BINDING SERVICES
Tel +44 (0)29 2087 4949
Fax.+44 (0)29 2037 1921
E-Mail Bindery@Cardiff.ac.uk

An *in silico* Study to Understand the Factors which are Important for the Rational Design of Novel Inhibitors of dUTPase, an Antimalarial Drug Target.

A Thesis submitted to Cardiff University for Degree of

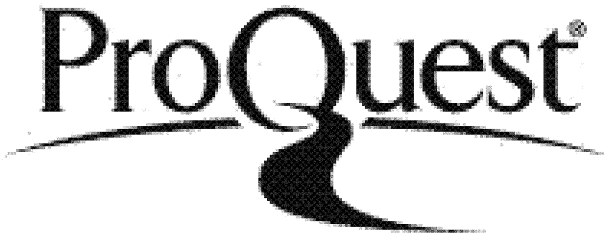
Doctor of Philosophy

by

Alessandro Schipani

January 2007

Supervisor: Professor Ian H. Gilbert



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Inhibition of deoxyuridine triphosphate (dUTPase), a ubiquitous enzyme and 'first line' of defence against misincorporation of uracil into DNA, is lethal in *E. coli*, *S. cerevsiae* and is also probably in a variety of other organisms including, *Plasmodium* spp., the causative agent of malaria. dUTPase is a potential chemotherapeutic target.

An assortment of molecular modelling tools was used to take advantage of the available structural information to identify differences between the human and protozoan dUTPase enzymes. A major objective of this study was to develop, rationally, the available lead compounds into antimalarial drug candidates.

Starting from the crystal structure of a cyclic nucleoside derivative bound to *P. falciparum* dUTPase, structural models for the interaction of a library of analogues with the dUTPase receptor were generated by automated computational docking. The obtained complexes were evaluated for their consistency using quantitative structure-activity relationship analysis. A comparison between the calculated interaction free energies and experimental biological activities was also made. All the possible interactions of the investigated compounds at the active site and probable ligand binding conformations provided an improved basis for structure-based rational ligand design. Based on docking results, a successful (R^2=0.80) model for prediction of binding affinities was developed using the linear interaction energy method (LIE). In addition, three-dimensional Quantitative Structure-Activity Relationship studies (3D-QSAR) generated using the CoMFA methodology yielded highly predictive models (R^2=0.99 and q^2=0.74).

Extensive use of this set of molecular modelling methodologies (docking, GRID, LIE, 3D-QSAR) has allowed for the development of models, which were employed successfully to introduce the available structural information early into the development of antimalarial drugs. This has resulted design of new inhibitors and has opened up a variety of new chemistry projects.

In addition, some work was carried out for the development of a predictive in silico model (Volsurf) for the passive blood brain barrier (BBB) permeation of compounds.

Declaration

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date 10/05/07

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of ....Ph.D.....(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed ..... (candidate) Date 10/05/07

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources are acknowledged by explicit references.

Signed ..... (candidate) Date 10/05/07

STATEMENT 3 - BAR ON ACCESS APPROVED

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access** approved by the Graduate Development Committee.

Signed ..... (candidate) Date 10/05/07

Acknowledgments

First of all, I would like to thank my supervisor Professor Ian H. Gilbert for giving me the opportunity to start this PhD project and also for his guide, his suggestion and his support during the whole duration of this work.

Thanks to all the partners who collaborated to this research project: Dr. D. Gonzalez-Pacanowska and all the collaborators at the Instituto de Parasitologìa y Biomedicina, Consejo Superior de Investigationes Cientìficas in Granada. Dr. Reto Brun and co-workers at the Swiss Tropical Institute in Basel. Dr. Keith Wilson at Structural Biology Laboratory of the University of York. Nils Gunnar Johannson at Medivir AB, Stockholm.

I would like to thank also all the past and present members in Professor Gilbert's group at the Welsh School of Pharmacy and at Dundee University: Corinne, Gian Filippo, Orlagh, Fede, Salvo, Ludo, Alessandro, Shane, Shahienaz, Huaqing, Dimitri, Andrea, Ruth and specially Chido for the enjoyable time spent in and outside the lab.

Publications

Part of this thesis has been published in the following journals:

- Mc Carthy, O. K.; Schipani, A.; Buendia, A. M.; Ruiz-Perez, L. M.; Kaiser, M.; Brun, R.; Pacanowska, D. G.; Gilbert, I. H. **Design, synthesis and evaluation of novel uracil amino acid conjugates for the inhibition of Trypanosoma cruzi dUTPase** *Bioorganic & Medicinal Chemistry Letters* **2006,** 16, (14), 3809-3812.

- Nguyen, C.; Ruda, G. F.; Schipani, A.; Kasinathan, G.; Leal, I.; Musso-Buendia, A.; Kaiser, M.; Brun, R.; Ruiz-Perez, L. M.; Sahlberg, B. L.; Johansson, N. G.; Gonzalez-Pacanowska, D.; Gilbert, I. H. **Acyclic nucleoside analogues as inhibitors of Plasmodium falciparum dUTPase** *Journal Of Medicinal Chemistry* **2006,** 49, (14), 4183-4195.

The full articles are referred in the Appendix 3.

I would like to dedicate this work to my parents who gave me love, constant support, and taught me the important things in life.

Voglio dedicare questo lavoro a mio padre e mia madre che mi hanno sempre dato il loro amore e appoggio, e mi hanno insegnato le cose importanti della vita.

Abbreviation & Acronysms

BBB	Blood brain barrier
Boc	N-tert-butoxycarbonyl
cAPK	Cyclic AMP-dependent kinase
CDK2	A-cyclin-dependent kinase 2
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative Molecular Similarity Indices Analysis
ddCTP	Dideoxycytidine 5'-triphosphate
DHFR	Dihydrofolate reductase
DHPS	Dihydropteroate synthetase
EBV	*Epstein-Barr* virus
EIAV	Equine infectious anemia virus
FAD	Flavin Adenine Dinucleotide
FEP	Free energy perturbation
FIV	Feline immunodeficiency virus
HMC	Hybrid Monte Carlo
HSV1	Herpes simplex virus
IC_{50}	Inhibitory concentration
LCR	Low complexity region
LIE	Linear interaction energy
LRA	Linear response approximation
MC	Monte Carlo
MD	Molecular Dynamics
MIFs	Molecular interaction fields
NAD	Nicotinamide adenine dinucleotide
OPLS	Optimized Potential for Liquid Simulations
PLS	Partial Least Squares
PPi	Pyrophosphate
QSAR	Quantitative Structure Activity Relationship

RMS	Root Mean Square
RMSD	RMS deviation
ROI	Reactive oxygen intermediates
SAR	Structure activity relationship
SGB	Surface generalized Born
SI	Selectivity index
TBDMS	Tert-butyl-dimethylsilyl
TBDPS	Tert-butyldiphenylsilyl
TBDPSO	O-tert-butyldiphenylsilyl
TI	Thermodynamic integration
TIPS	Triisopropylsilyl
TPS	Triphenylsilyl
TPSO	O-triphenylsilyl
Tr	Triphenylmethyl (trityl)
ugn	Uracil-DNA-glycosidase gene
WHO	World Health Organization
WSP	Welsh School of Pharmacy

1 Chapter: Background

1.1 Malaria

According to the World Health Organization (WHO), between 300 million and 500 million clinical case of malaria occur every year[1]. More recent estimations have suggested the total could run as high as 660 million[2]. Malaria is estimated to kill more than 1 million people annually and possibly as many as 3 million, with most deaths amongst children under age six living in sub-Saharan Africa. The health and the economic toll are enormous. Political, economic, and social factors affecting health care and government involvement in malaria prevention and control are significant factors. Despite decades of fighting malaria, the disease is gaining ground as the parasite's resistance to drugs and the parasite-carrying mosquito's resistance to insecticides expand.

Human malaria is caused by four different species of *Plasmodium*: *Plasmodium vivax, Plasmodium malariae, Plasmodium ovale* and *Plasmodium falciparum*. *P. vivax* is the most widespread parasite, occurring largely in the topics and throughout Asia, but the most severe form of the disease is caused by the *P. falciparum* (Figure 1.1).

About 90% of malaria-related deaths occur in Africa and are caused by *P. falciparum*. People, who survive multiple infections in childhood, acquire a natural immunity that limits the severity of the disease. This immunity will decrease without continued exposure to infection. Pregnant women have reduced immunity.

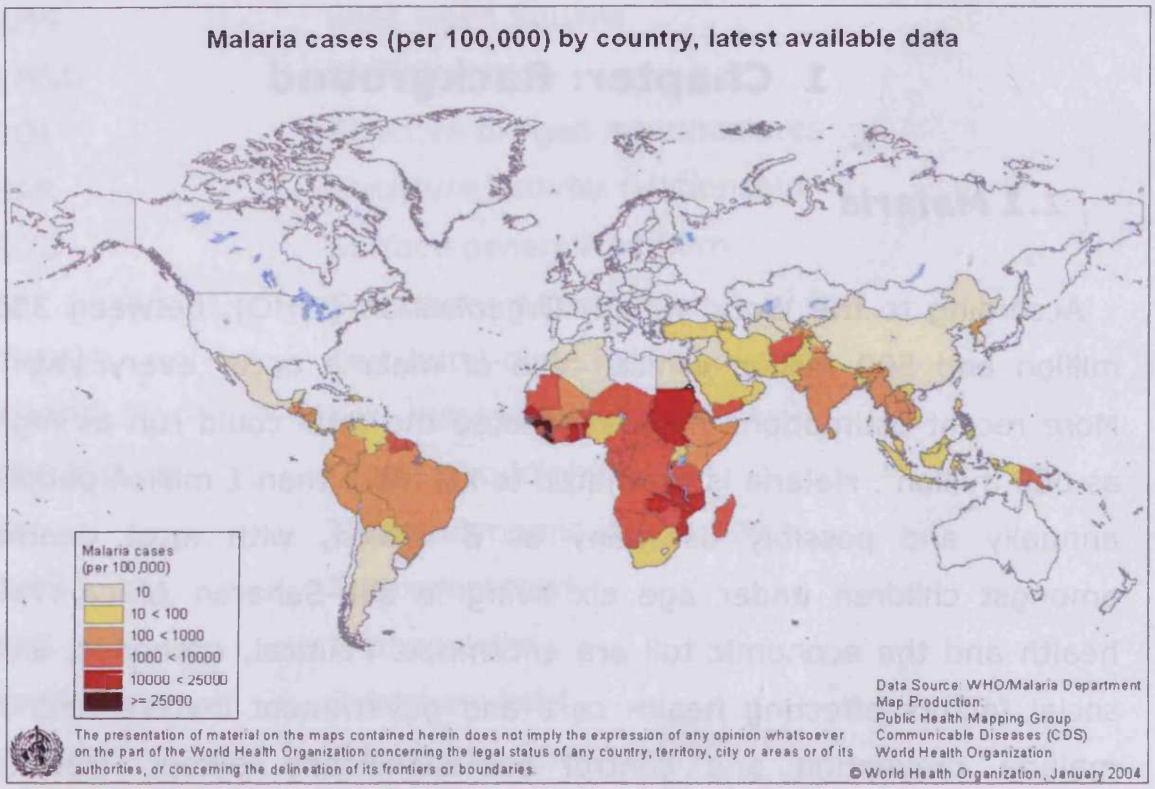


Figure 1.1 Malaria distribution in the world[1]

1.1.1 Life Cycle

The malaria parasite uses both a mosquito and human host. Malarial parasites enter a human host when an infected female mosquito feeds. Infectious sporozoites quickly migrate to the liver, where they develop and multiply. This asymptomatic incubation phase can last from a week to months. Then the parasites emerge from the liver cells as merozoites, burst from the cells, and infect red blood cells. At this stage the infection results in the fever, flu-like symptoms, and anaemia associated with malaria. The intermittent fever paroxysms are due to the lysis of the infected erythrocytes. Some of the parasites mature into reproductive gametocytes that are ingested by a feeding mosquito. In the insect gut, they develop into oocysts that grow, rupture, and release sporozoites, which make their

way to the mosquito's salivary glands, to complete the cycle (Figure 1.2).

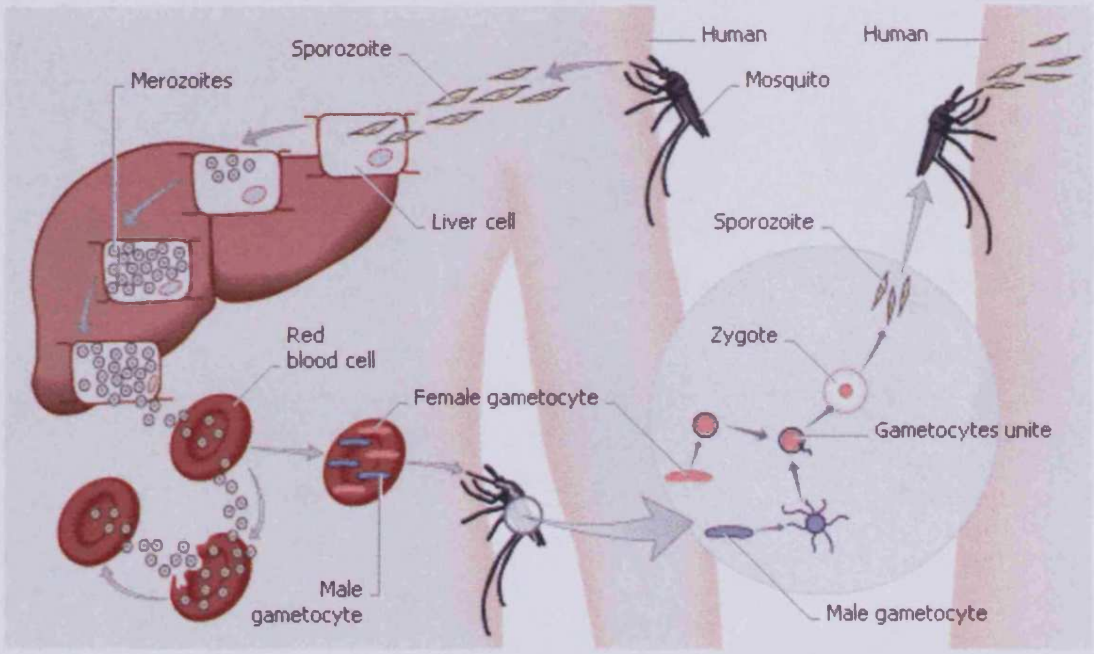


Figure 1.2 Life Cycle of *Plasmodium*[3]

Red blood cells are the principal sites of infection in malaria and all the clinical manifestations are primarily due to the infection of red blood cells (Figure 1.3).

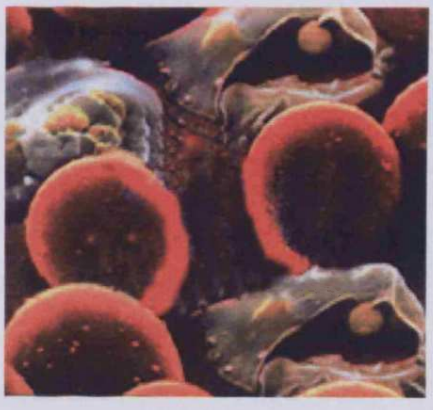


Figure 1.3 Red blood cells infected with *Plasmodium falciparum*[4].

The growing parasite consumes and degrades the intracellular proteins, mainly hemoglobin. The transport properties of the red cell

membrane are altered. Several severe complications can be associated with *falciparum* malaria with cerebral malaria being the most notable and a frequent cause of death (Table 1.1).

	vivax	*ovale*	*malariae*	*falciparum*
Initial Paraoxysm Severity	moderate to severe	mild	moderate to severe	severe
Average Parasitemia (mm^3)	20,000	9,000	6,000	50,000-500,000
Maximum Parasitemia (mm^3)	50,000	30,000	20,000	2,500,000
Symptom Duration (untreated)	3-8+ weeks	2-3 weeks	3-24 weeks	2-3 weeks
Maximum Infection Duration (untreated)	5-8 years	12-20 months	20-50+ years	6-17 months
Anemia	++	+	++	++++
Complications			renal	cerebral

Table 1.1 Disease Severity and Duration[5].

Cerebral malaria is characterized by an impaired consciousness. The presenting symptoms are severe headache followed by drowsiness, confusion, and ultimately, coma. Convulsions are also frequently associated with cerebral malaria. These neurological manifestations are believed to be due to the accumulation of the infected erythrocytes in the cerebral microvasculature. The accumulation might be caused by the *cytoadherence* of the infected erythrocytes to endothelial cells of deep vascular beds in vital organs, especially brain, lung, gut, heart and placenta. This accumulation provides several advantages for the parasite. The major advantage is the avoidance of the spleen and the subsequent elimination of infected erythrocytes. In addition, the low oxygen tensions in the deep tissues may provide a better metabolic environment.

Cytoadherence appears to be mediated by the protuberances on the surface of the infected erythrocyte. Histidine-rich parasite proteins are found under these "knobs"[6]. The "knobs" are expressed during the erythrocyte infection (trophozoite and schizont stages) and are formed as a result of parasite proteins exported to the erythrocyte membrane. Among human *Plasmodium* species, "knobs" are restricted to *P. falciparum.* Electron microscopy also shows that the "knobs" are contact points between the infected erythrocyte and the endothelial cell[7]. These knobs extrude a strain specific, adhesive variant protein of high molecular weight that mediates red cell attachment to receptors on venular and capillary endothelium, causing *cytoadherence*[6].

P. falciparum infected red cells also adhere to uninfected red cells to form aggregates. *Cytoadherence* and red cell aggregates are central to the pathogenesis of *P. falciparum* malaria. This further interferes with the microcirculation allows parasite development, away from the principal host defense, splenic processing and filtration. As a result, in *P. falciparum* malaria, only younger forms of the parasite are found in the peripheral circulation and the peripheral parasitemia is usually an underestimation of the true parasite loading. Mature forms of *P. falciparum* are rarely seen in the peripheral blood and when found, indicate severe infection. Sequestration does not occur in cases of *P. vivax* and *P. malariae* infections and therefore, all stages of the parasite can be seen in the peripheral blood and complications are rare[8].

1.1.2 Antimalarial treatments

In most cases, antimalarial drugs are targeted against the asexual erythrocytic stage of the parasite. The majority of antimalarial drugs

act by disturbing the polymerization (and/or the detoxification by any other way) of *heme*, thus killing the parasite with its own metabolic waste [10]. Few compounds are active against gametocytes, and also against the intra-hepatic stages of the parasite.

It is possible divide antimalarials drugs in four main classes based of their different target and mechanism of action:

- Food vacuole
- Antifolates
- Drugs involving redox mechanism
- Antibiotic therapy

1.1.2.1 Food vacuole drugs

The parasite degrades hemoglobin in its acidic food vacuole[9], producing free *heme* able to react with molecular oxygen and thus to generate reactive oxygen species as toxic by-products. A major pathway of detoxification of *heme* moieties is polymerization as the malaria pigment[10].

The food vacuole is a lysosome-like organelle in which the breakdown of hemoglobin and the detoxification of *heme* occurs. Quinine, along with its dextroisomer quinidine, has been the drug of last resort for the treatment of malaria, especially severe disease. Chloroquine is a 4-aminoquinoline analogue of quinine first synthesized in 1934 and has since been the most widely used antimalarial drug.

Chloroquine is found in high concentration in the food vacuole of the parasite. Possible mechanisms for this selective accumulation in the food vacuole are: 1) protonation and ion trapping of the chloroquine due to the low pH of the food vacuole; 2) active uptake

of chloroquine by a parasite transporter(s); and/or 3) binding of chloroquine to a specific receptor in the food vacuole.

The exact contributions of these three postulated mechanisms are not clear, but it is generally accepted that chloroquine exerts it toxic effect by interfering with the conversion of free heme to hemozoin. Large quantities of heme are released as a result of hemoglobin digestion in the food vacuole. The free heme can lyse membranes and lead to the generation of reactive oxygen intermediates. Heme is normally detoxified in the food vacuole via a biocrystallization process in which the heme is sequestered into large insoluble crystals called hemozoin or the malarial pigment. The exact mechanism by which chloroquine inhibits hemozoin formation is not known, but chloroquine can bind heme and this binding may prevent the heme from being incorporated into the hemozin crystal. Parasite killing is therefore a result of the accumulation of metabolic wastes (i.e., heme) associated with the digestion of hemoglobin[10].

Other quinoline containing anti-malarials, such as mefloquine and quinine, also appear to affect the food vacuole. However, it in not clear whether these drugs bind heme or affect the formation of hemozoin. Furthermore, these drugs are weaker bases than choroquine and may not exhibit the same degree of ion trapping within the food vacuole.

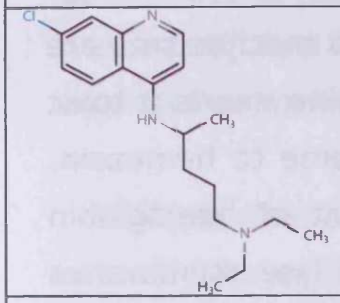
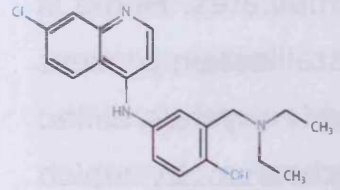
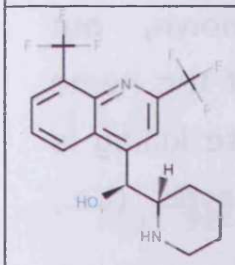
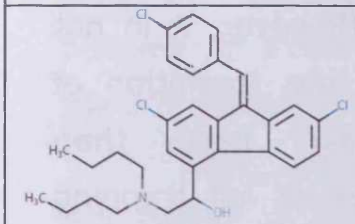
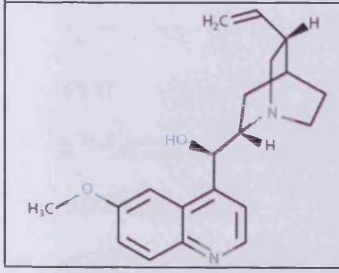
Structure	Name and Chemical class	Indications	Disadvantage	Ref.
	Chloriquine 4-aminoquinoline	Chemoprophylaxis and treatment of malaria	Widespread resistance.	11, 12, 13, 14, 15, 16
	Amodiaquine 4-aminoquinoline	Chemoprophylaxis. It is used in presence of some chloroquine-resistant strains of *P.falciparum*, although there is cross-resistance.	Risk of agranulocytosis	11
	Mefloquine 4-Methanolquinoline	Chemoprophylaxis and treatment of uncomplicated *falciparum* malaria and *vivax* malaria	Frequently nausea, vomiting, abdominal pain, anorexia, diarrhoea, headache, dizziness, loss of balance, dysphoria, somnolence and sleep disorders, notably insomnia and abnormal dreams	11,17
	Lumefantrine Aryl aminoalcohol	It is highly effective against multidrug resistant	Oral bioavailability is variable and is highly dependant on administration with fatty foods	11
	Quinine Alkaloid derivative	Treatment of *falciparum* malaria	Drug resistance. It is not suitable for children under 2 years, cardiotoxicity, very toxic in overdosage.	11, 17

Table 1.2 Principal antimalarial drugs

1.1.2.2 Antifolates

Folate metabolism is the target of several antimalarials as well as drugs used against other pathogens. Reduced folates serve as co-factors in a many one-carbon transfer reactions involved in the biosynthesis of amino acids and nucleotides. Due to its high rate of replication the malaria parasite has a high demand for nucleotides as precursors for DNA synthesis, and thus is particularly sensitive to antifolates. The two primary targets of antifolate metabolism are the de novo biosynthesis of folates and dihydrofolate reductase (DHFR).

The malaria parasite synthesizes folates de novo whereas the human host must obtain preformed folates and cannot synthesize folate. The inability of the parasite to utilize exogenous folates makes folate biosynthesis a good drug target.

DHFR is a ubiquitous enzyme that participates in the recycling of folates by reducing dihydrofolate to tetrahydofolate. The tetrahydrofolate is then oxidized back to dihydrofolate as it participates in biosynthetic reactions (e.g., thymidylate synthase). Inhibiting DHFR will prevent the formation of thymidylate and lead to an arrest in DNA synthesis and subsequent parasite death. Pyrimethamine and proguanil are the two most common DHFR inhibitors used as antimalarials. These drugs inhibit DHFR from the parasite to a greater degree than the host enzyme and thus show a selective toxicity towards the parasite.

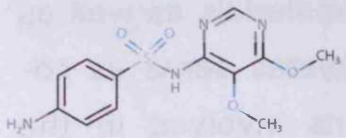
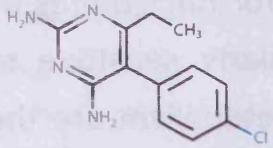
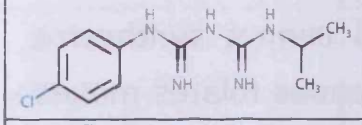
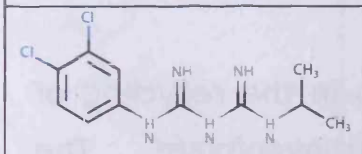
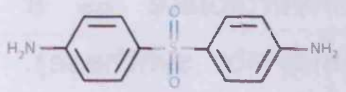
Structure	Name and Chemical class	Indications	Disadvantage	Ref.
	Sulfadoxine Sulfonamide	Treatment of *falciparum* malaria. Competitive inhibitor of dihydropteroate synthase. Use with pyrimetamine. Target: DHPS	Hypersensitivity reactions may affect different organ system.	11
	Pyrimethamine 2,4-Diaminopyrimidine	Treatment of *falciparum* malaria. It used in combination with a sulfonamide. It is active on Tissue Schizontocide and Hypnozoiticide. Target: DHFR	Widespread resistance. Can induce intravascular haemolysis not suitable for pregnant women and G6DP deficient patients	11, 17
	Proguanil Biguanide (Prodrug)	Chemoprophylaxis of malaria Target: DHFR	Drug resistance. nephrotoxicity	11
	Chlorproguanil Biguanide (Prodrug)	Chemoprophylaxis of malaria. It used in combination with dapsone. Target: DHFR	Drug resistance. Nephrotoxicity	11
	Dapsone Sulfone	Chemoprophylaxis of malaria. Target: DHPS	Haemolysis and mathaemoglobinaemia	11

Table 1.3 Principal antimalarial antifolates.

1.1.2.3 Drugs involving redox mechanism

Several anti-protozoal drugs are believed to act via oxidative stress. Metabolic processes will produce reactive oxygen intermediates (ROI), which can damage cellular components such as lipids, proteins, and nucleic acids. The high metabolic activity of most protozoan pathogens will result in production of even higher levels of ROI. This is exemplified by the malaria parasite, which produces ROI as a consequence of hemoglobin digestion and the release of free heme. All cells have mechanisms by which the ROI can be detoxified. Drugs, which specifically increase the levels of oxidative stress in the parasite, may overwhelm these ROI defense mechanisms and lead to parasite death. Levels of oxidative stress can be increased by drugs that are direct oxidants, as well as by drugs, which participate in oxidation-reduction cycling, sometimes called futile redox cycling[18].

An important sesquiterpene compound has been discovered from the plant *Artemisia annua*. Artemisinin and various synthetic derivatives and analogues represent a very important class of antimalarials. The most important artemisinin derivatives are artesunate, artemether, arteether and dihydroartemisinin. These compounds are used for treatment of severe malaria and have shown very rapid parasite clearance times and faster fever resolution than occurs with quinine. They are becoming more and more commonly used throughout the world.

Artemisinin's structure is unlike those of any other known antimalarial and is thus likely to have a different mechanism of action. It is likely that the endoperoxide briege is necessary for the antimalarial activity. Since peroxides are a known source of reactive oxygen species such as hydroxyl radicals and superoxide[18], this observation suggested that free radicals might be involved in the

mechanism of action. A number of other laboratories have subsequently shown that artemisinin treatment of membranes (especially in the presence of heme) can cause lipid peroxidation, a typical product of free radical damage[18, 19, 20]. A possible mechanism of action might be that artemisinin interacts with intraparasitic heme, suggesting that intraparasitic heme or iron might function to activate artemisinin inside the parasite into toxic free radicals[18]. The malaria parasite is rich in heme-iron, derived from the proteolysis of host cell hemoglobin[21]. This could explain why artemisinin is selectively toxic to parasites. Once formed, the artemisinin-derived free radicals appear to damage specific intracellular targets, possibly via alkylation.

Adverse effects are rare in patients treated with artemisinin derivatives. However, in several animal studies, artemisinin derivatives have been clearly shown to cause neurotoxicity at high doses[18, 22]. There have been two case reports of patients suffering neurological problems after receiving artemisinin treatments. However, the neurological problems in these two isolated cases may have been unrelated to their treatment. In light of the very large number of patients treated safely with artemisinin, neurotoxic adverse effects, if they occur, are quite rare[18].

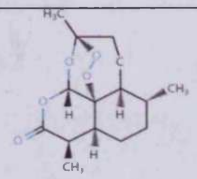
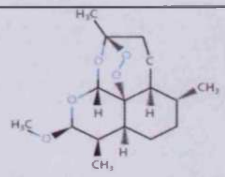
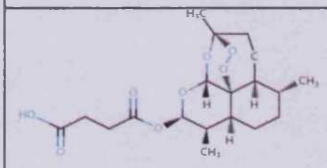
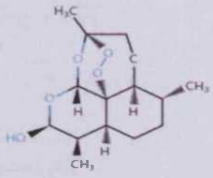
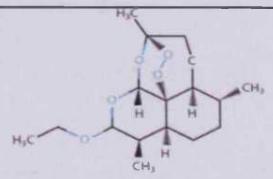
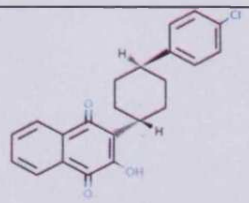
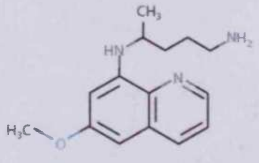
Structure	Name and Chemical class	Indications	Disadvantage	Ref.
	Artemisinin Sesquiterpene lactone	Active against schizontocide of all *Plasmodium* species. It also kills the gametocytes	Probable neurotoxicity, low solubility	11, 12, 13
	Arthemether Sesquiterpene acetal with endoperoxide	Treatment of severe chloroquine resistant malaria	Probable neurotoxicity	11, 12, 13
	Artesunate Sesquiterpene ester endoperoxide	Treatment of severe chloroquine resistant malaria	Low stability at pH7, intravenous administration	11, 12, 13
	Dihydroartemisinin Sesquiterpene derivative	Dihydroartemisinin is the main active metabolite of the artemisinin derivatives, but can also be given orally and rectally as a drug in its own right	Probable neurotoxicity	11, 12, 13
	Artemotil Ethyl ether of artemisinin	It is oil-based so water insoluble. It is given by intramuscular injection only	Slow absorption, low solubility. Probable neurotoxicity	11, 12, 13
	Atovaquone Hydro-naphthoquinone	Active against all *Plasmodium* species, Atovaquone interferes with cytochrome electron transport	Poorly absorbed from the gastrointestinal tract	11
	Primaquine 8-Methanolquinoline	Adjunct in the treatment of *P. ovale* and *P. vivax* malari. Probable interference with the parasite's energy metabolism in the mitochondria	Widespread resistance. Can induce intravascular haemolysis; not suitable for pregnant women	11, 17

Table 1.4 Principal antimalarial drugs

1.1.3 Antibiotic therapies

Tetracyclines are often used in conjunction with other drugs to combat chloroquine resistant falciparum malaria. Plasmodium protein synthesis appears to be eukaryotic, and is insensitive to chloramphenicol, but affected by cycloheximide. It has been suggested that antibiotics such as tetracycline act on the mitochondrial ribosomes of the parasite, inhibiting protein synthesis. Macrolides such as erythromycin seem to inhibit autophagic vacuole formation, thus potentiating the action of chloroquine. Resistance to these compounds is not a current problem[23].

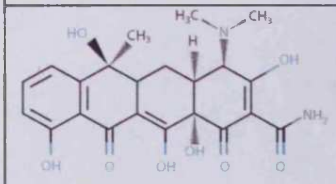
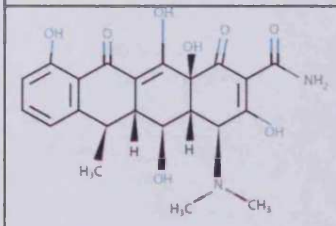
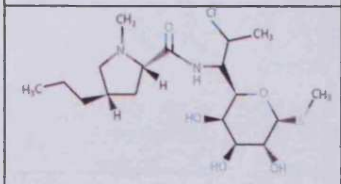
Structure	Name and class derivative	Indications	Disadvantage	Ref.
	Tetracycline Antibiotics originally derived from certain *Streptomyces* species	Inhibitors of aminoacyl-tRNA binding during protein synthesis. Administered orally or intravenously	Gastrointestinal side effects. Hypersensitivity reactions. Deposited in calcifying areas	11
	Doxycycline Tetracycline derivative	It may be preferred to tetracycline because of its longer half-life, more reliable absorption and better safety profile in patients with renal insufficiency	Gastrointestinal effects	11
	Clindamycin Lincosamide antibiotic	It inhibits the early stages of protein synthesis by a mechanism similar to that of the macrolides	Diarrhea, pseudo-membranous colitis may develop during or after treatment, which can be fatal	11

Table 1.5 Principal antimalarial drugs

1.1.4 Drug Resistance

The emergence of drug resistance severely limits the arsenal of available drugs against protozoal pathogens. Parasites have evolved numerous ways to overcome the toxicity of drugs (mutations in target gene, increase production of target, decrease drug accumulation, drug inactivation). Quite often drug resistance involves mutations in the drug target so that the drug does not bind or inhibit the target as well. Drug resistance can develop quickly in situations where a single point mutation can confer resistance. Another mechanism of drug resistance involves expressing higher levels of the target. This can be accomplished either through increased transcription and translation or gene amplification. This results in a requirement for higher levels of drugs to achieve the same level of inhibition. Decreasing drug accumulation or metabolizing the drug to non-toxic products will result in less drug reaching the target and can also contribute to drug resistance. Drug resistance can also involve the accumulation of mutations in the same or different targets, which will have additive or synergistic effects. Parasites with mutations or genetic polymorphisms, which confer a decrease in drug sensitivity, will be selected under drug pressure[24].

A strategy that has received much attention recently is the combination of antimalarial drugs, such as mefloquine or amodiaquine, with an artemisinin derivative. Artemisinin drugs are highly efficacious rapidly active, and have action against a broader range of parasite developmental stages. This action apparently yields two notable results. First, artemisinin compounds, used in combination with a longer acting antimalarial, can rapidly reduce parasite densities to very low levels at a time when drug levels of the longer acting antimalarial drug are still maximal. This greatly reduces both the likelihood of parasites surviving initial treatment and the

probability that parasites will be exposed to suboptimal levels of the longer acting drug.

Second, the use of artemisinins has been shown to reduce gametocytogenesis by 8- to 18-fold. This reduces the probability that gametocytes carrying resistance genes are passed onwards and potentially may reduce malaria transmission rates. Use of combination therapy has been linked to slowing of the development of mefloquine resistance and reductions in overall malaria transmission rates in some parts of Thailand and has been recommended for widespread use in sub-Saharan Africa[24].

Name	Commercial Name	Indication	Disadvantages	Ref.
Pyrimethamine with Sulfadoxine	Fansidar®	Adjunct to quinine in treatment of *falciparum* malaria	Same as pyrimethamine, severe side-effects on long term use.	11, 25
Proguanil with Atovaquone	Malarone®	Prophylaxis and treatment of acute uncomplicated *falciparum* malaria	Nephrotoxicity, not administrable during pregnancy and during breast feeding	11
Arthemeter with Lumefantrine	Riamet®	Treatment of acute uncomplicated malaria	Not suitable in case of cardiac problems (arrhythmias, bradycardias, etc.) and during breast feeding.	11
Chloroproguanil with Dapsone	LapDap®	Treatment of uncomplicated *falciparum* malaria	Not suitable during breast feeding	11, 25

Table 1.6. Main combinations used for the treatment of malaria according to the BNF N.46

2 Chapter: The enzyme dUTPase

The enzyme deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase, E.C. 3.6.1.23) is involved in nucleotide metabolism, it catalyses the hydrolysis of deoxyuridine triphosphate (dUTP) to deoxyuridine monophosphate (dUMP), an essential precursor in the biosynthesis of thymidine nucleotides, and inorganic pyrophosphate in the presence of magnesium ions (Figure 2.1).

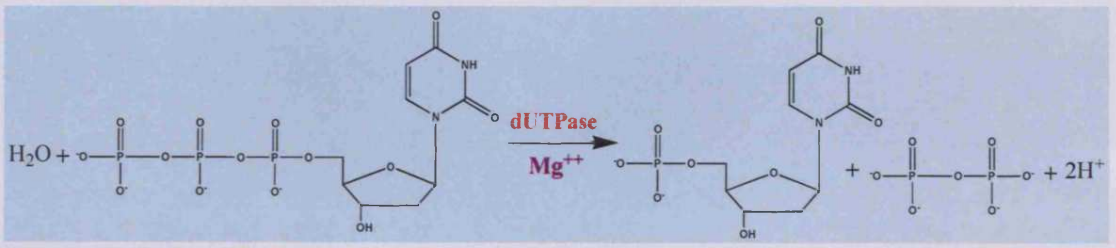


Figure 2.1 Hydrolysis reaction of dUTP into dUMP and PPi

The dUTPase family is part of group of enzymes (NTPase), which have been defined as "house cleaning", their function is to cleanse the cell of potentially dangerous metabolites and modulate the accumulation of intermediates in biomedical pathways.

For a house-cleaning enzyme, the protection of the genome DNA integrity can be considered the most important role. Despite the high capability of DNA polymerases to discriminate between nucleotides some mutagenic nucleotides and non-canonical nucleotides (Figure 2.2) can be incorporated into the chromosomal DNA, causing mispairing that can lead to serious DNA damage such as genome mutations[26].

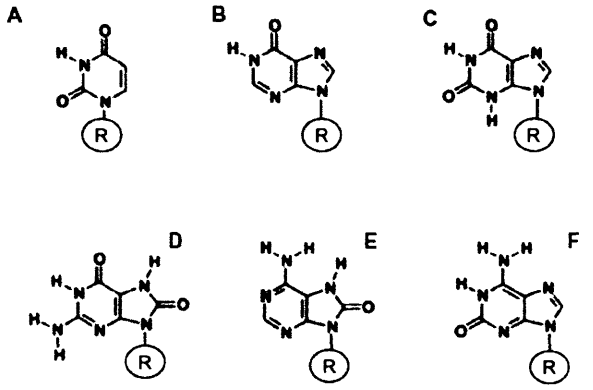


Figure 2.2 Structures of non-canonical NTPs discussed in the text. A, dUTP; B, dITP; C, dXTP; D, 8-oxo-dGTP; E, 8-oxo-dATP; F, 2-oxodATP. R in all cases indicates deoxyribose triphosphate[26].

Uracil in DNA results from either (i) deamination of cytosine (Figure 2.3), creating a premutagenic U:G mispair, or (ii) from misincorporation of deoxyuridine instead of deoxythymidine by the enzyme DNA polymerase, using dUTP instead of dTTP as substrate, hence creating a U:A pair[27].

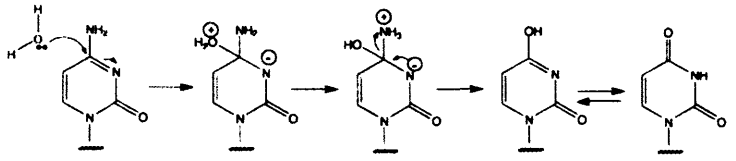


Figure 2.3 Mechanism for the deamination of cytosine to uracil, this process is catalysed by the enzyme cytosine deaminase.

If not replaced, the U:G mispair can, after replication, lead to an A:T transition thus undermining the integrity of the cell genome.

dUTPase is thought to be crucial to DNA integrity in two ways. It prevents the build up of dUTP and ensures the provision of dUMP, the substrate for thymidylate synthase in the biosynthesis of deoxythymidine triphosphate (dTTP). As a result, a low dUTP/dTTP ratio is maintained, which greatly reduces uracil incorporation into DNA because DNA polymerases do not discriminate between dUTP and dTTP. Under normal circumstances, following dUTP misincorporation into DNA, uracil is excised and replaced by thymine

through a repair process catalyzed by uracyl-DNA glycosylase (Figure 2.4). However, an abnormally high incorporation of dUTP into the DNA triggers an attack by uracil-DNA glycosilase that can lead to double-strand breaks, which compromise the DNA integrity. The currently available data indicate that uracil glycosylase-mediated DNA repair is central to this process[27].

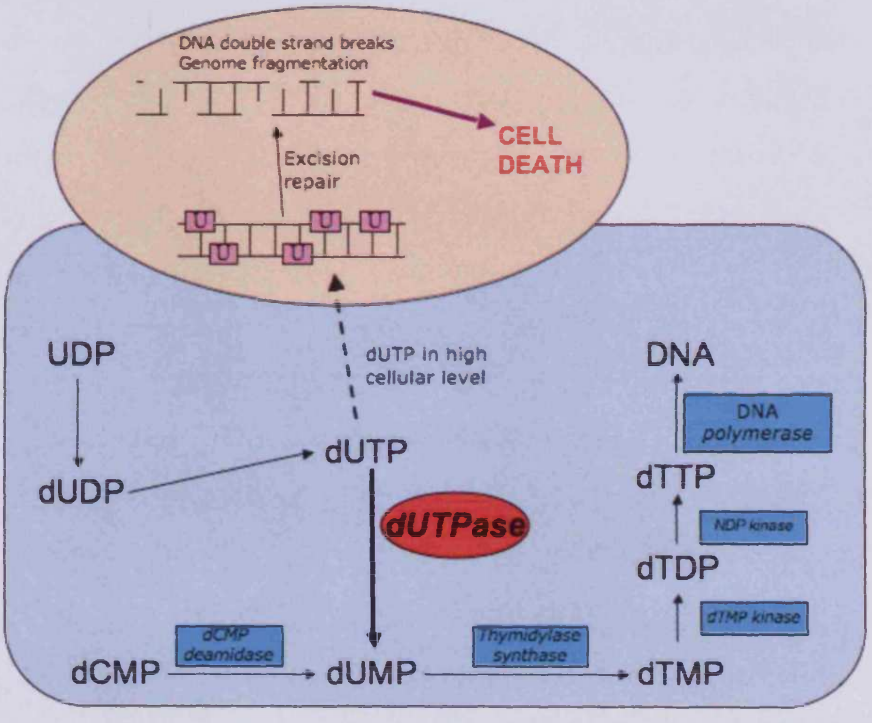


Figure 2.4 Metabolic pathways of nucleotides

Some experiments on *E.coli* showed that mutations of the dUTPase gene (*dut*) are lethal in the presence of a functional uracil-DNA-glycosidase (*ung*) gene[28]. The deletion of the gene *ung* can suppress this lethality and lead to the incorporation of uracil into DNA. This effect can be explained by the elevated incorporation of uracil into DNA triggering a high recombination frequency and an abnormal mutation rate, and leading to the DNA fragmentation and the cell death[29, 30]. Another explanation of this lethality, could be that this excision process simply delays the completion of DNA which eventually results in cell death. In both of cases, the absence of

dUTPase leads to excessive incorporation of dUTP in the DNA, which is lethal for the cell[27].

The dUTPase enzyme is encoded in virtually every bacterial, archaeal or eukaryotic genome (some viruses too). The enzyme is even found in organisms, like human, that do not encode a dCTP deaminase and produce dUMP by deamination of dCMP[26].

Although uracil appears very similar to the thymine, it has been proved that its extensive incorporation into DNA is lethal in bacteria, yeast[31] and likely all cellular systems[17].

Drugs targeting enzymes of *de novo* thymidylate biosynthesis, *e.g.* fluorouracil against thymidylate synthase or methotrexate against dihydrofolate reductase, are widely used in anticancer and anti-infective chemotherapy. Their molecular mechanism of action is to induce grave perturbances in the cellular dUTP/dTTP pools that result in thymine-less cell death preferentially killing cells with actively ongoing DNA synthesis, *e.g.* tumor cells and pathogens[27]. dUTPase inhibition should result in similar effects. Opportunities exist for designing inhibitors not only for human dUTPases, but also selective inhibitors for *P. falciparum*, *T. brucei* or *T. cruzi* dUTPases, in order to find a new remedy against parasitic diseases.

2.1 Structure of dUTPases

Structurally, dUTPases can be classified into three families according to their subunit composition as monomers, dimers, or trimers. Trimeric dUTPases form the largest and most varied group (Table 2.1).

Subunit composition	Sources
Monomers	Herpes virus
Dimers	Leshmania, Trypanosomes, Campylobacter
Trimers	Human, Bacteria, Fungi, Equine Infectious, Anemia virus, *Plasmodium falciparum*

Table 2.1 Structures of dUTPase form several sources

Currently, X-ray structures of seven homotrimeric enzymes have been published: *E. coli*[32, 33], *Homo sapiens*[34], *equine infectious anemia virus* (EIAV)[35], *feline immunodeficiency virus* (FIV)[36], *Methanococcus jannaschii*[37], *Mycobacterium tuberculosis*[38], and *Plasmodium falciparum*[39]. Two crystal structures of homodimeric enzymes are available: *Trypanosoma cruzi* and *Campylobacter jejuni*. Recently, the crystal structure of the monomeric form of *Epstein-Barr virus*[40] has been published (Table 2.2).

	Unliganded	dUMP	dUDP	dUTP	α, β-Imido dUTP	Inhibitor
P.falciparum						1vyq
M.tuberculosis	1mq7	1snf(Mg^{2+})	1slh(Mg^{2+})	1smc 1sn8(Cr^{3+})	1six(Mg^{2+}) 1sjn(Mg^{2+})	
E.coli	1dup 1eu5 1euw	1seh	1dud	1syl	1rn8	
Human	1q5u	Not in PDB	1q5h	Not in PDB		
EIAV	1dum		1duc(Sr^{2+})			
FIV	1dut 1f7d 1f7o	1f7k 1f7n	1f7p 1f7r	1f7q		

Table 2.2 Accession codes of dUTPase crystal structures deposited in PDB. When metals are included in the crystal structure, they are indicated in parenthesis next to the PDB ID code.

