound		Unbound	Bound
Ile^{7}	4	–	Thr^{56}	18	9
Val^{8}	7	2	Ser^{59}	29	19
Ala^{9}	19	4	Ile^{60}	15	–
Gly^{20}	60	60	Pro^{61}	61	52
Asp^{21}	54	54	Asn^{64}	40	24
Leu^{22}	34	16	Pro^{66}	40	40
Trp^{24}	6	6	Leu^{67}	7	–
Arg^{28}	52	51	Lys^{68}	67	66
Asn^{29}	50	50	Arg^{70}	4	2
Glu^{30}	10	3	Val^{115}	10	–
Phe^{31}	52	15	Gly^{116}	4	3
Arg^{32}	57	51	Tyr^{121}	4	–
Tyr^{33}	2	2	Phe^{134}	3	3
Phe^{34}	22	–	Val^{135}	1	1
Gln^{35}	45	29	Phe^{179}	9	9
Thr^{39}	32	32			

Residues having an absolute value of solvent accessibility lower than 2 Å^2 have not been considered.

human structures (31.6%). Similar situations, where residues of the same position should have different classification in different DHFRs, occur also in the positions 33, 60 and 179.

The change in RSA of residues on ligand binding was investigated for the human DHFR–folate complex (Table 6). There is little change in RSA for the non-conserved residues on binding folate, except for positions 31, 34, 35 and 64. This lack of contact with non-conserved residues could be part of the reason for the non-selectivity of MTX.

Conformational change on ligand binding

It is important to establish if there is significant change in conformation of the active site on binding ligands (DHFR inhibitors or NADPH cofactor) as this would reduce the utility of the models and structures in defining differences in structure between enzymes from different species which may be exploitable in selective drug design. From the study of the rmsd over the fitted *E. coli* DHFR structures (4dfr, 5dfr, 6dfr and 7dfr), there appears to be little conformational change at the active site on binding MTX (Table 7). However, some of these data were further analysed (4dfr, complexed with MTX resolved at 1.7 Å; 5dfr, free protein resolved at 2.3 Å) by looking at the rmsd of individual residues (Table 8).

Table 7. Rms deviations calculated over the superimposed structures

Superimposition	Rmsd of main-chain atoms	Rmsd of all atoms
4dfr/5dfr	0.68	0.85
4dfr/6dfr	0.48	0.57
4dfr/7drf	0.48	0.56
5dfr/6dfr	0.45	0.67
5drf/7dfr	0.75	0.67
6dfr/7dfr	0.55	0.60

Superimposition was performed fitting all the atoms of the residues which are known to be involved in the binding (positions 7, 8, 9, 24, 30, 31, 32, 33, 34, 35, 38, 56, 59, 60, 64, 67, 68, 70, 114, 115, 116, 121, 135 and 136, human numbering). Main chain defined by the atoms N,C$^{\alpha}$,C,O.

Table 8. Rmsd calculated over individual residues of the active site of sturctures 4dfr and 5dfr fitted on all atoms

Position	Rmsd		Position	RMSd	
	Main chain	All atoms		Main chain	All atoms
7	0.376	0.324	60	1.532	1.410
8	0.386	0.383	64	0.998	2.219
9	0.512	0.532	67	0.408	0.338
24	1.346	0.973	68	0.413	0.549
30	0.633	0.766	70	0.197	0.277
31	0.719	1.136	114	0.207	0.229
32	0.677	0.803	115	0.203	0.304
33	0.422	0.424	116	0.548	0.591
34	0.298	0.362	121	0.336	0.468
35	0.275	1.066	134	0.154	0.206
38	0.168	0.185	135	0.155	0.232
56	0.826	0.832	136	0.302	0.352
59	1.632	1.689			

The same set of residues as in Table 7 was used for the superimposition.

It was found that the observed rmsd values in the active sites were not evenly distributed. Main-chain (N,C$^{\alpha}$,C,O) deviations >1.0 Å were seen for residues 24, 59 and 60, while all-atom deviations >1.0 Å indicating side chain displacements were seen for residues 31, 35, 59, 60 and 64.