2.1.1 Homotrimeric

In general, the homotrimers, formed from three identical subunits, are rich in β-strands. Usually, the subunit also contains one or two short α-helices. The β-strands (eleven to thirteen) are connected by loops, some of which are regular β-turn. Each β-pleated subunit resembles the canonical jellyroll fold of β-barrel proteins, with the C-terminal β-strand in contact with the neighboring subunit. (Figure 2.5)

Despite the limited sequence identity prevailing between the trimeric dUTPases analyzed so far, there is an important degree of similarity in the three-dimensional geometry of the core monomers, reflecting the high frequency of well conserved amino acid substitutions from bacteria to man and even retrovirus (Figure 2.6). In particular it is possible identify five conserved sequence motifs (Figure 2.7)[41].

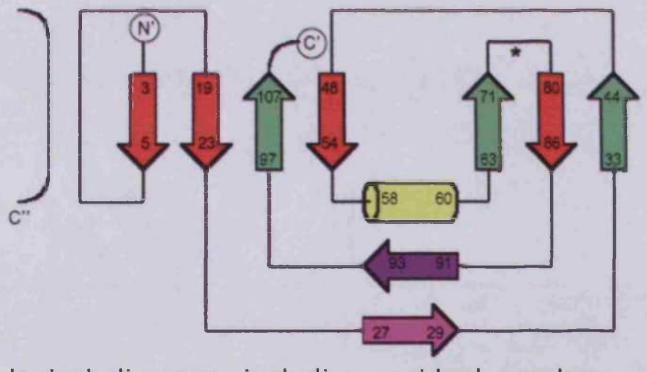


Figure 2.5. Topological diagram, including residual numbers, of the fold of the EIAV dUTPase monomer. N' and C' denote the N-terminus and C-terminus, respectively, while C'' indicates the carboxy-terminus of an adjacent subunit. The asterisk indicates the position of the β-hairpin loop that is involved in base and sugar recognition[41].

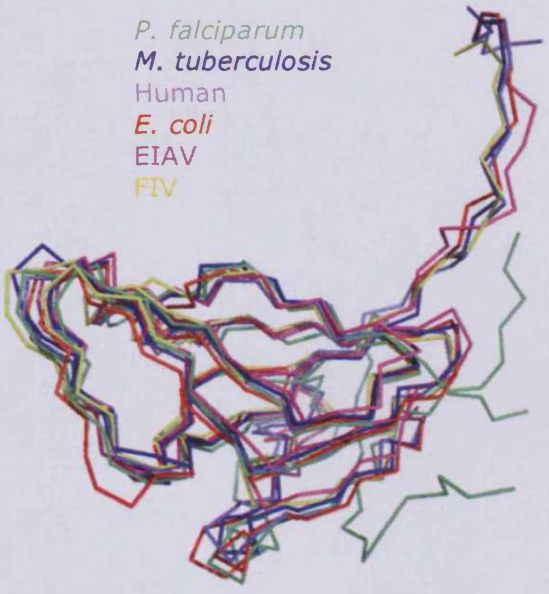


Figure 2.6. Structure of the dUTPase monomer. α-Carbon traces of six dUTPase homologs are shown superimposed.

The only major structural variations between the monomeric subunits are in surface loop regions and in the length of the polypeptide. Given the structural similarity between the dUTPase so far published, it can be expected that all trimeric dUTPases will have similar overall folds. It is also likely that this family of dUTPases operate by the same catalytic mechanism.[41]

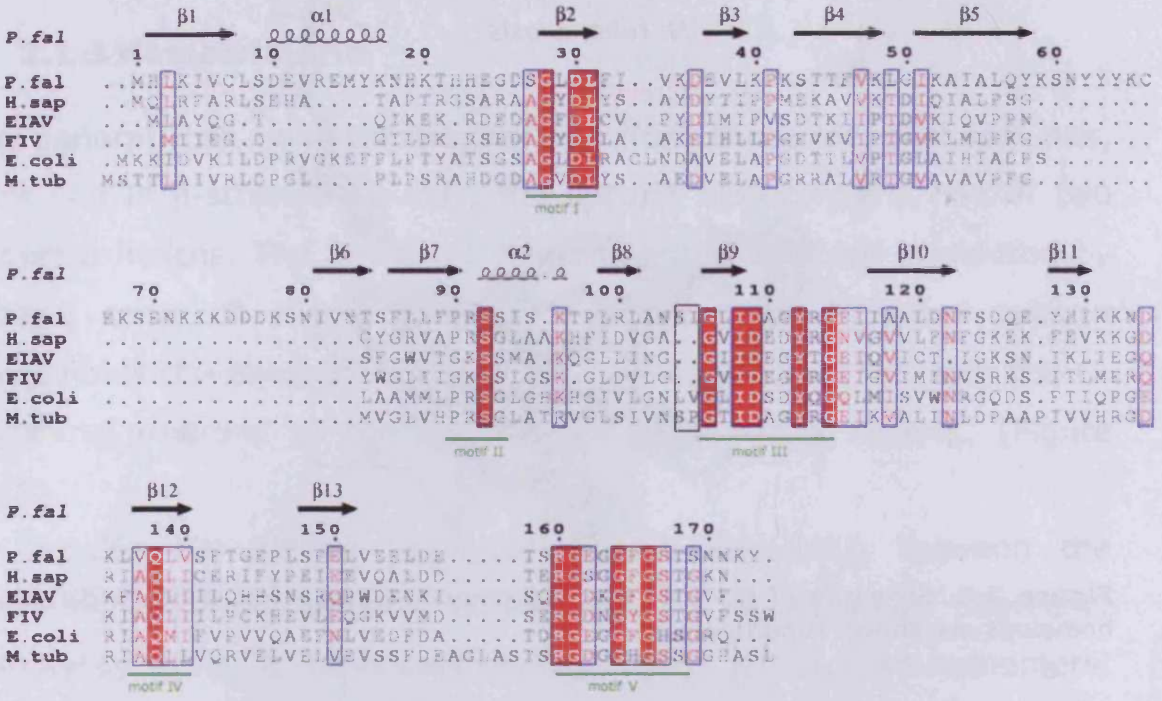


Figure 2.7 Alignment of the Sequences of Trimeric dUTPases. Known structures from: *P. falciparum* (P. fal), human (H. sap), *equine infectious anemia virus* (EIAV), *feline immunodeficiency virus* (FIV), *E. coli* (*E. coli*), and *M. tuberculosis* (M. tub) are shown.
The sequences were initially aligned using the program ESPript[42] and then adjusted manually in the light of structural data. The numbering and the secondary structure elements above the alignment refer to the *Pf*dUTPase structure. Strictly conserved residues are highlighted in red blocks while largely conserved residues are outlined by blue boxes. The secondary structure of the *P. falciparum* dUTPase is indicated above the sequence. The insertion of two residues is in the black box. Locations of the five conserved sequence motifs are indicated below the sequences[39].

The interactions between the monomeric subunits in the dUTPase homotrimer fall into three classes[41] (Figure 2.8).

1. Firstly, two neighboring subunits create hydrophobic contacts.
2. The C-terminal β-strand of the first subunit becomes integrated into the β-stranded jelly-roll fold of the second subunit. The C-terminal arm continues in crossing over the surface of the second subunit to reach the active site that receives conserved sequence motifs from the third subunit.
3. Residues from all subunits contribute to form the central channel of the homotrimer. The three contact types were termed pairwise, arm crossing and threefold interaction, respectively. Pairwise and β-strand swapping contacts are of highly similar character in all

dUTPase homotrimers[41].

The active site architecture also follows a common pattern where the three active sites of the homotrimer, located in clefts between neighboring subunits, are built in a fashion of basically 3-fold symmetry utilising conserved motifs from all subunits[32, 34, 41].

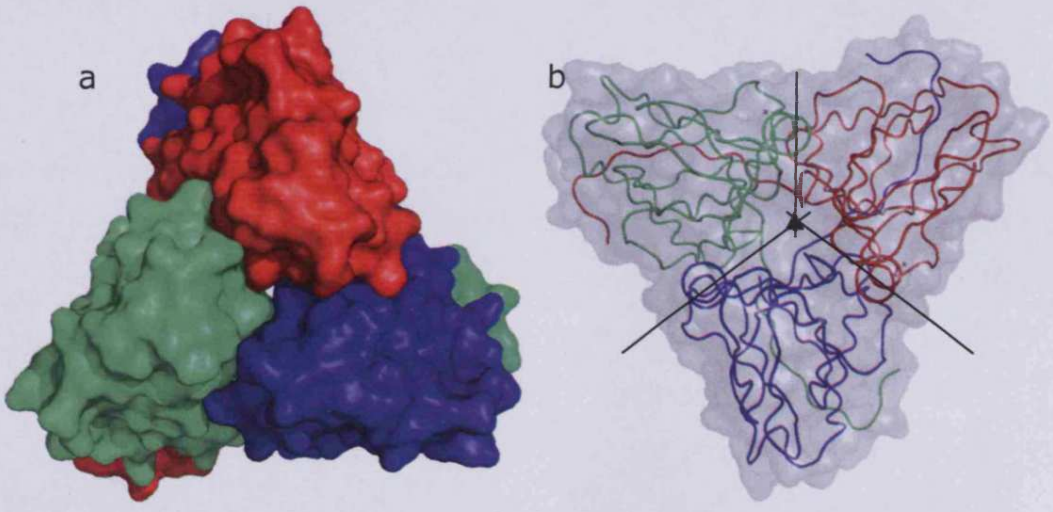


Figure 2.8 a) Surface representation of the *E. coli* dUTPase trimer. The individual subunits are coloured red, blue and green; b) Ribbon representation (and transparent surface) of the *E. coli* dUTPase trime viewed approximately along the three-fold axis (black lines), the black triangle shows the position of central channel of the homotrimer. The three individual subunits are colored in red, blue and green.

The arm-crossing contacts are very similar for all the crystal structures known (except for the *P. falciparum* dUTPase); the crossing arm runs in the same plane as the neighboring subunit, and contributes a β-strand to form a β-barrel. These interactions are very well defined for the bacterial proteins, but less pronounced for the EIAV dUTPase, which does not define any precise secondary structure, forming only a few contacts.

In the *P. falciparum* enzyme, the C-terminal part of the chain executes a sharp turn and the last β-strand packs alongside the first β-strand in the core of the molecule. The redirection of the chain is facilitated by the presence of residue Gly 144, which is replaced by a

bulkier residue in all the other trimeric dUTPases (Figure 2.7). As a result, each folding domain in *P. falciparum* dUTPase is made up of a single polypeptide chain and many of the inter-subunit interactions involving the C-terminal are absent[39].

2.1.1.1 *Eukaryotic and prokaryotic homotrimeric dUTPases*

Despite the overall similarities in quaternary structure, the available data suggest the existence of marked differences in the mode in which the residues, from all subunits, contribute to forming the central channel of the homotrimer. These differences are considerable between prokaryotic and eukaryotic dUTPases.

- Firstly, structural superimpositions and sequence alignments have indicated that the 3-fold inner channel of the trimer is highly polar character in human dUTPases, whereas it is constructed by apolar residues in the prokaryotic enzymes[34, 43]. In EIAV and FIV enzymes the interactions are of mixed hydrophobic and hydrophilic character.
- Secondly, in agreement with the altered hydropathy of the inner 3-fold interaction surfaces, a divalent metal ion (Mg^{2+}) has been discovered at the base of this inner channel in the crystal structure of the human enzyme[34], 18 Å away from the active site. There is no presence of the metal ion in the structure of the prokaryote *Escherichia coli* dUTPase where the 3-fold channel is densely packed with hydrophobic side chains. However, *M. tuberculosis* dUTPase was crystallized with the presence of a molecule of *tris* (Figure 2.9), which is relatively hydrophilic, at the base of the inner channel[38]. Despite the fact that this is a prokaryotic organism, the presence of the *tris* molecule indicates that the channel in *M. tuberculosis* dUTPase is not very hydrophobic.

- Thirdly, comparing the composition of the residues in eukaryotic and prokaryotic enzymes, the latter has some insertions (Figure 2.7) (Ser, Pro in *M. tuberculosis* and Leu, Val in *E. coli*) that constrict the diameter of the trimer interface channel relative to the human enzyme, reducing the volume of the channel to approximately half of the size of the human enzyme[38].
- Fourthly, based on the crystal structure data of human and *E. coli* dUTPases, the crucial conformational change, leading to the formation of the catalytically competent closed enzyme conformer, may occur with an altered timing in these two enzymes. In the bacterial enzyme, detailed functional investigations[43], in agreement with the crystal structure in complex with the nucleotide dUDP[32], indicate that the γ-phosphate of the substrate is indispensable in inducing the closed conformer, and after catalytic cleavage the active site pops open. In the case of the human enzyme, the crystal structure suggests the retaining of the closed conformer even with the product dUMP[34]; however there is no functional data published in the human enzyme. This difference in active site conformational rearrangements, dependent on subunit interactions, suggests that the catalytic cycle may be altered in the bacterial and human dUTPases. It may be a general difference between bacteria and eukaryotes[38].

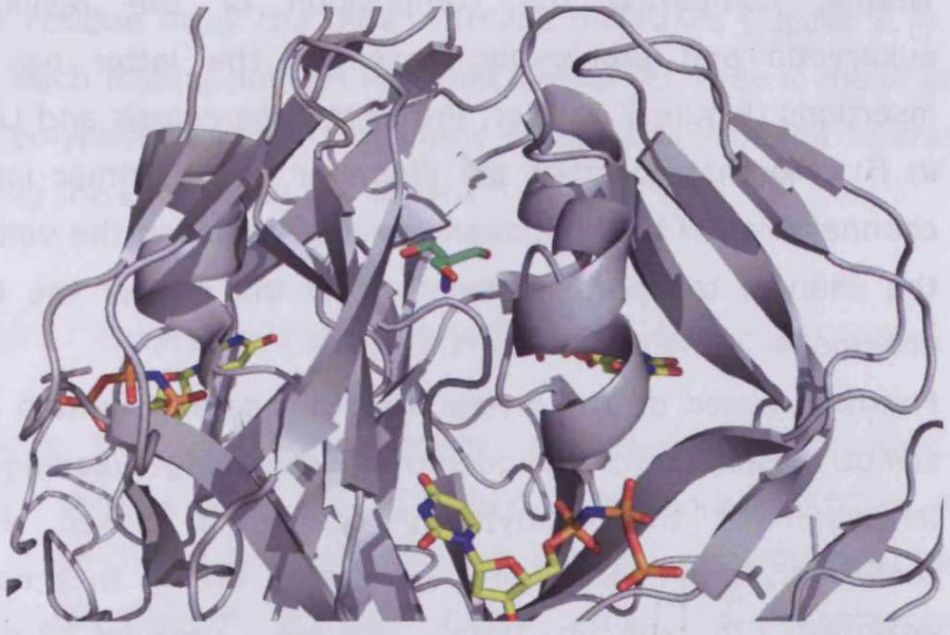


Figure 2.9 Cut-away view of the trimer interface channel of *M. tuberculosis* dUTPase. The Tris molecule is colored in green. A non-nucleoside inhibitor targeted to bind in this channel would likely interfere with dUTP binding by distorting the conformation of residues lining the uracil binding pocket and/or the trimer assembly[38].

The *P. falciparum* dUTPase is a eukaryote enzyme; however it does not present the typical eukaryotic features that distinguish eukaryotic dUTPase from prokaryotic dUTPase.

- Firstly, analyzing the crystal structure, it can be noticed that the crossing arm has a rather atypical conformation. The C-terminal executes a steep turn interacting only with its subunit, whilst all the other known dUTPases show a crossing C-terminal arm in contact with a neighboring subunit.
- Secondly, the 3-fold inner channel in the human enzyme is highly hydrophilic; however in the *P. falciparum* dUTPase this channel has a remarkably hydrophobic character, as reported also in prokaryotes.
- Thirdly, the *Plasmodium* enzyme contains an insertion of two residues (Ser, Pro), which are part of the trimer interface channel. This insertion is present also in *M. tuberculosis* and *E. coli* (Figure 2.7).

For all these reasons, the *P. falciparum* dUTPase may be

considered closer to the prokaryotic kingdom than to the eukaryotic.

2.1.1.2 *Active site*

The dUTPase active sites, as defined by crystal structures of numerous substrate and substrate analogue complexes, are positioned in the depressions at the subunit interface.

Two successive β-strands and their connecting turn constitute the core of the active site that accommodates the uracil and deoxyribose rings of the substrate dUTP. Phosphate recognition is provided by conserved motifs of the neighbouring subunit. More specifically, a single pocket is composed of five conserved sequence motifs (Figure 2.10).

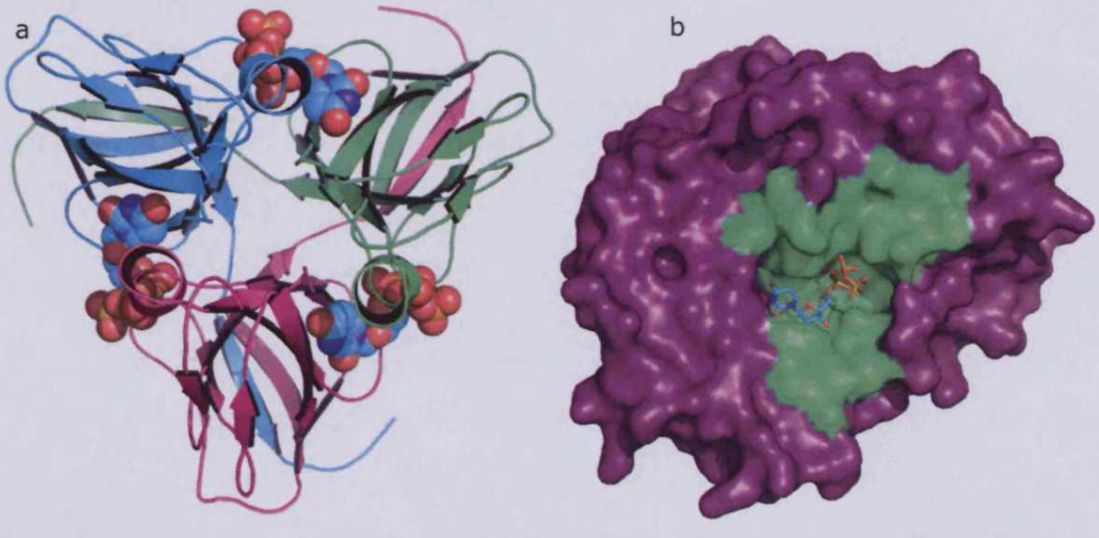


Figure 2.10 a) Ribbon representation of the *E. coli* dUTPase trimer in complex with the substrate analogue dUDP viewed approximately along the three-fold axis; b) Surface model of one active site region of the *E. coli* dUTPase with dUDP bound. dUDP is coloured by atom while the conserved motifs, which are all clustered around the active site, are coloured in green

Motif I, (Ser/Ala)-Gly-(Pho)-Asp-Leu; motif II, (Pro/Gly)-(Arg/Lys)-Ser-(Ser/Gly) and motif IV (Arg/Lys)-(Pho)-(Val/Ala)-Gln-(Leu/Met) from one subunit contribute to binding the triphosphate moiety. Motif

III, Gly-(Pho)-Ile-Asp-(Nnn)-(Gly/Asp)-Tyr-(Nnn)-Gly, from the adjacent subunit contributes primarily to binding the nucleoside moiety. Motif V, Arg-Gly-$(Nnn)_2$-Gly-(Nnn)-Gly-(Ser/His)-(Thr/Ser)-(Ser/Gly), from the remaining subunit is a domain swapped element that contributes an arginine residue involved in a bifurcated hydrogen bond with the ligand's phosphate moiety (the amino acid in three letter code: Pho, hydrophobic residue; Nnn, any residue). Motif V has been reported to contain sequence similarity to the P-loop motifs commonly employed in nucleotide binding enzyme to bind phosphate group[33] (Figure 2.11).

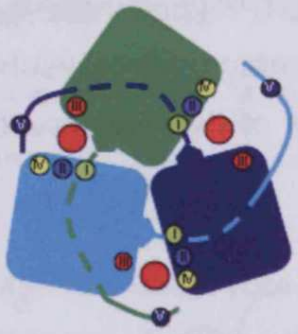


Figure 2.11. Schematic organization of the trimeric dUTPases and localization of the conserved sequence motifs I–V around the active sites represented by red dots.

2.1.1.3 Recognition of the Uracil group

Recognition of the uracil base is mainly accomplished by two anti-parallel β-strands, which are connected together by a β-harpin loop. The open end of the β-hairpin is narrow, since the two strands are in contact by hydrogen bonds, involving the backbone and a molecule of water. The bottom of the active site, where the uracil group is accommodated, overall has a hydrophobic character. The β-hairpin widens towards the turn loop, as there are not hydrogen bonds between the two strands in this region (Figure 2.12).

Uracil binding involves hydrogen bonds with the backbone of the active site and a conserved buried molecule of water. This water

molecule is conserved in each complex and it is hydrogen bonded to O4 of uracil. The N3, O4 and O2 of the uracil are also bonded to bottom of the cleft. These interactions are all recognised specifically by the enzyme in a mode that mimics base-pair hydrogen bonding.

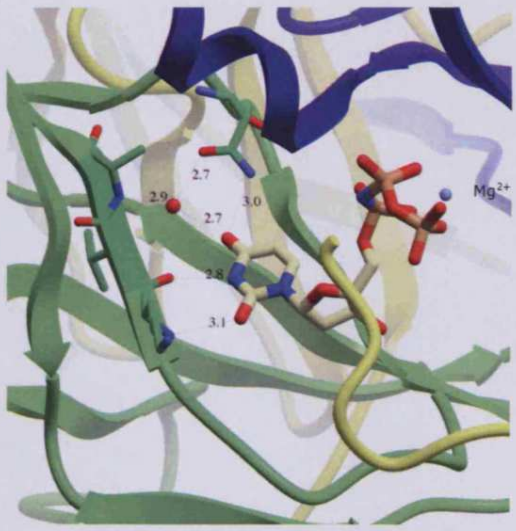


Figure 2.12. An overall view of the active site of *M. tuberculosis* dUTPase with bound α, β-imido dUTP and Mg^{2+} illustrating the structural basis of substrate specificity and the hydrogen-bonding pattern between the enzyme and the uracil moiety. The structurally conserved water molecule participating in uracil recognition is shown in red.

The residues in the ligand-binding pocket are well conserved in all the known dUTPases. However, there is one significant difference in the architecture of the uracil-binding pocket of prokaryotic dUTPases that sets them apart from dUTPases from other source[41]. A residue of one of the two β-strands of the hairpin provides a hydrogen bond donor for O4 of uracil. This donor is a backbone or side chain atom depending on the enzyme source. In the non-prokaryotic dUTPases, the hydrogen bond donor is a backbone amide nitrogen atom (e.g. in the human enzyme Gly 76). In the prokaryotic dUTPases there is a two-residue β-bulge insertion in β-strand, in *M. tuberculosis* Ser 78 and Pro 79, in *E. coli* Leu 85 and Val 86. This places the Nδ2 atom of an asparagine side chain (Asn 77 in *M. tuberculosis* and Asn 84 in *E. coli*) in the position of hydrogen bond donor to O4. A similar insertion

is present in the *Plasmodium* enzyme (Ser 104 and Ile 105), and the hydrogen bond donor to O4 in this case is the Nδ2 atom of an asparagine (Asn 103). This emphasizes the peculiarity of this enzyme, which presents a remarkable number of characteristics typical of prokaryotic dUTPases. It might be expected that the insertion of two residues would significantly change the volume of the nucleotide binding pocket. However, because the β-bulge is directed toward the trimer axis and away from the uracil-binding pocket, the volume of the pocket changes only slightly. Hence, the structural complementarities between ligands and enzyme are remarkably similar in all dUTPase homologues[38].

2.1.1.4 dUTPase selectivity

As already mentioned, the dUTPase "house cleaning" has to be highly selective with regard to the base moiety. Binding and hydrolysis of nucleotides containing bases other than uracil needs to be avoided. This in particular applies to the structurally most similar nucleotides dCTP and dTTP, as unspecific hydrolysis of those would be highly damaging to DNA synthesis, repair and cause futile energy consuming cycle. Specificity for uracil over cytosine is mainly a consequence of differences in the number of hydrogen bonds the base recognition site is able to form with the two bases. Uracil forms four hydrogen bonds with the polypeptide backbone and a structurally conserved water molecule. In contrast, cytosine would form an alternate hydrogen-bonding pattern, presumably involving fewer bonds. Kinetic studies on various microorganisms demonstrate that dUTPase hydrolyses dCTP much less efficiently than dUTP, and the results indicate that this is due mainly to poorer binding of this nucleotide[26, 44] (Table 2.3). The selectivity is in particular derived from a strong hydrogen bond between a main-chain carbonyl oxygen in the enzyme and the N3 of the base, which is protonated in uracil

but not in cytosine. Thymine exclusion from the β-hairpin is based on steric hindrance. The binding site is narrow and binding of thymine would cause a steric clash between the polypeptide and the 5'-methyl group.

EIAV dUTPase shows slightly less specificity especially for the descrimination of dTTP. This might be due to the presence of two glycine residues (Gly 68 and Gly 69), which can adopt double conformation by flipping the peptide bond back and forth allowing to some flexibility, which might permit a low degree of dTTP binding as well[41].

Nucleotide	*T. cruzi*	*L.major*	*E. coli*	*EIAV*	*HSV-1*	*MMTV*
dUTP:						
K_m	0.53	2.11	0.2	1.1	0.3	0.8
(k_{cat}/K_m)	($5.2x10^6$)	($2.3x10^7$)	($4.0x10^3$)	($2.0x10^7$)	($2.0x10^7$)	($2.0x10^6$)
dTTP						
K_m	>800	1514	>2000	260	400	nd
(k_{cat}/K_m)	($2.5x10^3$)	($5.5x10^3$)	(34)	(<2000)	(1000)	
dCTP						
K_m	Nd	>2500	4000	3000	1000	nd
(k_{cat}/K_m)		($5.0x10^3$)	(<100)	(1000)	(2000)	
UTP						
K_m	Nd	>2500	2500	nd	1000	nd
(k_{cat}/K_m)		($2.0x10^3$)	(<100)		(200)	
dUDP						
K_m	6.27	62.7	K_i=15	K_i=3.6	K_i=17	nd
(k_{cat}/K_m)	($1.7x10^6$)	($1.0x10^6$)				
dUMP: K_{ip}	18.4	13.05	1500	130	170	nd

Table 2.3 Michaelis-Menten (K_m), specificity (k_{cat}/k_m) and inhibition constants (K_i) of viral, bacterial and trypanosomatid dUTPase for different nucleotides. Units for K_m, K_i and K_{ip} are mM and for k_{cat}/k_m is $M^{-1}s^{-1}$. EIAV: equine specific anaemic virus, HSV-1: herpes simplex virus-type 1, MMTV: mouse mammary tumour virus[44].

2.1.1.5 Sugar moiety recognition

The nucleotides dUTP and UTP are structurally closely related. As UTP is a metabolite required for RNA synthesis it is essential that the trimeric dUTPases maintain exquisite specificity with regard to the pentose moiety of the substrate. The reason for this specificity is due

to a Tyr residue. This residue is at the turn of the β-hairpin and at van der Waals distance from the sugar moiety; it sterically excludes ribonucleotides. The tyrosine side-chain is anchored through a hydrogen bond to the opposite strand of the β-hairpin and thus contributes to maintaining its architecture. Another highly conserved residue, an isoleucine, is positioned close to the C2' edge of the sugar and together the tyrosine and the isoleucine sandwich the deoxyribose. This rigid packing sterically precludes ribose binding to dUTPase, as the ribose group of UTP would penetrate the van der Waals' radii of the tyrosine side-chain atoms[41] (Figure 2.13).

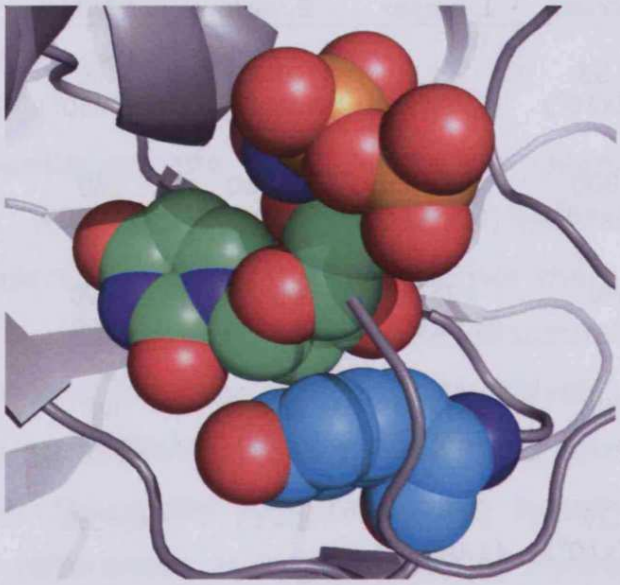


Figure 2.13 Space-filling model of α, β-imido dUTP (green) and the conserved tyrosine (cyan) in the *M. tuberculosis* dUTPase active site.

2.1.1.6 Phosphate chain recognition

While the interactions for the uracil base and the sugar moiety are very conserved in all crystal structures, the interactions of the phosphate chain with the active site vary significantly in each protein ligand complex. Probably, this diversity depends on the type of nucleotide used for crystallization of the complexes, the degree of disorder of the motif V (therefore the number of visible residues in

the "crossing arm"), and the presence of the Mg^{2+} ion in the crystal structure. All these differences in the crystal complexes have generated confusion regarding the interactions and the conformation of the phosphate tail.

The latest research on the crystal structures of *M. tuberculosis* and *E. coli* dUTPases showed that previous structural data might have provided only a partial view of the phosphate binding of the enzyme-substrate complex. It is likely that these two latest structures offer a clearer picture of the dUTPases complexed with Mg^{2+} and non-hydrolyzable substrate analogue, α, β-imido-dUTP, compared with the other X-ray data complexes[33, 38]. The amino analogue of nucleotide phosphates was used as an isosteric replacement, which is structurally extremely similar to the natural substrate dUTP (1.63 Å and 1.68 Å, for P–O and P–N moieties bond lengths respectively, and bond angles 128.7° and 127.2° for P–O–P and P–N–P respectively). The lower electronegativity of the nitrogen atom compared to the oxygen, results in less reactivity of the phosphate ester amino analogues[38].

Crystallization of the *M. tuberculosis* dUTPase-dUTP complex in the absence of metal ions revealed two conformations of the 5'-triphosphate tail (Figure 2.14).

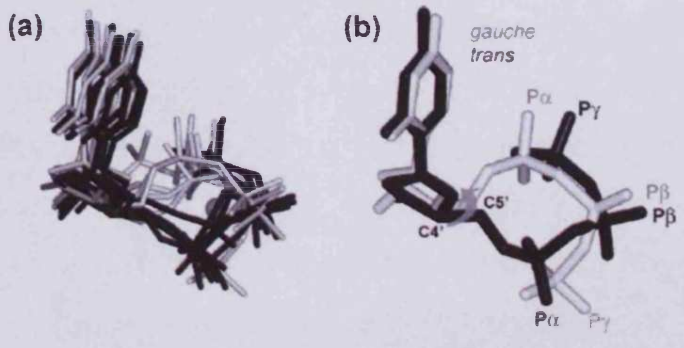


Figure 2.14 a) Superposition of ligands bound to dUTPases from all five sources. The structures may be divided into two groups, differing primarily in the torsion angle about the C4'–C5' bond. One group binds with trans geometry (black). The other group binds with gauche geometry (gray). b) *Trans* (black) and *Gauche* (gray) geometry of dUTP bound to *M. tuberculosis* dUTPase in the absence of metal ions. The distance between αP positions of the two conformations is approximately 4.5Å[38].

The two conformations differ in the position of the α-phosphate of the triphosphate tail. The difference can be roughly described by a 120° rotation about the C4′–C5′ bond of the deoxyribose moiety.

The predominant *Gauche* conformation of the triphosphate tail closely resembles the α, β, γ-tridentate conformation described in the previously for the Mg^{2+}·α, β-imido–dUTP complex and maintains the same set of interactions with the enzyme residues, even though the metal ion is absent[38]. In addition, the complex Mg^{2+}·α, β-imido–dUTP *E. coli* dUTPase was found with a *Gauche* conformation.

The conformational variability (*Gauche* and *Trans*) of the C4′–C5′ bond was reported in crystallographic studies of dUTPases from other sources. The *Gauche* conformation was found in the *E. coli*–dUDP complex and FIV–dUMP complexes, the *Trans* conformation was found in human, FIV, and EIAV dUTPases complexed with dUDP. It seems that a low energy barrier separates the two conformations *Gauche* and *Trans*. It is likely that the metal ion (Mg^{2+}) stabilizes the *Gauche* conformation of the triphosphate binding geometry in *M. tuberculosis* and *E. coli* dUTPases[33, 38]. Therefore, the latest crystal structures suggest that the *Gauche* conformation is the most likely for the natural ligand dUTP in the presence of magnesium ion.

In general, considering the *Gauche* conformation as a correct model, the conserved residues from motif II and IV form the predominant interactions with the triphosphate chain. Motif II, at the beginning of an α-helix, contains a highly conserved serine, which interacts by hydrogen bond with β-phosphate. This serine is generally preceded by a basic residue (arginine or lysine), which is also involved in the interaction with β-phosphate by its side-chain. The γ-phosphate is in contact with another basic residues from motif IV. The α-phosphate forms a hydrogen bond with a conserved water, which is likely to be important in the mechanism of catalysis.

2.1.1.7 The "crossing arm"

In most of the crystal structures the crossing arm (C-terminal) is almost completely disordered, which is due to its high degree of flexibility. Hence it is consistently not present in the structural information.

Motif V, which is part of the C-terminal, is glycine-rich. This feature seems to be common in the nucleotide-binding proteins, such as the dehydrogenases that utilize NAD or oxidoreductase that use FAD[34]. A glycine-rich tail has been also found as a component of protein kinases cAPK and CDK2[34]. This glycine-rich polypeptide segment is very flexible. It seems to be required for the formation of backbone dihedral angles, which are sterically unfavourable for other amino acid residues.

The role of the glycine residues is not completely clear: it might provide a high flexibility to the crossing arm, due to their ability to form dihedral angles disallowed for other residues, it may simply be sterically favorable, or capable to direct the crossing arm backbone amides at the nucleotide phosphates[41].

The glycine-rich peptide in dUTPases has a sequence composed three highly conserved glycine residues, Arg-Gly-X-X-Gly-Phe-Gly (X denote any amino acid; the Phe is replaced by Tyr in the FIV structure and His in the *M. tuberculosis* sturcture). It appears to have some features in common with the glycine-rich loops in other nucleotide binding proteins. In general, the dUTPase nucleotide binding site incorporates features from other nucleotide binding proteins, including the glycine-rich loop and an α-helix whose dipole is directed at the phosphate groups[34].

The two conserved non-glycine residues, an arginine and an aromatic residue appear to be fundamental in the process of catalysis. In an investigation of the catalytic properties of EIAV dUTPase, most of the residues included in motif V were studied

separately to investigate their roles[41]. The substitution of the arginine for a lysine results in total loss of activity. This residue might be involved in stabilizing the developing negative charge on the α-phosphate during cleavage of the scissile bond. A lysine residue is not capable of forming electrostatic interactions as extensively as an arginine and so may be less efficient in charge neutralization. The conserved arginine may also be involved in a bifurcated hydrogen bond with the γ-phosphate, which is ordered by the divalent metal ion (Mg^{2+}), as suggested for the *M. tuberculosis* structure[38].

Substituting the conserved aromatic residue for a non-aromatic counterpart also renders the enzyme completely inactive. The aromatic ring of this residue has been found to stack with the uracil ring in complexes of the human and FIV enzymes. This stacking may be essential for ordering the crossing arm[41].

Crystal structures solved so far suggest that the ordering of the C-terminal loop requires both metal and nucleotide triphosphate. However, there is the exception; in the structure of the human dUTPase (pdb code: 1q5h) the loop that closes the active site is clearly visible. In these two crystal structures, which lack both the metal and the triphosphate, the nucleotides in the complexes are in *Trans* binding conformation instead of *Gauche*[38].

2.1.1.8 Role of the metal ion

The dUTPase activity is dependent on the presence of a divalent metal ion, Mg^{2+}. The role of the metal ion has not yet been completely elucidated, but recently crystal structures of the complexes *E. coli* dUTPase α, β-imido-dUTP and *M. tuberculosis* dUTPase α, β-imido-dUTP were solved with the presence of Mg^{2+}, giving a clue about the function of magnesium and the mechanism of catalysis. The structures reveal the triphosphate moiety chelating the

metal ion through the α, β and γ phosphates; each phosphate contributes one oxygen to coordination with magnesium (Figure 2.15)

This type of coordination is very similar to the coordination geometry of DNA polymerase beta, where the magnesium is coordinate to dideoxycytidine 5'-triphosphate (ddCTP)[45].

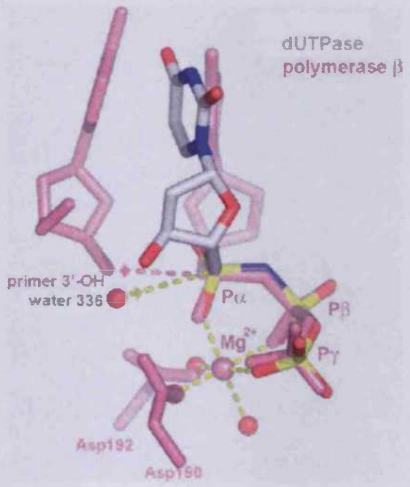


Figure 2.15 Comparison of conformation of Mg^{2+}·α, β imido dUTP bound to dUTPase with the conformation of the incoming ddCTP in DNA polymerase β (1bpy). (The 30-hydroxyl of the primer was modeled based on the position of the dideoxyribose ring.) Both enzymes use a metal ion to coordinate the triphosphate moiety with α,β,γ tridentate geometry. Both enzymes catalyze the in-line nucleophilic attack on the α-phosphorus of a dNTP. Even the nucleophiles in both enzymes have the same position with respect to the α-phosphorus. The difference is that the nucleophile in the polymerase reaction is the 30-hydroxyl of the primer strand rather than a water molecule as it is in dUTPase[38].

The major difference is in the nucleophile agent, which in the polymerase is the 3'-OH of the primer strand, whilst in dUTPase the nucleophile is a molecule of water (the mechanism of catalysis will be discussed later).

The formation of this tridentate coordination with the magnesium was confirmed by kinetic experiments[41] where a racemic mixture of 2'-deoxyuridine-5'-(α-thio)-triphosphate (dUTPαS, see Figure 2.16) was used. The result of the study revealed that in the catalytic reaction, the dUTPase enzyme hydrolyzed only one of the enantiomer analogues. It is likely that, the enantiomer hydrolyzed was the *R* enantiomer. The *S* enantiomer would force the sulfur group to

interact with magnesium ion; this is less favourable than the Mg^{2+}-oxygen interaction.

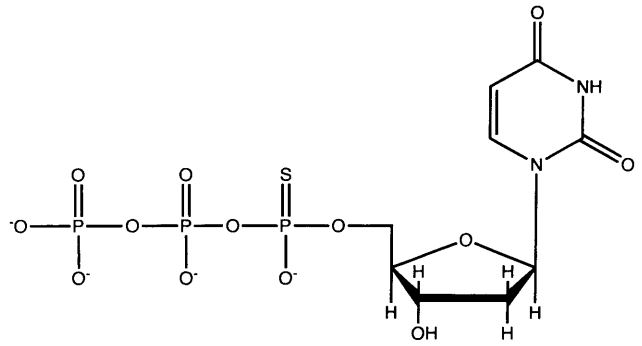


Figure 2.16. 2'-deoxyuridine 5'-(α-thio)-triphosphate.

The metal ion might also be involved with the coordination of three conserved water molecules. These water molecules are hydrogen bonded to two conserved aspartate residues (motif I) (Figure 2.7, see also Figure 2.17), and a conserved arginine residue (motif V)[38].

The X-ray structure of the Cr^{3+}·dUTP *M. tuberculosis* complex showed that the C-terminal remains disordered. This might be because the Cr^{3+}-dUTP decreases the electrostatic attraction between the conserved arginine in motif V and the γ-phosphate due to the extra positive charge carried by the chromium ion.

In general, the ion might have a purely structural role or it might also serve to shield the negative charges on the incoming substrate (dUTP) during the catalysis[38] and so giving to the phosphate a "better leaving group". However, kinetic studies on *E. coli* dUTPase[38] reported that the enzyme can hydrolyze the phosphate chain in absence of divalent metal ions. This implies that the magnesium ion only enhances the catalytic action and, therefore, does not act as an obligatory cofactor (Table 2.4).

Metal ion	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$ x 10^6)
No metal ion	4.5±0.1	0.49±0.1
Mg^{2+}	14.7±2.2	25.1±7.4
Mn^{2+}	14.5±2.2	Not determinated
VO^{2+}	27.3±4.1	2.8±0.9
$VO(acac)_2$	24.2±3.6	2.4±0.7

Table 2.4 Comparison of the influence of divalent metal ions on steady-state kinetic parameters governing hydrolysis of dUTP catalyzed by recombinant *E. coli* dUTPase[46].

2.1.1.9 Catalytic mechanism

In a kinetic studying on the trimeric enzyme, using a water molecule labeled with O^{18}, it was established that the nucleophilic attack is on the α-phosphate. The labeled oxygen was detected bonded to the dUMP and not to the pyrophosphate[41]. It was also demonstrated that a highly conserved aspartate residue, which is part of the β-hairpin (motif III), is involved in the mechanism of hydrolysis, activating the water molecule for the nucleophilic attack (Figure 2.17).

The analysis of the crystal structures of *E. coli* and *M. tuberculosis* dUTPases complexed with α, β-imido-dUTP, suggested a potential mechanism of catalysis. Along the axis of the α-phosphorus-imido nitrogen bond there is a single possible candidate for in-line nucleophile attack: a water molecule coordinated to a conserved aspartate residue. Either carboxyl oxygen of the aspartic acid residue (residue 83 in the *M. tuberculosis* structure) may participate in hydrogen-bonding to this water. Asp-O is also within H-bonding distance of the deoxyribose 3′-OH group[38].

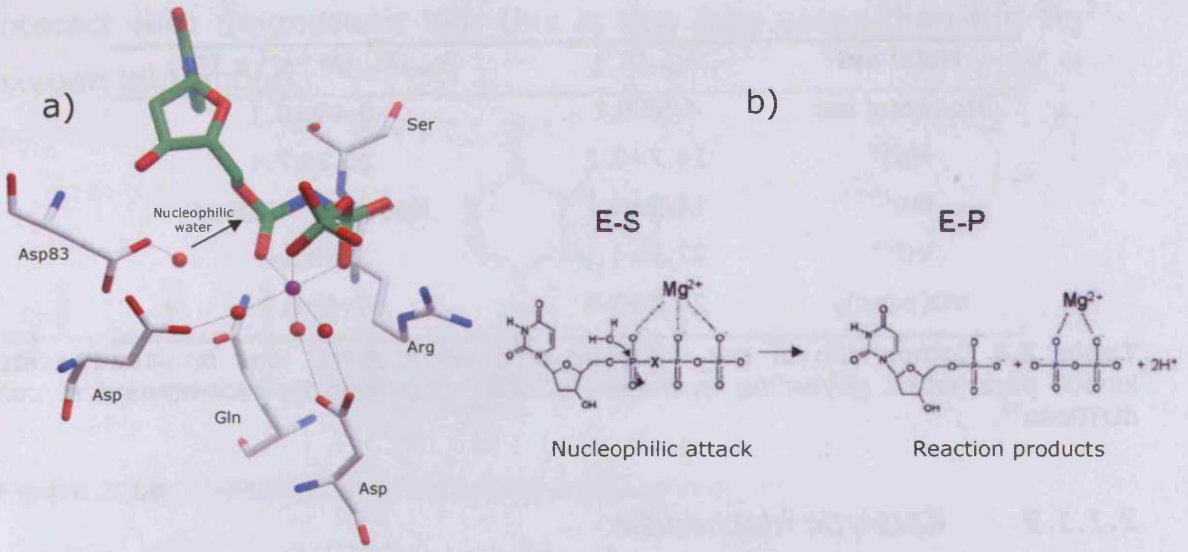


Figure 2.17 a) Simulated annealing showing the ligand α, β-imido-dUTP, magnesium, and coordinated water molecules; b) theoretical reaction schemes of dUTP (*X*=O) or α, β-imido-dUTP (*X*=N) hydrolysis. The catalytic water molecule initiating nucleophile attack and the Mg ion coordinated to the triphosphate chain are shown as determined in the crystal structure.

The mutation of this conserved Asp to an Asn in *E. coli* dUTPase drastically compromised the catalytic efficiency of the enzyme[33]. Hence, this Asp residue is very important for hydrolysis of dUTP. Two roles have been associated with this residue: (1) coordination of deoxyribose 3'-OH and (2) H-bonding to catalytic water. The mutation of Asp to Asn (*i.e.* Asp-O, Asn-N) did not affect the interaction with 3'-OH. However, the attenuated catalytic efficiency of the mutant enzyme is probably due to the lack of proper coordination to the catalytic water molecule in the physiological reaction[33].

The β and γ phosphate groups might facilitate nucleophilic attack at the α-phosphate by sterically forcing the α-phosphorus closer to the nucleophilic water molecule. Crystallographic studies of the complexes of *M. tuberculosis* dUTPase[38], revealed that the distance between of the catalytic water and the α-phosphate dUTP in presence of the magnesium was decreased in the order dUMP>dUDP>α, β-imido–dUTP respectively. This observation might suggest that the inability of dUTPase to catalyze the hydrolysis of dUDP may, at least in part, be due to the increased distance between the α-phosphate

and the nucleophilic water molecule. The γ-phosphate appears to be important for locating the α-phosphate in the optimum position for catalysis[38] (Figure 2.17).

2.1.2 Dimeric dUTPase

In the genomes of several bacteria and eukaryotes, including important pathogens such as *Campylobacter jejuni*, *Staphylococcus aureus, Leishmania major* and *Trypanosoma cruzi*, there is not a gene encoding for a trimeric dUTPase. Instead, these organisms encode an alternative, dimeric form of dUTPase that has an entirely different 3D structure[37, 47] and is an analogue, not a homologue, of the trimeric dUTPase[26].

Structural characterization of two members of this family, the dUTPases from *T. cruzi* and the gastric pathogenic bacterium *C. jejuni*, revealed a homodimeric enzyme with a novel all-α fold, consisting of 11 α-helices with long connecting loops[37] (Figure 2.18). Their structures do not have any similarity with the trimeric enzymes. Additionally, the protozoan dimeric dUTPases display slightly broadened substrate specificity; dUDP, which is a potent inhibitor of trimeric dUTPases, is a substrate for the dimeric enzymes. All the other physiological nucleotides except dUMP have negligible binding affinities and do not compete significantly with the physiological substrate.

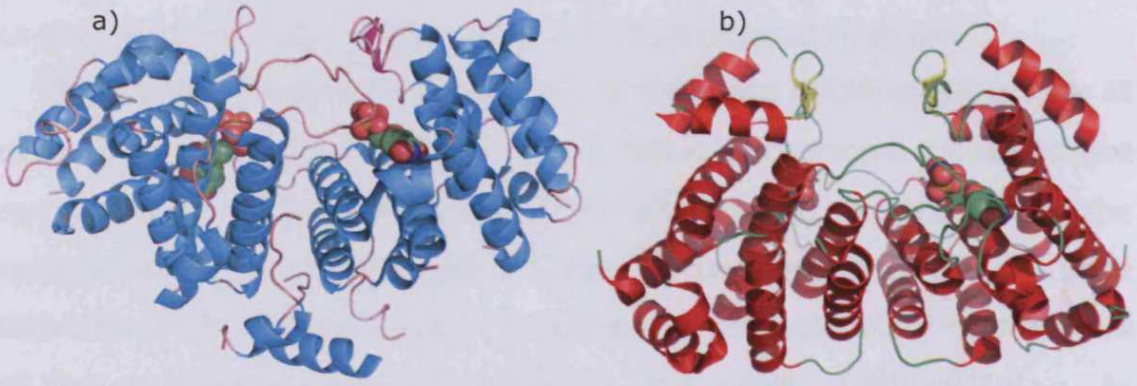


Figure 2.18 a) Ribbon representation of *T. cruzi* dUTPase dimmer in complex with the substrate analogue dUDP. b) Ribbon representation of *C. jejuni* dUTPase dimer in complex with the substrate analogue dUpNHp.

The first protein of this family to be structurally characterized was the *T. cruzi* dUTPase. This dimeric dUTPase showed an elaborate array of four residues (Gln 14, Asn 18, Asn 22 and His 58) that co-ordinate O2, N3, and O4 atoms of uracil, which ensures the strict substrate specificity of this enzyme[37]. A steric hindrance due to tight packing of buried substrate, gives poor binding affinity of the extra 5-methyl group of thymine and 2' hydroxyl group of ribose.

As with the trimeric forms of dUTPase, the dimeric requires a metal ion as cofactor. In the absence of Mg^{2+} ions the *L. major* dUTPase is completely inactive. No enzymatic activity was observed when the Mg^{2+} was substituted with Ca^{2+} or Cu^{2+}. However in the presence of Co^{2+} and Mn^{2+} the dimeric dUTPase retains or increases its activity[44].

The enzyme dUTPase has been confirmed to be essential for the viability of *L. major* and *T. brucei*. In the former antisense RNA oligonucleotide[44] and in the latter RNAi[48], caused down regulation of the protein expression and affected the normal growth of the parasites in culture cell. These experiments demonstrated that the enzyme is indispensable for the viability of *T. brucei* in both stages of the parasites (the blood form in the mammalian host and the procyclic form in the insect)[48].

Intracellular location analyses studies revealed[48] that the dimeric enzyme is localized only in the nucleus of the *T. brucei* cell, with no presence in the mitochondria. These findings suggest further differences with the corresponding human enzyme, which has two isoforms (mitochondrial and nuclear).

2.1.2.1 *Trypanosoma cruzi dUTPase*

The *T. cruzi* crystal structure has been solved recently, in both native form and as a complex with the substrate dUDP[47]. The amino acid sequences of *T. cruzi* and *L. major* dUTPases share 50% identity and 75% similarity[49], which suggest that the two enzymes may have similar structures. The structure is formed of mainly α helices; each monomer is arranged in two distinct domains termed mobile and rigid, with an angle of approximately 60° to each other. The rigid domain of one monomer interacts with the corresponding domain of the other monomer, forming the rigid core of the enzyme (Figure 2.19).

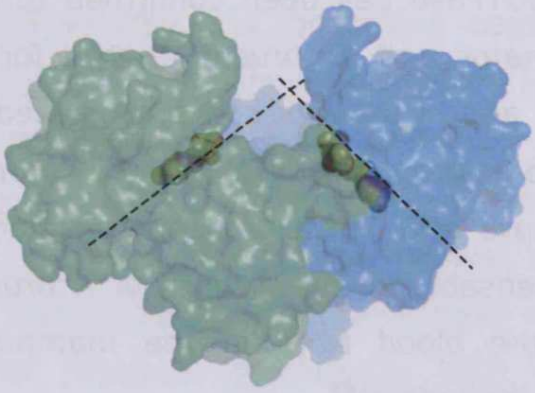


Figure 2.19 Surface representation *T. cruzi* dUTPase dimer in complex with the substrate analogue dUDP. The monomer is arranged in two distinct domains (colored in green and cyan) termed mobile and rigid, with an angle of approximately 60° to each other.

The main feature of this enzyme is the difference between the two forms. The enzyme dramatically changes its geometry when the substrate is bound in the active site. The active sites lie in the groove formed at the interface of the mobile and rigid domains within individual subunits. In the native form the overall shape is an "open" structure where the active sites are exposed to the environment. When the nucleotides bind to the enzyme, the mobile domains undergo structural changes "closing" on themselves engulfing the

nucleotides in the active sites (Figure 2.20).

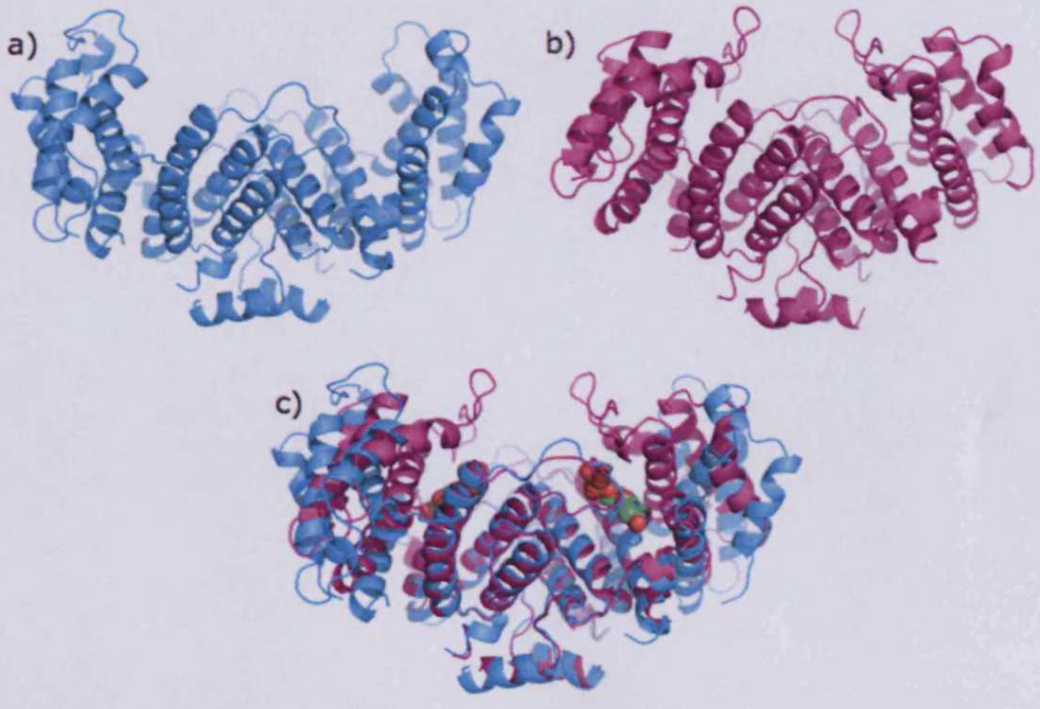


Figure 2.20 a) *T. cruzi* dUTPase with both active sites empty in "open" form. b) *T. cruzi* dUTPase with both active sites empty in "close" form. c) *T. cruzi* dUTPase dimer native open form, superimposed on the its closed ligand-bound form.

The interactions of the substrate in the active sites are showed in Figure 2.21, where the uracil base is involved in a network of hydrogen bonding with the side chain atoms of the polypeptide (Asn 26, Gln 22 and Trp 61). The uracil is kept in the site by interaction with the side chain of Met 29. The sugar moiety is recognized by two hydrogen bonds with the Asn 201 and two interactions with the aromatic rings of Phe 84 and His 83. The phosphate groups are surrounded with several polar residues, mainly glutamates, which probably also bind the metal ion (Mg^{2+}).

Recently, a structure-base approach to design a specific inhibitor for this enzyme was made[50]; unfortunately the search for lead inhibitors of the *T. cruzi* dUTPase remains unsuccessful.

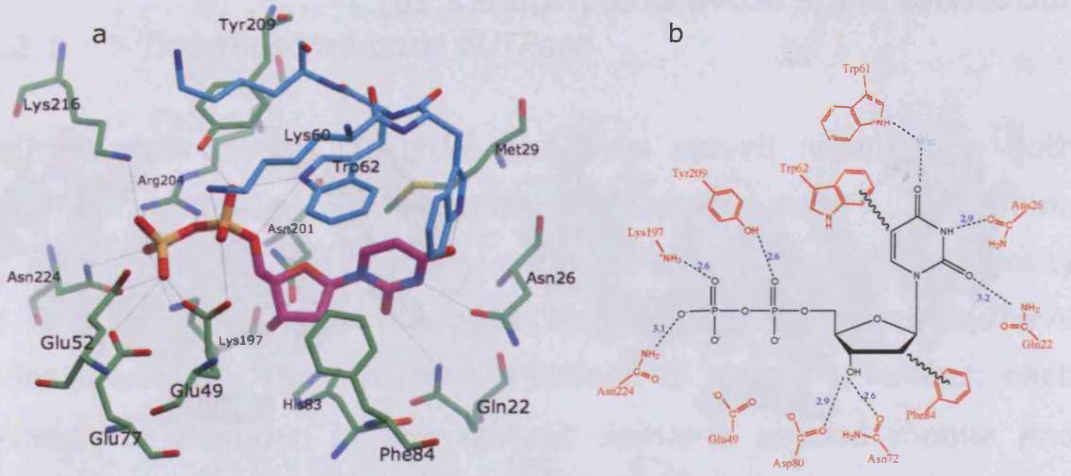


Figure 2.21 a) An overall view of the active site of *T. cruzi* dUTPase with bound dUDP illustrating the structural basis of substrate specificity and the hydrogen-bonding pattern between the enzyme and the substrate, showing the different subunits colored in green and cyan; b) Schematic of dUTPase–substrate interactions showing the structural basis for the exquisite specificity for dUTP that requires all three subunits. Hydrogen bonds (<3.5 Å) are shown (dashed lines) with the donor–acceptor atom distance beneath the protein atom label. Hydrophobic interactions are also shown (wavy lines).

2.1.3 Monomeric dUTPase

Monomeric dUTPases are encoded by mammalian and avian herpesviruses. These enzymes have a limited homology to the trimeric dUTPases: the five motifs forming the active site are present, but they are re-ordered and spread out over a single polypeptide that is twice as long as the sequence of the subunit in trimeric enzymes. It had been proposed that these enzymes originated from a duplication of a gene coding for a trimeric enzymes[40].

The herpesviruses are divided into four families: three of these, α, β, and γ herpesviruses. The α and β families infect mammals and birds, whereas the last family infects only cold-blooded animals. The division into α, β, and γ-herpesviruses is based on their biological properties as well as on gene location and homology. All the three families herpesviruses have a gene coding for a monomeric dUTPase. These genes have been shown to be functional for HSV1, an α-

herpesvirus, and for *Epstein-Barr* virus (EBV), a γ-herpesvirus[40]. Recently EBV has been crystallized in complex with dUMP -Mg^{2+}, and α, β-imido-dUTP and Mg^{2+}.[40]

The EBV dUTPase is almost entirely a β strand structure, with 21 strands and only 3 short α-helices. The structure can be subdivided into three domains, each occupying a position corresponding to a subunit of the trimeric form of the enzyme (Figure 2.22).

The EBV dUTPase has only one active site located at the interface between domain I and domain II (Figures 2.22A). It contains motifs 1–4 of the classical set of five conserved motifs described for the trimeric dUTPases, whereas motif 5 is disordered. Motifs 1, 2, 4 and 5 are contributed by domain II, and motif 3 is contributed by domain I (Figure 2.22B). The interactions of the base and deoxyribose groups are similar to the one seen in the trimeric enzymes. It is remarkable how well the interactions in the active site are conserved despite the low overall sequence homology.

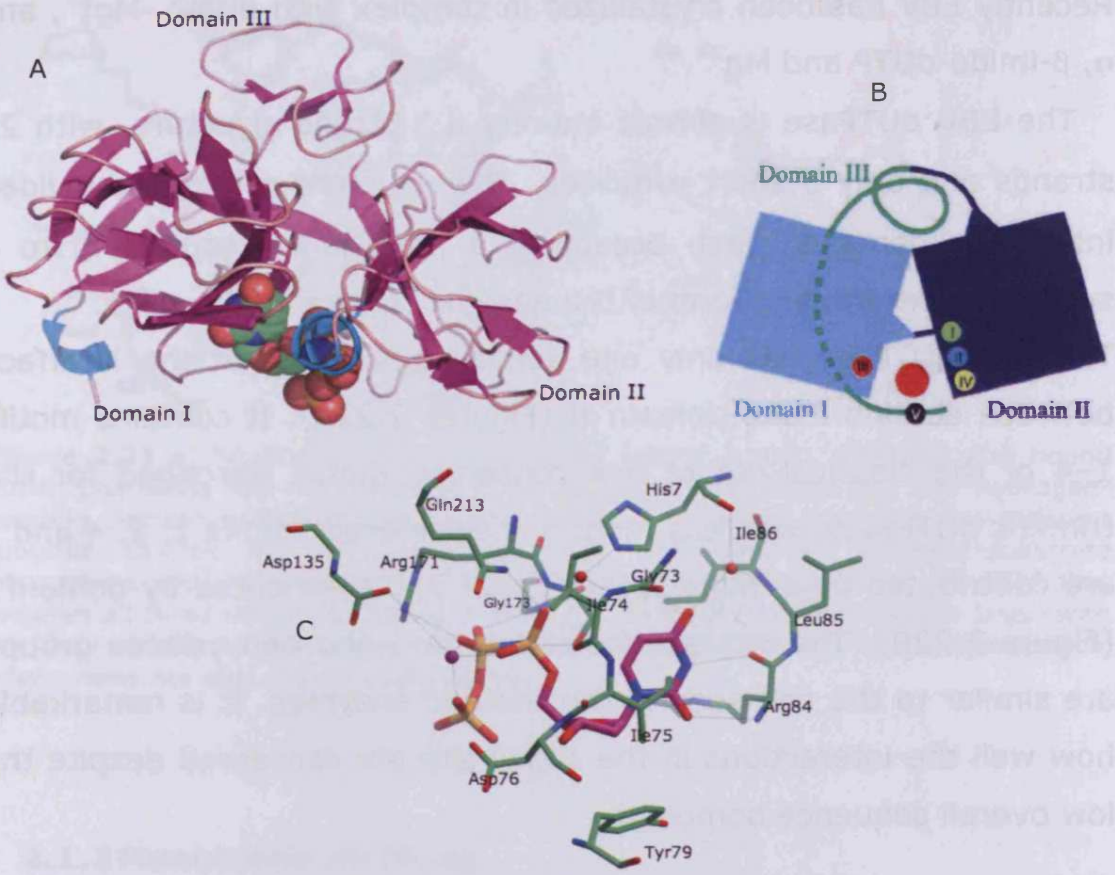


Figure 2.22 A) Ribbon representation of the EBV dUTPase. B) Schematic organization of the monomeric dUTPases and localization of the conserved sequence motifs I–V around the active sites represented by red dots. C) An overall view of the active site of EBV dUTPase with bound α, β-imido-UTP illustrating the structural basis of substrate specificity and the hydrogen-bonding pattern between the enzyme and the substrate.

After the protease[51], dUTPase represents the second structure with an enzymatic function that has been determined for this virus. It might be a rare example of an evolution from a multimeric to a monomeric protein. It shows a surprising degree of local conservation in the active site that is virtually identical to that in the trimeric dUTPase family despite a divergent evolution from the parent enzyme reflected by a globally low degree of homology. Nevertheless, the structure reveals differences in some details of the active site that may be exploited to design specific inhibitors.

2.1.4 Plasmodium falciparum dUTPase

Although *Plasmodium falciparum* is a protozoa family such as the *L. major* or *T. cruzi,* it has been found that the *P. falciparum* dUTPase is not dimeric enzyme, but a trimeric enzyme; hence it is structurally related to the human dUTPase. This finding may suggest a different phylogenetic evolution for this parasite compared to the trypanosomatid protozoa[39].

This enzyme has been recently crystallized[39]; in each of the three subunits, the polypeptide fold is the same, consisting of thirteen segments of β strand and two α helices (Figure 2.23), which is a unique feature, since all the other trimeric enzymes crystallized so far, have only one distinctive α-helix. The strands are arranged in three discrete β sheets that pack against one another enclose, forming a hydrophobic core, representing a variation of the jelly-roll.

Figure 2.23 Ribbon representation of the *P. falciparum* dUTPase trimer in complex with the ligand WSP869. The β-sheets and α-helices are colored in magenta and cyan respectively.

Over all, the complex shows in total a homotrimer consisting of 429 residues, three inhibitor molecules (WSP869 see Figure 2.24),

and 121 water molecules.

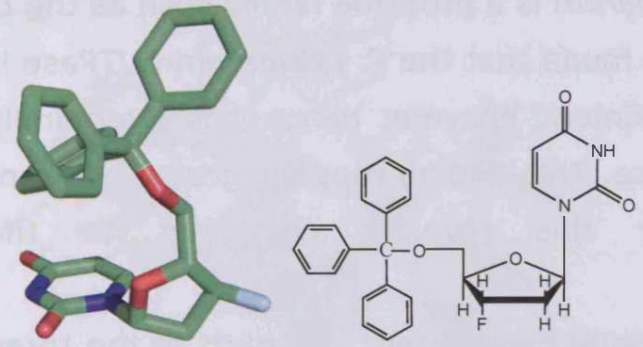


Figure 2.24 ligand WSP869.

There are a number of residues not resolved in the structure; the correspond to disordered regions of the *P. falciparum* dUTPase trimer or low complexity region (LCR), such as the C-terminal region (residues 156–173) and the LCR of 15 amino acids (residues 64–79). The latter LCR is part of an insertion of 25 amino acids (Figure 2.7) present only in *P. falciparum* dUTPase. Many *Plasmodium* proteins have insertions, which usually contain distinctive homopeptide repeats of hydrophilic amino acid residues[39]. These insertions[52] seem to be a distinctive characteristic of *P. falciparum* proteins. The role (if any) of these insertions is unknown. The functions of these insertions have been suggested to be: 1) they might affect the overall shape/charge of the enzyme, changing the mechanism of the substrate recognition compared to other trimeric enzymes. 2) They might act as a "smokescreen" against the host immunogenic response[52]. 3) These insertions may only be the by-product of the antigenic variability concomitant to repetitive elements, perhaps a consequence of the oxidative stress generated by the *Plasmodium* metabolism[52].

Studies with isoforms of human dUTPase have shown a nuclear and mitochondrial distribution in agreement with the major role of the enzyme in DNA replication. In contrast, immuno-fluorescence studies and intracellular localization analyses suggested that the *P.*

falciparum dUTPase[39] does not appear to be associated with organelles such as the nucleus or the apicoplast[53].

Despite the fact that *Plasmodium* and human dUTPases fall in the same structural class and both are eukaryotic, the *Plasmodium* enzyme has some features that resemble the prokaryotic dUTPases. Firstly, the 3-fold inner channel of the trimer has a highly polar character in human dUTPases, whereas in the prokaryotic enzymes and in *Plasmodium* it is constructed of non-polar residues. Secondly, the *Plasmodium* has an insertion of two residues that forms the trimer interface channel, which seems typical of prokaryotic dUTPases. Thirdly, the conformation of the crossing arm of the *P. falciparum* dUTPase is surprisingly unusual compared with other trimeric enzymes including the human. In the *Plasmodium*, it executes a steep turn interacting only with its core subunit, while in the human the crossing arm is in contact with neighboring subunit. Nevertheless, despite these differences and the low sequence identity 28.4%, the active sites of these two enzymes, as for all the trimeric dUTPases, are remarkably similar. The small differences of the binding pockets have presented a difficult challenge to overcome. However, the ligand co-crystallized with the *P. falciparum* dUTPase was found to be a potent and highly selective inhibitor for the *Plasmodium* enzyme[39].

Each of the three inhibitors makes an extensive set of interactions with conserved motifs in the protein (Figure 2.25 F). The uracil moiety forms hydrogen bonds to the side chain of the residue Asn 103 and the main chain of the residue Ile 117, and is further stabilized by hydrophobic contacts with the side chain of the highly conserved residue Ile 108. Additionally, the main-chain of Ala 119 is in hydrogen bond contact with the uracil O4 via a conserved water molecule placed in the bottom of the active site.

The sugar moiety is maintained in position by a parallel ring stacking interaction with strongly conserved side chain of Tyr 112.

However, the fluorine atom attached to C3′ of the sugar ring makes no direct interactions with protein. Finally, the aromatic rings of the trityl group form van der Waals interactions with the side chain of the residues Phe 46, Lys 96, and Ile 117 (Figure 2.25).

Comparison of the mode binding of inhibitor WSP869 with dUDP in human dUTPase showed that the uracil base binds very similarly in both the enzymes, while the sugar moiety of the ligand WSP869 is displaced to allow the trityl group to interact with a different, more hydrophobic part of the binding pocket. In fact, the phosphate tail of dUDP, as in all substrate analogs, is placed in a completely different and hydrophilic part of the binding site. This comparison revealed that the selectivity of the inhibitor WSP869 for *P. falciparum* over human dUTPase is probably due to the way the trityl group interacts with the side chains of residues Phe 46 and Ile 117, which are equivalent to the smaller and less hydrophobic residues Val 42 and Gly 87 in the human enzyme (Figure 2.25). In particular, the side chain of residue Phe 46, which is unique for the *P. falciparum* dUTPase, forms π–π interaction with trityl group. The replacement of this aromatic residue with Val 42 in human dUTPase probably accounts for the loss of activity for the human enzyme, probably due to the lack of the π–π interaction with the Phe residue. In addition, the Lys 96 in the *P. falciparum* enzyme might also interact with the triphenyl group forming cation-π interaction[54] (Figure 2.26). These diversities of the active sites could be exploited for selective inhibition of the *Plasmodium* enzyme.

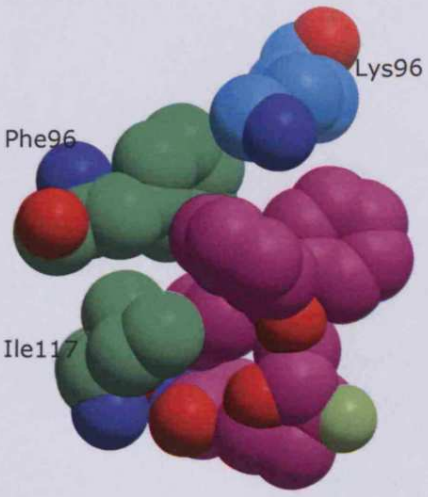
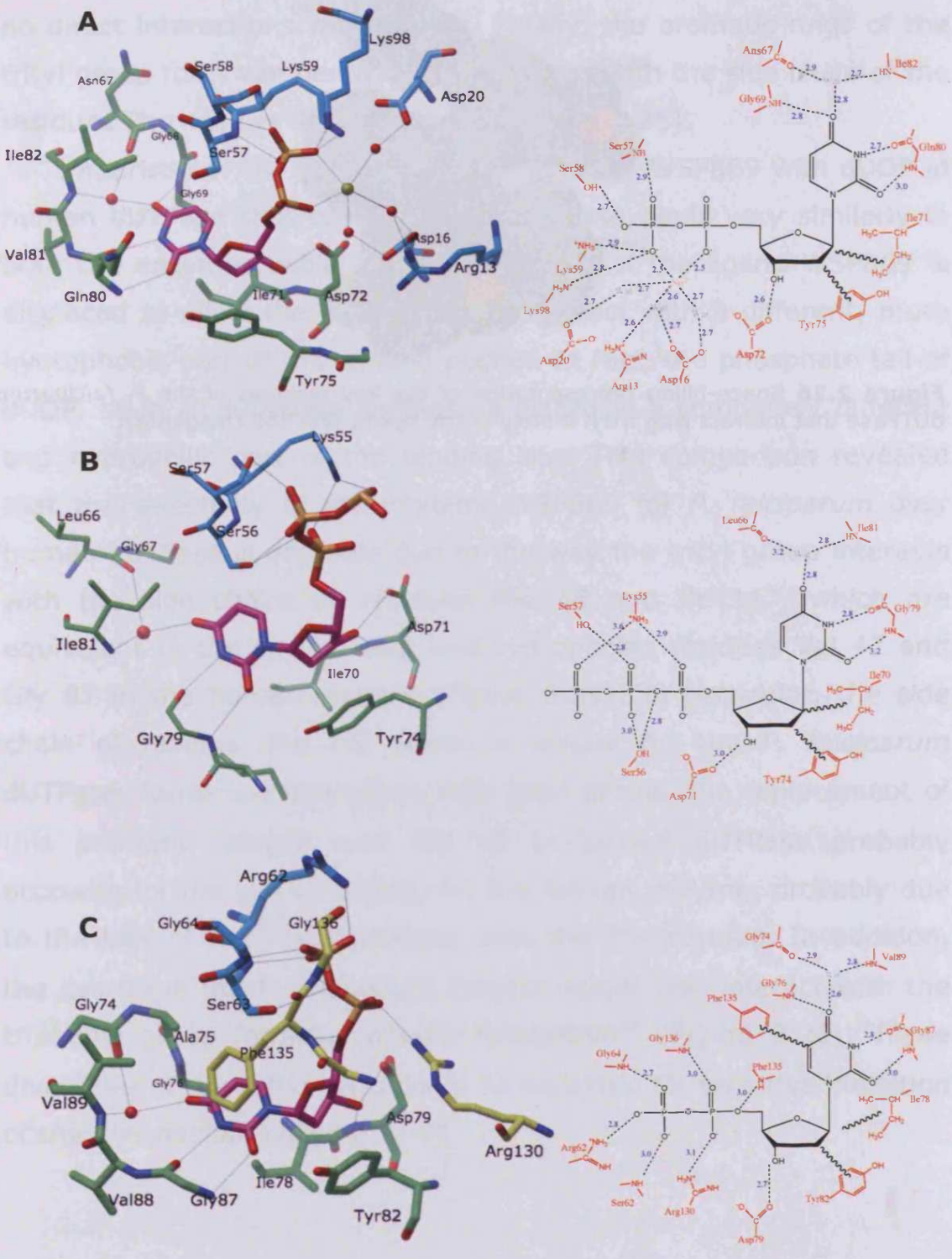


Figure 2.26 Space-filling representation of the key residues of the *P. falciparum* dUTPase that interact with trityl moiety of the ligand WSP869 (magenta).


A
Lys98
Asn67
Ser58
Lys59
Asp20
Gly66
Ile82
Ser57
Gly69
Asp16
Val81
Arg13
Gln80
Ile71
Asp72
Tyr75
B
Lys55
Ser57
Leu66
Ser56
Gly67
Asp71
Ile81
Ile70
Gly79
Tyr74
C
Arg62
Gly64
Gly136
Gly74
Ser63
Ala75
Phe135
Gly76
Val89
Asp79
Arg130
Ile78
Val88
Gly87
Tyr82

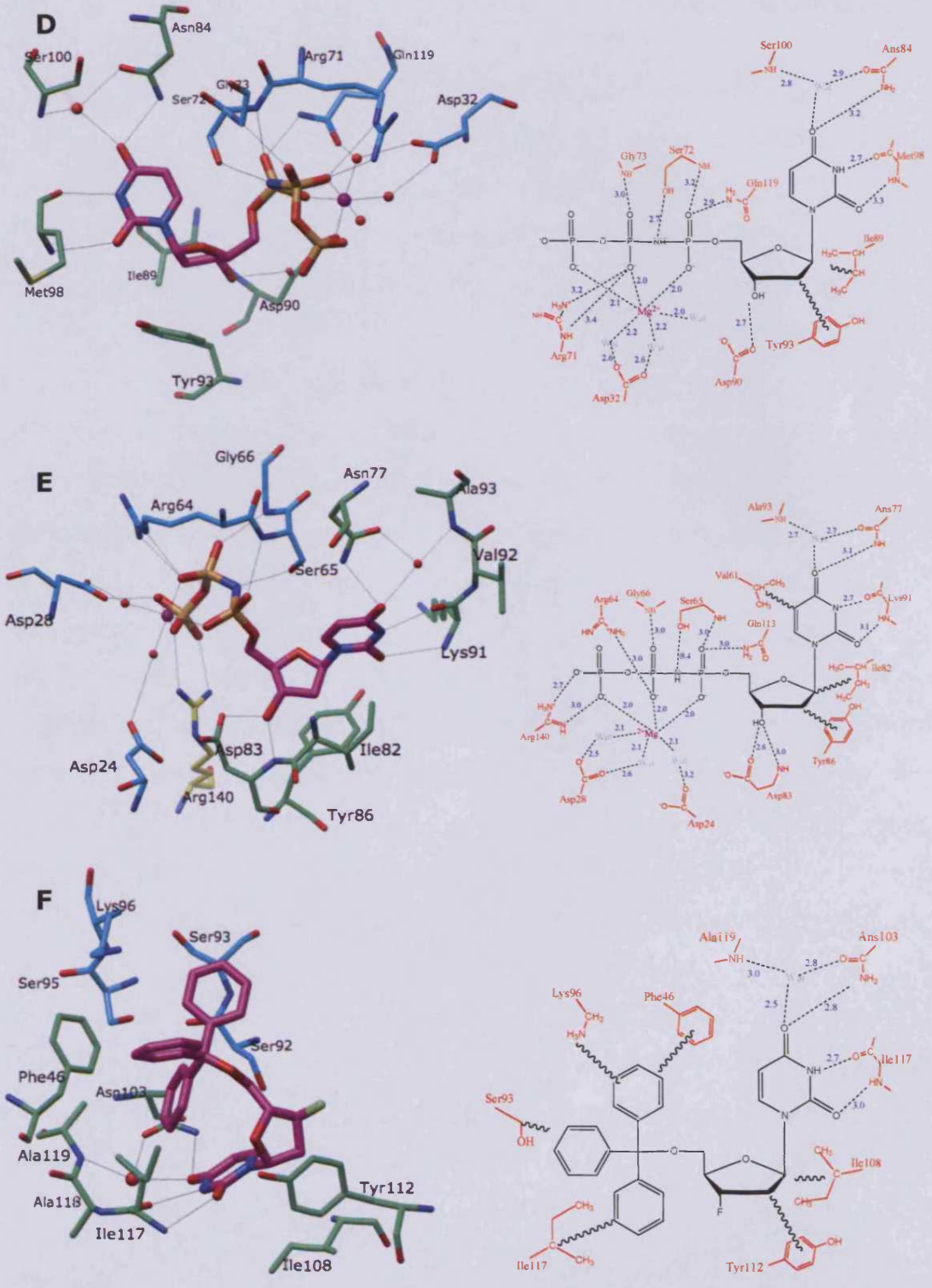


Figure 2.25 An overall view of the active site of the six dUTPase enzymes, showing the different subunits colored in green cyan and yellow: A) EIAV dUTPase bound dUDP; B) FIV dUTPase bound dUTP; C) human dUTPase bound dUDP; D) *E. coli* bound α, β-imido dUTP; E) *M. tuberculosis* bound α,β-imido dUTP; F) *P. falciparum* bound WSP869.

2.2 Conclusion

The six trimeric dUTPase structures described in this chapter, display similar overall β-sheet folds and features. The homotrimeric layout is not unique in the protein world. However, the prerequisite observed in the trimeric dUTPases for all three subunits to supply essential amino acid residues to each active site has not been observed before[34, 41].

dUTPases are required to sustain impeccable specificity in choice of substrate for polymerization. The homotrimeric variety does so by using two motifs commonly found in proteins: the β-hairpin loop and the glycine-rich motif, where the latter is often observed in phosphate binding proteins. The former is, however, utilized in a novel way by the trimeric dUTPases, *i.e.* to promote interactions between the polypeptide and the nucleotide base and sugar moieties. By mimicking base pairing the enzyme efficiently excludes binding of non-uracil containing nucleotides, while the sugar moiety is selected based on size and shape. The glycine-rich motif in the flexible arm of the trimeric dUTPases contributes to specificity by requiring a tri-phosphate containing substrate before closing down on the active site and promoting catalysis[55].

Due to its essential viability, the enzyme is regarded as a potential drug target for treatment of infectious diseases and cancer. Molecular and structural studies have provided some insight into the structure and function of the enzyme originating from different sources[41].

3 Chapter: dUTPase inhibitors

dUTPase is a very selective enzyme which can discriminate between the different natural nucleobases and the sugar moiety. In fact none of the other nucleotides present in the nucleotide pool are substrates for the enzyme. This should not be surprising as non-productive hydrolysis of nucleotide triphosphates would be a futile cycle for cellular metabolism.

Few inhibitors of dUTPase have been reported[27]; examples are given in the Table 3.1.

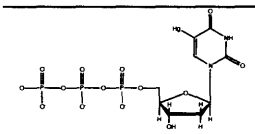
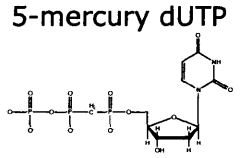
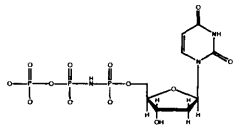
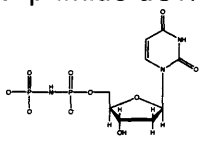
Compound	*L. major*	*T. cruzi*	*E. coli*	EIAV	HSV-1	HSV-2
 5-mercury dUTP	–	–	–	–	*Ki*=27[56]	*Ki*=21[56]
 M-dUTP	–	–	*Ki*=2.5[57]	–	–	–
 BM-dUTP	–	–	*Ki*=0.3[57]	–	–	–
 α-β-imido dUTP	*Ki*=0.89[44]	–	–	–	–	–
 α-β-imido dUDP	–	*Ki*=0.24[58]	–	–	–	–

Table 3.1 Inhibition constants (*Ki*) for some known inhibitors of dUTPases (all values are μM)

These inhibitors are nucleotide triphosphate analogues in which the bond between the α and β phosphate is non-hydrolyzable (imino or methylene) or they contain a heavy metal like mercury in position 5

of the uracil ring. These compounds are not suitable drug candidates as they have multiple charges which would prevent the penetration across membranes; they are enzymatically and chemically unstable and toxic. However, they have been useful for the characterization of the enzyme. For example, the nucleotide analogue α–β-imido dUTP can inhibit the *L. major* dUTPase with a *Ki* of 0.89 μM, indicating that the enzyme can bind nucleotides analogues, which lack a hydrolysable bond between the α and β phosphate groups. The same α–β-imido dUTP is also a good inhibitor of the trimeric dUTPase from *E. coli*[33]. *T. cruzi* dUTPase is inhibited by the diphosphate analogue α–β-imino dUPD, and also this compound has shown the capacity to inhibit the trypanosomal enzyme with a *Ki* of 0.24 μM.

Recently, in our laboratories two series of cyclic and acyclic (Figure 3.1) nucleotide derivatives have been prepared. Some of this has been published[59, 60]. The cyclic series resembles the 2'-deoxyuridine compound, which was set out to initially focus on the positions 3' and 5' of the 2'-deoxyuridine. In particularly, this focused library probed these positions with a range of lipophilic and hydrophilic substituents of varying steric bulk. The acyclic series of compounds was built in an attempt to optimize the distance between the base (uracil) and a trityl moiety.

Some of the compounds of this focus library showed a promising inhibition activity for the dUTpase enzyme, and in particular the *P. falciparum* dUTPase (appendix 1). These series of inhibitors were developed in the Welsh School of Pharmacy in collaboration with York University, Medivir AB Sweden, *Instituto de Parasitologıa y Biomedicina,* Granada, and *Swiss Tropical Institute*, Basel.

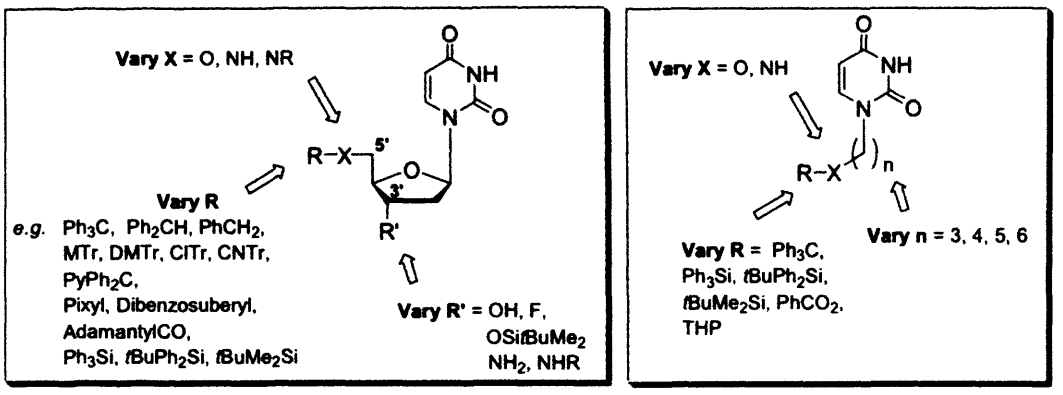
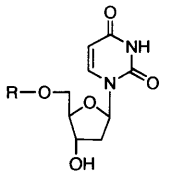


Figure 3.1. Generic target compounds.

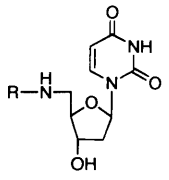
3.1 Cyclic series

It is possible divide the cyclic series into six groups of compounds that were made to investigate structure-activity relationships (SARs) (Figure 3.2, see also Appendix 1 for the complete series of cyclic derivatives).



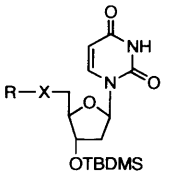
WSP Group 1

487 R = $(PhO)_2P(O)$
659 R = $(EtO)_2P(O)$
784 R = $(EtO)_2P(O)CH_2C(O)$
662 R = $CH_3C(O)$
669 R = $CH_3(CH_2)_3C(O)$
660 R = PhC(O)
666 R = Ts
790 R = H_2NSO_2
789 R = $H_2NSO_2N(-)C(O)$
483 R = Ph_3C
951 R = TIPS
950 R = TBDPS
949 R = TPS
1059 R = (*para*-$CH_3O)Ph_3C$
1243 R = (*para*-$CH_3O)_2Ph_3C$
1244 R = (*ortho*-Cl)Ph_3C
1245 R = Dibenzosuberil
1247 R = Pixyl
1288 R = (*para*-NC)Ph_3C
1329 R = Trybenzylsilyoxy



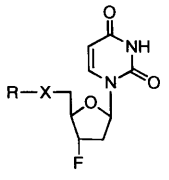
WSP Group 2

665 R = $(PhO)_2P(O)$
785 R = $(EtO)_2P(O)CH_2C(O)$
958 R = AdamantylC(O)
663 R = $CH_3C(O)$
668 R = $CH_3(CH_2)_3C(O)$
661 R = PhC(O)
667 R = Ts
680 R = CH_3SO_2
952 R = Ph_3CH_2
871 R = Ph_3C
804 R = H
1068 R = (*para*-$CH_3O)Ph_3C$
1149 R = 4-$PyPh_2C$
1289 R = (*para*-NC)Ph_3C
1290 R = (*ortho*-Cl)Ph_3C
1291 R = Pixyl
1292 R = Dibenzosuberil



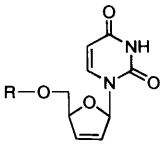
WSP Group 3

486 X= O, R = $(PhO)_2P(O)$
659 X= O, R = $(EtO)_2P(O)$
784 X= O, R = $(EtO)_2P(O)CH_2C(O)$
789 X= O, R = $H_2NSO_2NHC(O)$
790 X= O, R = H_2NSO_2
485 X= O, R = H
484 X= O, R = Ph_3C
953 X= O, R = Ph_3C
704 X= NH, R = $(PhO)_2P(O)$
933 X= O, R = TBDMS



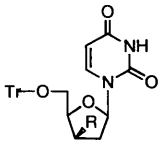
WSP Group 4

868 X= O, R = TBDMS
948 X= O, R = TBDPS
947 X= O, R = TPS
1191 X= O, R = H
869 X= O, R = Ph_3C
946 X= NH, R = Ph_3C
1345 X= O, R = Trybenzylsilyoxy



WSP Group 5

831 R = $(PhO)_2P(O)$
875 R = $H_2NSO_2NHC(O)$
866 R = $CH_3(CH_2)_{14}C(O)$
865 R = AdamantylC(O)
830 R = H_2NSO_2
828 R = TBDMS
867 R = TBDPS
1073 R = TPS
829 R = H
864 R = Ph_3C



WSP Group 6

1361 R = OH
1362 R = OAc
1362 R = N_3

Figure 3.2 Compound structures

In Group 1, a variety of substituents were placed on the 5'-

position. Compounds WSP487 and WSP659 represent phosphate esters of dUMP with bulky or small substituents. Compounds WSP662 and WSP660 are simple esters of varying lipophilicity. Compounds WSP666 and WSP790 are sulfonates, which should be isosteres for the phosphate group. WSP483 is a simple ether, whereas WSP951 and WSP949 bear a silyl ether moiety. WSP784 and WSP789 are isosteres of the diphosphate dUDP. A similar group of compounds was prepared in which the 5'-oxygen is replaced with an amine. WSP1059 and WSP1243 contain one *para*-methoxy and two *para*-methoxy groups on the trityl respectively, in order to probe steric and electronic effects at the active site. WSP1244 has chlorine atom at the *ortho* position of a phenyl, which may create additional steric hindrance. WSP1288 contains a 4-cyanotrityl variation might provide the molecule with an additional hydrogen bond acceptor and reduced electron density in the aromatic ring. Compounds WSP1245 and WSP1247 contain a variation of the trityl group, pixyl or 9-phenylxanthen-9-yl moiety respectively. These have a highly rigid and planar backbone compared with the trityl, not locked in a propeller shape (as in the trityl). Finally, compound WSP1329 has in 5'-position a tribenzylsilyl group, which represents a "mimic" of a triply "extended" trityl[59].

In the case of the acyl derivatives, the amide analogues are likely to be biologically more stable than are their ester counterparts, which may be subject to hydrolysis by cellular esterases. The conversion from ester (O) to amide (NH) also changes a hydrogen-bond acceptor to a hydrogen-bond donor and may be useful to provide more information on the interactions required at the enzyme active site. Group 3, 4, 5 and 6 were designed to probe the 3'-position and comprise a selection of compounds in which the 5'-position is varied while the 3'-hydroxyl is either blocked as a *tert*-butyl dimethylsilyl ether (Group 3), replaced by a fluorine isostere (Group 4), or removed altogether (Group 5), and finally changed its

stereochemistry (Group 6). These modifications aimed to investigate the effects of increasing lipophilicity and steric hindrance, and of altering possibilities of hydrogen bonding at the 3'-position. In addition, in the case of the didehydrodideoxy analogues (Group 5), the presence of the double bond at the 2'-3' position provides rigidity for the system[59].

In general, it seemed that the enzyme inhibition (*Ki* < 10 μM) appears to be related to the presence, at the 5'-position, of a bulky lipophilic substituent containing two or three phenyl groups as well as an amine or ether linkage (Table 3.2). Thus, preferred moieties were found to be either a trityl amine or ether or a triphenylsilyl (TPS) or *tert*-butyl diphenylsilyl (TBDPS) ether. SAR analysis showed that the presence of at least two aromatic ring substituents at the 5'-position is required for enzyme inhibition. Changing the tritylamino moiety WSP871 to a benzylamino group WSP952 results in a drastic loss of activity (more than 500-fold). The 5'-adamantylamido derivative WSP958 is inactive against *P. falciparum* dUTPase (*Ki* 426 μM), indicating that it is unlikely that steric bulk alone is sufficient for activity. However, compound WSP958 bears an amide rather than an amino moiety at the 5'-position, which could also be significant. The data from the silyl group (compounds WSP951-949, WSP868-947, and WSP828-1073) support the hypothesis that a bulky lipophilic group bearing an aromatic moiety is required at the 5'-position for the inhibition of *P. falciparum* dUTPase. Similar activity is observed for the 5'-triphenylsilyloxy (TPSO) and 5'-*tert*-butyldiphenylsilyloxy (TBDPSO) derivatives (WSP949 and WSP950, *Ki* 2.8 and 4.2 μM, respectively; WSP1073 and WSP867, *Ki* 1.3 and 1.2 μM, respectively), whereas replacing more than one phenyl by alkyl groups (even bulky isopropyls) gives a negative effect. This is evidenced by the poor activity of the TBDMS derivative WSP828 (*Ki*

10 μM) and the triisopropylsilyl (TIPS) derivative WSP951 (*Ki* 227 μM) compared to that of their TPS and TBDPS analogues (WSP1073 *Ki* 1.3 μM, WSP867 *Ki* 1.2 μM, WSP949 *Ki* 2.8 μM and WSP950 *Ki* 4.2 μM, respectively)[59].