Some of these residues are known to be directly involved in MTX binding: the serine in position 59 gives a hydrophobic interaction with the N10 methyl of MTX, the isoleucine in position 60 and the leucine in position 31 make contacts with the *p*-aminobenzamide moiety whilst the backbone of the lysine in position 35 is in contact with the glutamate moiety. In

addition, the tryptophan in position 24, although indirectly, is involved in binding MTX through an H-bond from N^{ϵ_1} to a conserved water molecule [30]. Despite these differences, none of the conformational changes which occur on complexation are large, and the unliganded crystallographic and modelled structures are thus reasonable starting points for drug design.

The change in the conformation of the active site of *E. coli* DHFR on binding ligands has been discussed in more detail by Champness et al. [31]. They see little variation in the conformation of the active site except for a mobile region from 17 to 23 (human numbering). This flexible region moves towards the folate binding site on binding NADPH.

Electrostatic maps of the active site

The set of 35 residues proposed as being involved in the active site (see Table 3) includes a number of charged residues, varying from 4 in *E. coli* to 7 in *T. cruzi* structures. The resulting net charge ranges from +3 (*T. cruzi*) to zero (*L. major*). A more detailed examination of the structures, however, shows that the charged residues in positions 29 and 135, while close to the pteridine sub-site, have their side chains oriented towards the surface away from the active site (residue 29) and the core (residue 135) of the protein. Therefore, these charges are unlikely to contribute to the overall charge distribution in the active site area or to be important in direct ligand–protein interactions. Ignoring these residues results in a net charge within the active site ranging from +1 (*L. major*) to +3 (*T. cruzi*).

In general, it can be seen (Figure 5) that all the negatively charged residues are localised in the area of the pteridine sub-site while the positively charged ones are almost exclusively localised in the glutamate sub-site. This corresponds to the nature of the natural substrate, dihydrofolate, where N1 can interact with the conserved negative charge at position 30 and the two glutamate carboxyl groups interact with the conserved residue Arg^{70} and with the more variable positions 32, 35, 64 and 68. Despite this high degree of similarity of the charge distribution in the DHFR structures of different species, it is possible to note some interesting points:

(1) The portion of the active site that is shared with the NADPH active site (positions 20, 21 and 22) shows a negative charge in human and *L. major* structures but a positive one in *T. cruzi* DHFR and it is uncharged in *T. brucei* and the bacterial enzymes.

(2) The human positively charged position 32 is conserved in *T. cruzi* DHFR only.

(3) Position 35, glutamine in human, is a positively charged residue in both protozoan and bacterial structures.

(4) Bacterial DHFRs are the only ones to have a positive charge in position 64.

(5) *L. major* and human structures are the only ones to have a positive charge in position 68.

Size of the active site

It has been suggested by Champness et al. [32] that the size of the active site is important in determining the selectivity of inhibitors. They proposed that the selectivity of trimethoprim for bacterial DHFR is a result of the smaller active site when compared with mammalian (mouse) DHFR. We measured the active site at a number of points (data not shown). All the protozoan structures show similar-sized active sites to one another, to bacterial and to human DHFR. It appears that the mouse DHFR studied by Champness et al. has an unusually large active site and this is not a general difference between the DHFRs of mammals and lower species.

SURFNET active site maps

Analysis of the differences in shape of the active site between human and other DHFR structures should provide useful information about the regions of the active site that could account for selectivity. For each of the structures, the surface of the active site cavity was generated by SURFNET using probes with radii ranging from 1.4 to 4.0 Å. SURFNET is a programme which generates molecular surfaces and gaps between surfaces from 3D coordinates supplied in a PDB format. The differences between the human and other DHFR active site SURFNET maps were observed by superimposition of the maps. The residues listed in Table 3 were used for the superimposition.

Comparing the human DHFR cavity with the corresponding protozoan cavity, it is possible to observe some important differences despite a generally high degree of similarity (Figures 6a–c).

In *T. cruzi* DHFR:

(1) Two major differences occur in the glutamate sub-site and *p*-aminobenzamide regions, giving rise to a larger hydrophobic binding pocket in the case of *T. cruzi*. The residues 31 and 64 in *T. cruzi* are Met and Phe, whilst in the human structure they are Phe and Asn.

(2) There is a deep, narrow gap, not present in the human structure, under the pteridine ring. This is

caused by a different conformation of the bulky side chain of Trp^{24} in the two structures.