WSP	R	X	R'	*P. f.*[a]	Human	SI *P. f.*[b]
483	Ph3C	O	OH	1.8	18	10
484	Ph3C	O	OTBDMS	515	>1mM	>1.9
485	H	O	OTBDMS	648	119	0.2
486	(PhO)2P(O)	O	OTBDMS	597	>1mM	>1.7
487	(PhO)2P(O)	O	OH	238	135	0.6
659	(EtO)2P(O)	O	OH	316	>1mM	>3
659	(EtO)2P(O)	O	OTBDMS	340	>1mM	>3
660	PhC(O)	O	OH	268	>1mM	>3
661	PhC(O)	NH	OH	178	469	2.6
662	CH3C(O)	O	OH	294	>1mM	>3
663	CH3C(O)	NH	OH	>1mM	>1mM	nd
665	(PhO)2P(O)	NH	OH	26	>1mM	>38
666	Ts	O	OH	142	>1mM	>7
667	Ts	NH	OH	>1mM	>1mM	nd
668	CH3(CH2)3C(O)	NH	OH	189	>1mM	>5
669	CH3(CH2)3C(O)	O	OH	228	>1mM	>4
680	CH3SO2	NH	OH	178	>1mM	>5
704	(PhO)2P(O)	NH	OTBDMS	67	232	3.5
784	(EtO)2P(O)CH2P(O)	O	OH	123	>1mM	>8
784	(EtO)2P(O)CH2P(O)	O	OTBDMS	217	nd	nd
785	(EtO)2P(O)CH2P(O)	NH	OH	324	>1mM	>3
789	H2NSO2N(-)C(O)	O	OH	468	>1mM	>2
789	H2NSO2NHC(O)	O	OTBDMS	244	>1mM	>4
790	H2NSO2	O	OH	>1 mM	>1mM	nd
790	H2NSO2	O	OTBDMS	88	324	3.7
804	H	NH	OH	>1mM	>1mM	nd
828	TBDMS	O	_d	10	56.4	5.6
829	H	O	_d	298	>1mM	>3
830	H2NSO2	O	_d	254	>1mM	>4
831	(PhO)2P(O)	O	_d	99	>1mM	>10
864	Ph3C	O	_d	1.9	157	83
865	adamantylC(O)	O	_d	58	298	0.5
866	CH3(CH2)14C(O)	O	_d	256	350	1.4
867	TBDPS	O	_d	1.2	>1mM	>833
868	TBDMS	O	F	628	>1mM	>1.6
869	Ph3C	O	F	5	457	91

871	Ph3C	NH	OH	0.2	46	230
875	H2NSO2NHC(O)	O	_d	82	799	9.7
946	Ph3C	NH	F	12	>1mM	>83
947	TPS	O	F	975	>1mM	nd
948	TBDPS	O	F	89	808	9.1
949	TPS	O	OH	2.8	909	325
950	TBDPS	O	OH	4.2	806	192
951	TIPS	O	OH	227	>1mM	>4
952	PhCH2	NH	OH	111	>1mM	>9
953	Ph3C	NH	OTBDMS	313	>1mM	>3
958	adamantylC(O)	NH	OH	426	>1mM	>2
1059	(*para*-CH3O)Ph3C	O	OH	5.2	>1mM	>192
1068	(*para*-CH3O)Ph3C	NH	OH	o.72	374.3	520
1073	TPS	O	_d	1.3	>1mM	>769
1149	4-PyPh2C	NH	OH	0.23	>1mM	4345
1191	H	O	F	nd	nd	nd
1243	(*para*-CH3O)2Ph3C	O	OH	2.17	>1mM	>460
1244	(ortho-Cl)Ph3C	O	OH	2.56	>1mM	>390
1245	Dibenzosuberil	O	OH	31.55	>1mM	>31
1247	Pixyl	O	OH	3.11	>1mM	>321
1288	(*para*-NC)Ph3C	O	OH	1.14	816	715
1289	(*para*-NC)Ph3C	NH	OH	0.4	231.5	579
1290	(*ortho*-Cl)Ph3C	NH	OH	3.36	238	70.8
1291	Pixyl	NH	OH	27.6	>1mM	>36
1292	Dibenzosuberil	NH	OH	4.06	705	173.6
1329	Trybenzylsilyoxy	O	OH	103.2	>1mM	>9.6
1345	Trybenzylsilyoxy	O	F	>1 mM	>1mM	nd

Table 3.2 Inhibition Constants *Ki* (μM) for Compounds against *P. falciparum,* and human dUTPases and selectivity indexes for *P. falciparum.* [a]*Plasmodium falciparum.* [b] SI *P. f.*) selectivity index for *P. falciparum* defined as (*Ki* Human)/(*Ki P. falciparum*); nd) not determined; [d] compounds from the Group 5.

The introduction of a silicon atom (Ph_3C/Ph_3Si replacement) appears to produce as much as a 32-fold increase in enzyme selectivity while the same level of activity is retained (compare WSP483 *Ki* 1.8 μM and WSP949 *Ki* 2.8 μM; WSP864 *Ki* 1.9 μM and WSP1073 *Ki* 1.3 μM). An exception to this is the 3'-fluoro analogues (WSP869 and WSP947) in which the introduction of the TPS moiety resulted in a total loss of activity (*Ki* 5 and 975 μM, respectively). However, in this case, it may be that a premature chemical degradation of the molecule might explain the loss of inhibitory effect[59].

The replacement of the 5'-oxygen with a 5'-nitrogen generally leads to an increase in enzyme inhibition.

- In the 3'-OH and 3'-OTBDMS deoxyuridine Group (compare Group 1 and 3 with Group 2), this generally results in as much as a 10-fold increase in inhibitory activity, as evidenced by compounds WSP483 (1.8 μM) and WSP871 (0.2 μM).
- In the case of the 3'-fluoro compounds, WSP869 has a *Ki* value of 5 μM, whereas its amino analogue WSP946 shows a *Ki* value of 12 μM.
- As far as selectivity is concerned, data are limited, but in the case of compounds WSP483 and WSP871, the O/NH replacement produces a striking 25-fold increase in selectivity.
- Regarding the modification of the trityl group, the overall best modification appeared to be the introduction of a pyridyl group in place of a phenyl group (WSP1149). Indeed, the compound WSP1149 with a *Ki* 0.23 μM, inhibited the plasmodial dUTPase as effectively as the tritylamino lead WSP871 and is one of the most potent inhibitors of the series.
- Furthermore, the introduction of the pyridyl unit resulted in a significant 20-fold increase in selectivity (WSP1149, SI>4345; WSP871, SI 230 μM) directly related to a loss of affinity for the human enzyme. These results are all the more interesting in that such a modification is also expected to improve properties such as water solubility and stability in acidic media, which are important factors in the search for a drug-like lead.
- The other modifications considered produced variable effects among the amino analogues. The introduction, on one phenyl of a cyano group (WSP1289) or a methoxy group in *para* position (WSP1068), led to a selectivity increase (from 230 for WSP871 to over 500 for WSP1068) while the activity was

more or less retained (*Ki* values varied from 0.2 μM for WSP871 to 0.4 μM for WSP1289 or 0.7 μM for WSP1068). However, replacing the trityl group by a 2-chlorotrityl (WSP1290) or a dibenzosuberyl unit (WSP1292) produced a noticeable 15-20 fold loss of activity along with a possible loss of selectivity comparing the compound WSP871 with WSP1068 and WSP1292 respectively).

- Finally the introduction of a pixylamino group (WSP1291, *Ki* = 28 μM) did not promote additional interactions within the enzyme active site as hoped for (through the heterocyclic oxygen) but resulted instead in an over 100-fold loss of activity (comparing entries WSP871 and WSP1291, Table 3.3)[59].

In the case of the ether derivatives (Table 3, Group 1), little variation was observed in *Ki* (less than 3 fold) was observed as a result of introducing a substituent on the trityl group, or replacing it by a pixyl group (comparing WSP483 with WSP1247) but there was an increase (20-70 fold in selectivity). On the other hand, the dibenzosulberyl analogue (WSP1291) showed poor affinity for the *P. falciparum* dUTPase (*Ki* 32 μM), and the tribenzylsilyl derivatives (entries WSP1329 and WSP1345) were inactive. The latter result could suggest that "extending the trityl from within" with an additional methylene unit (the tribenzylsilyl group is the silicon isostere of the tribenzyl methyl group) is to be avoided, possibly for steric reasons. However, an alternate explanation could be that the tribenzylsilyl derivatives may lack stability under the assay conditions and were degraded to inactive entities (possibly 2′-deoxyuridine and 3′-fluoro- 2′,3′-dideoxyuridine)[59].

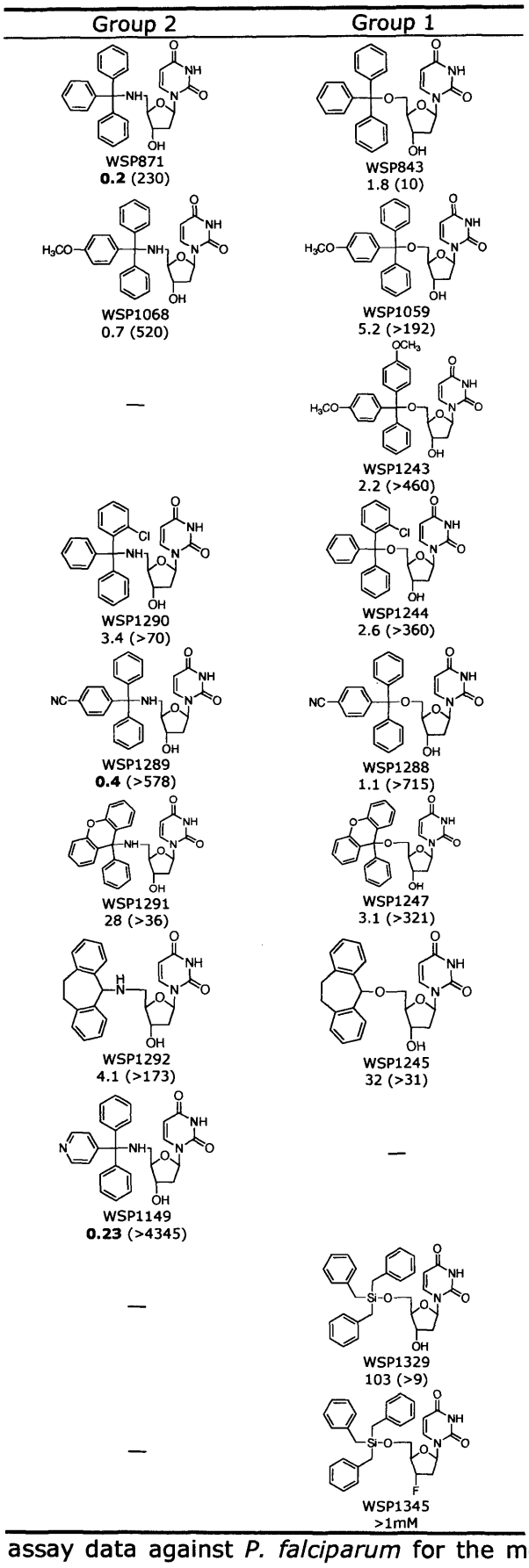


Table 3.3 Enzyme assay data against *P. falciparum* for the modified trityl series. For each entry, the first figure corresponds to the *Ki* in μM, followed by the selectivity index given in brackets). The most potent values are shown in bold.

The requirements of the substituents at the 3'-position are less clear. The comparison of the 3'-OH and Group 5, in which there is no 3'-substituent (Group 1 and 5), shows essentially no significant effect on activity, although the *Ki* values obtained for the analogues of the Group 5 might tend to be slightly lower.

The comparison of the 3'-OH and 3'-F Group (Group 1, 2, and 4) does not provide any obvious trend. Although the presence of a fluorine instead of the hydroxyl at the 3'-position seems to lower the activity, the extent of this effect varies considerably with the nature of the group at the 5'-position: from an insignificant decrease (5'-Ph_3CO derivatives WSP483 and WSP869), to a 20-fold decrease (5'-TBDPSO derivatives WSP950 and WSP948) or 60-fold decrease (5'-Ph_3CNH derivatives WSP871 and WSP946), to possibly even more, as in the case of the 5'-TPSO derivatives WSP949 and WSP947 (*Ki* 2.8 and 975 μM, respectively). However, in this latter case a more complex scenario involving the possible hydrolysis of the TPS moiety may have to be considered. The silylation of the 3'-OH produces a total loss of activity against *P. falciparum* dUTPase (compare Group 1 and 2 to Group 3), particularly for the 5'-Ph_3CNH (WSP871 and WSP953) and 5'-Ph_3CO (WSP483 and WSP484) derivatives for which the *Ki* values increase from 0.2 to 313 and 1.8 to 515 μM, respectively. Overall, the above results suggest that a bulky lipophilic group (such as TBDMS) is not tolerated at the 3'-position and a more polar substituent may be improving the inhibition potency[59].

Compounds with the substituent at the 3'-position in the "up" (xylo) position might promote interactions with the side-chain of the Ser 92 residue within the enzyme active site or alternatively with a conserved water molecule. This is a feature that will be discussed in chapter 4. The biological results obtained for the three key compounds (WSP1361, WSP1362 and WSP1363) prepared to probe

this hypothesis are given in Table 3.4 along with comparative data from the corresponding deoxyuridine analogues (3'-substituent "down" or deoxylo).

The most interesting compound was by far the analogue with a 3'-hydroxyl in the "up" position (compound WSP1361). This compound was one of the most potent and selective inhibitors of *P. falciparum* dUTPase seen so far, with a *Ki* value of 0.3 μM and a selectivity index over 3000. Compared to its deoxyridine analogue (3'-OH "down"), it showed improved inhibition of the *P. falciparum* enzyme (6-fold lower Ki value) as well as greatly increased selectivity (over 300-fold). Thus, while the derivative with a 3'-OH "down" was a weak inhibitor of the human dUTPase (*Ki* = 18 μM), its 3'-OH "up" analogue showed no inhibitory effect (*Ki* > 1mM). These results are reflected in the cellular assay, with compound WSP1361 ("up" analogue) showing better inhibition of parasite growth (IC_{50} = 1.7 μM compared to 6 μM for the 3'-OH "down" analogue).

The two other xylofuranosyluracil derivatives (compounds WSP1362 and WSP1363) were poor inhibitors of *P. falciparum* dUTPase (*Ki* = 350 and 49 μM). However, both compounds showed significant activity *in vitro* (with IC50 values of 2-4 μM). The lack of correlation observed between enzyme and *in vitro* assay results could be due to the compounds acting on a molecular target other than dUTPase or, in the case of the acetate analogue WSP1362, behaving as a pro-drug.

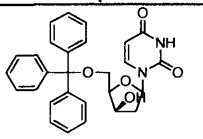
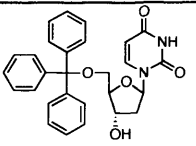
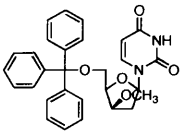
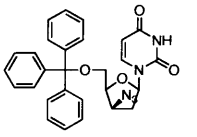
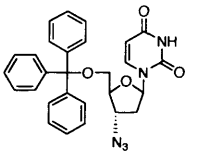
"Up"	"Down"
 WSP1361 **0.3** (SI > 3333)	 WSP483 1.8 (SI = 10)
 WSP1362 350 (SI > 2)	—
 WSP1363 49 (SI >20)	 WSP1205 48 (SI > 20)

Table 3.4 Biological assay data against *P. falciparum* for the xylofuranosyluracil series. *Ki* values are given in μM. SI is the selectivity index defined as (*Ki* against human dUTPase)/(Ki against *P. falciparum* dUTPase). The most potent values are shown in bold.

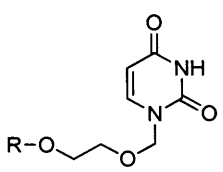
3.2 Acyclic Series

The acyclic series were designed in order to optimize the distance between the base (uracil) and trityl group and to allow further alterations and optimization of drug structure[60].

Acyclic nucleoside derivatives are successful antiviral agents; e.g. aciclovir for the treatment of herpes infections. Initially, an acyclic derivative in which there was an oxygen atom in the chain was prepared (Figure 3.3, compound WSP870). However, this particular analogue proved to be metabolically unstable, poorly selective, and fairly cumbersome to purify. Therefore a group of compounds with no oxygen in the chain was prepared (Figure 3.3, Group 2 to 5). In addition, to investigating the effect of adding rigidity to the side-chain and to possibly improve the interaction with a tyrosine residue (Tyr 112) within the enzyme active site, compounds, in which there was both an *E* and *Z* double bond in the acyclic chain (compounds WSP1315 and WSP1317), were prepared (Group 6). Finally, some branched acyclic derivatives were prepared: γ-branched analogues of the penciclovir type, and the β-branched analogues (Group 7)[60].

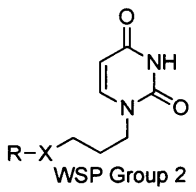
Such branched structures were designed to partially mimic the backbone of 2'-deoxyuridine and possibly increase affinity for the enzyme active site (Figure 3.4).

Compound WSP870 was the closest single chain acyclic analogue of 5'-trityloxy-2'-deoxyuridine WSP483 (Figure 3.3); it retained similar or even slightly improved (2-fold) enzyme activity and selectivity (WSP870, *Ki* 0.7 μM, SI 25; WSP483, *Ki* 1.8 μM, SI 10). Replacing the trityl (Tr) ether by a *tert*-butyldimethylsilyl (TBDMS) ether, a benzoate ester or an adamantyl amide group resulted in loss of activity, suggesting that lipophilicity and steric bulk alone did not account for enzyme inhibition. This is in agreement with conclusions previously reported for deoxyuridine derivatives[60].



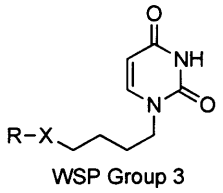
WSP Group 1

870 R = Ph_3C
873 R = *t*$BuMe_2Si$
872 R = PhC(O)
874 R = adamantylC(O)



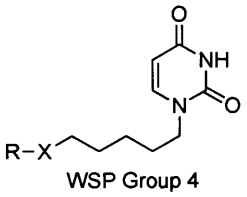
WSP Group 2

961 R = Ph_3C, X = O
1230 R = Ph_3C, X = NH
1072 R = Ph_3Si, X = O
1069 R = H, X = O
962 R = *t*$BuMe_2Si$, X = O
1064 R = PhC(O), X = O



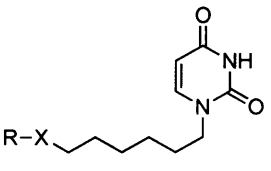
WSP Group 3

1000 R = Ph_3C, X = O
1060 R = Ph_3C, X = NH
1001 R = Ph_3Si, X = O
1003 R = H, X = O
1002 R = PhC(O), X = O
1065 R = $PhCH_2OC(O)$, X = O
1066 R = *t*-BuOC(O), X = O



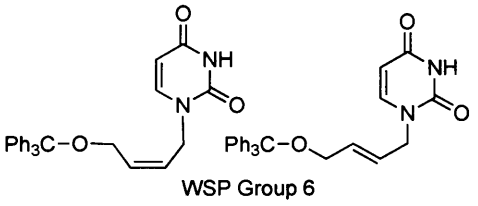
WSP Group 4

1142 R = Ph_3C, X = O
1224 R = Ph_3C, X = NH
1228 R = Ph_3Si, X = O
1208 R = H, X = O



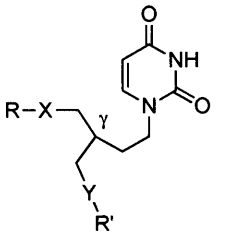
WSP Group 5

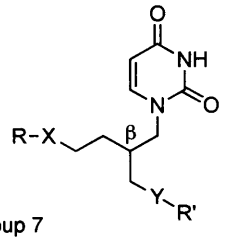
1004 R = Ph_3C, X = O
1241 R = Ph_3C, X = NH
1067 R = Ph_3Si, X = O
1005 R = *t*$BuPh_2Si$, X = O
1062 R = H, X = O
1063 R = PhC(O), X = O



WSP Group 6

1315 Z **1315** E





WSP Group 7

1318 R = Ph_3C, X = NH, R' = CH_3CO, Y = O
1320 R = Ph_3C, X = NH, R' = H, Y = O
1319 R = Ph_3C, X = O, R' = $CH_3O(CO)$, Y = NH
1328 R = Ph_3C, X = O, R' = H, Y = O
1321 R = H, X = O, R' =Ph_3C, Y = O
1322 R = Ph_3C, X = O, Y = N_3

Figure 3.3 Compound structures

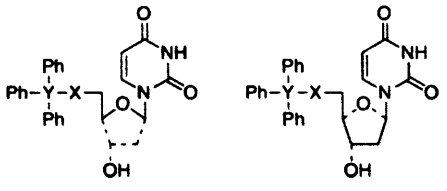


Figure 3.4. Analogy between acyclic and deoxyuridine derivatives[60].

Compound WSP1000 gave rise to a similar inhibition of the *P. falciparum* enzyme compared to compound WSP870 (WSP870, *Ki* 0.7

μM; WSP1000, *K*i 1.6 μM), which demonstrated that while having an oxygen in the side-chain probably is a closer mimic of the nucleoside, this is not a key feature for activity. Furthermore, WSP1000 showed a much lower affinity for the human enzyme than WSP870 (Table 3.5; WSP1000, *Ki* > 1000 μM; WSP870, *Ki* 17 μM) and consequently was at least 25 times more selective (selectivity index SI = [*Ki* human]/[*Ki P. falciparum*)]; Table 3.5). Hence, it can be postulated that while the oxygen does not play a significant role in *P. falciparum* dUTPase inhibition, it undergoes interactions that are more important within the human enzyme active site. In addition, the 2-(trityloxyethoxy) methyl acyclic derivative WSP870 is prone to rapid degradation by human liver microsomes. Therefore, Group 2 was not progressed further and more analogues with varying side-chain length (*n* 3 to 6) in combination with a selection of functional groups at the end of their aliphatic chain were assayed (Table 3.5). Functional moieties considered (Figure 3.2, Table 3.5) included trityl ether, tritylamine, various silyl ethers, alcohol, benzoate ester, and carbamates (Cbz and Boc) for varying steric bulk, lipophilicity and electronic characteristics[60].

3.2.1 Effect of Side-Chain Length

No general correlation between activity and side-chain length was revealed. Within the trityloxy derivatives (first compounds of the Groups 2, 3, 4 and 5), the lowest *Ki* values were obtained for WSP1000 and WSP1142. Within the tritylamino derivatives (second compounds of the Groups 2, 3, 4 and 5), compound WSP1230 was the most active, although the least selective as well. Within the triphenylsilyloxy derivatives (third compounds of the Groups 2, 3, 4 and 5), activity was retained with no significant change in *Ki* when the chain length was varied from three to six carbons. However, the chain length of three carbon atoms correlated with a lower selectivity for the parasite over the human enzyme. Possibly there is also a slight reduction in selectivity for the chain length of six carbon atoms. Compound WSP1060 (the acyclic analogue of lead WSP871) appeared as the best overall inhibitor (*Ki* 0.9 μM, SI > 1111)[60]. In addition, when focusing on the effect of the O/NH switch at the end of the side-chain within a series (Group 2 to 5 considered individually), few variations are observed as regards activity, with the clear exception of the C3 analogues; indeed while the trityl ether WSP961 did not show significant inhibition of *P. falciparum* dUTPase (*Ki* 87 μM), the corresponding tritylamine WSP1230 showed a greatly improved activity (over 400-fold) with a *Ki* 0.2 μM. A 5-fold increase in activity along with a 10-fold improvement in selectivity can also be noted when the trityloxy is replaced by a tritylamino group in the C6 series (Table 3.2; WSP1004, *Ki* 9.6 μM, SI 50; and WSP1241, *Ki* 1.8 μM, SI > 543). Very similar observations can be made when considering a C/Si switch within the trityl group itself (that is when replacing the trityloxy by a triphenylsilyloxy group). Indeed, (i) the silyl ether WSP1072 was much more active than its analogue WSP961 (*Ki* 1.6 and 87 μM, respectively); (ii) within Group 5, the silyl ether WSP1067

showed both increased activity and greater selectivity compared to WSP1004 (Table 3.2; WSP1067, *Ki* 1.7 μM, SI 226; and WSP1004, *Ki* 9.6 μM, SI 50); and (iii) little variation was noticeable within Series 3 and 4[60].

3.2.2 Unsaturated Acyclic Analogues WSP1315 and WSP1317.

The two unsaturated isomers WSP1315 (Z) and WSP1317 (E) (Group 6) showed similar levels of inhibition of the *Plasmodium* dUTPase, also comparable to that of their saturated analogue WSP1000, with the *E* isomer WSP1317 being possibly slightly more active (Table 2; WSP1000, Ki 1.6 *μM*; WSP1315, *Ki* 1.2 μM; WSP1317, *Ki* 0.6 μM). On the other hand, a great difference in activity against the human dUTPase was observed: WSP1315 gave rise to some inhibition of the enzyme (*Ki* 5.2 μM) whereas WSP1317 was totally inactive (*Ki* > 1 mM). Consequently, while WSP1315 showed very poor selectivity (SI 4), its *E* isomer WSP1317 was very selective toward the *P. falciparum* enzyme (SI > 1666). This striking observation suggests that introduction of rigidity through a double bond in position 2,3 of the side-chain can be beneficial (WSP1000 [*Ki* 1.6 μM; SI > 617] and WSP1317 [*Ki* 0.6 μM; SI > 1666]); however, the choice of conformation of the double bond is crucial if selectivity is to be retained[60].

3.2.3 Branched Acyclic Analogues (Group 7).

One of the key SAR features established so far for both acyclic and cyclic analogues is the requirement for a group bearing two or even three phenyl rings, trityl and triphenylsilyl being the groups of choice. Therefore, the branched acyclic derivatives prepared to further our

investigations all contained one such group (either TrO or TrNH), as well as another functional group in the branched chain. Furthermore, it was decided that one chain would be four carbons long as a mimic of the nucleoside analogues[60].

The γ-branched derivative WSP1320 represented the closest analogue to lead WSP871 and was found to be as potent an inhibitor of *P. falciparum* dUTPase, with a *Ki* value of 0.2 μM. However this result was partially contrasted by the 10-fold loss in selectivity observed for WSP1320 (Table 3.2). This selectivity loss was a direct consequence of an increased inhibition of the human dUTPase (Table 3.5 and Figure 3.3; WSP871, *Ki* 46 μM; WSP1320, *Ki* 5.7 μM). Thus, increasing the flexibility of the inhibitor seemed to allow enhanced interactions within the human dUTPase[60].

Blocking the free hydroxyl in WSP1320 as an acetate ester in WSP1318 resulted in a noticeable 7-fold loss of activity against the *P. falciparum* dUTPase. However, a slight increase in selectivity was also observed. This suggests that the loss of the OH hydrogen bond donor has a greater effect on interactions within the human dUTPase than with the *P. falciparum* enzyme. Compound WSP1319, that possesses a trityl ether at the end of one C4 chain and a Boc carbamate at the end of the other C4 chain, gave very similar results to the unbranched analogue WSP1000.

Similar inhibition of the *Plasmodium* dUTPase was seen for the analogue with a β-hydroxymethyl compared to the unbranched analogue WSP1000, although with a slight loss of selectivity. Replacing the hydroxyl group with an azide group (WSP1322) led to a 10-fold loss in activity. Compound WSP1321 has a three carbons chain with a β-hydroxyethyl branch. Comparison of WSP1321 with its unbranched C3 trityloxy analogue WSP961 shows that the introduction of a hydroxyethyl β-branch enhanced antiplasmodial activity (40-fold) as well as selectivity (6-fold). However, the latter remains in the low range (SI 23)[60].

WSP	*Ki P. f.*[a]	*Ki* human	SI *P.f.*[b]
870	0.7	17	25.3
872	98	311	3.2
873	44	63	1.43
874	49	>1000	6.8
961	87	313	0.58
962	73	73	1
1000	1.6	>1000	>617
1001	2.2	>1000	>446
1002	224	496	2.21
1003	>1000	>1000	nd
1004	2.3	476	49.7
1005	8.5	>1000	<118
1060	0.9	>1000	>1111
1062	>1000	636	<0.6
1063	32	>1000	>31
1064	9.2	898	>98
1065	17	>1000	>60
1066	112	>1000	>9
1067	1.7	383	226
1069	368	>1000	>2
1072	1.6	25	16
1142	2	>1000	>500
1208	>1000	>1000	nd
1224	4.3	>1000	>233
1228	3.8	>1000	>265
1230	0.2	1.4	6.75
1241	1.8	>1000	>543
1315	1.2	5.2	4.3
1317	0.6	>1000	>1639
1318	1.3	74	55.5
1319	2.5	>1000	>403
1320	0.24	5.7	23.9
1321	0.9	21	23.7
1322	23	>1000	>43.2
1328	2.2	260	120.9

Table 3.5 Inhibition Constants *K*i (μM) for Compounds against *P. falciparum,* and human dUTPases and selectivity indexes for *P. falciparum.*
[a]*Plasmodium falciparum.* [b] SI *P. f.*) Selectivity index for *P. falciparum* defined as (*K*i Human)/(*K*i *P. falciparum*); nd) not determent.

4 Chapter: Comparison of ligand-binding sites derived from grid-mapping

The active sites of the trimeric dUTPase are generally very similar. The common structural features include: two anti-parallel β-strands; a connecting β-harpin loop; and a short α-helix (Figure 4.1). All trimeric dUTPase structures contain five conserved motifs, which contribute to the interactions with the nucleotide and a magnesium ion. Despite the high similarity and identity of the dUTPase active sites there are subsite differences, which it may be possible to utilise for selective inhibitor design. A molecular modelling study should facilitate this process.

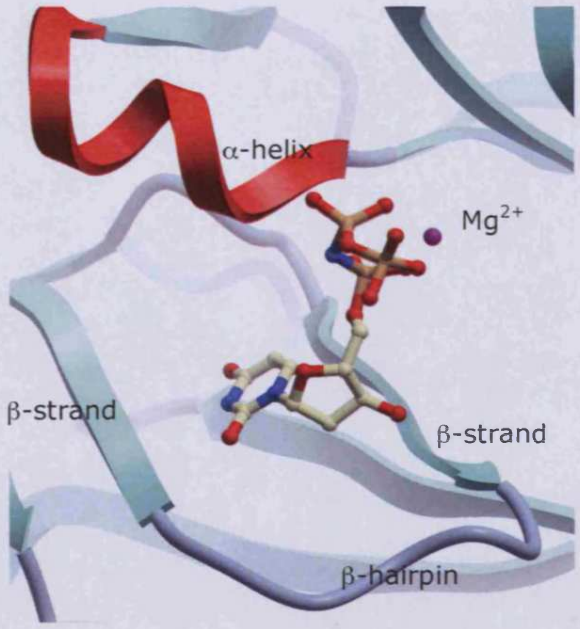


Figure 4.1 *M. tuberculosis* dUTPase active site, showing key structural features.

The design of the inhibitor WSP869 for the *P. falciparum* dUTPase was based on hit discovery by a library of nucleoside analogues. In this inhibitor there is a hydrophobic group on the 5'-position which binds to a hydrophobic pocket adjacent to the substrate binding pocket. Exploiting this new hydrophobic binding site could be a way

to obtain selective inhibitors for *P. falciparum* dUTPase. All the other dUTPase enzymes analysed in this chapter contain a similar hydrophobic pocket, although not as hydrophobic as that in *P. falciparum*, which might accommodate a relatively wide range of functionalities. Differences in the residues surrounding this site, and other sub-sites, might be used for selectivity purposes. The examination of this feature is the main goal of this study by using GRID program.

4.1 Materials and methods

The calculations were performed with version 22 of the GRID software[61]. This program is a computational procedure originally developed to identify region of energetically favourable interactions between a molecule and a probe. The probes reflect the chemical characteristics of a binding partner, or fragments of it. The interaction energies are calculated at each node of a three-dimensional grid in terms of electrostatic, hydrogen bond and Lennard-Jones potentials. Hydrogen atoms were added by the program GRIN (part of GRID package). The GRID box dimensions were chosen to encompass all of the important parts of the active site. The grid spacing was set at 1 Å, and the molecular interaction fields (MIFs) for 8 different GRID probes (Table 4.1) were calculated using GRID program. The MOVE option in GRID was used in order to take the flexibility of side chains into account. In this way multiple side chain rotamers are considered simultaneously in a single GRID calculation. Model substrates (α, β-amide dUTP; dUDP and WSP869) were used for defining the important regions in the active site. To facilitate the analysis, each active site was divided into 4 subsites (Figure 4.2), subsite S2 was further divided into S2 and S2', leading to five regions which were

examined individually (Figure 4.3 to 4.6, see also Table 4.2 to 4.5). It was necessary to treat the sub-sites individually because weak interactions in one region could be obscured by stronger interactions in other region. (The amino acids numeration is based on the sequence aliment, Figure 2.7).

Name	Chemical group
DRY	Hydrophobic probe
C3	Methyl CH_3 group
C1=	sp^2 CH aromatic or vinyl
N2+	sp^3 amine NH_2 cation
N1:	sp^3 NH with lone pair
O	sp^2 carbonyl oxygen
O::	sp^2 carboxy oxygen atom
O1	Alkyl hydroxy OH group
F	Organic fluorine atom

Table 4.1 Overview of GRID Probes Used in the GRID analysis

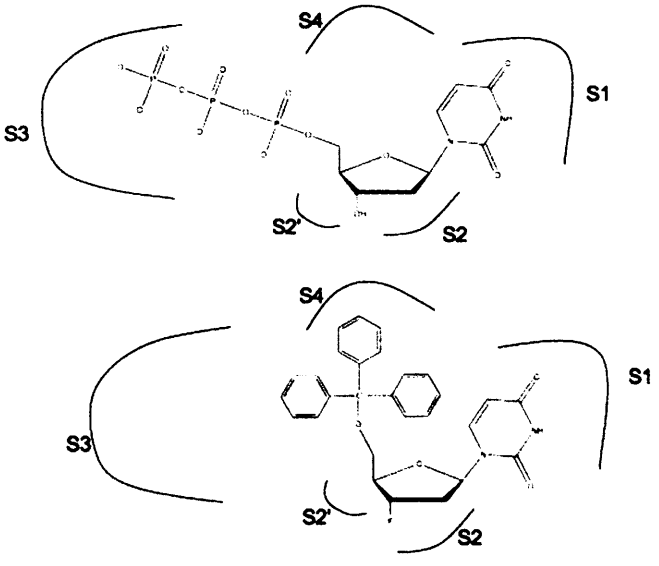


Figure 4.2 Schematic representation of the five sub-sites

Figure 4.3 Comparison of sub-sites S1. A) Superimposition of the *P. falciparum* sub-site S1 (green) with the human sub-site S1 (magenta). B) Superimposition of the *P. falciparum* sub-site S1 (green) with the *E. coli* sub-site S1 (cyan). C) Superimposition of the *P. falciparum* sub-site S1 (green) with the *M. tuberculosis* sub-site S1 (yellow)

P. falciparum	**Ile117**	**Ala118**	**Ala119**	**Asn103**	**Leu88**	**Gly106**	**Leu107**	**Ile108**
Human	Gly87	Val88	Val89	Ala75	Val59	Gly76	Val77	Ile78
M.tuberculosis	Lys91	Val92	Ala93	Asn77	Leu107	Gly80	Thr81	Ile82
E. coli	Met98	Ile99	Ser100	Asn84	Met68	Gly87	Lue88	Ile89

Table 4.2 Variation of the residues in the subsite S1 of the *P. falciparum*, human, *M. tuberculosis* and *E. coli* dUTPase active sites.

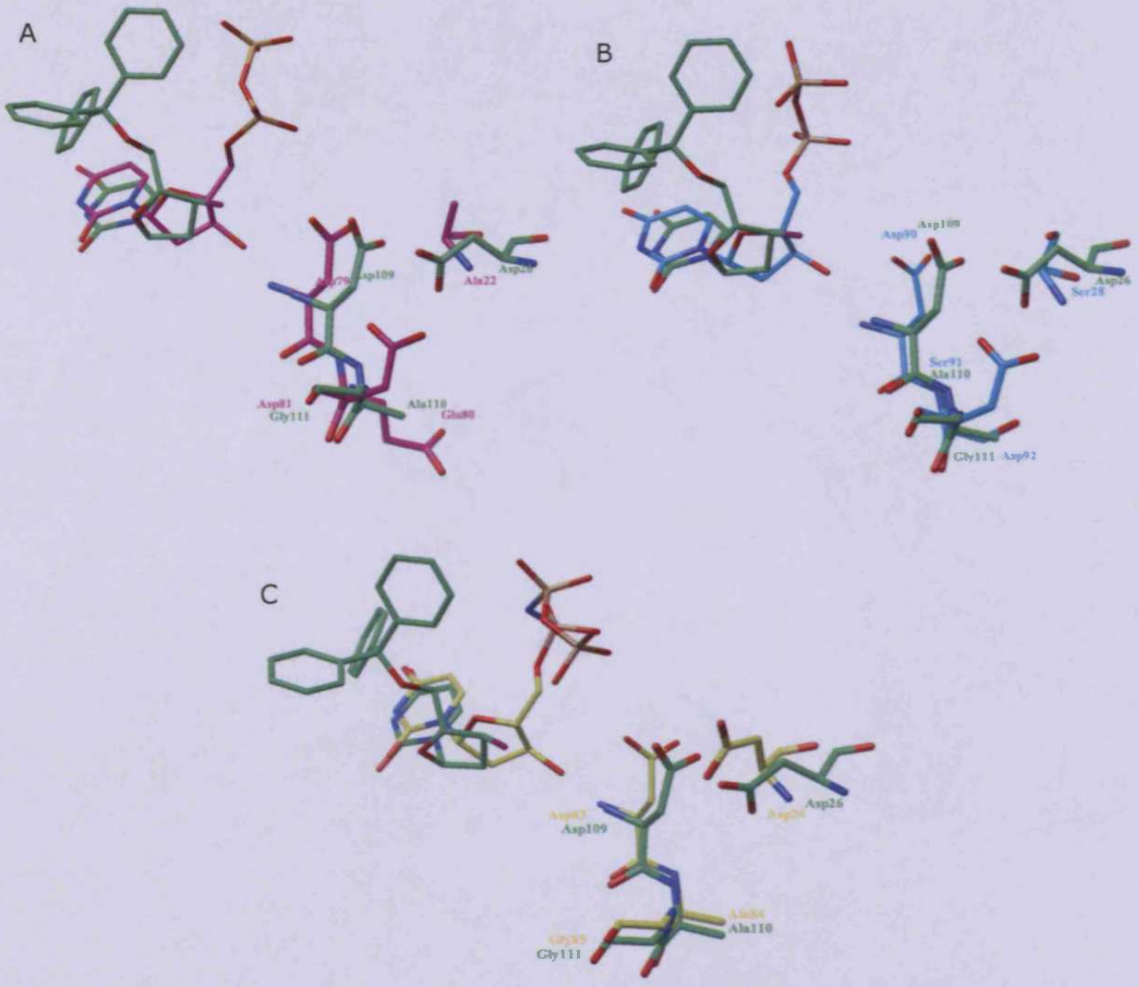


Figure 4.4 Comparison of the sub-sites S2'. A) Superimposition of the *P. falciparum* sub-site S2' (green) with the human sub-site S2' (magenta). B) Superimposition of the *P. falciparum* sub-site S2' (green) with the *E. coli* sub-site S2' (cyan). C) Superimposition of the *P. falciparum* sub-site S2' (green) with the *M. tuberculosis* sub-site S2' (yellow)

P. falciparum	**Asp109**	**Ala110**	**Gly111**	**Asp26**
Human	Asp79	Glu80	Asp81	Ala22
M. Tuberculosis	Asp83	Ala84	Gly85	Asp24
E. coli	Asp90	Ser91	Asp92	Ser28

Table 4.3 Variation of the residues in the subsite S2' of the *P. falciparum*, human, *M. tuberculosis* and *E. coli* dUTPase active sites.

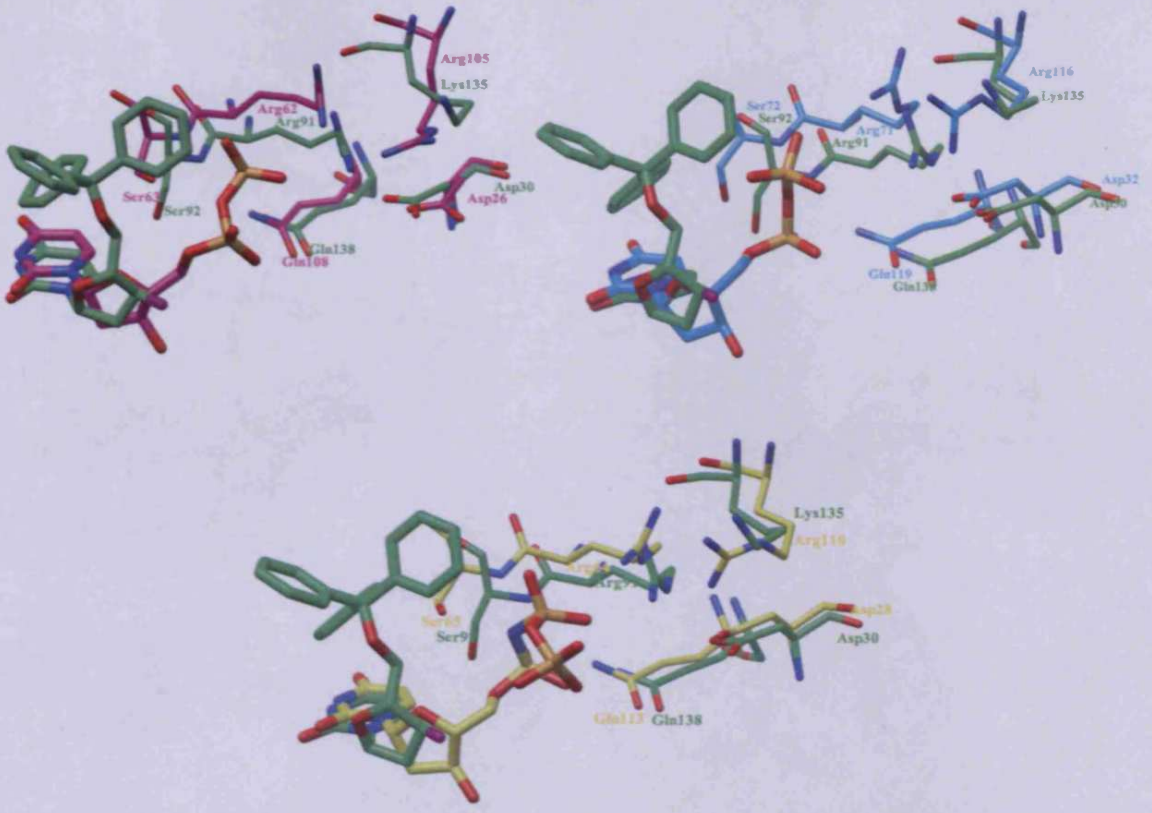


Figure 4.5 Comparison of sub-sites S3. A) Superimposition of the *P. falciparum* sub-site S3 (green) with the human sub-site S3 (magenta). B) Superimposition of the *P. falciparum* sub-site S3 (green) with the *E. coli* sub-site S3 (cyan). C) Superimposition of the *P. falciparum* sub-site S3 (green) with the *M. tuberculosis* sub-site S3 (yellow).

P. falciparum	**Ser92**	**Arg91**	**Lys135**	**Asp30**	**Gln138**
Human	Ser63	Arg62	Arg105	Asp26	Gln108
M. Tuberculosis	Ser65	Arg64	Arg110	Asp28	Gln113
E. coli	Ser72	Arg71	Arg116	Asp32	Gln119

Table 4.4 Variation of the residues in the subsite S3 of the *P. falciparum*, human, *M. tuberculosis* and *E. coli* dUTPase active sites.

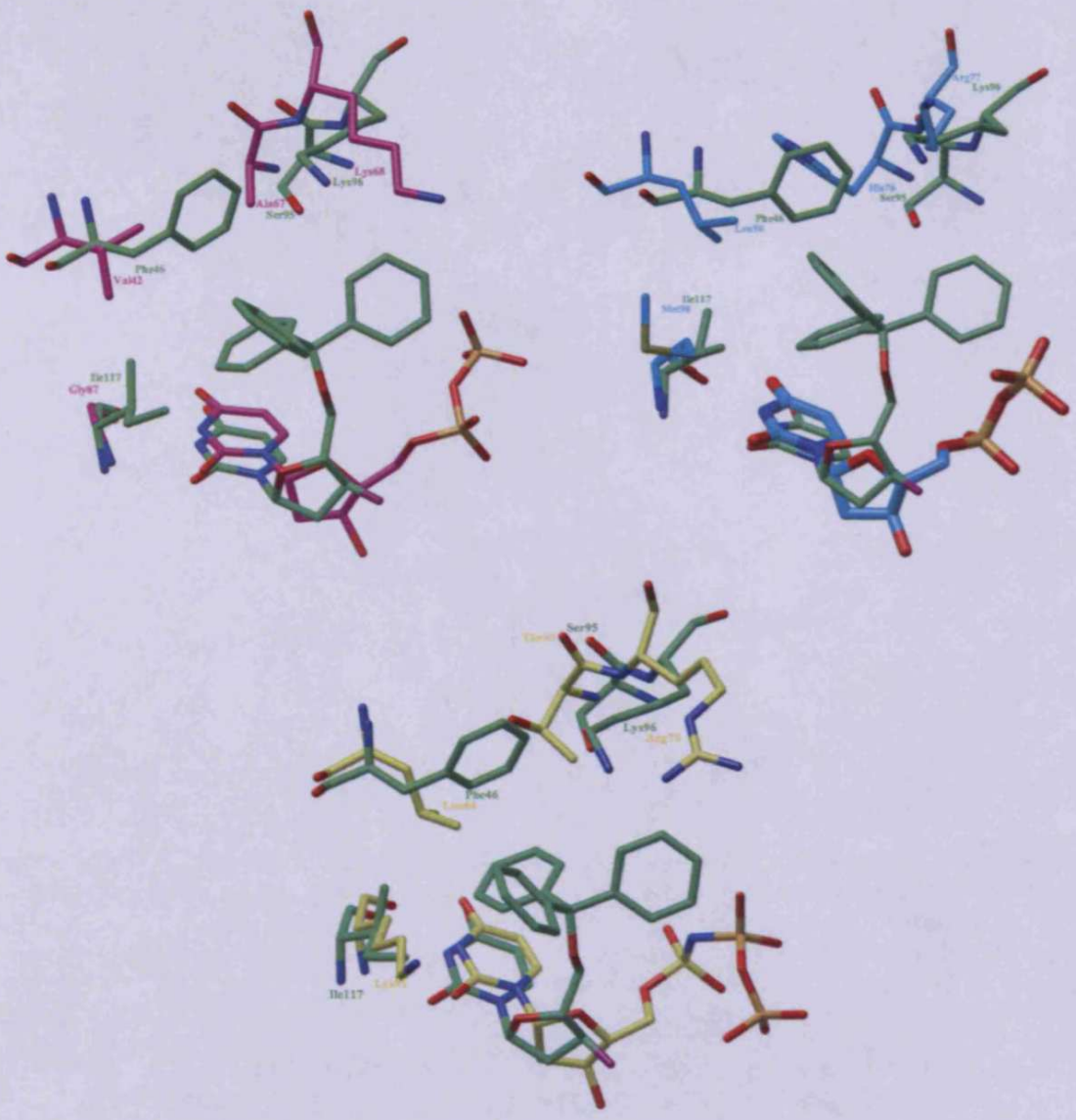


Figure 4.6 Comparison of sub-sites S4. A) Superimposition of the *P. falciparum* sub-site S4 (green) with the human sub-site S4 (magenta). B) Superimposition of the *P. falciparum* sub-site S4 (green) with the *E. coli* sub-site S4 (cyan). C) Superimposition of the *P. falciparum* sub-site S4 (green) with the *M. tuberculosis* sub-site S4 (yellow).

P. falciparum	**Ile117**	**Phe46**	**Ser95**	**Lys96**
Human	Gly87	Val42	Ala67	Lys68
M. Tuberculosis	Lys91	Leu44	Thr69	Arg70
E. coli	Met98	Leu50	His76	Arg77

Table 4.5 Variation of the residues in the subsite S4 of the *P. falciparum*, human, *M. tuberculosis* and *E. coli* dUTPase active sites.

4.2 Results and discussions

Four dUTPase structures (*P. falciparum*, human, *E. coli*, *M. tuberculosis*) were evaluated, with emphasis on differences and similarities as expressed by interaction energies with different GRID probes. In the following, each sub-site is described individually, and an overview of the results is presented in Table 4.6. The C-terminal crossing arm, which contains the conserved motif 5 is a part of the low complexity region (LCR); thus in some of the crystal structures this region was not solved. Only in the human and *M. tuberculosis* is this arm partially visible and "close" over the active site (Figure 4.7), interacting with the ligand (dUDP). For this reason, all the protein active sites were analysed first without considering the crossing arm[34, 38, 41]; subsequently the human and *M. tuberculosis* enzymes, were studied considering this region of the enzyme.

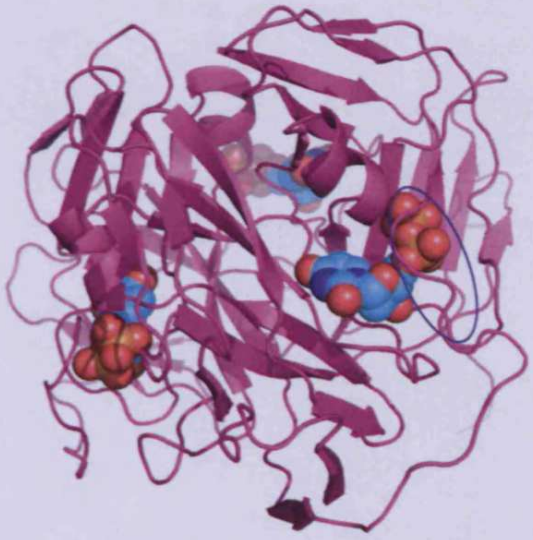


Figure 4.7. Human dUTPase with the crossing arm that closes over the active site circled.

The probes (Table 4.1) were selected considering the different characteristics of the ligands and the nature of the amino residues that form the active sites. The hydrocarbon probes (DRY, C3 and

C1=) were chosen to show favourable hydrophobic interactions. In particular, the interpretation of the fields in the S4 binding site might highlight the differences between the enzymes, which can explain the high selectivity of the trityl derivatives for the *Plasmodium* dUTPase. The trityl moiety mostly interacts with the S4 pocket. The main characteristic of the pocket in the *Plasmodium* is the high hydrophobicity. It is likely that S4 pocket contains the key amino acids that influence the selectivity of these compounds.

Enzyme	S1	S2	S2'	S3	S4
P. falciparum	-	Hydrophobic	Positive charged/ H-bond donator	Polar	Small H-bond acceptor /aromatic
Human	Small H-bond donator	Hydrophobic	H-bond donator	Polar	-
E. coli	-	Hydrophobic	Positive charged	-	H-bond acceptor
M.Tuberculosis	Small H-bond donator	Hydrophobic	H-bond donator	-	-

Table 4.6 Overview of favourable ligand properties at the 4 dUTPases, as revealed from the grid analysis

S1 Pocket. In all dUTPases the S1 pocket is relatively neutral and partially solvent-exposed. The main favourable interaction with the hydrophobic probe is due to a conserved isoleucine 108, which probably interacts with the sugar moiety and partially with the uracil group (Figure 4.8).

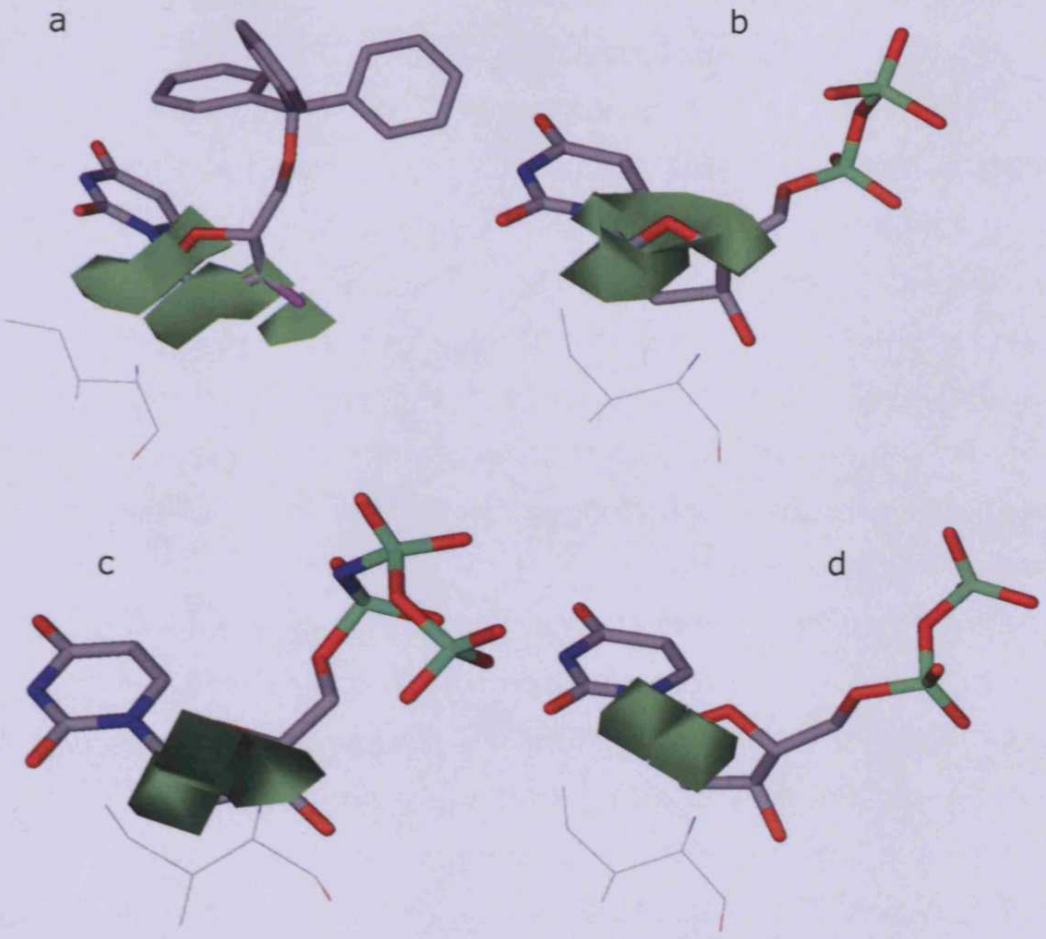


Figure 4.8. Areas of favourable interaction with the GRID DRY probe relative to the conserved isoleucine (108 in *P. falciparum* dUTPase) for a) *P. falciparum*; b) E. coli; c) M. tuberculosis; d) human dUTPase. The co-crystallised ligands are shown for clarity: a) WSP869; b) dUDP; c) α, β-imino-dUTP; d) dUDP.

The aromatic probe and the methyl probe form relatively strong interactions, though they do not distinguish between any of the active sites. However, the human dUTPase seems to have slightly better affinity with all the probes compared with the other proteins, although the dUTPase active sites are structurally very similar, especially in the pocket S1. The slightly different affinities for the probes in the human enzyme might simply be a consequence of the difference in sizes of the pockets. This is probably affected by the residue in position 103 (see sequence, Figure 2.7). In this position, the human dUTPase contains Ala, which is relatively smaller

compared to asparagine conserved in the others proteins, making the human pocket slightly more spacious.

The analysis of the positively charged probe (N2+) suggested that human, *E. coli* and *M. tuberculosis* dUTPases may have more affinity with this probe, compared with the *Plasmodium* pocket. This might be due to the generally higher hydrophobicity of the *P. falciparum* pocket. However, a common low affinity for the positive probe can be observed in all the enzymes. In fact, the probe N1: (resembling a secondary amine with lone pair) showed more favourable interactions than the N2+ probe, confirming a low affinity for positively charged probes in all the enzymes.

The alkyl hydroxyl group (O1) displays favourable interactions with the human and *M. tuberculosis* enzymes. In comparison the O1 probe has generally weaker interactions *P. falciparum* and *E. coli*. This might be explained by the fact that the human and *M. tuberculosis* enzyme have less hydrophobic character then the *P. falciparum* and *E. coli* enzymes. The best interactions with this probe are displayed in the *M. tuberculosis* dUTPase, which may be a consequence of the more solvent-exposed active site and the difference of the residue 107 (Figure 4.9). In this position the *M. tuberculosis* contains a threonine, which increases the hydrophilic character of the sub-site, while the *Plasmodium* and *E. coli* contain a leucine and the human enzyme a valine, which are purely hydrophobic residues. This might be an interesting feature in order to design selective ligands for the *M. tuberculosis* dUTPase. The fluorine probe displays the best interactions with the *M. tuberculosis* dUTPase, confirming, once again, the affinity of the *M. tuberculosis* enzyme to polar probes.

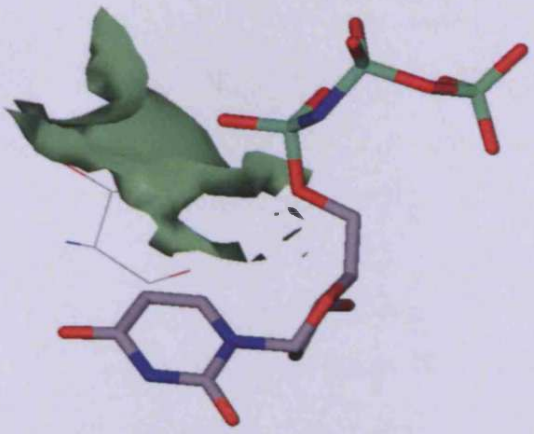


Figure 4.9. Area of favourable interaction of the GRID O1 probe within the S1 pocket. The threonine residue in position 88 *M. tuberculosis* is shown. The co-crystallised ligands α,β-amido dUTPase (grey) is shown for clarity

S2 Pocket. The S2 pocket is very similar in all structures because it is partly solvent-exposed. The proteins are all favoured by the dry probe since the sub-site is lined by a conserved tyrosine (residue 112 *P. falciparum* numbering). For similar reasons, favourable interactions are clearly displayed with aromatic and methyl probes (Figure 4.10).

S2' Pocket. This sub-site normally interacts with part of the sugar moiety, specifically with the hydroxyl group of the deoxyribose. All the proteins interact more favourably with polar probes than with hydrophobic probes in this pocket. The probe N2+ shows favourable interactions in all the proteins, with *Plasmodium* and *E. coli* clustering together. Probably this may be explained by the presence of two aspartate residues, especially with the conserved aspartate residue in position 109, which is only solvent-exposed to the small extent. In contrast, the N2+ probe displayed less affinity for the *M. tuberculosis* and human enzymes; this might be explained by the presence of histidine and arginine residues respectively, which could affect the distribution and also the diminution of the negative change in the S2' pocket. The lack of the positively charged residues in the sub-site of the *Plasmodium* and *E. coli* enzymes increased the affinity for this positive probe. The neutral secondary amine probe (N1:) displays

more affinity for human and *M. tuberculosis* than with the other proteins, which could be due to differences in the distribution of negative charge (Figure 4.11) The probes O::, O and F have low affinity with all the proteins, which is simply a consequence of the presence of aspartate residues in the sub-site. On the other hand, these residues are capable of interacting with OH probe, displaying favourable interactions with all the proteins.

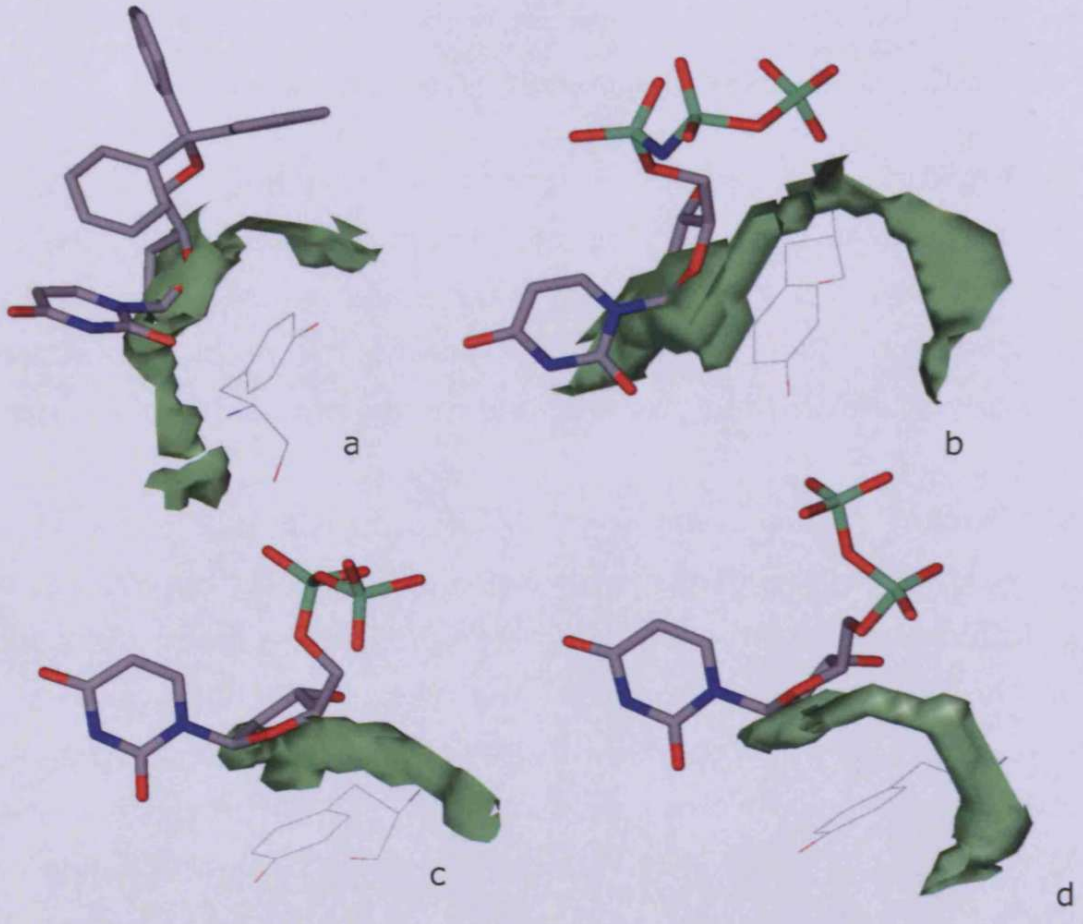


Figure 4.10 Areas of favourable interaction with the GRID DRY probe relative to the conserved tyrosine (112 *P. falciparum* dUTPase numbering) for a) *P. falciparum*; b) M. tuberculosis; c) E. coli; d) human dUTPase. The co-crystallised ligands are shown for clarity: a) WSP869; b) α, β-imino-dUTP; c) dUDP; d) dUDP.

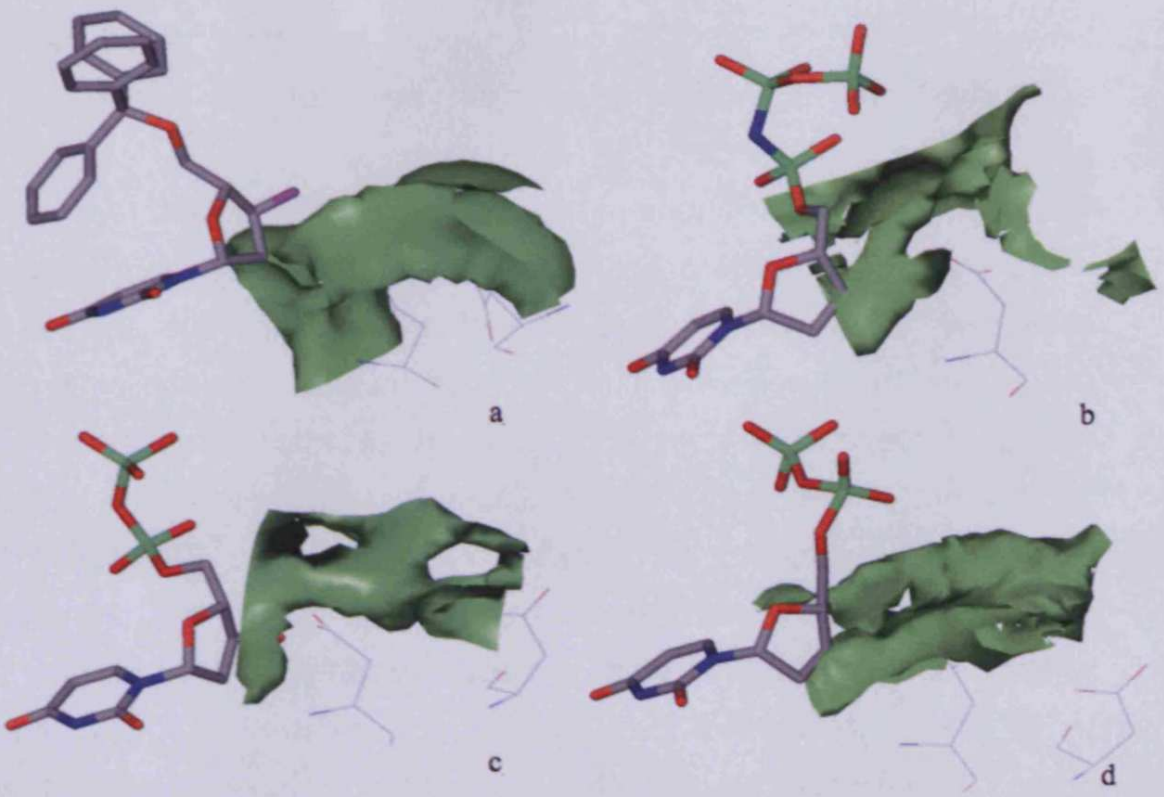


Fig. 4.11 Areas of favourable interaction with the GRID N2+ probe relative to the conserved aspartic acid residues for a) *P. falciparum*; b) M. tuberculosis; c) E. coli; d) human dUTPase. The co-crystallised ligands are shown for clarity: a) WSP869; b) α, β-imino-dUTP; c) dUDP; d) dUDP.

S3 Pocket. The S3 pocket is thought to interact mostly with the phosphate chain of the dUTP ligand. It is relatively wide and the main difference between the enzymes is the variation in size and shape, which are well defined in all the enzymes except for the *Plasmodium*. This is due to the difference in length of the loop formed approximately by the residues 20 to 30 that defines part of the S3 pocket. In the *Plasmodium* this loop is shorter than in the other proteins, which creates a pronounced difference in size and shape of the site between the *Plasmodium* and the other enzymes (Figure 4.12-13). The *Plasmodium* pocket is more elongated on one side of the pocket and more solvent-exposed. It cannot be excluded that this feature might be due to the lack of phosphate chain in the ligand WSP869, consequently the protein assumes a different rearrangement of the active site, compared to the other enzymes.

However, this difference could open up the possibility for accommodating ligands and it might be exploited for selectivity proposes.

Independent of variations in size and shape, the pocket is generally characterised as highly polar in all the structures; hence it does not show any noteworthy interactions with the hydrophobic probes. The top of the site is generally characterised by the presence of arginine residues, hence N2+ does not show any significant interactions. At the bottom of the site, the N2+ displays weak interactions, it could be due to the presence of aspartate residues in the S2' pocket. The probe N1: shows weak interactions with the top of the sub-site, it might be due to the presence of a conserved aspartate (residue 30). In all the proteins, the best interactions mostly are displayed by probes O, O:: and O1, which are capable of interacting well with arginine and aspartate residues, which are abundant in the pocket.

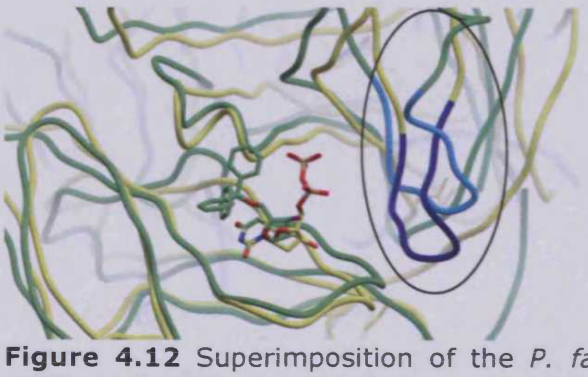


Figure 4.12 Superimposition of the *P. falciparum* dUTPase (green) with the human dUTPase (yellow). In the circle are highlighted the loops that affect the size of the sub-site S3.

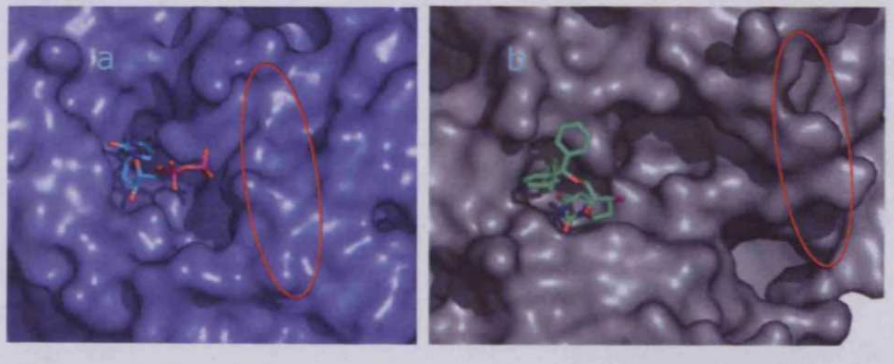


Figure 4.13 a) Human active site (open form) b) *P. falciparum.* This surface representation shows clearly the shape changes around sub-site S3. In circle is indicated the loop that affects the shape of the pocket

S4 Pocket. The S4 pocket is a hydrophobic pocket in all dUTPases. The most pronounced difference between the enzymes at this site in the position 46 (*P. falciparum* numbering). In this position either valine or leucine are contained in all the enzymes except in the *Plasmodium* dUTPase, where a phenylalanine residue is present. This particular feature probably explains the selectivity of the tritylated compounds. The most favourable interactions of the dry probe displayed for the *Plasmodium* dUTPase due to the possibility of this probe to interact with isoleucine 117 and phenylalanine 46 (Figure 4.14). In position 117 the human enzyme contains a glycine, which does not display any favourable interaction with the hydrophobic probe. The *E. coli* and *M. tuberculosis* enzymes, in position 117 contain methionine and lysine respectively, which do not form any significant interaction with the dry probe. However, the dry probe shows some favourable interaction in human, *E. coli* and *M. tuberculosis* dUTPases at the residue corresponding to position 46 of the P. falciparum dUTPase. Thus is also interacting with the alkyl chain of the conserved lysine at position 96. The *Plasmodium* pocket with the presence of phenylalanine and isoleucine residues has the most hydrophobic character. The C1= and C3 probes display the best interaction with *Plasmodium* due mostly to the presence of the phenylalanine 46. Interestingly, the *Plasmodium* enzyme is distinguished by interacting more favourably with the N1: probe compared with the other dUTPases. This is due to the serine in position 95. Human and *M. tuberculosis* enzyme, contain alanine, and the *E. coli* contains a histidine. The *Plasmodium* and *E. coli* sites display more interactions with O and O:: probes (H-bond acceptor) compared with the other proteins. This might be due to the presence of polar residues in positions 95 and 96, which are in the *Plasmodium* serine and lysine respectively, and histidine and lysine in *E. coli*.

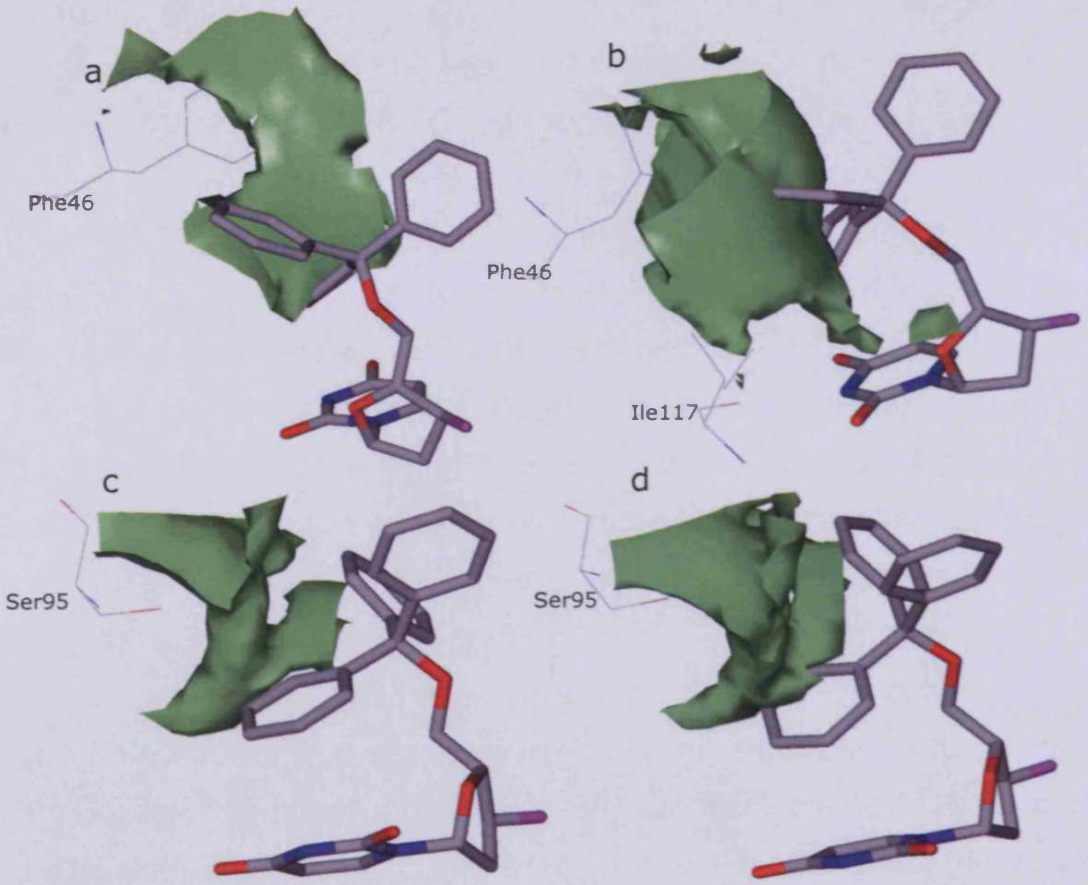


Figure 4.14 Areas of favourable interaction in the *P. falciparum* with the probe a) GRID C1= probe b) DRY probe c) O probe d) O:: probe

4.2.1 "Closed form" and "Open form"

The human and *M. tuberculosis* enzymes were crystallized with V motif partially solved in two of the three active sites of the protein. In general, the active sites in the closed form in both the structures are naturally less solvent-exposed, which affects the interaction fields of the pockets (Figure 4.15). The presence of the V motif, has the most effect on the fields of the pocket S2 and S2'.

The hydrophobic probe displays some favourable interactions with pocket S2' of the *M. tuberculosis* dUTPase, in the open form. In the closed form the dry probe does not show any interactions; this is due to the presence of the arginine residue contained in the V motif, which increases the polar character of the pocket. The C3 and C1= probes as the dry probe, do not display favourable interactions with

the pocket S2' in the closed form. The N2+ and N1: display similar interactions comparing the closed form with the open form. The probe O1 showed the most significant difference between the interaction fields in the two forms of the *M. tuberculosis*. In the closed form the interaction is more favourable then in the open form. It might be due to the increasing polar character of the pocket 2' or might be simply due to the double nature of the O1 probe (alkyl hydroxyl group) being a donator and acceptor of hydrogen bonds at the same time.

The human dUTPase was crystallised with the V motif almost completely solved, with two of the three active sites in the closed form. Most of the differences between the two forms are displayed in the pocket S2 and S2', with minor differences in the pockets S1 and S3. The dry, C3 and C1= probes, interact less favourably with the pocket S2 in the closed form compared with open form. Probably, this is due to the presence of the conserved arginine 160, which affects the hydrophobicity of the pocket 2. However, hydrophobic interactions are more favorable at the top of the pocket S1, which might be due to the presence of the residue 165 in the V motif, a phenylalanine, which close over the subsite S1, interacting with dry probe. The N2+ probe shows similar interactions with both forms, on the other hand N1: displays better interaction in the close form than in the open one, with the pocket S2'. This is probably because N1: is capable of interacting with the arginine 160. The most significant variations in the interactions fields are displayed between by the probes O, O:: and OH, with the pockets S2' and S3. This might be due to the limited solvent-exposed of the pockets and the presence of the arginine residue.

The flexibility of this loop might open up possibilities for accommodating ligands by an induced-fit mechanism. However, predictions of this feature are extremely difficult to make for two main reasons. First, in this GRID study the rigid enzyme structures

have been considered, and secondly the role of V motif (crossing arm) has not yet been fully understood.

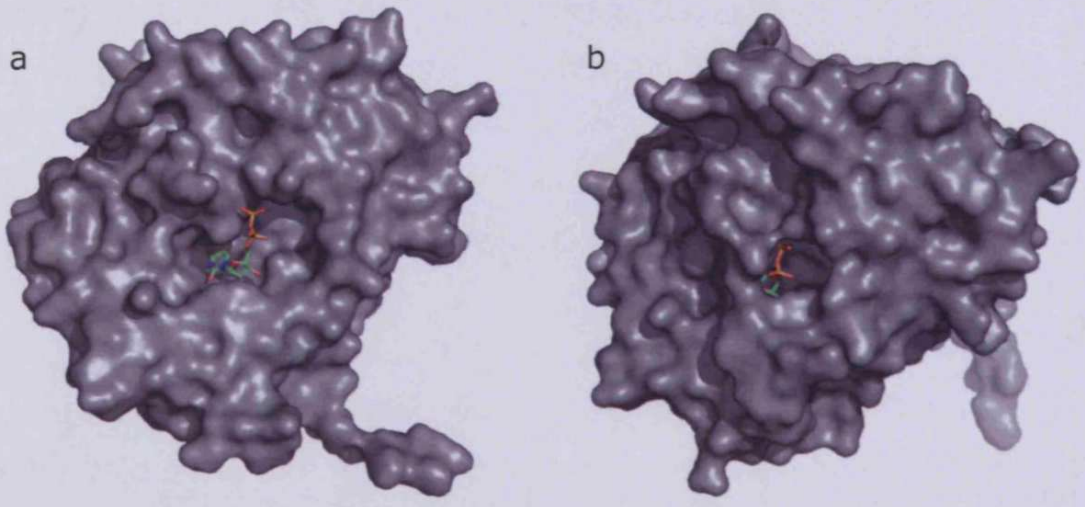


Figure 4.15 a) human dUTPase open form, b) human closed form

4.3 Conclusion

The GRID approach was used successfully to examine and highlight differences in four dUTPases, which are considered the most interesting as therapeutic targets. In particular, at the S2', S3 and S4, there are some differences among the structures, and these sites are predicted to be important in the design of selective ligands. Even though the most obvious differences are observed at the S4 site, it must be concluded that there are other possibilities for obtaining selectivity by using the differences observed at the other sites. The described differences in the S2' and S3 pockets should make it possible to distinguish between these four dUTPases, and in combination with the S4 pocket, might be possible to obtain more selective inhibitors.

5 Chapter: Docking study

Structure-based drug methods utilize knowledge of the three dimensional structure of a receptor complexed with a lead molecule in an attempt to optimize the bound ligand.

Structure-based design has become an integral part of medicinal chemistry. Although the knowledge about molecular recognition and its foundation on structural principles is still far from being complete, it has already fueled significant advances and contributed to many success stories in drug discovery[62, 63]. Evidences have been accumulated for a large number of targets that the protein 3D structure could be used to design small molecule ligands with high affinity for the protein. Several marketed compounds can indeed be attributed to successful structure-based design[64, 65].

Structure-based drug design is an iterative process[62, 63, 66]. It requires as starting point the crystal structure or a reliable homology model of the target protein, preferentially complexed with a ligand. The first step of the process is a detailed analysis of the binding site and a compilation of all aspects possibly responsible for binding affinity and selectivity.

These data are then used to generate new ideas how to improve existing ligands or to develop alternative molecular frameworks. Computational methods and molecular modelling play an essential role in this phase of hypothesis generation. They help to exploit information about the binding site geometry by constructing new molecules *de novo*, by analyzing known molecules with respect to their affinity and binding geometry, or by searching compound libraries for potential hits to suggest new leads. Hits that are commercially available or synthetically accessible are then experimentally tested and their binding properties examined by biochemical, crystallographic, and spectroscopic methods. The 3D

structure of new complexes together with the acquired activity data is subsequently used to start a new cycle of ligand design to improve the hypotheses stated in the previous round[62].

In this chapter, the automated docking was mainly used in two critical issues: prediction of the correct orientation (pose) of the bioactive conformation in the target receptor, and the estimation of the binding affinity of the complex target-ligand (score energy). The calculations were carried out using FlexX program. This program allows torsional flexibility in the ligand while protein is required to be rigid.

The nucleotide derivatives described in the chapter three were docked in the *P. falciparum* dUTPase enzyme. As presented in the chapter three, the ligands were divided in two groups, acyclic and cyclic. For some of the less active or inactive parent compounds, manual docking was carried out in an attempt to elucidate the reasons behind the reduced biological activity. Moreover, the correlation between the biological activities of the compounds and calculated ligand-protein interaction energies was investigated, in order to understand the observed SAR analysis; thus it should allow more informed design of the inhibitors. In this approach structural information of the substrate-binding pocket of the target enzyme was used from the experimental crystal structure. This data should assist in the design of modifications to lead molecules to improve their affinity with the active site.

5.1 Materials and methods

The ligands under study were built employing the SKETCH module of SYBYL6.9 and minimized using Tripos Force Field. Docking studies were carried out using the FlexX program[67] interfaced with SYBYL 6.9. FlexX is a fast-automated program based on an incremental construction procedure. In this method the flexibility of the ligands is considered while maintaining a rigid structure for the protein. All the molecules were docked into the active sites of the *P. falciparum* dUTPase. The 3D coordinates of the active sites were taken from the X-ray crystal structures of the *P. falciparum* dUTPase reported as complexes with the inhibitor WSP869, deposited in the Brookhaven Protein Databank (PDB codes: 1VYQ). Molecules of water were deleted. While creating the receptor description file (rdf), the active site was defined as the area within 6.5 Å around the cocrystallized ligand. Formal charges were assigned to all the molecules and FlexX run was submitted.

5.1.1 FlexX program

The FlexX program keeps the receptor as rigid, and the ligand as flexible. Torsion angles at acyclic single bonds are based on the MIMUMBA database[68]. The goal of the program MIMUMBA is to enumerate a large number of low-energy conformations of a small organic molecule. It is done by a set of preferred torsion angles at acyclic single bonds, and multiple conformations for ring systems. The torsion angles are taken from the MIMUMBA database containing about 900 molecular fragments with a central single bond. This method models the conformational flexibility. Therefore, reasonable minimized geometry should be used. The number of ring atoms is computed with the program SCA, which is part of the FlexX package.

It computes alternative conformations for polycyclic systems of five- to seven-membered rings.

The model molecular interactions used in FlexX has been created from the work of Bohm[69] and Klebe[70]. The intermolecular interactions can be classified by the strength of their geometric constraint. Geometrically restrictive interactions are used in FlexX when placing the ligand into the active site. Such interactions are mainly hydrogen bonds, but some hydrophobic interactions, for example those between phenyl rings and methyl groups, can be use for this propose as well.

Because most of the geometrically restrictive interactions have a favourable interaction distance, the molecular interactions are described by spherical surfaces. The program assigns an interaction type and an interaction geometry to each interacting group of the molecule. For algorithmic reasons, the interaction surfaces on the receptor side are approximated by finite sets of points, called the interaction points[71] (Figure 5.1).

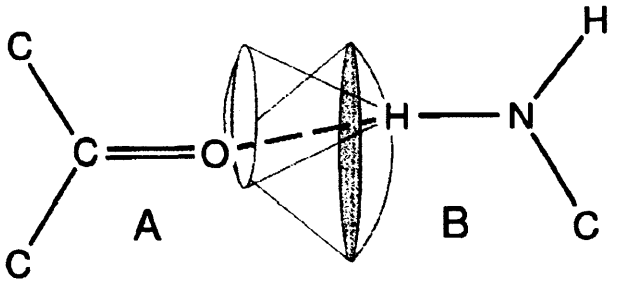


Figure 5.1 Condition for the formation of interactions: a hydrogen bond between the carbonyl oxygen and the nitrogen. The interaction centers are the oxygen and the hydrogen atom forming the hydrogen bond. They have to fall mutually on the surrounding interaction surfaces[71].

The docking algorithm in FlexX is based on an incremental construction strategy, which consists three phases:

(1) Base selection. The first phase of the docking algorithm is the selection of a connected part of the ligand, the base fragment.

(2) Base placement. In the second phase, the base fragment is placed into the active site independently of the rest of the ligand.
(3) Complex construction. In the last phase, called the construction phase, the ligand is constructed in an incremental way, starting with the different placements of the base fragment.

In the current version of FlexX, the base selection is performed interactively. In this phase the ligand is fragmented into components by severing at all acyclic single bonds. Then FlexX automatically forms a set of alternative base fragments by selecting single components or combinations of them.

The base fragments are the first parts of the ligand that are placed into the active site. Actually two algorithms[71] are in use. The first one called pose clustering superposes triples of interaction centers of a base fragment with triples of compatible interaction points in the active site (Figure 5.2). If a base fragment has fewer than three interaction centers or if the number of placements is too low, the second algorithm, called line matching, is started. This matches pairs of interaction centers with pairs of interaction points. Because of geometric ambiguity, multiple placements are generated by rotation around the axis defined by the interaction points and centers. Both base placement algorithms typically generate a large number of solutions. A reduction by clash-tests and clustering follows.

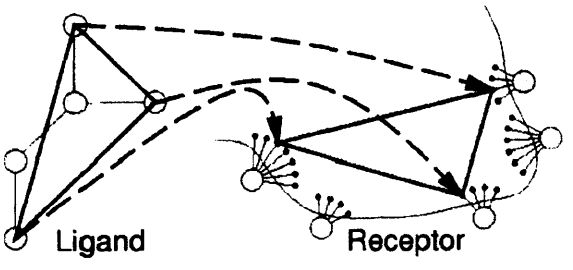


Figure 5.2 The fragment placing algorithm: mapping three interaction centers (grey spheres) of the ligand onto three discrete interaction points in the active site (black dots) defines a unique transformation of the ligand into the active site[71].

Once a set of favorable placements for the base fragment has been

computed, the program starts incremental construction. From the different base placements the complete ligand is constructed by linking the remaining components in compliance with the torsional database step by step. After adding one component new interactions are searched and the scoring function is used to select the best partial solutions which are used for the next extension step.
Ranking of the docking results is done with a scoring function developed by Bohm[72].

5.2 Docking results

Docking of the acyclic and cyclic compounds, in general revealed a consistent set of recurring interactions. Although the predicted energy score of binding is a useful descriptor of the ligand-receptor complementarity, the choice of the best docking result was ultimately dictated also by its agreement with structure-activity relationships (SARs) and the crystal structure information. Finally, for each ligand a comparison with some less active parent compounds was made. The results are described in the following section.

Firstly, to validate the modelling, compound WSP869 was docked in to the *P. falciparum* dUTPase active site, in an attempt to reproduce the binding conformation found in the crystal structure. There was a root mean square deviation value (RMSD) of 1.14 Å (the highest RMSD value was 8.37 Å) between the docked and the crystallographically determined conformation of the WSP869, suggesting that the docking was reliable (Figure 5.3). Hence, the two subsets to inhibitors (acyclic and cyclic) were docked into the active site.

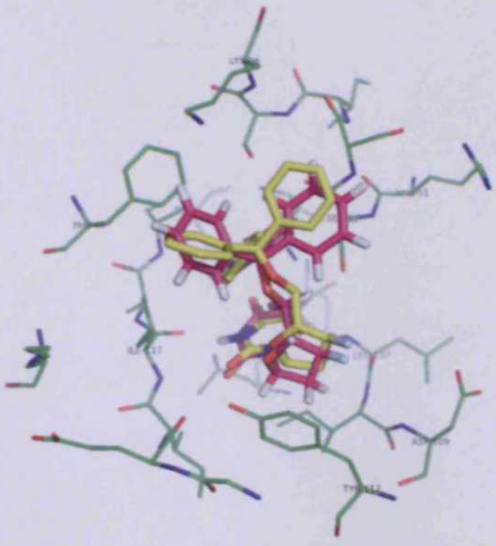


Figure 5.3. Superimposition of the binding conformation of compound WSP869 obtained crystallographically (yellow) and by docking calculation (magenta).

5.2.1 Acyclic nucleotide derivatives

In general, it can be observed that the acyclic inhibitors bind in a manner similar to the crystallographically determined conformation of WSP869. The best docked conformation for the ligands was used to determinate the interaction energy (score). The energy scores obtained from FlexX program showed a correlation with the experimentally determined inhibition constant (Table 5.1). Compounds with very poor experimental binding affinity (*Ki* > 500) also showed poor scores with FlexX, which is consistent with the model possibly showing predictive potency. There appears to be a correlation between observed best *Ki* values and docking score, which gave a linear correlation with an R^2 of 0.62 (Figure 5.4B, see also Figure 5.4A).

The acyclic nucleotides derivatives were clustered in seven Groups (see also chapter 3). These compounds have a characteristic structure with two distinctive extremities, formed by the uracil group and a trityl moiety. The extremities are linked together by a spacer of variable length.