In *T. brucei* DHFR:

(1) There is a larger hydrophobic cavity in the *T. brucei* structure adjacent to the benzamide binding site than is found in the human DHFR. This results from replacement of Phe^{31} in the human structure with Met in *T. brucei* and due to different orientations of the side chains at position 60.

(2) There is a similar but larger gap under the pteridine ring. Again this results from a different conformation of the Trp^{24}.

In *L. major* DHFR:

(1) Again there is a larger hydrophobic pocket in the protozoan enzyme than the human enzyme adjacent to the benzamide binding site.

(2) The gap under the pteridine ring observed in the other protozoan structures is much smaller and it is not relevant from a drug design perspective.

(3) The pteridine sub-site is larger than the corresponding human sub-site as a result of the different residue in position 21 (Asp in human and Ser in *L. major*).

A similar comparison of the bacterial structures with the human structure shows a very high degree of similarity between human and bacterial enzymes with only some minor differences (Figures 6d and e). In both the *E. coli* and *L. casei* structures, it is possible to observe three regions, one in each sub-site, which are slightly smaller than the human ones:

(1) In the pteridine sub-site, for the *L. casei* enzyme, this seems to be the result of a different position of the backbone of residues 25 (Pro in human and His in *L. casei*) whilst for *E. coli* the smaller sub-site is the result of the different orientation of Leu at position 27.

(2) In the *p*-aminobenzamide sub-site, both the bacterial structures have different orientations of the side chains at positions 50 and 60 compared with the human structure.

(3) In the glutamate sub-site, the difference is caused, for *L. casei*, by the different conformation of the side chain of the Leu in position 67 and, for *E. coli*, by the different position of the backbone of the residue in position 68 (Pro in bacterial and Lys in human) compared with the corresponding human residues.

Discussion

Comparison of the sequence data between human, protozoan and bacterial DHFRs suggests that it may be possible to design selective inhibitors of the protozoan enzymes, as the sequence homology is low to the human enzymes. In addition, the protozoan enzymes have high sequence homology to one another, suggesting that it may be possible to design an inhibitor which is selective for all three enzymes: this would be of significant advantage when trying to develop potential drugs for 'orphan' diseases such as the trypanosomiases and leishmaniasis. Therefore, enzyme active sites of the protozoan, human and bacterial enzymes were compared in order to identify differences which may be exploitable for drug design. A number of features may give rise to exploitable differences. These include the differences in the active site in terms of residues, solvent accessibility, charge and shape, which will all give rise to different interactions between the enzymes and inhibitors.

Firstly, differences in residues present in the active site were examined. Non-conserved residues are tabulated in Table 9. Some of these residues have a very small surface area that is exposed to solvent and hence are unable to interact with potential inhibitors. We defined these residues as buried (those having relative solvent accessibilities of less than 15%). Buried hydrophobic residues will not have significant interactions with ligands and can be discounted. This included residues 7, 8, 52, 115, 134 and 179. In addition, 114 has an absolute solvent accessibility of less than 2 $Å^2$ and cannot interact with any ligand. The size, polarity and charge of the remaining residues were then compared using Table 4. Residues showing high similarity were then discounted: this included residues 22, 30, 60 and 61. Residue 33 is not discounted because of the large variation in solvent accessibility between species, suggesting the residue could give rise to different interactions with ligands.

This leaves a list of residues (20, 21, 28, 29, 31, 32, 33, 35, 64, 68 and 135) with significant differences which may be useful for drug design. The structures were then visually examined to ascertain their disposition in the active site and to see how significant their interaction with potential ligands might be. The disposition of residues compared to folate in human and MTX in *L. major* was studied. Whilst a potential inhibitor need not be based on the structure of folate or MTX, both these ligands occupy most of the active site. Residues which are not in very close proximity to the folate or MTX are unlikely to be important in ligand binding in the folate binding site. The following residues were then discarded on this basis (28, 29, 32, 33, 68, 114 and 135). Further details are given below.