WSP	*Ki* (μM)	Score (kJ/mol)
1230	0.2	-54.73
1317	0.61	-55.98
870	0.67	-54.62
1060	0.9	-61.73
1315	1.2	-51.26
1072	1.56	-52.69
1000	1.62	-56.01
1067	1.69	-52.22
1241	1.84	-54.46
1142	2	-53.18
1001	2.24	-48.91
1004	2.3	-54.37
1228	3.76	-54.17
1224	4.29	-52.83
1005	8.47	-46.62
1064	9.2	-42.48
1065	17	-49.21
1063	32	-51.20
873	44	-41.73
874	49	-48.88
962	73	-37.93
961	87	-53.28
872	98	-44.32
1066	112	-45.27
1002	224	-41.17
1069	368	-30.27
1003	1000	-32.78
1208	1000	-35.64
1062	1000	-38.88
1211	1000	-37.02
1294	1000	-34.56

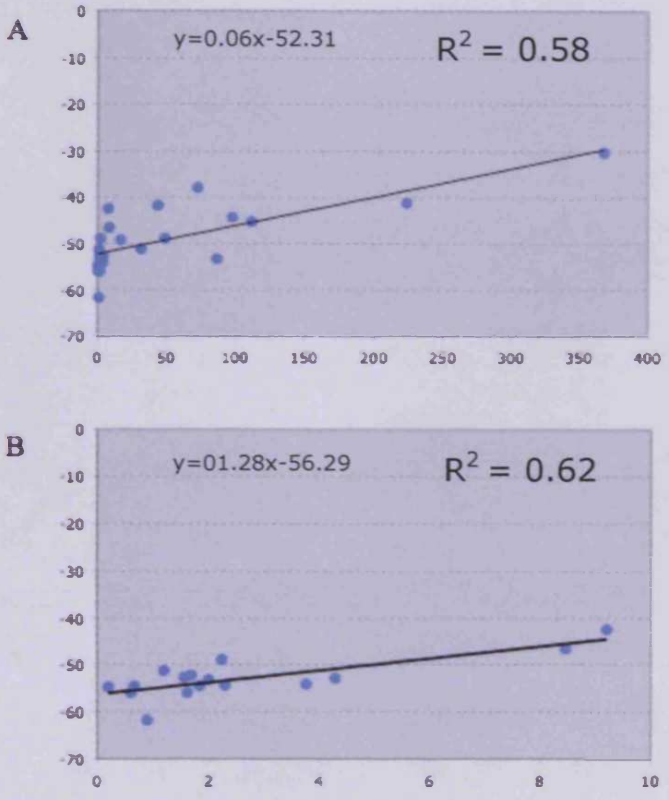
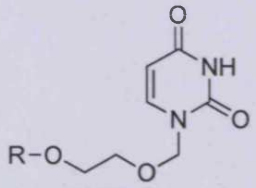


Table 1 Energy scores obtained from FlexX program and the experimentally determined inhibition constant.

Figure 4 (A) Correlation between *Ki* (*x*-axis) and FlexX score (*y*-axis) for compounds with a *Ki* < 400 μM. (B) Correlation between *Ki* (*x*-axis) and FlexX score (*y*-axis) for compounds with a *Ki* < 10 μM against *P. falciparum* dUTPase.

Group 1. This group was composed by four inhibitors characterised by the equal spacer length with a presence of a heteroatom in it (Figure 5.5). All the compounds docked in *P. falciparum* dUTPase resembling the binding mode of the experimental crystal structure of WSP869.



WSP Group 1

		Ki	Score
870	R = Ph_3C	0.67	-54.62
873	R = *t*$BuMe_2Si$	44	-41.73
872	R = PhC(O)	98	-44.32
874	R = adamantylC(O)	49	-48.88

Figure 5.5 Structures of compounds assayed of the group 1; Ki (μM), score (kJ/mol)

Compound WSP870 is the most active of this group (*Ki* 0.6 μM). The docking program was able to reproduce, with lowest energy in the ranking-score (-54.62 kJ/mol), a similar binding mode for the uracil and the trityl groups as seen in the co-crystal structure of WSP869 (Figure 5.6), with a "folded" conformation.

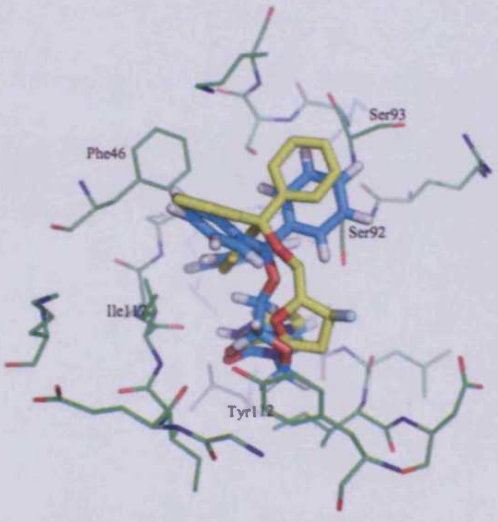


Figure 5.6 Docking result of the compound WSP870 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

The "folded" conformation is the shape assumed by the co-crystal structure of WSP869. In the experimentally found binding mode of WSP869, the base lies in the bottom of the active site forming a characteristic network of hydrogen bonds with the protein (see chapter 2). The base is closer to the Ile 108 (3.2 Å) than in the experimental structure (4.3 Å). The spacer makes hydrophobic interaction with Tyr 112 and Ile 117; the oxygen in the spacer seemed to make no interactions with protein.

The trityl group is held in place by the hydrophobic pocket at the top of the active site. Phe 46 might play the most important role in the interaction with the trityl, forming π-stacking interactions (see chapter 2). The experimental crystal structure showed that the inhibitor WSP869 interacted only with two phenyl rings; the third ring had virtually no interaction with the protein.

The "folded" conformation of WSP869 can be distinguished from the conformation assumed by the co-crystal structure of dUTP analogues found in all the dUTPase enzymes crystallised so far (Figure 5.7).

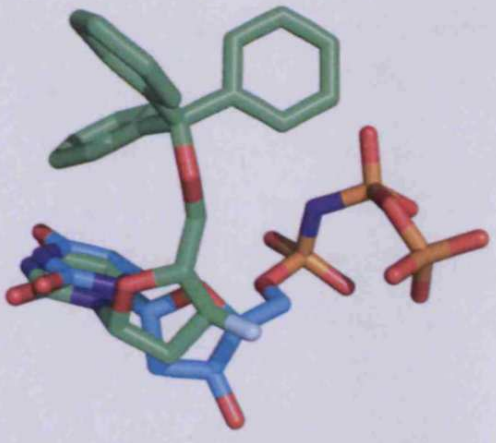


Figure 5.7 Superimposition of the co-clystallyzed sructure of WSP869 (green) and the α, β amido-dUTP (cyan) in *P. falciparum* and *M. tuberculosis* enzyme, respectively (the proteins were removed). WSP869 is showing in "folded" conformation and α, β amido-dUTP in "unfolded" conformation

In contrast with WSP869, these dUTP analogues assume an "unfolded" conformation, where the uracil and sugar moieties bind in

a similar manner to WSP869, but the phosphate chain is in a different conformation compared with the trityl moiety of WSP869.

In the docking result for WSP870, only one ring seemed to be involved in T-shaped interaction with Phe 46. The second ring did not seem to have a favourable interaction with residue 46. The third ring possibly interacted with the side chain of Ser 92 (α C) (3.7 Å). Additionally, it might interact with the amide bond of the Ser 92-Ser 93 backbone, making a stacking interaction (3.4 Å).

WSP872 contains, at the end of the spacer, an ester of benzoic acid instead of the trityl moiety. This substitution decreases dramatically the activity of the ligand (98 μM). It might suggest that the trityl group is essential for inhibition of the enzyme. As was mentioned above, the trityl is held in place by the hydrophobic and π-stacking interactions; only one of the phenyl rings seemed to form a clear interaction with the hydrophobic pocket according with crystal structure[39]. However, FlexX program was able to dock WSP872 in the *P. falciparum* in a conformation similar to the experimental structure, although the scoring function predicted a weak interaction (-44.32 kJ/mol). The spacer interacts with Tyr 112 and Ile 117; it seems that the benzoic ester does not interact with Phe 46, though it might interact with the amide bond of Ser 92-Ser 93 (4 Å). An alternative binding mode, with higher scoring energy, was found in an "un-folded" conformation (Figure 5.8).

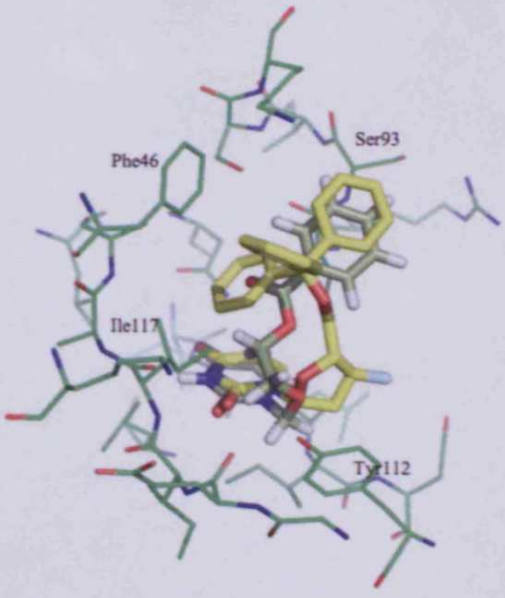


Figure 5.8 Docking result of the compound WSP872 (grey) superimposed with WSP869 obtained crystallographically (yellow).

In both the compounds WSP873 and WSP874, the trityl moiety was replace with bulky but not aromatic group, the tert-butyldimethylsilyl and adamantyl respectively (Figure 5.5). These compounds have a comparable activity (43.7 μM, 49.1 μM respectively). The result of the docking calculations showed a folded conformation for both the compounds, resembling the crystal structure (Figure 5.9A-9B). It is possible that a no aromatic, bulky, hydrophobic group could still interact with hydrophobic pocket of the enzyme. However, these interactions appear to be weaker than trityl moiety; this was also confirmed by a relatively higher scoring energy (Figure 5.5). The π-stacking interaction between the trityl group and the Phe 46 appears to be extremely important in the design of inhibitors.

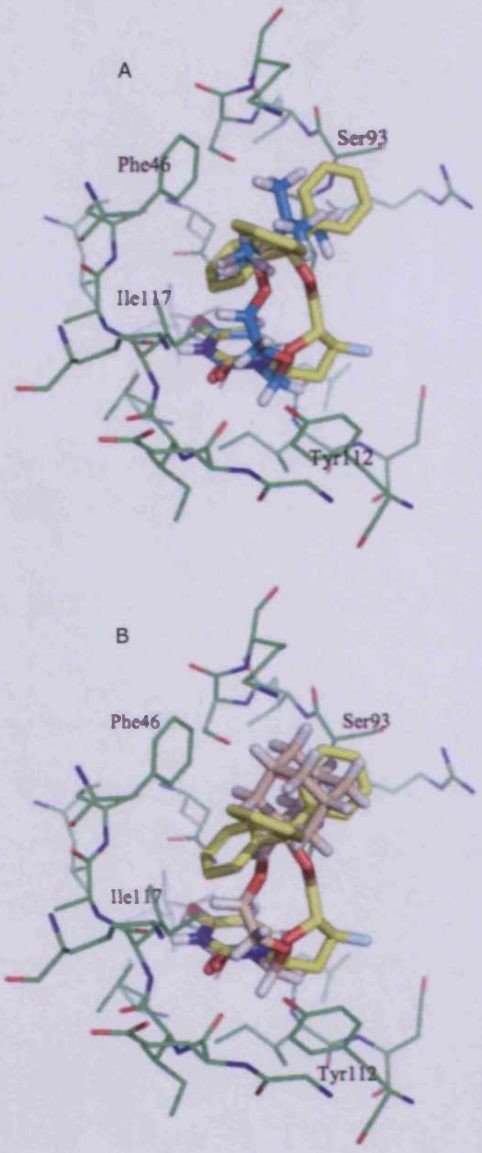
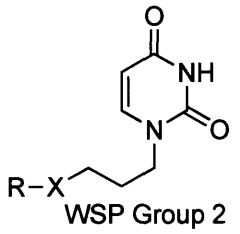


Figure 5.9 A) Docking result of the compound WSP873 (cyan) superimposed with WSP869 obtained crystallographically (yellow). B) Docking result of the compound WSP874 (pink) superimposed with WSP869 obtained crystallographically (yellow).

Group 2. Six compounds were included in this group (Figure 5.10); the feature of the compounds of this group was a short aliphatic spacer formed of three carbon atoms (3C), which should influence their degree of freedom.



WSP Group 2

		Ki	score
961	R = Ph_3C, X = O	87	-53.28
1230	R = Ph_3C, X = NH	0.2	-54.73
1072	R = Ph_3Si, X = O	1.56	-52.69
1069	R = H, X = O	368	-30.27
962	R = *t*$BuMe_2Si$, X = O	73	-37.93
1064	R = PhC(O), X = O	9.2	-42.48

Figure 5.10 Structures of compounds assayed of the group 2; Ki (μM), score (KJ/mol)

The FlexX program docked successfully WSP961 in *P. falciparum* with a folded conformation reproducing the binding mode of the crystal structure (Figure 5.11). The spacer formed hydrophobic interaction with Tyr 112, and no interaction with Ile 117, which might be due to the limited flexibility of its short spacer. The trityl moiety was linked to the spacer by an oxygen group, which seemed making no interactions with enzyme at all. The conformation of the trityl group was similar to the experimental structure. Two of the phenyl rings were involved in a T-shaped interaction with Phe 46, the third ring made hydrophobic interaction with αC of Ser 92 (3.05 Å). Although the docking program reproduced the crystal structure conformation, with one of the lowest scores for the group of acyclic inhibitors (see Table 5.1), WSP961 did not show good activity (*Ki* 87 μM) compared to WSP869 (*Ki* 5 μM), and poor selectivity for the *P. falciparum* enzyme (Appendix 1). The poor inhibition constant of

WSP961 might be due to a problem in the experimental assay, such as an unexpected instability of the molecule.

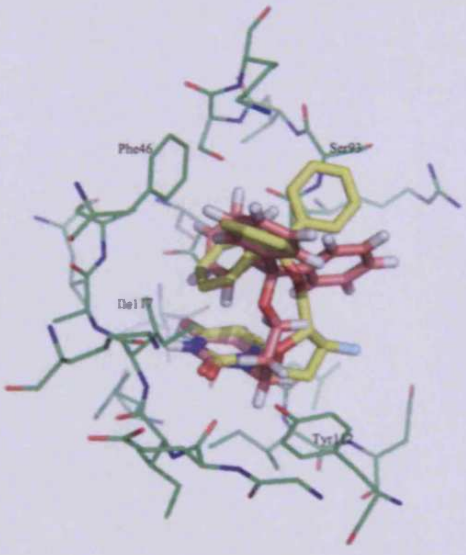


Figure 5.11 Docking result of the compound WSP961 (pink) superimposed with WSP869 obtained crystallographically (yellow).

The trityl moiety, in the compound WSP962, was substituted with non-aromatic group tert-butyldimethyl silyl. The FlexX program reproduced successfully the experimental binding mode of WSP869 (Figure 5.12). However, the poor activity of this inhibitor (*Ki* 73 μM) seems to confirm that the trityl group should be important for the activity and selectivity of these inhibitors.

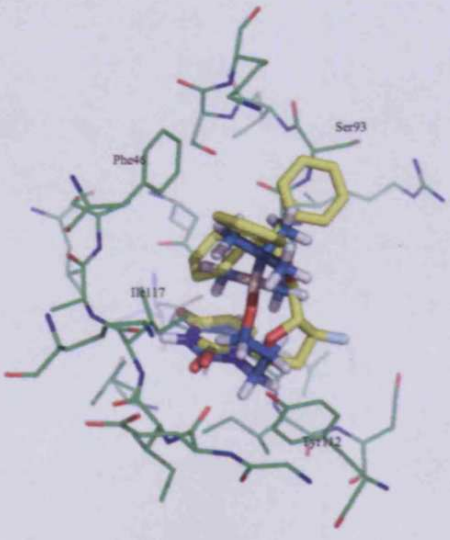


Figure 5.12. Docking result of the compound WSP962 (blue) superimposed with WSP869 obtained crystallographically (yellow).

Docking of WSP1064 (Figure 5.13) showed the characteristic "folded" conformation that resembles the experimental binding mode, although the trityl moiety was substituted with the ester of the benzoic acid (Figure 5.10). Despise the presence of only one aromatic ring, the compound shows a relatively good activity (*Ki* 9.2 μM). The spacer was in van der Waals contact with Tyr 112 and additionally the aromatic ring made a T-shape interaction with Phe 46. This might indicate that the presence of phenyl rings, such as the trityl moiety, is not absolutely necessary for a significant inhibition of the *P. falciparum* dUTPase.

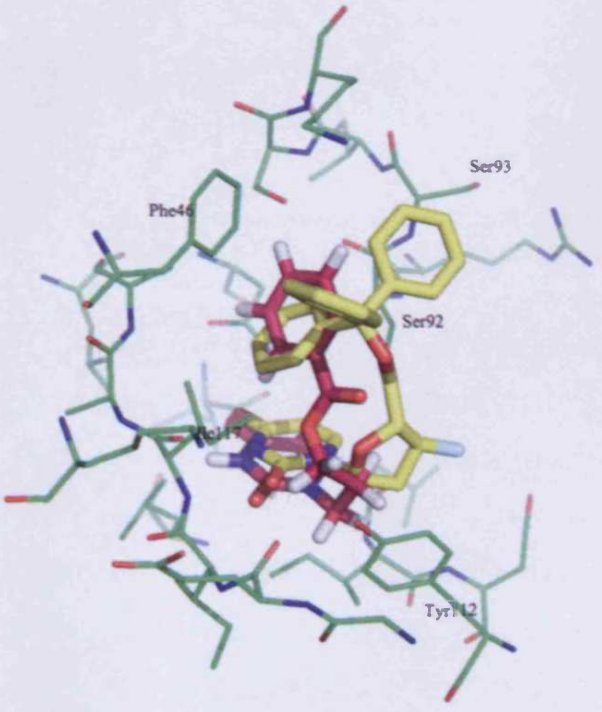


Figure 5.13 Docking result of the compound WSP1064 (magenta) superimposed with WSP869 obtained crystallographically (yellow).

WSP1069 had very low activity (*Ki* 368 μM), which was not surprising since the molecule had only a hydroxyl group at one of the extremity. However, FlexX program was able to dock WSP1069 with a binding mode similar to the crystal structure (Figure 5.14), and the high score (less negative) partially reflect the experimental data (Figure 5.10).

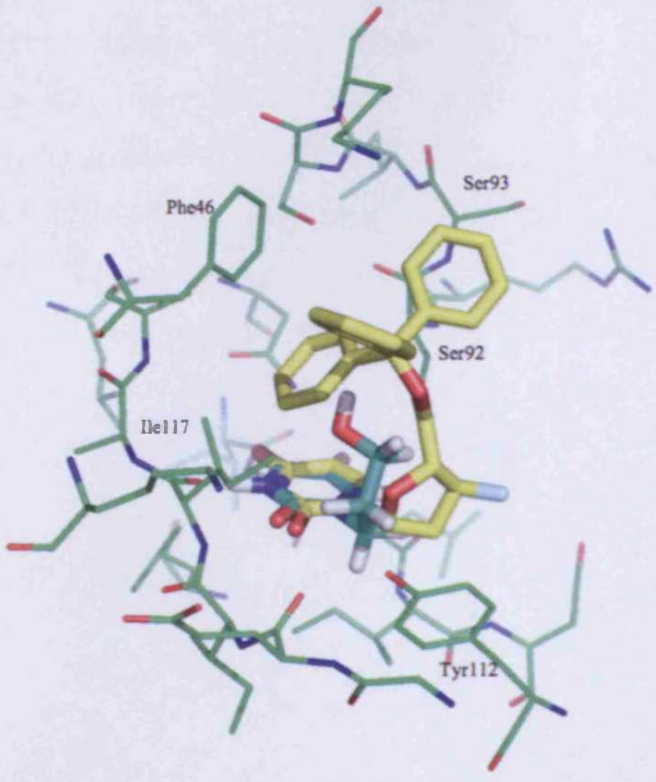


Figure 5.14 Docking result of the compound WSP1069 (light-green) superimposed with WSP869 obtained crystallographically (yellow).

WSP1072 docked in the folded conformation (Figure 5.15). In this compound, the trityl was substituted with triphenyl silyl, which assumed a conformation similar to the experimental structure. One aromatic ring formed T-shape interaction with Phe 46. The second ring interacted with side chain of Ile 117 by hydrophobic interaction. Despite the structure similarity between WSP961 and WSP1072 derivatives, the latter has a better activity (*Ki* 1.6 μM) and selectivity than WSP961 (appendix 1). This is likely due to the presence of the Si atom in the trityl moiety, which is larger than carbon atom (the average lengths of the Si-C and Si-O bonds are 1.9 Å and 1.6 Å respectively, and the average lengths of the C-C and C-O bonds are 1.5 Å and 1.4 Å respectively). This feature might give more flexibility to the triphenyl silyl moiety compared to the trityl moiety, thus more capable to make interactions with the protein active site.

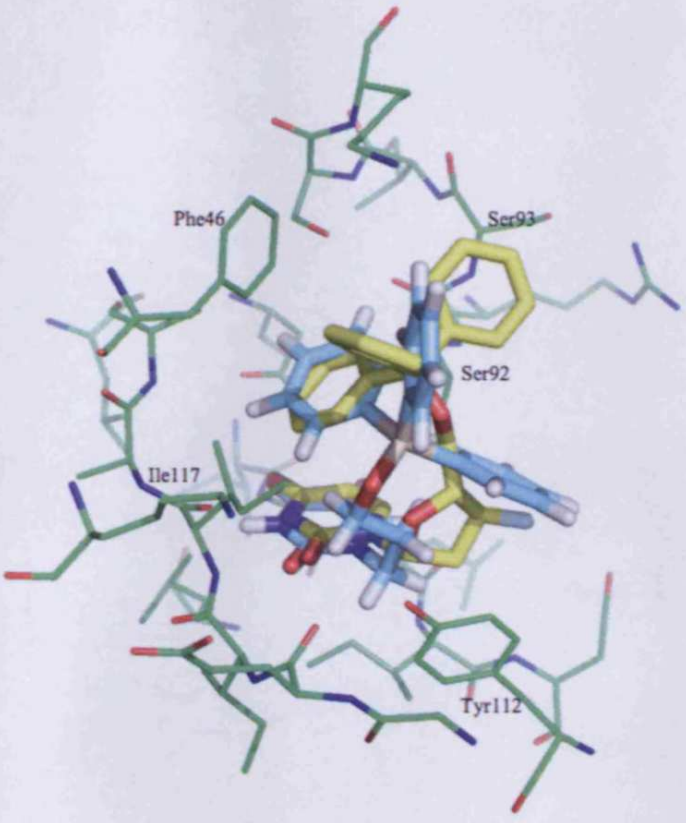


Figure 5.15 Docking result of the compound WSP1072 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

Docking of WSP1230, revealed a "folded" conformation characteristic of the experimental binding mode (Figure 5.16). The activity of this compound is the best of the acyclic inhibitors (*Ki* 0.2 μM); additionally the predicted score was one of the most negative in the series (-54.73 kJ/mol). The NH group, which connects the trityl moiety with the spacer, apparently had a little or no interaction with the protein. This can be also seen for the oxygen linker in the corresponded position of the derivatives WSP961 and WSP869 (oxygen atom in 5'-position of inhibitor complexed with *P. falciparum* dUTPase). In fact, they all have very similar binding modes. Despite this similarity, WSP1230 is approximately twenty-five and four hundred times more active than WSP869 and WSP961, rispectively. In some cases the presence of the NH group, which links the trityl and the spacer, increased significantly the inhibition (Appendix 1).

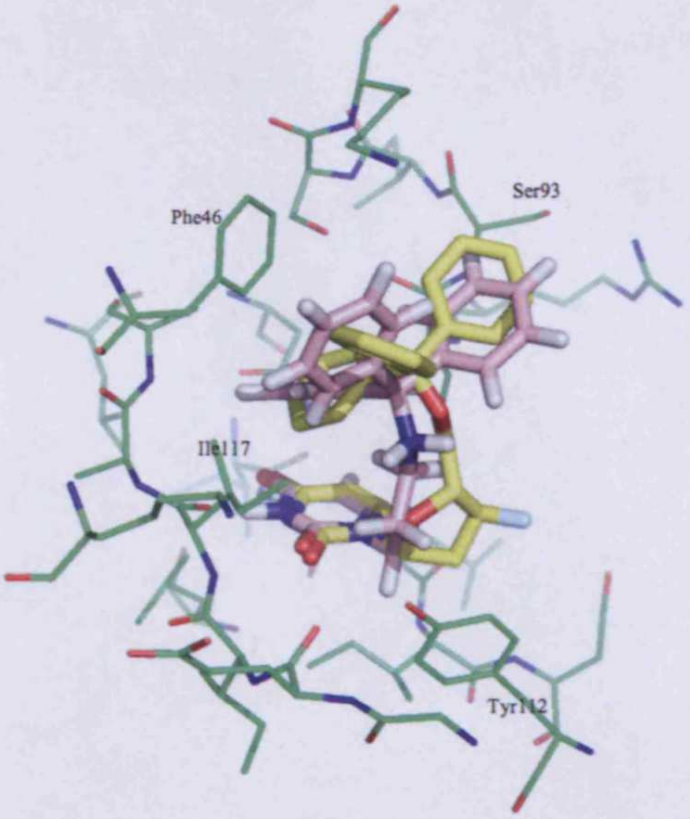
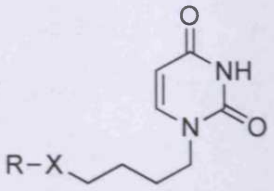


Figure 5.16 Docking result of the compound WSP1230 (pink) superimposed with WSP869 obtained crystallographically (yellow).

Group 3. This group had seven inhibitors (Figure 5.17), which were characterised by an equal length of the spacer (aliphatic chain of four C atoms).



WSP Group 3

	Ki	score
1000 R = Ph_3C, X = O	1.62	-56.01
1060 R = Ph_3C, X = NH	0.9	-61.73
1001 R = Ph_3Si, X = O	2.24	-48.91
1003 R = H, X = O	1000	-32.78
1002 R = PhC(O), X = O	224	-41.17
1065 R = $PhCH_2OC(O)$, X = O	17	-49.21
1066 R = *t*-BuOC(O), X = O	112	-45.27

Figure 5.17 Structures of compounds assayed of the group 3; Ki (μM), score (kJ/mol)

The docking of WSP1000 resembled the experimental binding mode (Figure 5.18).

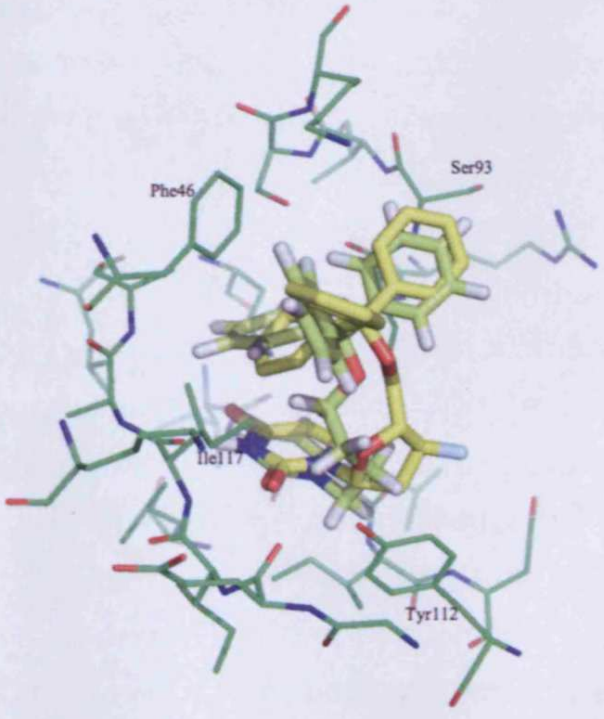


Figure 5.18 Docking result of the compound WSP1000 (lime) superimposed with WSP869 obtained crystallographically (yellow).

The base formed the characteristic network of hydrogen bonds with protein; the spacer makes hydrophobic interactions with Tyr 112 and Ile 117. The oxygen, which links the trityl with the spacer, does not seem to interact with protein at all. Two aromatic rings of the trityl moiety are involved in a T-shaped with Phe 46; the third may interact with amide linking Ser 92-Ser 91, forming stacking interaction. From the docking calculation an alternative second binding mode was found, which was less populated in the calculations and has a higher energy score of binding (more positive, less favourable; 6 kJ/mol higher, see appendix 1). The difference with respect to the top-ranking binding mode is in the orientation of the trityl, which occupies a more external and hydrophilic region in this seconding binding mode (S2' pocket, see chapter 4). In this second binding mode, the trityl makes virtually no significant interactions with protein, thus a certain ambiguity of the ligand orientation may be possible. However, using automated docking calculation, it is not surprising to obtain several conformational clusters, since that with the rising of number of the atoms in the spacer, also the flexibility of the inhibitor increases, which might lead to alternative conformations.

The docking of WSP1001 showed a folded conformation similar to the experimental binding mode (Figure 5.19). The spacer makes hydrophobic interaction with Tyr 112 and Ile 117. The triphenyl silyl is bulkier than the trityl group. This is due to presence of silicon atom, which is larger than the carbon. This difference contributes to the triphenyl silyl being in a different geometry compared with crystal structure. One of the aromatic rings formed a T-shaped interaction with Phe 46, the second ring had a hydrophobic interaction with Ile 117 and it could form T-shaped interaction with Tyr 112. The third ring probably formed cation-π interaction with Lys 96.

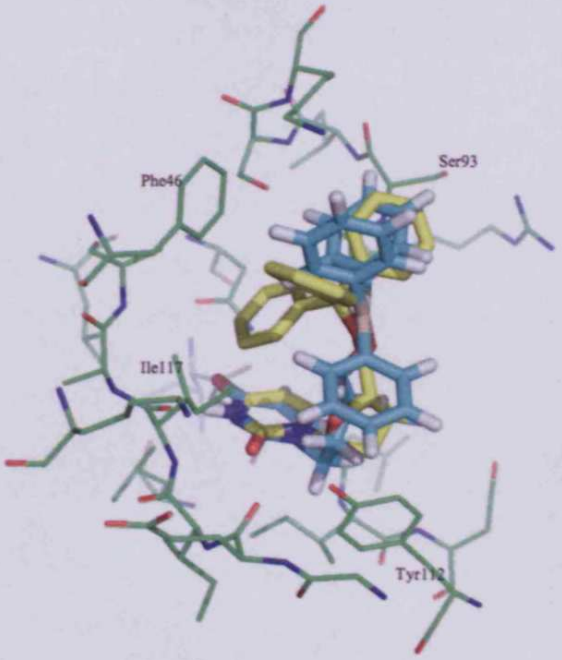


Figure 5.19 Docking result of the compound WSP1001 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

WSP1002 docked in folded conformation (Figure 5.20). It showed very poor activity (*Ki* 224 μM), which is consistent with the theory that a bulky and aromatic group as the trityl, is necessary to have good inhibition potency on the *P. falciparum*. In contrast, the compound WSP1064 from the Group 2 (Figure 2.13) has an interesting activity (9.2 μM), approximately 24 times more active than WSP1002. The only difference between these two compounds is the length of the spacer, which possibly plays an important role in orienting of the compounds in the active site. The docking result of WSP1064 showed the benzoic group being involved in a T-shaped interaction with Phe 46; on the other hand WSP1002 the same group appeared to make no interactions with protein. The possibility of having a second binding mode for WSP1002 cannot be excluded. In fact, WSP1002 is very flexible and the lack of the trityl moiety increases the possibility of obtaining a second binding mode, as there is likely to be a weaker interaction with the hydrophobic pocket. Thus an alternative "unfolded" conformation with a higher score energy was also predicted by the docking. In this conformation the uracil

formed the characteristic network of hydrogen bonds with bottom of the active site; the spacer lay on the Tyr 112; and the aromatic ring did not seem to interact with protein. However, the oxygen of the carbonyl ester made a hydrogen bond with Asp 109 backbone. The second binding mode could be an artefact of the program. Equally the lower binding energy (*Ki* 224 μM) might reflect these more weakly binding conformations.

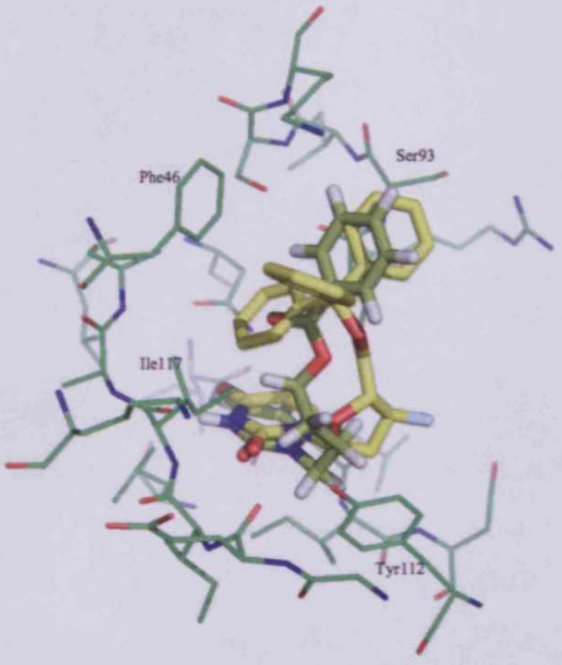


Figure 5.20 Docking result of the compound WSP1002 (lime) superimposed with WSP869 obtained crystallographically (yellow).

WSP1003 is characterised by presence, at one end of the spacer, of a hydroxyl group. The complete inactivity of this ligand was not surprising. However the docking program was able to accommodate the uracil of the WSP1003 in a conformation similar to the crystal structure (Figure 5.21), although its score was highest of the group.

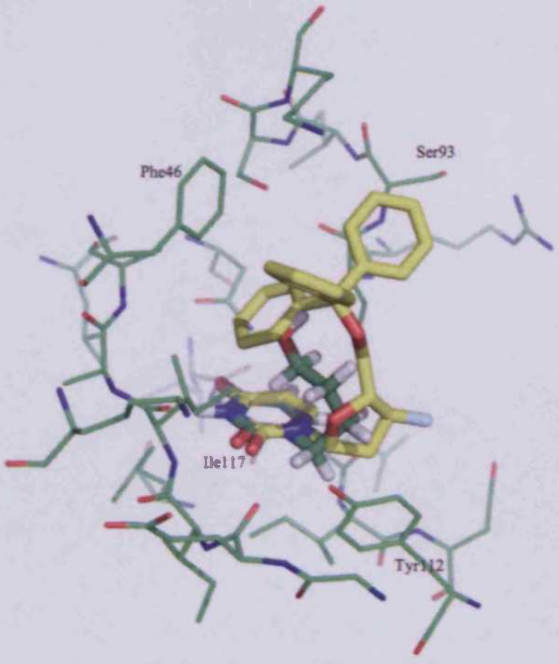


Figure 5.21 Docking result of the compound WSP1003 (dark-green) superimposed with WSP869 obtained crystallographically (yellow).

WSP1060 showed good inhibition potency (0.9 μM). FlexX docked the inhibitor in a folded conformation, which resembles the experimental structure (Figure 5.22), and the low score was in accord with the experimental data (Figure 5.17). It formed the usual network of hydrogen bond interactions with the bottom of the active site. The spacer interacted with Tyr112 and Ile117 making hydrophobic interactions, only one aromatic ring of the trityl formed a clear π-stacking interaction with Phe 46. The NH group seemed to make no contact with the enzyme. Comparing the compounds WSP1060 and WSP1000, structurally the only difference is the group that links the trityl moiety with the spacer (NH and O respectively, see appendix 1) and they form analogous contacts with protein. Despite this high similarity between the two inhibitors, WSP1060 has slightly better activity than the compound WSP1000 (1.6 μM), which would be consistent the generally observed improvement of the activity with the presence of amino group as linker.

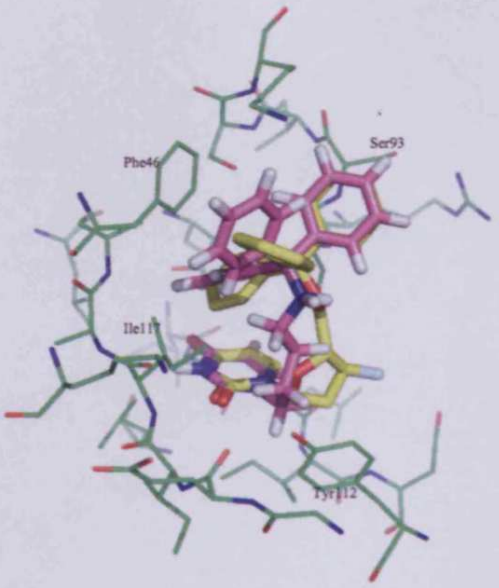


Figure 5.22 Docking result of the compound WSP1064 (light-violet) superimposed with WSP869 obtained crystallographically (yellow).

Compounds WSP1065 and WSP1066 have a *Ki* of 17 μM and 112 μM respectively. Both the inhibitors docked with binding modes that resemble the experimental structure (Figure 5.23A-23B). It suggested that the presence of an aromatic ring could affect positively the inhibition potency, because it might interact more efficiently with Phe 46.

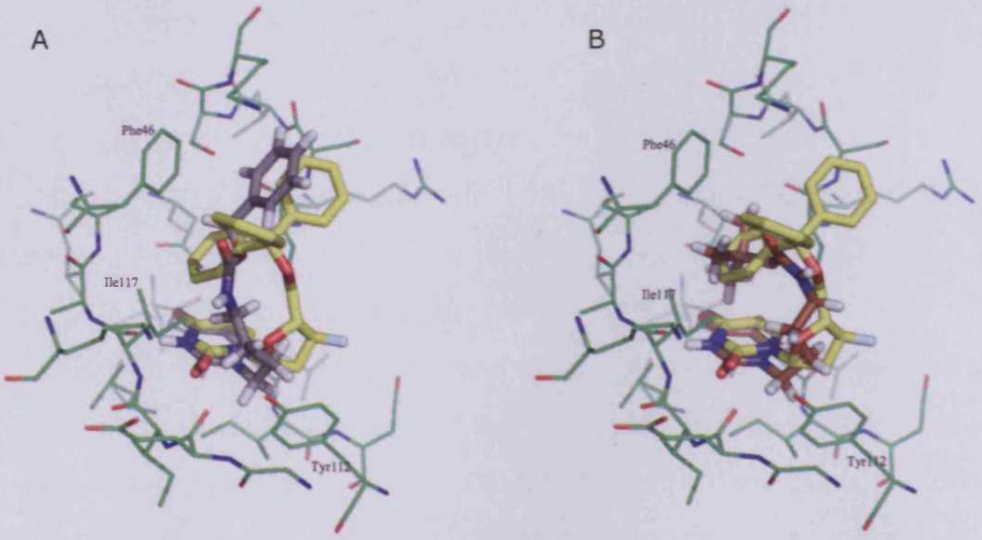
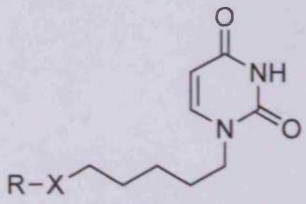


Figure 5.23 A) Docking result of the compound WSP1065 (dark-grey) superimposed with WSP869 obtained crystallographically (yellow). B) Docking result of the compound WSP1065 (brown) superimposed with WSP869 obtained crystallographically (yellow).

Group 4. This group contains four compounds. Each compounds has an aliphatic spacer of five C atoms (Figure 5.24). In general, increment of the length of the spacer increases the flexibility of the ligands, and consequently there are more possibilities of additional alternative binding modes.



WSP Group 4

	Ki	score
1142 R = Ph_3C, X = O	2	-53.18
1224 R = Ph_3C, X = NH	4.29	-52.83
1228 R = Ph_3Si, X = O	3.76	-54.17
1208 R = H, X = O	1000	-35.64

Figure 5.24 Structures of compounds assayed of the group 4; Ki (μM), score (kJ/mol)

WSP1142 was docked with the typical folded conformation, which resembles the binding mode of the crystal structure (Figure 5.25). However, the long spacer probably forced the trityl to assume a different geometry compared with the experimental structure and the majority of the inhibitors analysed so far. One aromatic ring made π-stacking and hydrophobic interactions with Phe 46 and Ile 117 respectively. A second phenyl ring formed interaction with chain of the Ser 95, while the third ring seemed to make no interactions with the protein. Although there was a different orientation of the trityl group, the compound WSP1142 shows a good activity for the *P. falciparum* dUTPase (2.0 μM). From the docking calculation an additional binding mode was found with higher score energy (8 kJ/mol higher, see appendix 1). The compound adopted an unfolded conformation, with the uracil group still in the bottom of the active site and the spacer lying on the Tyr 112. One ring of the trityl makes a T-shape interaction with the Tyr 112, the other rings did not form any interactions with the protein. The oxygen, which links the spacer

with the trityl, is at hydrogen bond distance from the oxygen of the Asp 109 carboxylate group (3.3 Å).

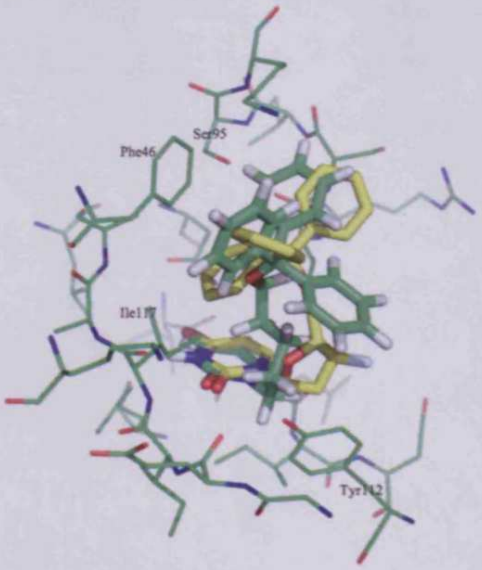


Figure 5.25 Docking result of the compound WSP1142 (green) superimposed with WSP869 obtained crystallographically (yellow).

WSP1208 did not show any activity (Table 5.1), which may suggest the importance of the trityl binding site in the drug design process. Also, it indicates that the length of the spacer does not influence the activity for this class of compounds without the trityl (see appendix 1). However, FlexX docked the uracil group in a conformation similar to the crystal structure (Figure 5.26).

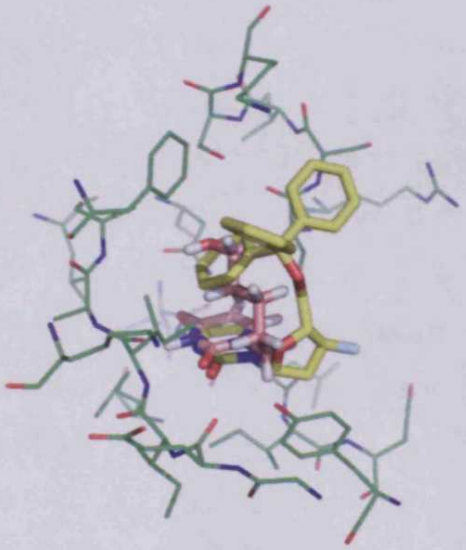


Figure 5.26 Docking result of the compound WSP1208 (pink) superimposed with WSP869 obtained crystallographically (yellow).

The docking of the compound WSP1224 showed a conformation similar to the experimental structure with a folded geometry (Figure 5.27). The spacer formed hydrophobic interactions with Tyr 112 and Ile 108. The trityl group interacted with the Phe 46 by two rings, forming face-to-face and T-shaped (edged-to-face) stacking interactions. The third ring might form hydrophobic interactions with the aliphatic side chain of the Lys 96 and additionally cation-π interaction with protonated nitrogen of the residue 96. The amino, which links the spacer and the trityl, seemed to make no interactions with the protein at all. From the docking calculations a second binding mode was found, which has higher score energy (10 kJ/mol higher). This additional conformation had an unfolded geometry. The spacer formed hydrophobic interactions with Tyr 112 and Ile 108, while the trityl might interact with the side chains of the Arg 91 and Leu 107. Additionally, the amino linker, is at hydrogen-bonding distance from Asp 109 (2.67 Å).

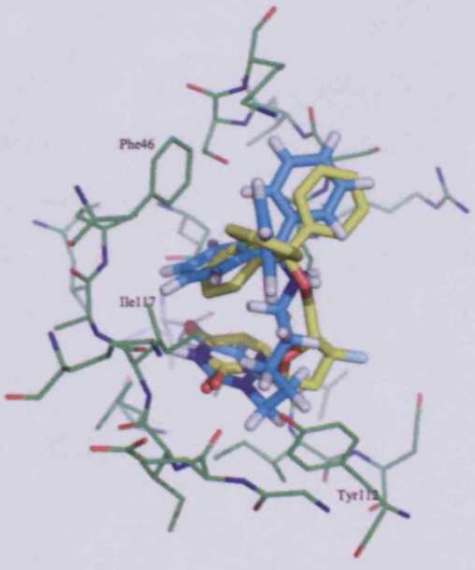


Figure 5.27 Docking result of the compound WSP1224 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

The inhibitor WSP1228 has the triphenyl silyl instead of the trityl. This difference seems make no substantial change in the activity (*Ki* 3.8 μM). The docking calculation result showed a folded conformation

similar to the experimental structure, giving the same network of interactions of the inhibitors with this series (Figure 5.28).

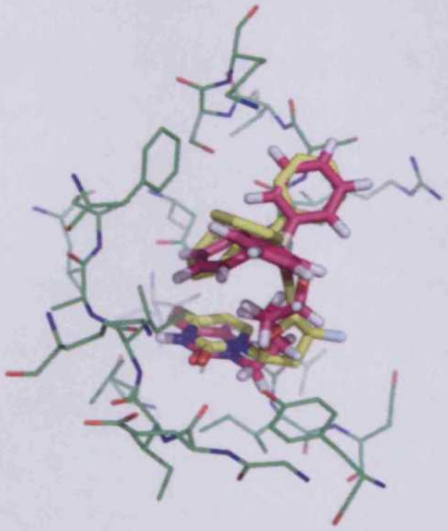
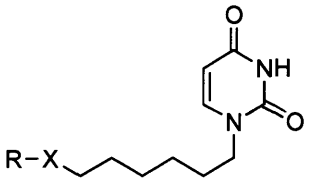


Figure 5.28 Docking result of the compound WSP1228 (magenta) superimposed with WSP869 obtained crystallographically (yellow).

Group 5. This series is composed from six inhibitors, which are characterised by an aliphatic spacer of six carbon atoms (Figure 5.29). The automated docking found as top-ranking conformation, for most of the ligand of the series, an unfolded geometry. However, the folded conformation, which resembles the crystal structure, was found as alternative binding mode with lower score energy. This docking result is in contrast with the calculation of the previous group, where the most populated cluster binding modes were in the folded conformation. This result might be due to the high flexibility of this series of inhibitors.