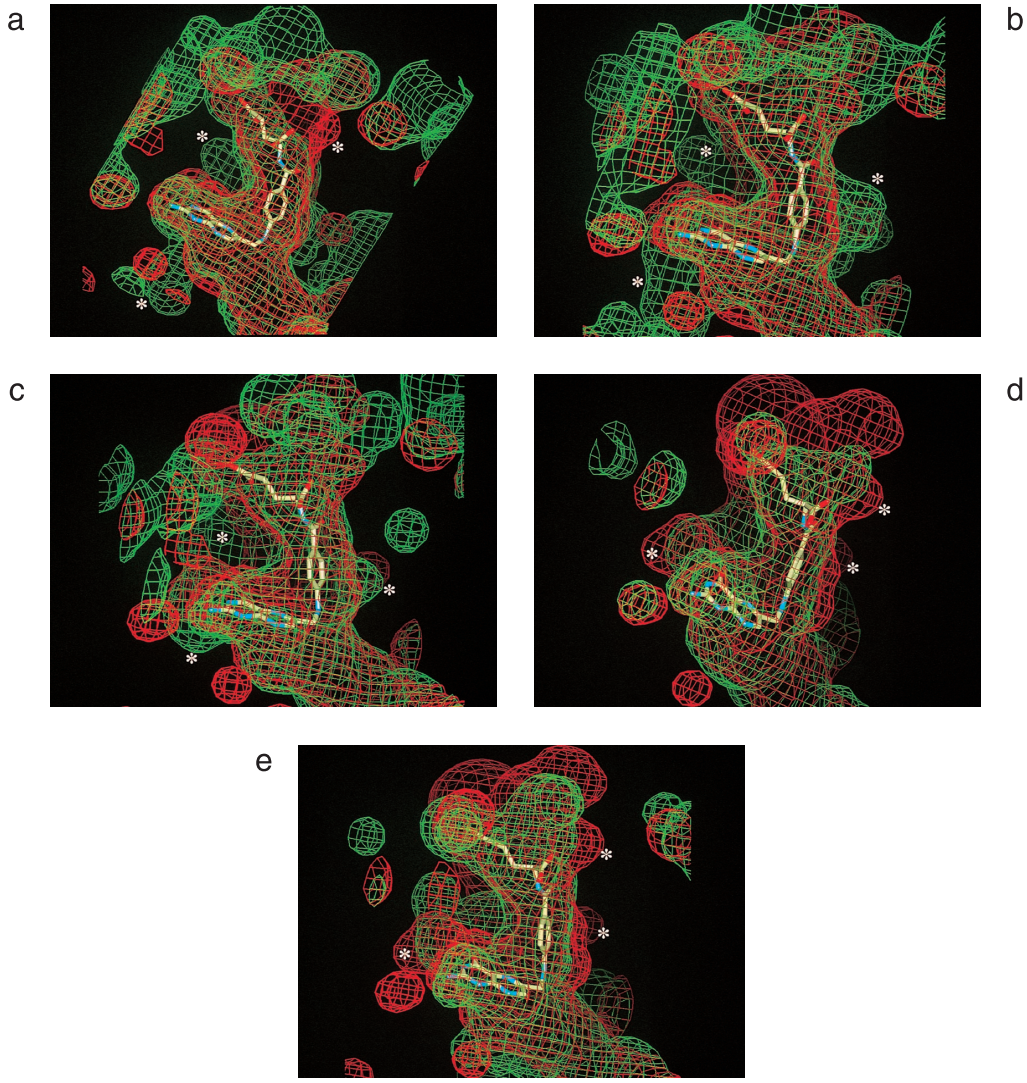

Figure 6. Comparison between dihydrofolate active site cavity shapes of human (red) and studied DHFRs (green): (a) *T. cruzi*; (b) *T. brucei*; (c) *L. major*; (d) *L. casei*; and (e) *E. coli*. White stars indicate the differences discussed in the text.

Residues 20 and 21 are located on the tunnel connecting the NADPH binding site to the folate binding site. Residue 20 is Gly for the human species and Gly, Arg and Glu for the protozoan species, whilst residue 21 is Asp for the human species and Thr or Ser for the protozoan species. This gives rise to a negative charge at this position for the human enzyme, neutral for *T. brucei*, positive for *T. cruzi* and negative for *L. major*. There is a possibility of exploiting this difference in the case of *T. cruzi* and possibly also *T. brucei* by introducing a negatively charged group in an inhibitor at this position.