WSP Group 5

	Ki	score
1004 R = Ph_3C, X = O	2.3	-54.37
1241 R = Ph_3C, X = NH	1.84	-54.46
1067 R = Ph_3Si, X = O	1.69	-52.22
1005 R = tBuPh$_2$Si, X = O	8.47	-46.67
1062 R = H, X = O	1000	-38.88
1063 R = PhC(O), X = O	32	-51.20

Figure 5.29 Structures of compounds assayed of the group 1; Ki (μM), score (kJ/mol)

WSP1004 and WSP1005 were docked into the *P. falciparum* enzyme with "unfolded" geometry, (Figure 5.30A-30B). In both the structures, the uracil interacted with the bottom of the active site by the characteristic network of hydrogen bonds. The spacers made hydrophobic interaction with the Tyr 112. The trityl of the WSP1004 and the tert-butyl diphenylsilyl of the compound WSP1005 seemed to form no interactions with protein. The oxygen-linker, was at hydrogen-bonding distance to Asp 109 (2.54 Å). However, a lower ranking conformation with a geometry that resembled the crystal structure was found for WSP1004.

The oxygen linker of the compound WSP1005 did not make any interaction with the protein at all. A folded binding mode with slightly higher score energy (0.6 kJ/mol in WSP1004 and 0.8 kJ/mol WSP1005 higher) was found. Nevertheless, this small score energy difference might suggest that the folded conformation is potentially possible.

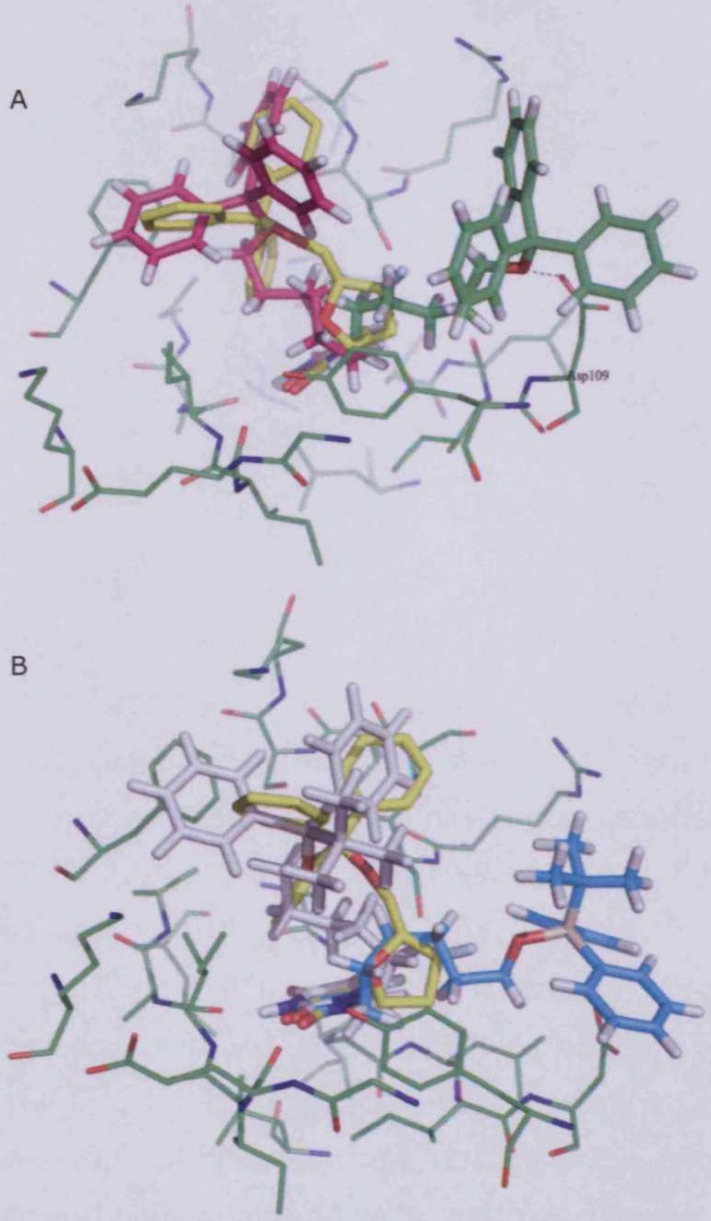


Figure 5.30. A) Docking result of the compound WSP1004 in folded (magenta) and unfolded (green) conformation, superimposed with WSP869 obtained crystallographically (yellow). B) Docking result of the compound WSP1005 in folded (white) and unfolded (cyan) conformation, superimposed with WSP869 obtained crystallographically (yellow).

FlexX was not able to find a folded conformation for the compound WSP1241. In fact, the automated docking result showed an unfolded geometry as the only possible binding mode. The nitrogen group, which links the trityl with the spacer, made hydrogen bond with the Asp 109, while the trityl interacted by hydrophobic interaction with side chains of Arg 91 and Leu 108. However, manual docking of

WSP1241 revealed that a folded conformation similar to the crystal structure is possible.

The compound WSP1062 has a hydroxyl group at one of the extremity of the spacer. The lack of the trityl group or of any bulky group capable of interacting with Phe 46, probably explains the inactivity of this compound (*Ki* > WSP1000 μM). The docking program found the uracil conformation similar the experimental structure. However it showed a low energy score (-38.88 kJ/mol), which is in good agreement with experimental data (see Figure 5.5).

WSP1063, which contains a benzoate substituent docked in the folded conformation (Figure 5.31). This benzoic group did not not seem form significant interaction with Phe 46, however phenyl ring close to the backbone of the Ser 92 (2.0 Å), possibly forming hydrophobic interaction with the side group of this residue.

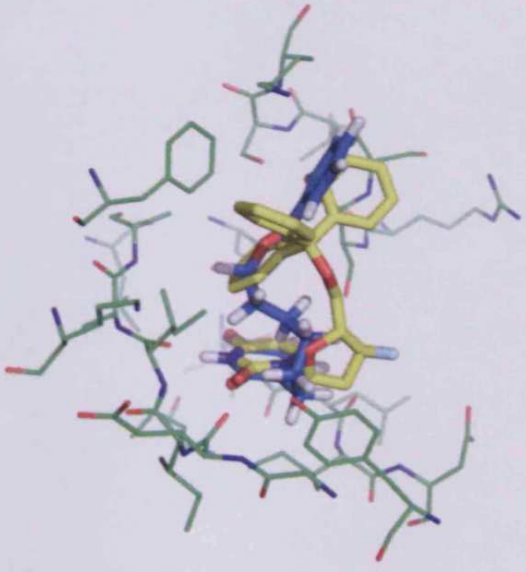


Figure 5.31 Docking result of the compound WSP1063 (blue) superimposed with WSP869 obtained crystallographically (yellow).

WSP1067 shows good inhibition and high selectivity compound to the *P. falciparum* enzyme (see appendix 1), which confirms the positive trend of this class of the inhibitors with a trityl moiety (Figure

5.32). The docking result showed a folded binding mode, which resembles the experimental structure of WSP869. One of the phenyl rings of the triphenyl silyl interacted by π-stacking with Phe 46; the second ring should interact with the side of Lys 96. The third ring seemed to make no interaction with protein at all.

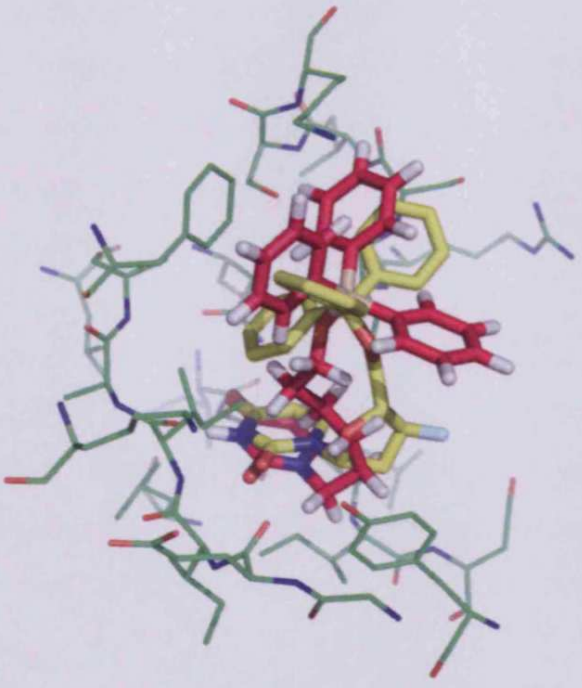
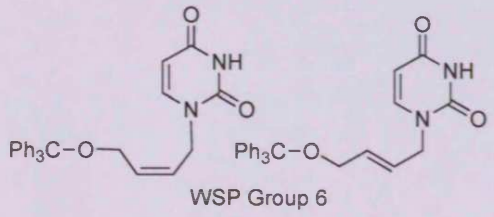


Figure 5.32 Docking result of the compound WSP1067 (magenta) superimposed with WSP869 obtained crystallographically (yellow).

Group 6. The compounds of this group are characterized by an unsaturation in the spacer (Figure 5.33). The double bond should gives, to the two isomers, more rigidity and improve the interaction between the spacer and the Tyr 112.



WSP Group 6

1315 Z		1315 E	
Ki	1.2	**Ki**	0.6
score	-51.26	**score**	-55.98

Figure 5.33 Structures of compounds assayed of the group 1; Ki (μM), score (kJ/mol)

The *Z* isomer (WSP1315) was docked in the *P. falciparum* enzyme in an "unfolded" conformation (Figure 5.34). The uracil group was in the bottom of the active site forming the typical network of hydrogen bond with the protein. The spacer interacted with on the Tyr 112 making stacking interaction, while the trityl moiety was not in contact with hydrophobic pocket that characterizes the *P. falciparum* enzyme[39]. It made contacts with the part of the active site that should normally interact with the phosphate chain of dUTP (see chapter 2). The trityl group formed hydrophobic interactions with side chains of the Arg 91 and Leu 107. The carbonyl group of the Leu 107 backbone forms a weak pseudo H-bond with the hydrogen of one of the trityl phenyl groups (2.2 Å). The docking calculation revealed a second binding mode, which had a folded conformation, similar to the crystal structure. This second binding mode had a 4 kJ/mol higher score energy of binding (see appendix 1). In the folded conformation one trityl ring made clearly stacking interaction with the Phe 46. The second trityl ring formed hydrophobic interaction with the side chain of the Ser 95, and the third ring seemed make no interaction with protein. The spacer might form hydrophobic interaction with the Tyr 112, but the unsaturation does not form any clear stacking interaction with the aromatic ring of the residue 112.

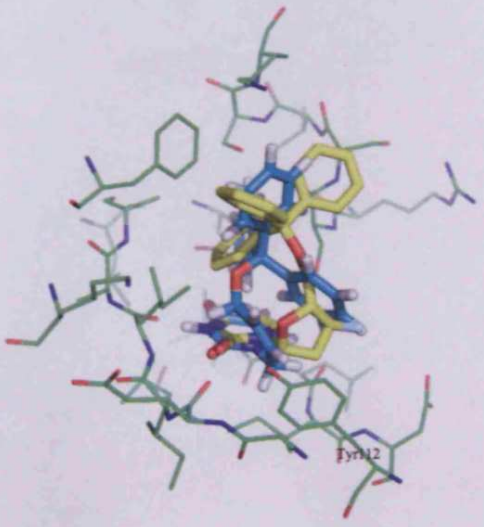


Figure 5.34 Docking result of the compound WSP1315 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

FlexX program was able to dock the *E* isomer (WSP1317) in a folded conformation, similar to the experimental structure (Figure 5.35). The unsaturated spacer interacted with the Tyr 112 side chain by forming stacking interaction. The trityl group was involved, with one aromatic ring, in a T-shape interaction with Phe 46. The other rings formed hydrophobic interactions with the side chains of the Ile 117 and Lys 96. The comparison between the inhibition potency of the compounds WSP1317 and the analogue WSP1000 (0.6 μM and 1.6 μM respectively), shows the marginal improvement of the activity for the ligand WSP1317. It may be due to the presence of the unsaturation in the spacer, which interacts with the Tyr 112. However there was essentially big no difference in activity between the E and Z isomers, although the modelling the predicted different binding conformation.

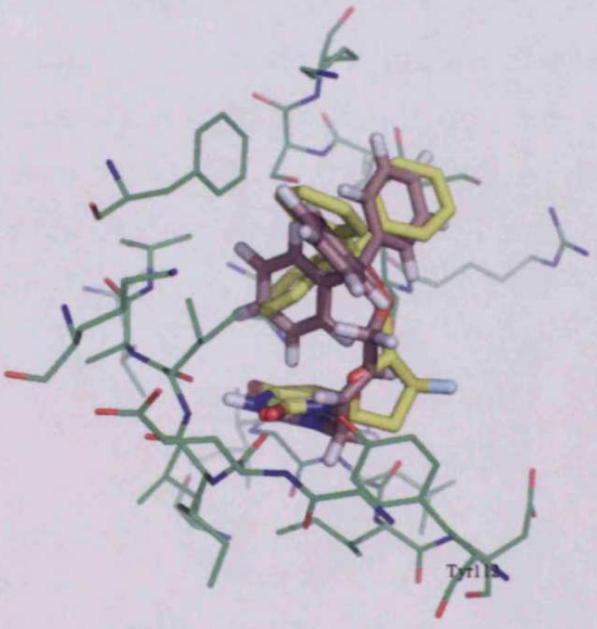
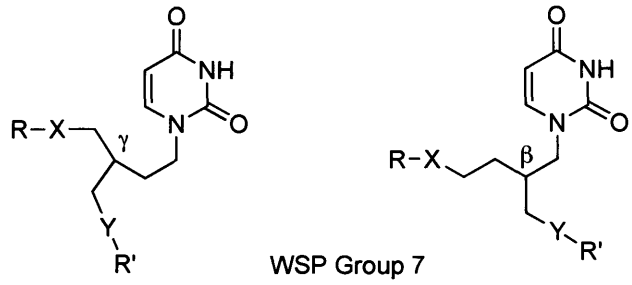


Figure 5.35 Docking result of the compound WSP1317 (dark-pink) superimposed with WSP869 obtained crystallographically (yellow).

Group 7. This group is characterized by six inhibitors (Figure 5.36, see also Table 5.2-3), which were divided in two series: γ-branched acyclic derivatives (analogues of the penciclovir) and β-branched

derivatives. There were prepared as the racemic mixtures, for the docking studies both the enantiomers were considered for each compound.



1318 R = Ph_3C, X = NH, R' = CH_3CO, Y = O
1320 R = Ph_3C, X = NH, R' = H, Y = O
1319 R = Ph_3C, X = O, R' = $CH_3O(CO)$, Y = NH
1328 R = Ph_3C, X = O, R' = H, Y = O
1321 R = H, X = O, R' = Ph_3C, Y = O
1322 R = Ph_3C, X = O, Y = N_3

Figure 5.36 Structures of compounds assayed of the group 1

WSP	Ki(μM)	Score(kJ/mol)
1318 *R*	1.33	-64.47
1318 *S*		-63.81
1319 *R*	2.48	-50.63
1319 *S*		-58.03
1320 *R*	0.24	-47.08
1320 *S*		None

Table 5.2 Experimental and score values of the γ-branched compounds

WSP	Ki(μM)	Score(kJ/mol)
1321 *R*	0.9	-57.45
1321 *S*		-56.97
1322 *R*	23.1	-65.74
1322 *S*		-60.05
1328 *R*	2.15	-60.17
1328 *S*		-57.95

Table 5.3 Experimental and score values of the β-branched compounds

WSP1318 *R* was docked in the *P. falciparum* active site and the result resembled the conformation of the crystal structure WSP869. The compound had a folded conformation with the trityl group placed in the hydrophobic pocket, interacting with Phe 46 by a T-shape interaction and with Ile 117, forming hydrophobic interaction (Figure 5.37A). The amino-linker seemed to make no interactions with the protein at all. The carbonyl group of the acetate formed one hydrogen

bond with the backbone amide of the Asp 109 (2.9 Å). The docking result showed an alternative binding mode, which had a 0.5 kJ/mol less favourable energy score. In this case, the acetate ester group is in contact with hydrophobic pocket on the top of the active site, while the trityl moiety is placed near the Asp 109. The methyl group of the acetate made a hydrophobic interaction with Ile 117, and the trityl moiety a hydrophobic interaction with side chain of the Leu 107.

The docking result of the compound WSP1318 *S* showed a conformation very similar to the top-racking binding mode of the enantiomer *R* (Figure 5.37B). The major difference between these two ligands is the position of the amino group (linker), which in the WSP1318 *S* seemed to be more expose to the solvent. However, it did not form any contact with the protein. Energetically, the stereoisomer *R* is slightly more favorable (WSP1318 *R* has -64.47 kJ/mol and WSP1318 *S* has -63.81 kJ/mol). It seems that both enantiomers might inhibit the enzyme. The docking is not conclusive enough to predict which compound will bind.

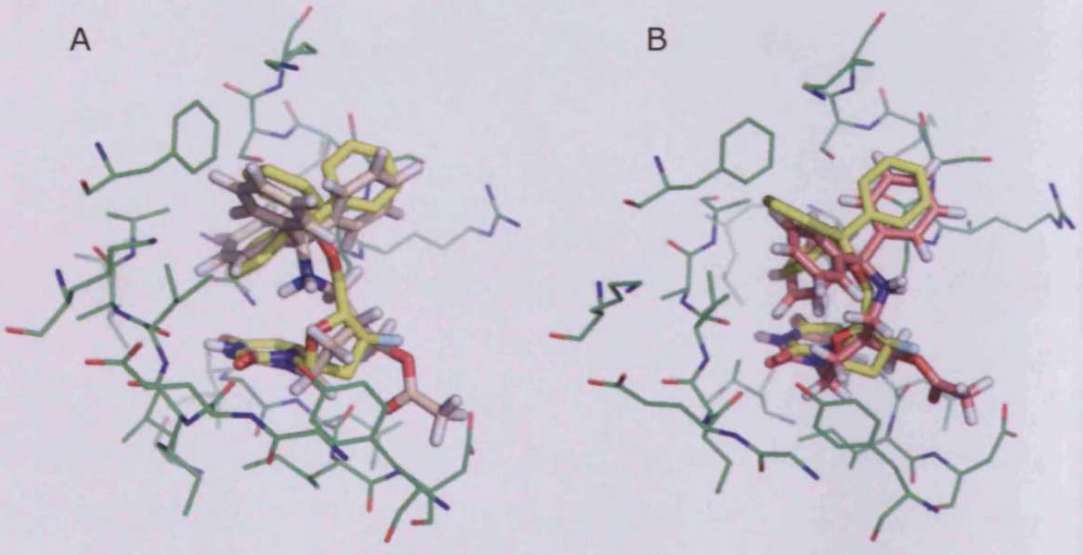


Figure 5.37 A) Docking result of the compound WSP1318 *S* (light-pink) superimposed with WSP869 obtained crystallographically (yellow). B) Docking result of the compound WSP1318 *R* (dark-pink) superimposed with WSP869 obtained crystallographically (yellow).

The automated docking of WSP1319 *S* showed only one possible conformation in the binding site. The ligand was found in a folded conformation as the crystal structure (Figure 5.38A). The carbonyl of the branch group made a hydrogen bond with the amide of the Asp 109 main chain (3.3 Å). Additionally, the hydrogen of the amide of the acetate group formed π-hydrogen bond with the Tyr 112 (3.0 Å).

FlexX program found an unfolded conformation for the compound 1319 *R* derivative (Figure 5.38A). This binding mode places the branch near the Asp 109, opposite the Tyr 112. A folded conformation might be sterically unfavorable because the acetate group would be too close to the Tyr 112, penetrating its van der Waals' radii. The unfolded conformation avoids the steric problems, but the acetate group and the trityl did not make any significant interaction with protein. It is possible that this alternative conformation is an artefact of the docking program.

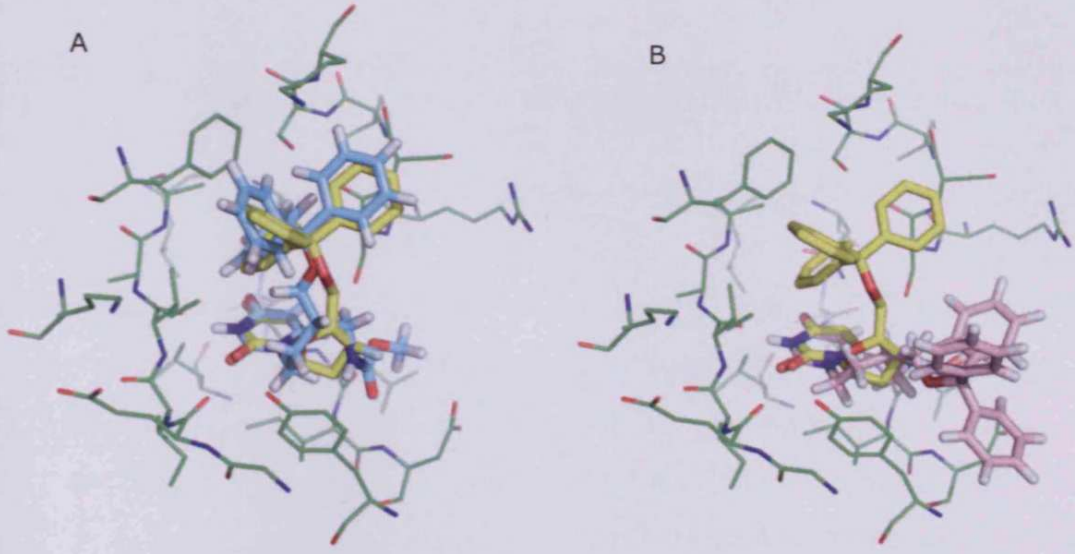


Figure 5.38 A) Docking result of the compound WSP1319 *S* (cyan) superimposed with WSP869 obtained crystallographically (yellow). B) Docking result of the compound WSP1319 *R* (light-magenta) superimposed with WSP869 obtained crystallographically (yellow).

FlexX was not able to dock the compound WSP1320 *S* derivative. However, a manual docking was performed on this enantiomer,

revealing that a binding mode similar to the crystal structure is possible.

Docking of WSP1320 *R* derivative showed an unfolded conformation (Figure 5.39) with a relatively high binding score energy (-47.08 kJ/mol), which is not consistent with the good inhibition potency (0.24 μM). This suggests that result may be an artefact of the docking program.

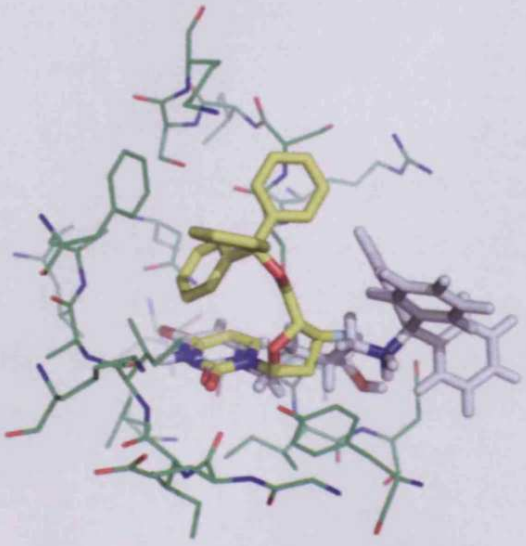


Figure 5.39 Docking result of the compound WSP1320 *R* (light-grey) superimposed with WSP869 obtained crystallographically (yellow).

Docking of WSP1321 *R* revealed a folded binding mode (Figure 5.40A). The trityl was in contact with the hydrophobic pocket, making T-shape interactions with the Phe 46 and, and additionally hydrophobic interaction with Ile 117. The hydroxyl group of the β-branch was hydrogen bond donator with carbonyl oxygen of the Leu 107 backbone (1.7 Å), and also, hydrogen bond acceptor with the amine group of the Asp 109 main chain (3.4 Å).

On the other hand, docking of the *S* enantiomer revealed an unfolded binding mode (Figure 5.40B). The trityl group was near the Asp 109, although it did not make any significant interaction with the protein. The hydroxyl group of the β-branch is a hydrogen bond donor, which formed a hydrogen bond with hydroxyl group of the Tyr 112 (2.3 Å). A "folded" conformation might be sterically unfavorable

because the β-branch would force its functional group to penetrate the van der Walls' radii of the Tyr 112. Comparing the binding score energies of the ligands, the *R* enantiomer (-57.45 kJ/mol) was slightly favourable over the *S* enantiomer (-56.97 kJ/mol).

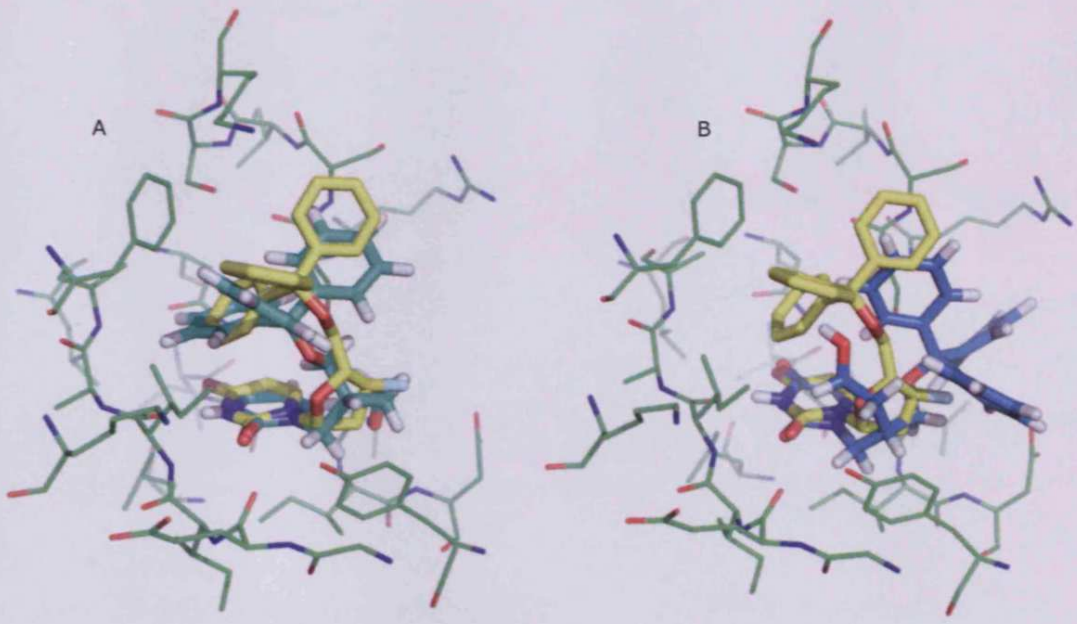


Figure 5.40 A) Docking result of the compound WSP1321 *R* (light-green) superimposed with WSP869 obtained crystallographically (yellow). B) Docking result of the compound WSP1321 *S* (blue) superimposed with WSP869 obtained crystallographically (yellow).

WSP1322 *R*, docked with an unfolded binding mode (Figure 5.41A); it might be sterically more favourable as described for previous molecules. The azide group seems to make no significant interaction with the protein.

The enantiomer WSP1322 *S*, was found in the same conformation as of the experimental structure WSP869 (Figure 5.41B), with the azide group of the β-branch forming an hydrogen bond with amide group of the Asp 109 backbone (2.8 Å). The high similarity of the docking result with the crystal structure could suggest that the enzyme had more affinity for *S* enantiomer than the *R* enantiomer. However the scoring energy of the WSP1322 *R* is lower than WSP1322 *S* (-65.74 kJ/mol and -60.05 kJ/mol respectively).

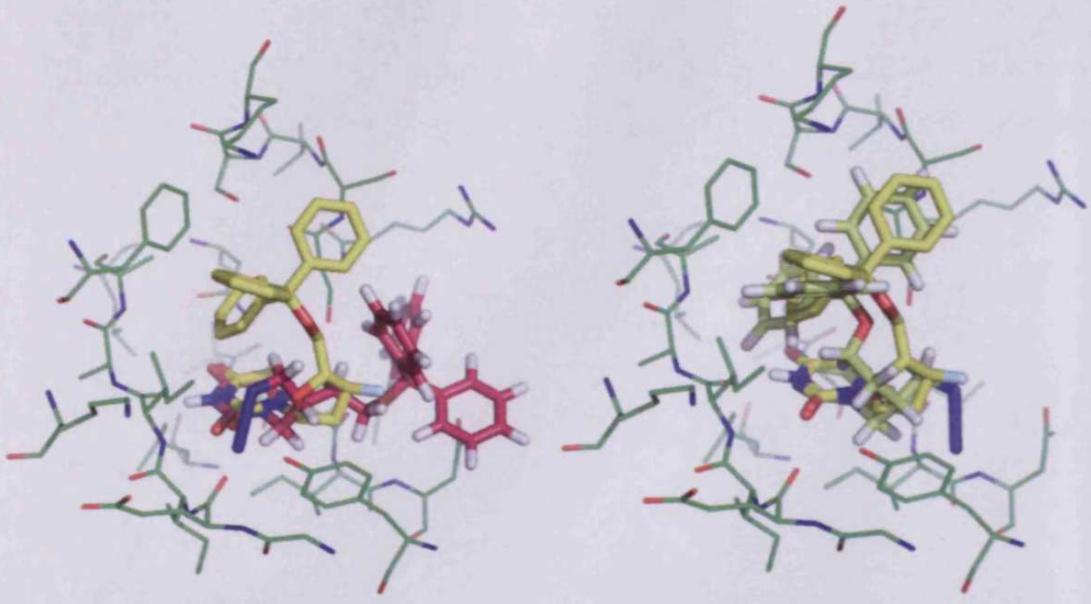


Figure 5.41 A) Docking result of the compound WSP1322 *R* (magenta) superimposed with WSP869 obtained crystallographically (yellow). B) Docking result of the compound WSP1322 *S* (lime) superimposed with WSP869 obtained crystallographically (yellow).

The automated docking for the enantiomers of the compound WSP1328 reproduce a result similar to *R* and *S* WSP1322. The WSP1328 R assumed an unfolded conformation (Figure 5.42A) while the binding mode of the enantiomer *S* highly resembled the pose of the crystal structure (Figure 5.42B). However, as in the previous docking result, the WSP1328 *R* has lower binding score energy than the WSP1328 *S* (-60.17 kJ/mol and -57.95 kJ/mol respectively). Hence, theoretically the R enantiomer forms more stable complex with protein.

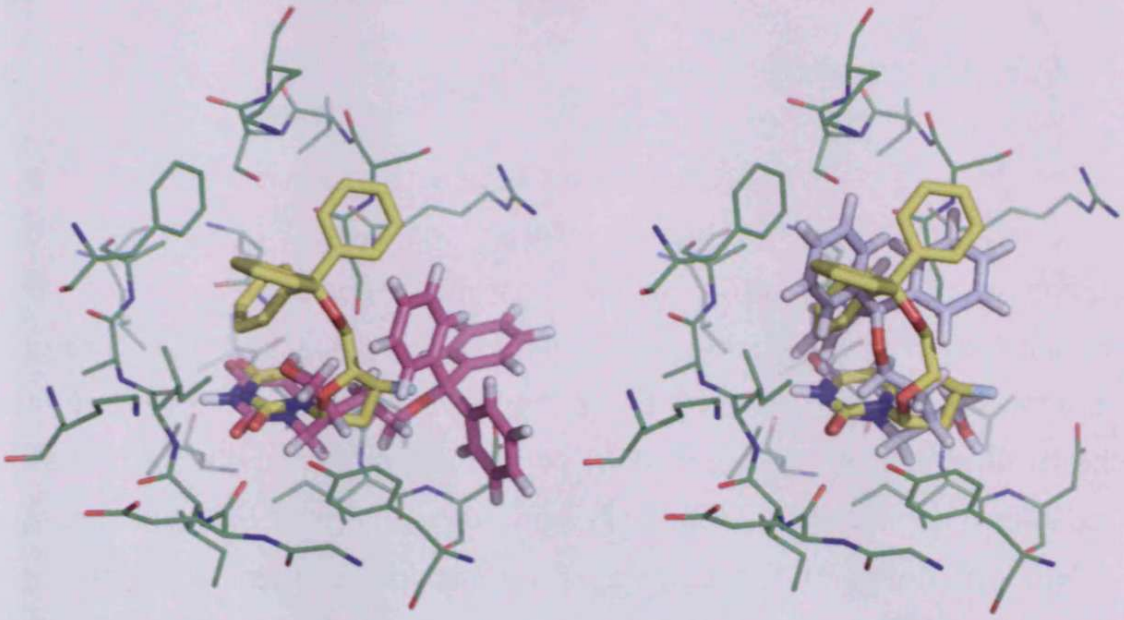


Figure 5.42 A) Docking result of the compound WSP1328 *R* (magenta) superimposed with WSP869 obtained crystallographically (yellow). B) Docking result of the compound WSP1328 *S* (light-grey) superimposed with WSP869 obtained crystallographically (yellow).

5.3 Discussion

The docking studies presented here provide a structural basis for a large range of dUTPase-ligand interactions, which might allow the design of next-generation compounds with enhanced activities. The consistency this model needs to be critically assessed. The docking results seem to be self-consistent, as the functional groups of similar chemical character are placed in similar ways for all docked ligands and show consistent interactions with the protein. Furthermore, the agreement between the calculated interaction free energies and the experimentally determined activities together with the reported SAR studies suggest that the binding modes of the examined ligands are reasonable.

5.3.1 Validation of the binding models on the basis of SAR.

The present study together with the recently published X-ray complex and the SAR made on the acyclic nucleotides derivatives, indicates no general correlation between the activity and the side-chain length. It was noticed that the presence of the amine group, which connects the trityl moiety and the aliphatic spacer, enhanced the activity of some of these compounds. From the crystal structure and the docking results it appears that no heteroatom (linker) in this position was involved in any direct interaction with the protein, especially when the ligand binds in a folded conformation as the X-ray complex.

The improvement of the inhibition for the compounds with an amine linker, compared with the compounds containing the oxygen linker, might be due to the presence of a water molecule which bridges the linker NH (5'- position) to the oxygen (hydrogen bond acceptor) of the Tyr 112 side chain by hydrogen bond.

A preliminary study[73] (Whittingham J. and Wilson K.S., personal communication) of the crystal structure of the *P. falciparum* complexed with the cyclic nucleotide derivative WSP871 (2.4 Å resolution) revealed the presence of this water molecule in proximity of the NH linker (Figure 5.43).

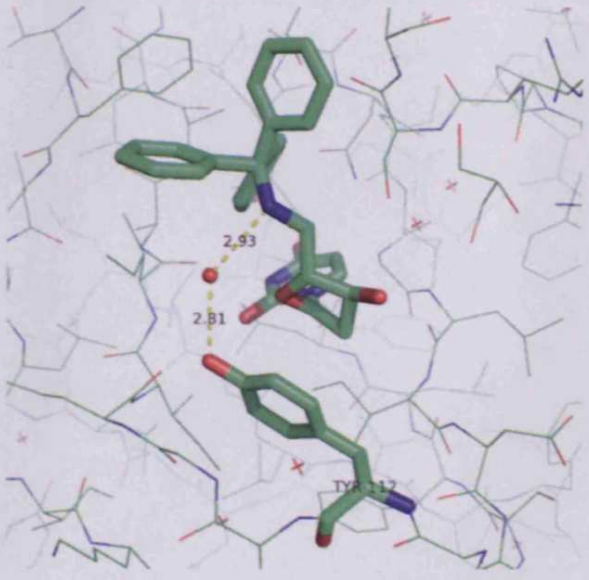


Figure 5.43 Crystal structure of the *P. falciparum* dUTPase complexed with the ligand WSP871 represented as green sticks. The Tyr 112 residue is also represented as green sticks. A molecule of water (red), forms a hydrogen bond with the OH of the Tyr and the secondary ammine of the ligand WSP871.

Further analysis of the local pKa of the NH group for the compounds WSP871 and WSP1060, using ACD/Labs[74], showed approximate values of *pKa* 6.72 and 7.69 respectively. Under physiological conditions these values would allow protonation of the NH linker group. A protonated NH group would produce conditions which favour solvation and hence the presence of the structured water molecule, this is not so for the oxygen linker group which is not protonated. However these are only speculations as the X-ray complex obtained in the preliminary study should be further investigated.

The inhibitor WSP1230, had a spacer of three carbon atoms, a trityl moiety and a NH group as linker. It was the most active against *P. falciparum* dUTPase (0.2 μM), although it was not very selective (human *Ki* 1.35 μM), in contrast with the other analogues compounds with longer spacer (appendix 1). This low selectivity might be explained by analysing the crystal structure of the human dUTPase and the docking results. FlexX program revealed two possible binding modes for WSP1230 in the human enzyme:

i. The result with the lowest scoring energy (better free energy of binding, -43.06 KJ/mol) showed the ligand WSP1230 in an unfolded conformation (Figure 5.44).

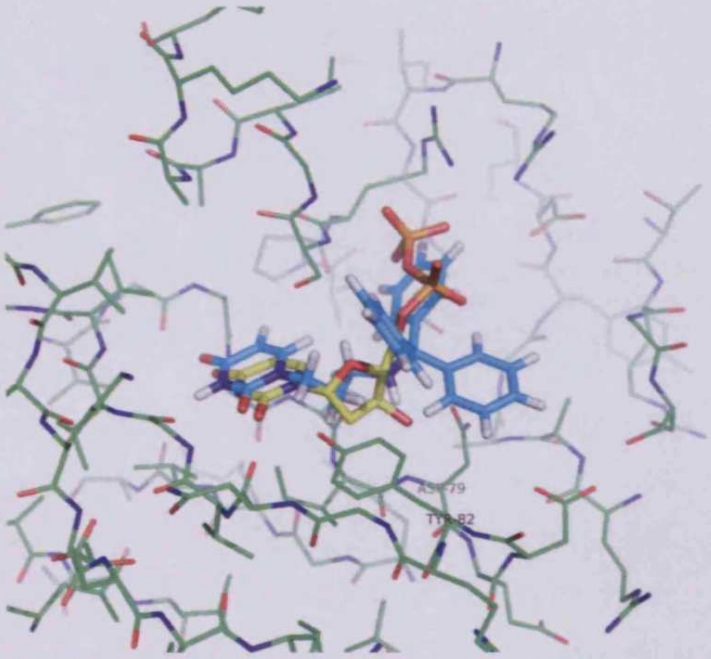


Figure 5.44 Docking result of the compound WSP1230 (cyan) in the human dUTPase superimposed with dUDP obtained crystallographically (yellow).

This is not surprising, since the human enzyme does not contain a hydrophobic pocket as described in the *P. falciparum* enzyme[39], which interacts with the trityl moiety. This binding mode showed the uracil interacting with the bottom of the active site and the spacer makes Van der Waals interaction with the conserved tyrosine (Tyr 82 in human enzyme). The trityl moiety seemed to make no significant

interactions with the protein, and the NH linker made a hydrogen bond with the conserved aspartate residue (Asp 79 in human enzyme, 1.5 Å). On the other hand, the compound WSP1060 (Figure 5.45), which has a spacer of four carbons, it is not active at all against the human, it might be simply because the spacer is too long, which makes impossible to accommodate the ligand in the catalytic pocket. However, FlexX program docked the ligand WSP1060 in the same manner of the WSP1230 (Figure 5.45), hence the docking results might suggest a certain ambiguity for this unfolded conformation.

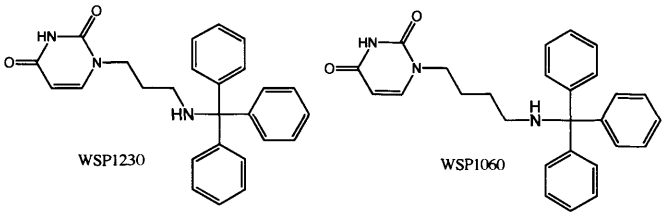


Figure 5.45 Schematic representation of the compounds WSP1230 and WSP1060

ii. The second binding mode has a higher scoring energy (6 KJ/mol higher) compared with the top-ranking score. The docking result revealed the possibility that the NH group of the WSP1230 interacts directly with the oxygen of the Tyr 82 in the human dUTPase, forming a hydrogen bond (Figure 5.46). This additional interaction might be possible because the spacer is sufficiently short to allow the NH group being at hydrogen-bonding distance from the Tyr 82 side chain (1.8 Å). The same type of interaction is unlikely in the *P. falciparum* dUTPase. In fact, it is sterically unfavourable due to the presence of Ile 117 which would undergo a steric clash with the trityl moiety; in the human enzyme is Ile 117 replaced by a smaller glycine residue. The inactivity of the compound WSP1060 against the human dUTPase might be due to the longer spacer, which cannot allow the NH group to reach the Tyr 82 side chain. In fact, for all the ligands

with more than three carbon atoms in the spacer, in the human enzyme FlexX found only unfolded conformation.

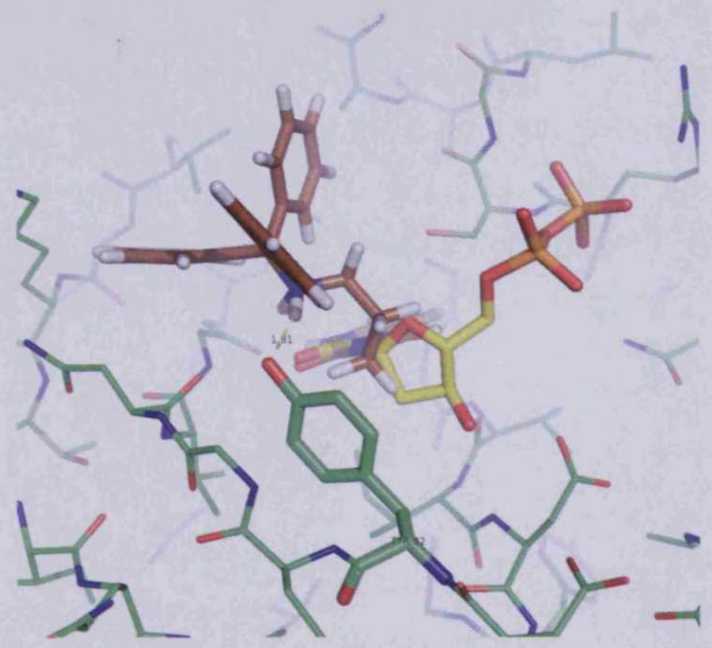


Figure 5.46 Docking result of the compound WSP1230 (cyan) in the human dUTPase superimposed with dUDP obtained crystallographically (yellow). This alternative binding mode showing possible interaction between the amino group and Tyr 82.
The secondary amine of the ligand WSP1230 forms a hydrogen bond with the OH of the Tyr 82 represented in stick.

In a similar way it might be explain the low inhibition of the compound WSP961 on the human dUTPase (Figure 5.47). The interaction between the oxygen linker and the hydroxyl of the Tyr 82 side chain is virtually impossible, because the OH group donates hydrogen bond to the protein backbone[34], therefore it can be only hydrogen-bond acceptor.

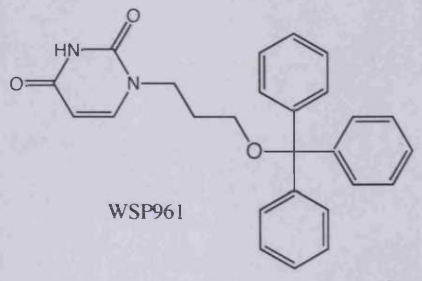


Figure 5.47 Schematic representation of the compound WSP961.

5.3.2 Triphenyl silyl derivatives

The triphenyl silyl derivatives were designed in order to take advantage of the fact that the silyl is larger than carbon atom, which might give more flexibility to the triphenyl silyl compared with the trityl group, hence providing more capability of forming interactions with the protein active site. However, the compounds with the triphenyl silyl (WSP1228, WSP1072, WSP1067 and WSP1001) showed a similar range of activity compared with the trityl derivatives (Figure 5.5, see also appendix 1). The docking results are consistent with this result. They revealed a conformation similar to the crystal structure, without any additional interaction with the active site.

5.3.3 Unsaturated compounds

Both WSP1315 and WSP1317 (Figure 5.33) showed reasonable inhibition of the *P. falciparum* dUTPase (*Ki* 1.2 μM and 0.6 μM respectively). However, only WSP1317 (*E* conformation) showed a significant selectivity for the parasite enzyme (WSP1315 *Ki* 5.16 μM for the human enzyme; WSP1317 *Ki* 1000 μM for the human enzyme). This difference in selectivity maybe a result of an alternate conformation adopted by WSP1315 in the human enzyme.

The analysis of the docking result of the human dUTPase showed an unfolded conformation for both the inhibitors, similar to the co-crystal structure of dUDP with the human enzyme. However, the top-ranking scoring energies were substantially different; for WSP1315 it was -56.17 KJ/mol and for WSP1317 it was -48.14 KJ/mol, which partially reflected the biological results.

In order to better understand the most energetically preferred conformations of the compounds WSP1315 and WSP1317, a conformational study was carried out, using a simulated annealing

technique. The calculations revealed that the lowest energy (42.30 Kcal/mol) conformation of WSP1315 in vacuum assumed a folded conformation similar to that one seen in the *P. falciparum* crystal structure of WSP869. However, an unfolded conformation, which resembles the geometry of the docking result in the human enzyme, had an energy only 0.5 Kcal/mol higher than the folded one. WSP1317 had the lowest energy conformation (38.54 Kcal/mol), with a folded conformation; the unfolded conformation was energetically 10 Kcal/mol higher than the folded one.

Although there is not experimental evidence, the docking study and the conformational analysis suggest that the loss of selectivity of the compound WSP1315 might be due to the fact that this inhibitor is able to assume an unfolded conformation in the human enzyme; this conformation is energetically accessible. On the other hand, an unfolded conformation seemed to be less favourable for the compound WSP1317, hence the binding interaction of WSP1317 is much weaker with the human dUTPase.