At position 28 there is Arg in the case of the human enzyme and Pro for the protozoan enzymes. The backbone carbonyl group in both cases appears to interact with the γ-carboxylate of the ligand. The guanidine group of the Arg in humans is pointing away from the cleft and towards the protein surface and is unlikely to interact with ligands. Therefore, despite the differences between residues these are unlikely to be significant from the perspective of selective drug design.

The residue 31 side chain contacts all three binding pockets (pteridine, benzamide and glutamate). In the case of humans it is Phe, and in the case of pro-

Table 9. Variable positions in the active sites of dihydrofolate reductase

Residue	Buried/exposed	Residue	Buried/exposed
7	buried	39	exposed
8	buried	52	buried
20	exposed	60	exposed
21	exposed	61	exposed
22	exposed	64	exposed
28	exposed	68	exposed
29	exposed	114	buried
30	buried	115	buried
31	exposed	121	buried
32	exposed	134	buried
33	buried/exposed	135	buried
35	exposed	179	buried

Residues have been classified as buried/exposed on the basis of the results obtained in the calculation of the accessible surface to water.

Table 10. Important target residues for the design of selective DHFR inhibitors

Position	Anti-protozoal	Sub-site
20	(*)	Pteridine/NADPH
21	*	Pteridine/NADPH
31	*	Pteridine/benzamide/glutamate
35	*	Glutamate
64	*	Glutamate

That in parentheses is useful for some species only.

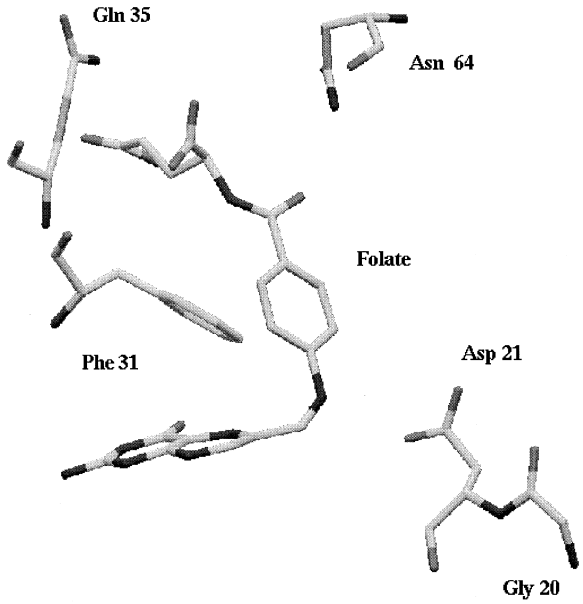

Figure 7. Position of the sites proposed to be useful in drug design in relation to the folate molecule in the human dihydrofolate active site.

tozoans it is Met. This difference probably gives rise to the larger binding site in the case of the protozoan parasites at the benzamide sub-site and will also give rise to different interactions with potential ligands. For example, aromatic compounds will have a π-stacking interaction in the case of the human enzyme but not the parasite.

Residue 32 is positively charged in the case of human (Arg) and *T. cruzi* (Lys), but neutral for *T. brucei* (Gln) and *L. major* (Thr). Whilst near to the glutamate binding site, these residues have no direct interaction with the ligand and probably are not significant for drug design.

Residue 33 is either Tyr or Phe. This has significant differences in solvent accessibility in the human and parasite enzymes. In the human the residue is essentially buried, but in the parasite it is quite exposed and may provide additional hydrophobic interaction with a ligand. However, this again has no direct contact with any of the binding sub-sites and probably is not a significant difference.

Position 35 is Gln for the human enzyme and either Arg or Lys for the parasite enzymes. Thus, the parasite enzyme will be positively charged at this position and ligands which can take advantage of this may show some selectivity. The side chains interact with the glutamate binding site.

Position 64 is Asn in the case of the human enzyme and Phe for the parasite enzymes. In the case of Asn the side chain has H-bonds to the carboxylates of the glutamate, whilst the Phe shows hydrophobic interactions with the glutamate moiety. Thus, a ligand with a hydrophobic moiety at this position should show selectivity for the parasite.

Positions 68 and 135 show no direct interaction with ligand and are unlikely to be significant for drug design.

This gives rise to the residues 20, 21, 31, 35 and 64 as being the most significant for selective drug design (Table 10). The locations of the residues showing significant variation are shown relative to dihydrofolate in Figure 7.

Therefore, non-conserved residues at the pteridine binding sites may allow the design of selective agents based on differences in charge. Differences at the benzamide sub-site may give rise to differences based on size and hydrophobic interactions, and differences at

the glutamate binding site may give rise to differences in hydrophobic and charge interactions.

Changes in solvent accessibility of the 24 non-conserved residues were examined in the case of the human on binding folate (Table 6). This should assess which residues have significant interaction with the folate ligand. Significant variation is only seen for residues 31, 34, 35 and 64, highlighting the importance of residues 31, 35 and 64 for selective drug design.

The SURFNET maps give information on the shapes of the active sites. It is clear from these that the benzamide binding pocket is larger in the case of the protozoan enzymes than in the case of the human and bacterial enzymes. Thus, ligands which are more bulky at this sub-site should show selectivity for the protozoan enzymes. The human enzyme has a phenylalanine residue at this position which is replaced by methionine in the protozoan enzymes. Hence, compounds which lack an aromatic group at this position which can π-stack with the phenylalanine should also show enhanced activity and selectivity for the protozoan enzymes.

The SURFNET maps also indicate a tunnel under the pteridine ring present in the *T. brucei* and *T. cruzi* enzymes and to a lesser extent in the *L. major* enzyme. In this portion of the active site, one of the major differences between the protozoan and non-protozoan enzymes is the conformation of the Trp^{24} side chain. However, this difference in conformation does not appear to be an artefact of the model as the conformation of the Trp is the same in the *L. major* structure and the *T. cruzi* model on overlay (data not shown).

Taking the information on residues in the active site and the information from the SURFNET maps, the following can be summarised as points for selective drug design:

(1) The benzamide binding sub-site is larger in the protozoan enzyme.

(2) The tunnel between the folate binding site is negatively charged for human and *L. major* but not for *T. cruzi* or *T. brucei.*

(3) The glutamate binding sites differ in locations of hydrophobic and charge interactions.

(4) There is a tunnel underneath the pteridine ring in the protozoan which is absent in the human enzyme.

Conclusions

Working models of *T. cruzi* and *T. brucei* DHFRs structures have been produced which have good stereochemical properties as judged by PROCHECK. They have been used in the comparison of protozoan, bacterial and human DHFRs. The crystal structures and models have been used to analyse where the principal differences between human, bacterial and protozoan enzymes lie with a view to future drug design. We have tried not to limit our prospect to possible modification of already known inhibitors and to highlight all those differences that could be useful for the possible design of a completely new class of selective compounds.

In general, there is a high degree of similarity between the active sites of all these enzymes. However, there are differences at the active site which may be exploitable for selective drug design. These are differences in both residue type and shape.

We propose a set of residues that could be exploited for selective binding on the basis of different ligand–protein chemical interactions (Table 10 and Figure 7). These occur principally in the region where DHFR binds the glutamate and benzamide moieties of MTX. We have also presented the different charge distributions within the binding cavities and the shape differences that could be used to generate selective binding. In this case, differences in pteridine and benzamide sub-sites are apparent between the protozoan and human structures. However, comparison of the human and bacterial structures displays only minor differences. Variations in the shape of the active site are highlighted in Figure 6. The most significant are a large binding pocket adjacent to the benzamide binding site in the protozoan enzymes and a small pocket under the pteridine binding site, also found in the protozoan enzymes but not in human.

In particular, the following differences provide possibilities for selective drug design for anti-protozoal or anti-bacterial agents:

(1) a larger binding pocket adjacent to the *p*-aminobenzamide sub-site (protozoan),

(2) a pocket found below the pteridine ring in the pteridine sub-site (protozoan),

(3) a different charge in positions 35 and 64 (protozoan and bacterial) and in the portion of active site that is shared with the NADP active site (*T. cruzi* only), and

(4) chemically important differences in amino acid sequence (see Table 10).

Further, the high level of structural similarity between the protozoal enzymes suggests that compounds which inhibit one enzyme selectively compared with human will probably inhibit the others. This is a very attractive feature as it may allow the development of broad-spectrum anti-protozoal drugs.

Although there is a very high degree of similarity in the structures of the enzymes, we believe that the data presented in this work show significant differences which may be exploitable for selective drug design. We are therefore currently working on the design and synthesis of possible lead compounds for selective protozoan DHFR inhibitors.

Acknowledgements

We thank D.A. Matthews for supplying coordinates for *L. major* DHFR prior to publication. The Welsh School of Pharmacy (F.Z.), SmithKline Beecham (F.Z.) and the MRC (A.C.R.M.) are thanked for funding. We also acknowledge the use of the EPSRC Daresbury facility.

References

1. Filman, D.J., Bolin, J.T., Matthews, D.A. and Kraut, J., J. Biol. Chem., 257 (1982) 13663.
2. Reyes, V.M., Sawaya, M.R., Brown, K.A. and Kraut, J., Biochemistry, 34 (1995) 2710.
3. Blaney, J.M., Hansch, C., Silipo, C. and Vittoria, A., Chem. Rev., 84 (1984) 333.
4. Knighton, D.R., Kan, C.C., Howland, E., Janson, C.A., Hostomska, Z., Welsh, K.M. and Matthews D.A., Nat. Struct. Biol., 1 (1994) 186.
5. Reche, P., Arrebola, R., Olmo, A., Santi, D.V., Gonzales-Pakanowska, D. and Ruiz-Perez, L., Mol. Biochem. Parasitol., 65 (1994) 247.
6. Gamarro, F., Yu, P.L., Zhao, J., Edman, U., Greene, P.J. and Santi, D., Mol. Biochem. Parasitol., 72 (1995) 11.
7. Ivanetich, K.M. and Santi, D.V., FASEB J., 4 (1990) 1591.
8. a. Li, Z., Chen, X., Davidson, E., Zwang, O., Mendis, C., Ring, C.S., Roush, W.R., Fegley, G., Li, R., Rosenthal, P.J., Lee, G.K., Kenyon, G.L., Kuntz, I.D. and Cohen, F.E., Chem. Biol., 1 (1994) 31.
   b. Toyoda, T., Brobey, R.K.B., Sano G., Horii, T., Tomioka, N. and Itai, A., Biochem. Biophys. Res. Commun., 235 (1997) 515.
9. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
10. Needleman, S.B. and Wunsch, C.D., J. Mol. Biol., 48 (1970) 443.
11. Martin, A.C.R., NW, unpublished data.
12. Felsestein, J., Cladistics 5, 164.
13. Sali, A. and Blundell, T.L., J. Mol. Biol., 234 (1993) 779.
14. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., J. Appl. Crystallogr., 26 (1993) 283.
15. Taylor, W.R., J. Mol. Evol., 28 (1988) 161.
16. CAMELEON, Oxford Molecular, Oxford, U.K.
17. McLachlan, A.D., Acta Crystallogr., A38 (1982) 871.
18. Martin, A.C.R., ProFit, unpublished data.
19. MACROMODEL, Colombia University, New York, NY, U.S.A.
20. QUANTA, Molecular Simulations, Burlington, MA, U.S.A.
21. Nicholls, A., Bharadwaj, R. and Honig, B., Biophys. J., 64 (1993) A166.
22. Sayle, R, RASMOL, Glaxo Wellcome.
23. Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
24. Hubbard, S., ACCESS, EMBL.
25. Laskowski, R., J. Mol. Graph., 13 (1995) 323.
26. Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M., Proteins, 12 (1992) 345.
27. Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V., J. Mol. Biol., 7 (1963) 95.
28. Martin, A.C.R., MacArthur, M.W. and Thornton, J.M., Proteins Struct. Funct. Genet., in press.
29. Oefner, C., D'Archy, A. and Winkler, F.K., Eur. J. Biochem., 174 (1988) 377.
30. Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
31. Champness, J.N., Stammers, D.K. and Beddell, C.R., FEBS Lett., 199 (1986) 61.
32. Champness, J.N., Achari, A., Ballantine, S.P., Bryant, P.K., Delves, C.J. and Stammers, D.K., Structure, 2 (1994) 915.