5.4 Conclusion

The binding mode of the acyclic series of novel nucleotide derivatives was explored using molecular docking software. A shallow pocket near known active pocket was found (pocket S3, see chapter 4). This new pocket has been partially explored by the compounds with a branched linker (Group 7). Some of these ligands showed a promising activity. In addition, a better understanding of the selectivity has been achieved by the docking analysis and a conformational search of compounds in this series.

Finally, this docking study offers certain guidelines for the design of high-affinity nucleotide derivative inhibitors that possibly present

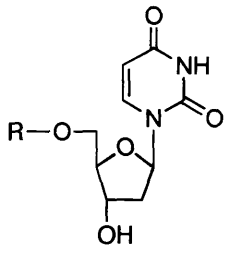
more drug-like characteristics. In fact, the docking analysis suggested that a three phenyl groups, as the trityl might not be necessary for the inhibition of the *P. falciparum* dUTPase. Two phenyl groups could be sufficient to keep a high binding affinity with the *P. falciparum* enzyme. This change will decrease the molecular weight, possibly increases the stability and solubility of these compounds.

6 Chapter: Docking study of the cyclic nucleotide derivatives

This series of cyclic nucleotide derivatives was divided in five groups. These ligands differ from the acyclic compounds due to the presence of the deoxyribose moiety; this series of cyclic compounds are structurally more related to the natural substrate (dUTP) than the acyclic groups. The docking calculations were performed using two different setups of the *P. falciparum* enzyme: (i) all the water molecules were removed; (ii) 5 conserved water molecules were included in the active site. With both the protein setting the docking calculations showed the similar results for all the compounds, except for the ligands of group 3. This aspect will be explored later in the discussion.

6.1 Docking results and discussion

Group 1. This group is characterised by a variety of substituents placed in 5'-position, while the rest of the structure is characterised by deoxyribose, which was used as scaffold (Figure 6.1). There were 20 compounds in this group.



WSP Group 1

487 R = $(PhO)_2P(O)$
659 R = $(EtO)_2P(O)$
784 R = $(EtO)_2P(O)CH_2C(O)$
662 R = $CH_3C(O)$
669 R = $CH_3(CH_2)_3C(O)$
660 R = PhC(O)
666 R = Ts
790 R = H_2NSO_2
789 R = $H_2NSO_2N(-)C(O)$
483 R = Ph_3C
951 R = TIPS
950 R = TBDPS
949 R = TPS
1059 R = (*para*-$CH_3O)Ph_3C$
1243 R = (*para*-$CH_3O)_2Ph_3C$
1244 R = (*ortho*-Cl)Ph_3C
1245 R = Dibenzosuberil
1247 R = Pixyl
1288 R = (*para*-NC)Ph_3C
1329 R = Trybenzylsilyoxy

Figure 6.1. Structures of compounds assayed of the group 1

FlexX program did not find any solution for WSP487 and WSP951. However, the other ligands were successfully docked in the active site of the *P. falciparum* enzyme. A careful inspection of the binding pocket indicated that the nucleosides derivatives docked in the same position as found in the co-crystal structure of *P. falciparum* with WSP869. All the molecules had the uracil group in the bottom of the active site. The sugar moiety is in van der Waals contact with the conserved residue Tyr 112, while the hydroxyl group in 3'-position virtually seems make no interaction with protein. Normally, the 3'-position of the dUTP is involved in a hydrogen bond with an aspartate residue (Asp 109). The substituents on the 5'-position were placed in

the hydrophobic pocket[39], which is mainly formed by the residues Phe 46 and Ile 117.

Eight compounds of the series showed a significant activity against the *P. falciparum* dUTPase (Table 6.1). Generally, the enzyme inhibition seems to be related to the presence, at the 5'-position, of a substituent containing two or three phenyl groups.

WSP	*Ki* (μM)
483	1.8
949	2.8
950	4.2
1059	5.2
1243	2.17
1244	2.56
1247	3.11
1288	1.14

Table 6.1. Values of the inhibition constant (*Ki*) of the most active compounds of the group 1

The automated docking placed WSP483 in the same location as the co-crystal structure with WSP869 (Figure 6.2). Two of the trityl rings made a strong π-stacking and T-shape interactions with Phe 46. The third ring probably interacted with a protonated amino group of the Lys 96 forming cation-π interaction[54]. This interaction might contribute to the high affinity of ligands containing trityl moiety in 5'-position with the protein. The oxygen of the trityl ether and the hydroxyl group in 3'-position seemed make no interactions with the protein; however they are at hydrogen bonding distance from each other (2.7 Å) possibly forming an intramolecular hydrogen bond.

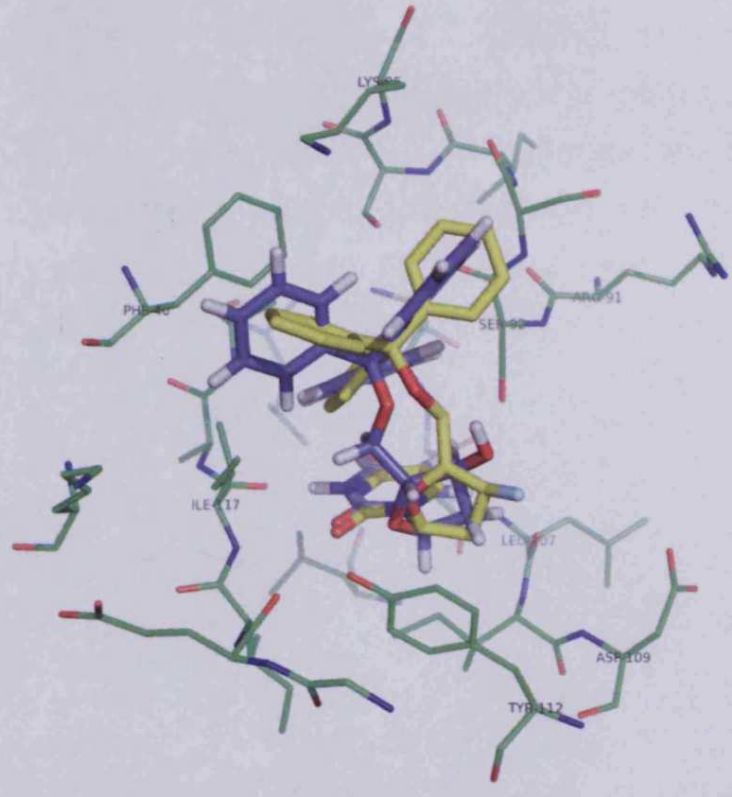


Figure 6.2 Docking result of the compound WSP483 (violet) superimposed with WSP869 obtained crystallographically (yellow).

The docking calculation revealed the possibility of an alternative binding mode, where the trityl group was not in contact with the hydrophobic pocket, but it was more exposed to the solvent and interacted with a more hydrophilic pocket of the enzyme, which is formed mainly from the residues of Ser 92, Arg 91, Leu 107, Asp 109 (see also chapter 4). Normally, these residues should be involved in general electrostatic interactions with the phosphate chain of the natural substrate dUTP[33]. In this second additional conformation, the oxygen in 5'-position might form a hydrogen bond with the Ser 92 side chain (2.9 Å). However, it is unlikely that in this second putative binding mode, the trityl moiety undergoes favourable interactions with the binding site.

WSP949 docked in a very similar conformation as the experimental structure of WSP483 and of the calculated structure of WSP483 (Figure 6.3). The main difference in respect to the binding mode of

the experimental structure was the position of the sugar moiety and in the orientation of the substituent in 3'-position. The displacement of the deoxyribose might be due to the hydroxyl interacting with the protein (Asp 109). However, the activity of this compound (*Ki* 2.8 μM) compared with the rest of the inhibitors (in particular WSP483 with a *Ki* of 1.8 μM) might indicate no additional interactions.

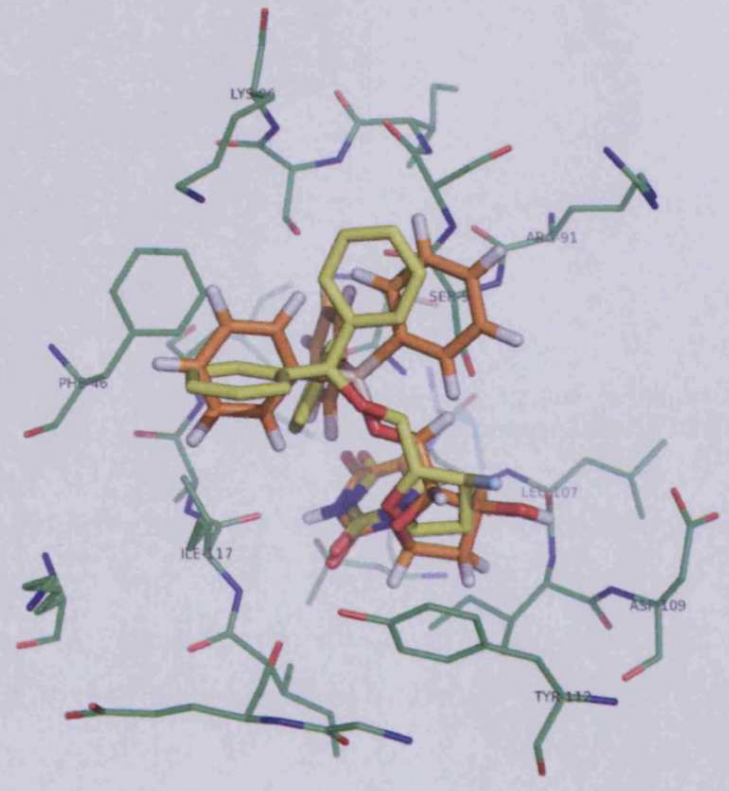


Figure 6.3 Docking result of the compound WSP949 (orange) superimposed with WSP869 obtained crystallographically (yellow).

The automated docking of the compound WSP950 found a binding mode that resembled the X-ray structure (Figure 6.4). The ligand had in 5'-position a *tert*-butyldiphenylsilyl, which was involved in a T-shape interaction with the Phe46. The *tert*-butyl group seemed to form hydrophobic interaction with the Phe 46. It may surprise that despite the absence of the usual trityl moiety, this compound still retains activity against the enzyme (4.2 μM). This suggests that all three phenyl rings in 5'-position might be not necessary to maintain

relatively good inhibition potency, and only two phenyl rings might be sufficient to retain the activity.

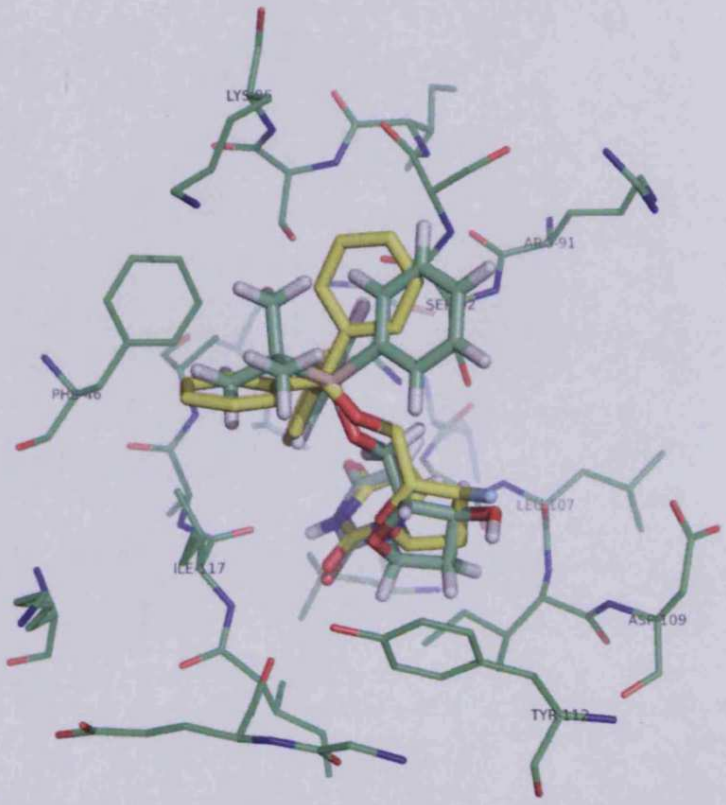


Figure 6.4. Docking result of the compound WSP950 (light-green) superimposed with WSP869 obtained crystallographically (yellow).

WSP1059 (Figure 6.5) and WSP1243 (Figure 6.6), were docked in a conformation similar to the crystal structure of WSP869. It seemed that the substituents on the phenyl rings of the trityl do not significantly affect the inhibitory activity (WSP1059 has a *Ki* of 5.9 μM and WSP1243 a *Ki* of 2.17 μM). In fact, they do not make any additional key interaction with the protein. It seemed likely that the rotation of the trityl ether on its axis allows the methyl ethers to be accommodated in a position, more exposed to the solvent and away from the core of the active site.

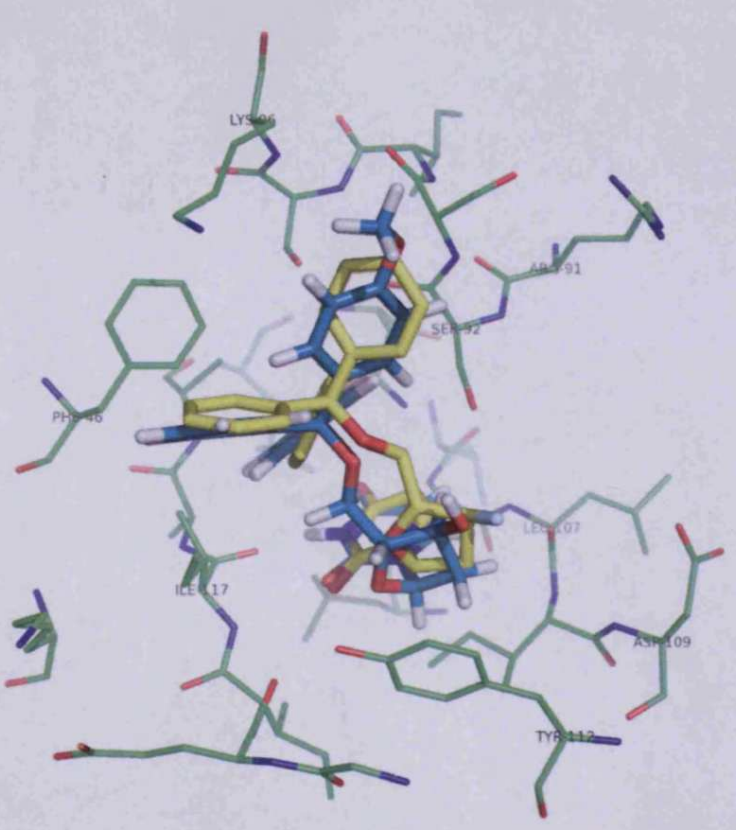


Figure 6.5 Docking result of the compound WSP1059 (blue) superimposed with WSP869 obtained crystallographically (yellow).

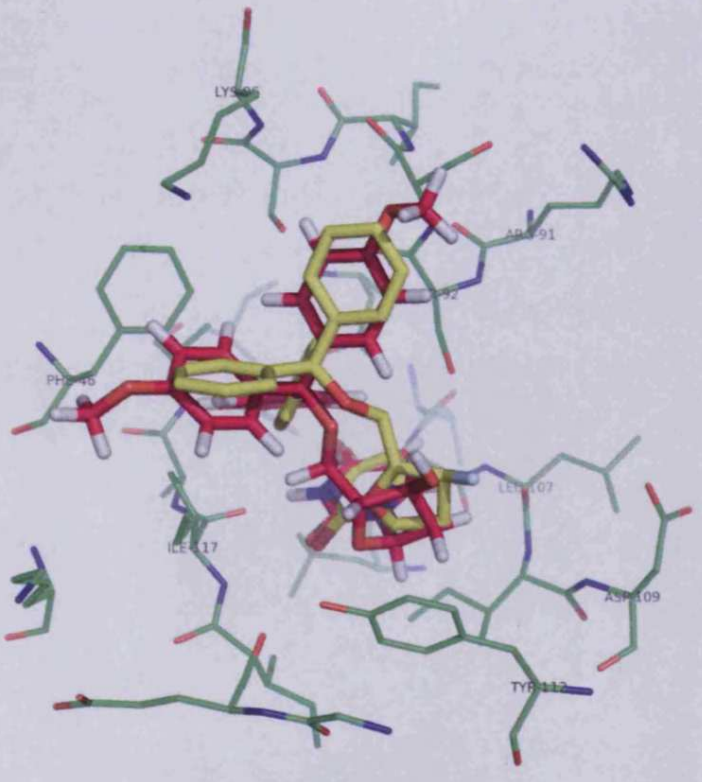


Figure 6.6. Docking result of the compound WSP1243 (magenta) superimposed with WSP869 obtained crystallographically (yellow).

Docking of WSP1244 and WSP1288 revealed a binding mode similar to the experimental structure. The substituent on the trityl moiety seemed make no interaction with the protein.

FlexX was able to dock the inhibitor WSP1247 in the same geometry of the crystal structure (Figure 6.7). This compound had in 5'-position a variation of the trityl moiety, in order to give more rigidity to the structure. However, the increased rigidity did not significantly affect the activity against the enzyme (*Ki* 3.11 μM).

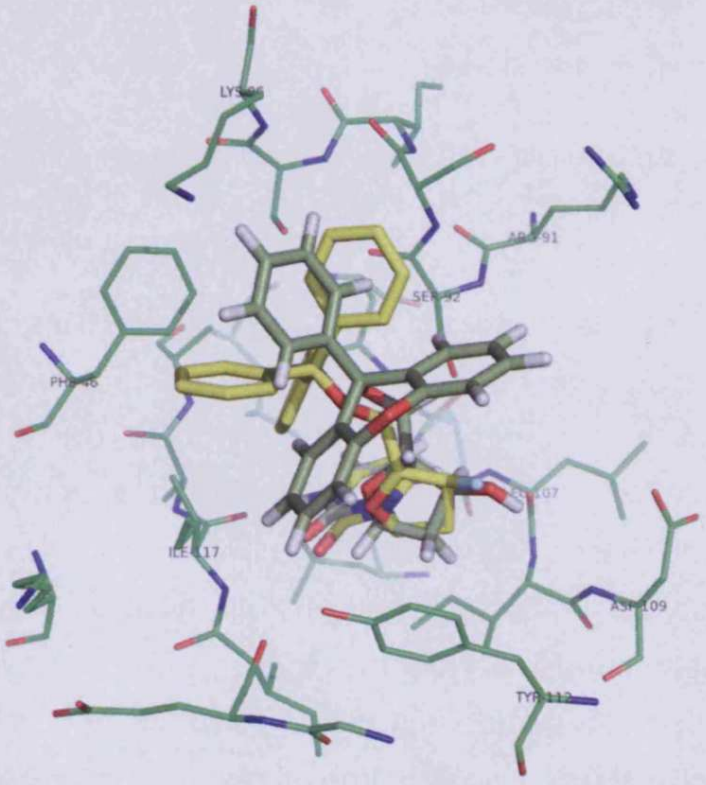
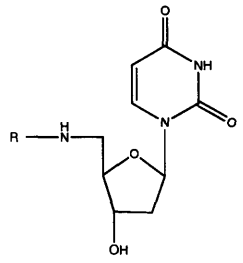


Figure 6.7 Docking result of the compound WSP1247 (dark-grey) superimposed with WSP869 obtained crystallographically (yellow).

Group 2. This second series is characterised by the presence of the amino group in 5'-position (Figure 6.8).



WSP Group 2

665 R = $(PhO)_2P(O)$
785 R = $(EtO)_2P(O)CH_2C(O)$
958 R = AdamantylC(O)
663 R = $CH_3C(O)$
668 R = $CH_3(CH_2)_3C(O)$
661 R = PhC(O)
667 R = Ts
680 R = CH_3SO_2
952 R = Ph_3CH_2
871 R = Ph_3C
804 R = H
1068 R = (*para*-CH_3O)Ph_3C
1149 R = 4-$PyPh_2C$
1289 R = (*para*-NC)Ph_3C
1290 R = (*ortho*-Cl)Ph_3C
1291 R = Pixyl
1292 R = Dibenzosuberil

Figure 6.8 Structures of compounds assayed of the group 2

It has been observed that the trityl amino group could enhance the activity of some of these inhibitors. The series is formed by seventeen compounds and for all of them FlexX program was able to reproduce a binding mode which resembled the X-ray structure. However, only six ligands showed significant activity against the *P. falciparum* dUTPase (Table 6.2)

The active compounds all had two or three aromatic rings at 5'-position. Compounds lacking the trityl group or analogues showed poor inhibition. This underlies the importance of this interaction for potent inhibition. Thus whilst some of the other structures lacking the trityl moiety, could be docked in the uracil and ribose binding pocket, exhibited weak theoretical binding score, which relate to the weak binding energy. These binding conformations of the compounds lacking of trityl moiety are very unlikely to be physiological relevant.

WSP	Ki (μM)
871	0.2
1068	0.72
1149	0.23
1289	0.4
1290	3.36
1292	4.06

Table 6.2 Values of the inhibition constant (*Ki*) of the most active compounds of the group 1

WSP871 is the most active of the cyclic nucleoside derivatives (0.2 μM). The docking result showed a good stacking interaction (face to face) and T-shape interactions between the trityl and the Phe 46 (Figure 6.9). The hydroxyl group in 3'-position and the amino group in 5'-position virtually seemed make no interactions with the protein. SAR analysis suggested that the presence of the trityl amino group might affect positively the inhibition potency of this compound. Unfortunately the docking result did not reveal any additional key interaction between the amino group and the protein that might explain the increment of the activity against the enzyme. However, as previously discussed, the crystal structure of the *P.falciparum* dUTPase complexed with this inhibitor[73] (see chapter 5) with a resolution of 2.4 Å, showing the presence of a molecule of water which might bridge the ligand and the protein by forming an hydrogen bond with the amino group in 5'-position and the side chain of the Tyr 112. This additional interaction could explain the high activity of this inhibitor, and in general of the compounds with the trityl amino group.

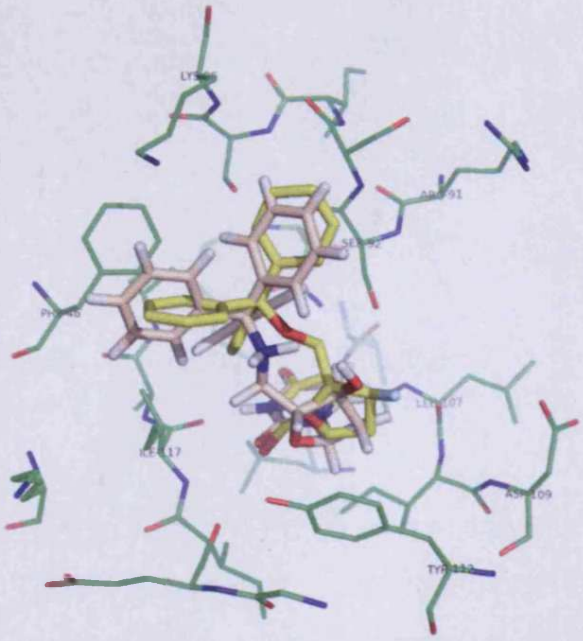


Figure 6.9 Docking result of the compound WSP871 (pink) superimposed with WSP869 obtained crystallographically (yellow).

The FlexX program showed similar results for WSP1068, WSP1149 and WSP1289 (Figures 6.10-11-12), which had a substituent on one phenyl ring of the trityl in the *para* position. WSP1289 had a pyridine ring in the trityl moiety (see also appendix 1). These compounds all have similar activities. The variations on the trityl seemed to make little affect on the activities.

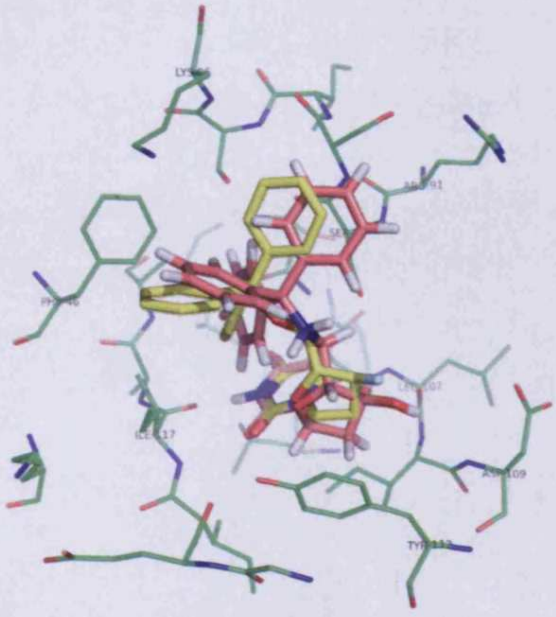


Figure 6.10 Docking result of the compound WSP1068 (dark-pink) superimposed with WSP869 obtained crystallographically (yellow).

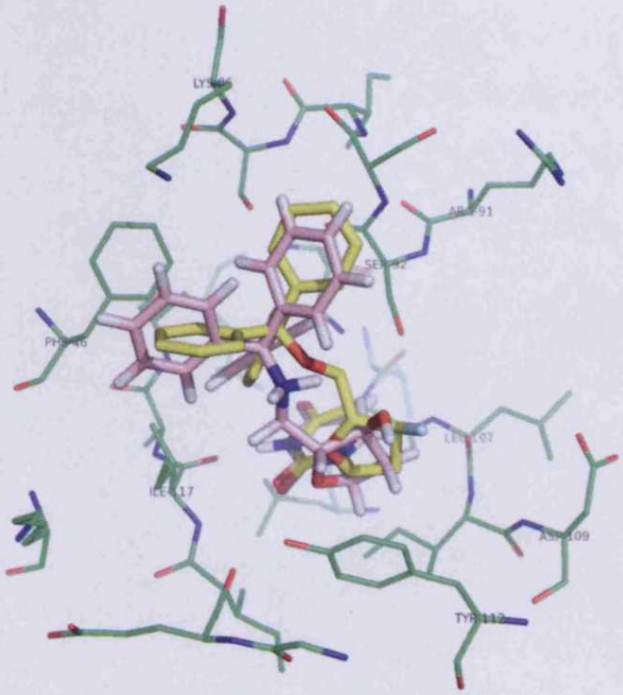


Figure 6.11 Docking result of the compound WSP1149 (pink) superimposed with WSP869 obtained crystallographically (yellow).

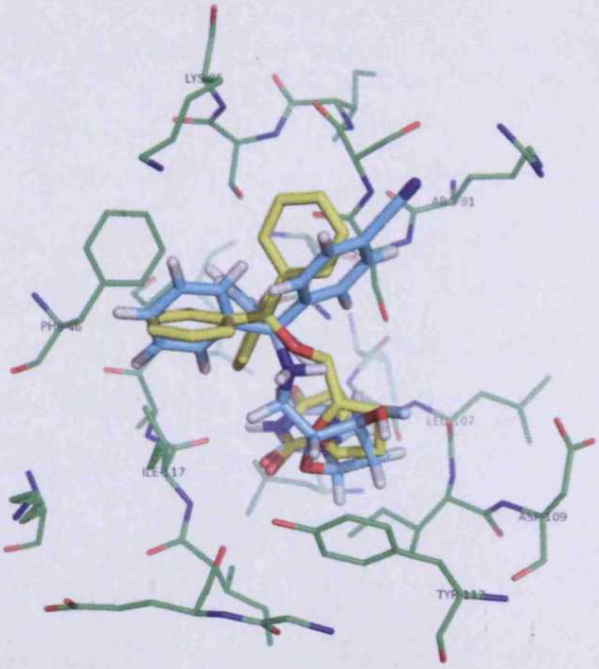


Figure 6.12 Docking result of the compound WSP1289 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

WSP1290, despite the presence of the trityl amino moiety in 5'-position, presented a diminution of the activity compared to the previous inhibitors of the series (3.36 μM). However, the automated docking was able to dock this compound in the same location of the crystal structure (Figure 6.13).

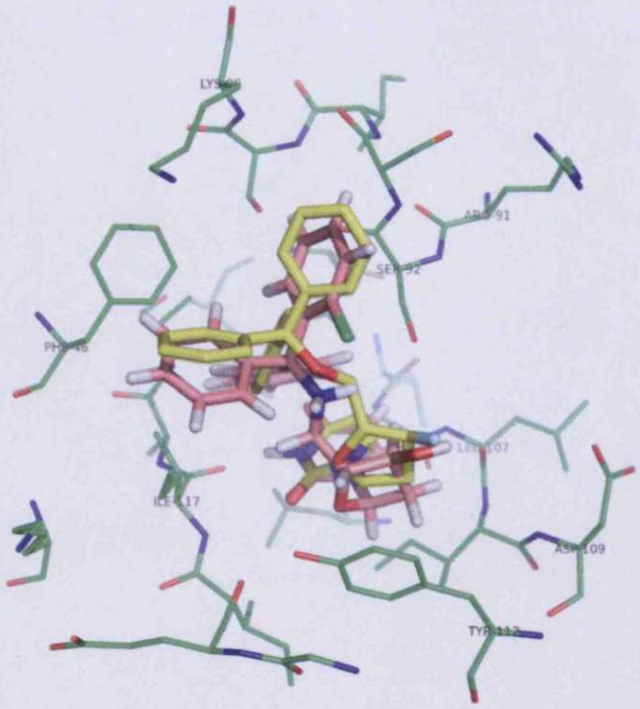


Figure 6.13 Docking result of the compound WSP1290 (pink) superimposed with WSP869 obtained crystallographically (yellow).

WSP1292 is the only compound of this group which showed a relatively good affinity for the enzyme (4 μM), although it lacks a trityl moiety the in 5'-position. However, it has a dibenzosuberyl substituent, which has two rigid phenyl rings (appendix 1). FlexX predicted a binding mode similar to the experimental structure of WSP869 (Figure 6.14), with the 5'-position substituent in contact with the hydrophobic pocket. However, the phenyl rings seemed to make no a strong interaction with the Phe 46.

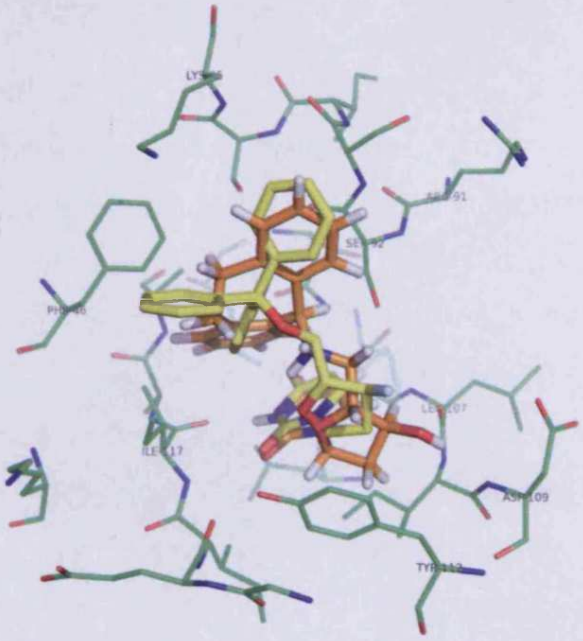
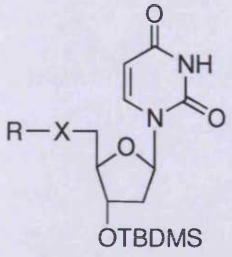


Figure 6.14 Docking result of the compound WSP1292 (orange) superimposed with WSP869 obtained crystallographically (yellow).

Group 3. This group contained ten compounds (Figure 6.15); each ligand was characterized by the presence of a bulky and hydrophobic substituent (TBDMS=tert-butyl dimethyl silyl) in 3'-position. None of these compounds showed a significant activity against the *P. falciparum* enzyme, the most active of the series was 704, with an activity of 67 μM. FlexX was able to dock the compound, reproducing the geometry of the crystal structure of WSP869, when this was carried out in the absence of the water molecules. In contrast, when the protein structure with the conserved water molecules was used, none of the structures docked as the crystal structure of WSP869.



WSP Group 3

486 X= O, R = $(PhO)_2P(O)$
659 X= O, R = $(EtO)_2P(O)$
784 X= O, R = $(EtO)_2P(O)CH_2C(O)$
789 X= O, R = $H_2NSO_2NHC(O)$
790 X= O, R = H_2NSO_2
485 X= O, R = H
484 X= O, R = Ph_3C
953 X= O, R = Ph_3C
704 X= NH, R = $(PhO)_2P(O)$
933 X= O, R = TBDMS

Figure 6.15 Structures of compounds assayed of the group 3

WSP484 (*Ki* 515 μM) was docked in the enzyme with the water removed, the calculation revealed a binding mode similar to the experimental structure (Figure 6.16). The substituent in the 3'-positon was placed between the Asp 109 and the Tyr 112, pointing towards the solvent and forming hydrophobic interactions with the tyrosine residue. In the presence of the water, the compound did not dock into the active site. It is probably due to the presence of the molecules of water, which are near the Asp 109. This might explain the weak activity of this group of compounds and all the other compounds of the group.

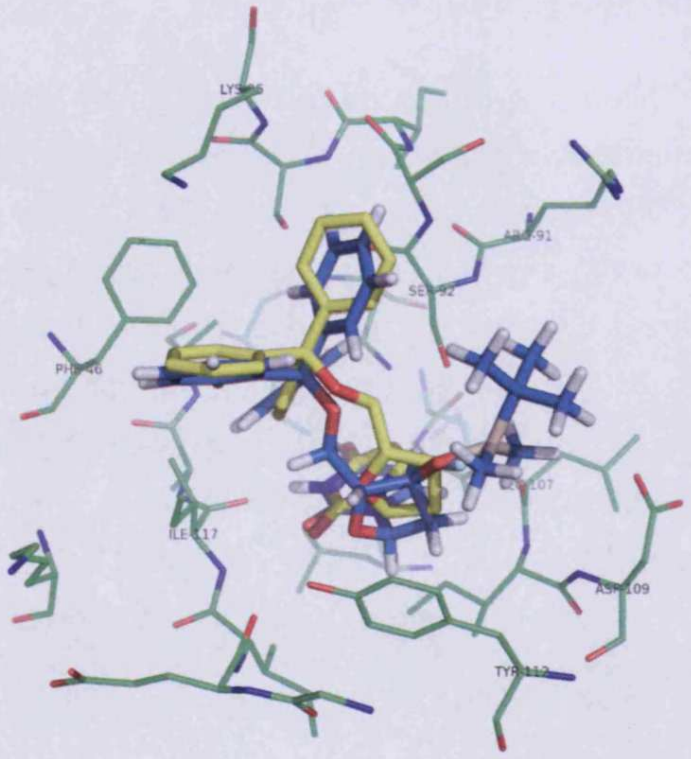
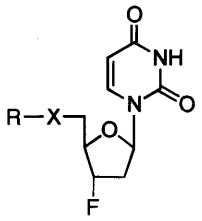


Figure 6.16 Docking result of the compound WSP484 (blue) superimposed with WSP869 obtained crystallographically (yellow).

Group 4. This group is characterized by the presence of the fluorine substituent in 3'position. The series is composed of seven

compounds (Figure 6.17); the most active ligand is 869 (*Ki* 5 μM), which was crystallised with the *P. falciparum* dUTPase[39].



WSP Group 4

868 X= O, R = TBDMS
948 X= O, R = TBDPS
947 X= O, R = TPS
1191 X= O, R = H
869 X= O, R = Ph_3C
946 X= NH, R = Ph_3C
1345 X= O, R = Trybenzylsilyoxy

Figure 6.17 Structures of compounds assayed of the group 4

The docking calculation of the compound WSP869 was used to test the ability of FlexX program to reproduce the binding mode of the crystal structure. The program was successful in reproducing the experimentally found binding mode (Figure 6.18). Both the X-ray complex and the calculated complex reveal that the uracil is inserted in a cavity formed by a β-harpin in the bottom of the active site while the trityl group is in contact with the hydrophobic pocket. The calculations, as well as the crystal structure, did not show any interaction between the fluorine in 3'-position with the protein, and no interactions between the oxygen in 5'-position with the protein.

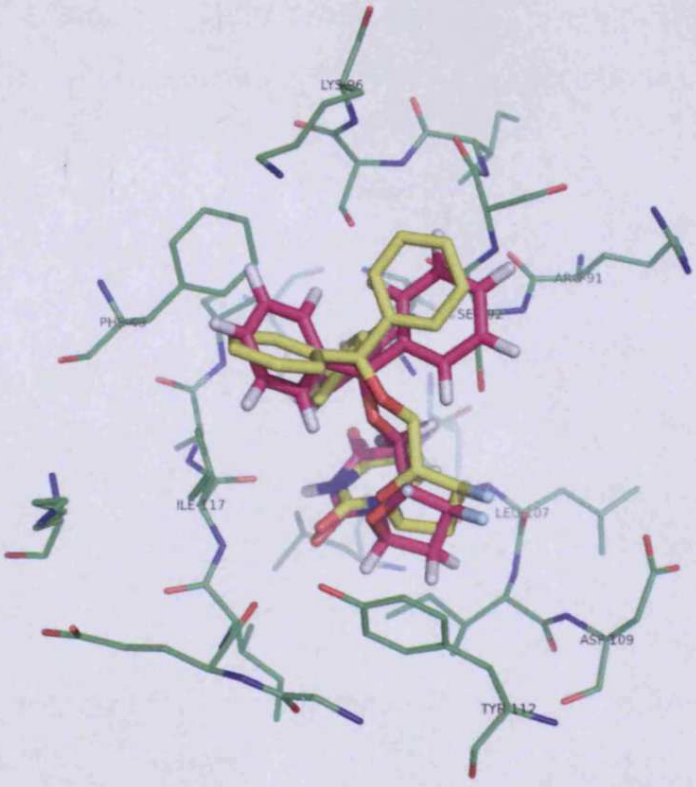


Figure 6.18 Docking result of the compound WSP869 (magenta) superimposed with WSP869 obtained crystallographically (yellow).

All the compounds docked in a conformation that resembled the experimental structure, except for WSP1345 (Figure 6.17, see also appendix 1). It might simply be due to the substituent in 5'-position was too big to be accommodated in the active site. However, this agrees with the experimental data. WSP1345 is essentially not an inhibitor of the enzyme.

Surprisingly, the compound WSP947 is completely inactive against the *P. falciparum* (*Ki* 1000 μM), although its structure is very similar to the WSP869, and the docking calculation revealed the same binding mode of the crystal structure (Figure 6.19). This inactivity may simply due to instability of the compound in the assay condition.

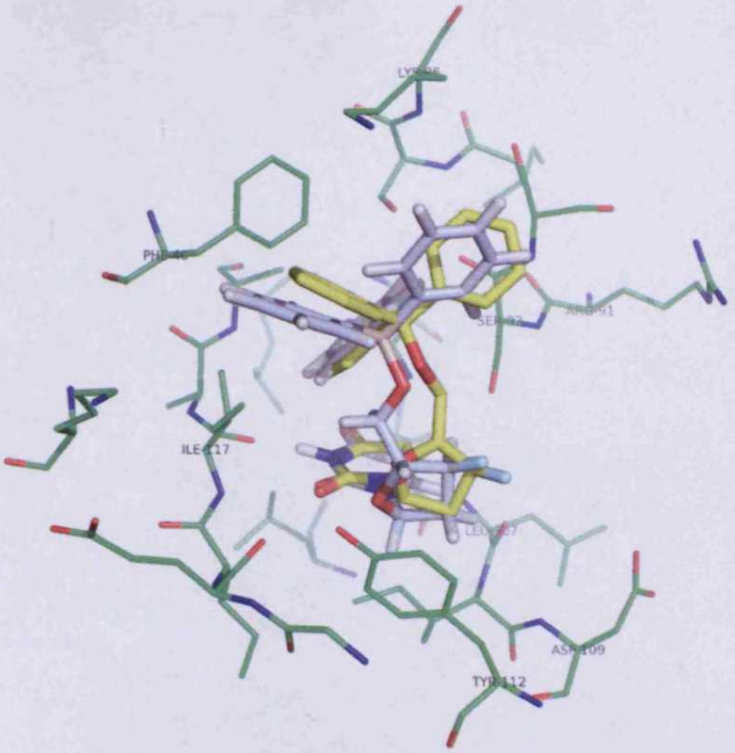


Figure 6.19 Docking result of the compound WSP1245 (silver) superimposed with WSP869 obtained crystallographically (yellow).

WSP946 had a relatively poor activity of 12.42 μM, which was in contrast with the trend that was observed with compounds that contained in 5'-position a trityl amino group (see Group II and also chapter 3). Usually, this feature improved the activity of the nucleoside derivatives. However, the automated docking was able to dock this ligand in the same conformation as the crystal structure of WSP869 (Figure 6.20). The reason for this is not clear.

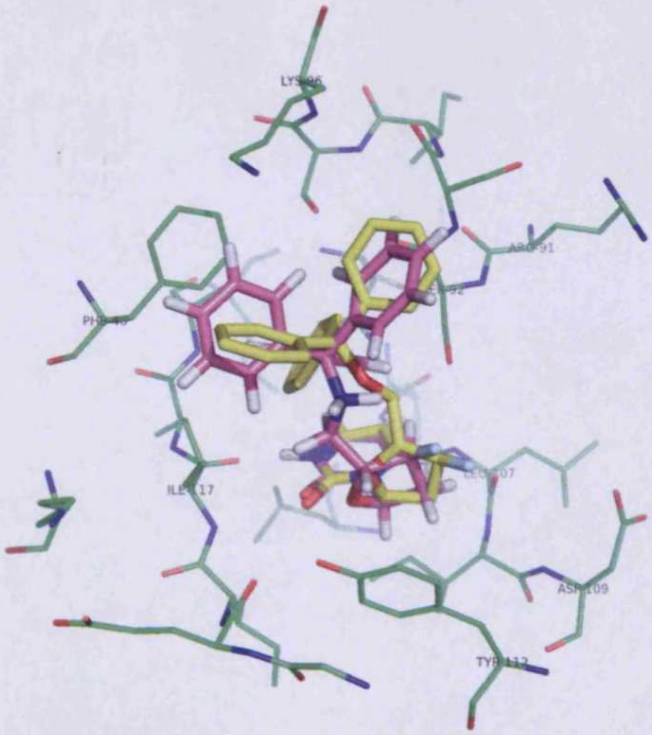
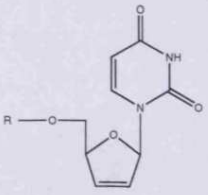


Figure 6.20 Docking result of the compound WSP946 (light-violet) superimposed with WSP869 obtained crystallographically (yellow).

Group 5. This series is composed from ten didehydrodideoxy nucleotide derivatives (Figure 6.21). The unsaturation of the sugar moiety might increase the interaction affinity with the aromatic ring of the Tyr 112 residue, and will probe the 3'-position.



WSP Group 5

831 R = $(PhO)_2P(O)$
875 R = $H_2NSO_2NHC(O)$
866 R = $CH_3(CH_2)_{14}C(O)$
865 R = AdamantylC(O)
830 R = H_2NSO_2
828 R = TBDMS
867 R = TBDPS
1073 R = TPS
829 R = H
864 R = Ph_3C

Figure 6.21 Schematic representation of the compounds of the group 5

All the ligands were docked successfully in the enzyme, except WSP831. This might be due to the larger size of the substituent in 5'-position (appendix 1), which was probably impossible to accommodate in the active site. However, this result is not surprising

since WSP831 did not show a significant activity against the *P. falciparum* enzyme (*Ki* 99 μM). In this group the most active compound are: WSP864, WSP867 and WSP1073 (1.3 μM, 1.2 μM and 1.3 μM respectively).

FlexX docked WSP864 in the same location as the experimental structure (Figure 6.22), with the sugar moiety forming a strong hydrophobic interaction with the Tyr 112.

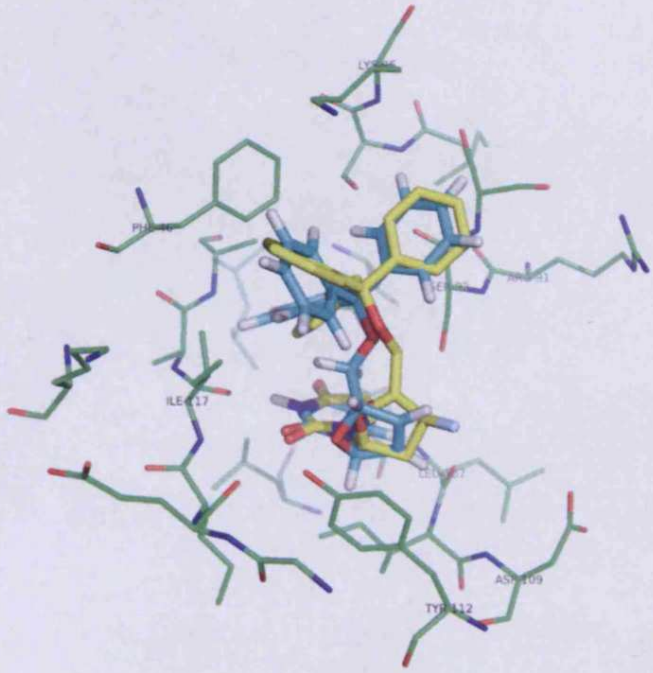


Figure 6.22 Docking result of the compound WSP864 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

Despite the fact that the compound WSP1073 was structurally very similar to WSP864, the substituent 5'-position substituent (triphenylsilyl) was found partially displaced from the hydrophobic pocket (Figure 6.23). Only one of the phenyl rings was involved in a T-shape interaction with the Phe 46. This might be due to the presence of the double bond, in the sugar moiety. The double bond undergoes a π-stacking interaction with the aromatic ring of the Tyr 112, which may alter the orientation of the 5'-position, giving less

possibility for the accommodation of the triphenyl silyl, in the hydrophobic pocket. Triphenylsilyl moiety is larger than the triphenylmethane moiety, probably accounting for the different result compared to WSP864

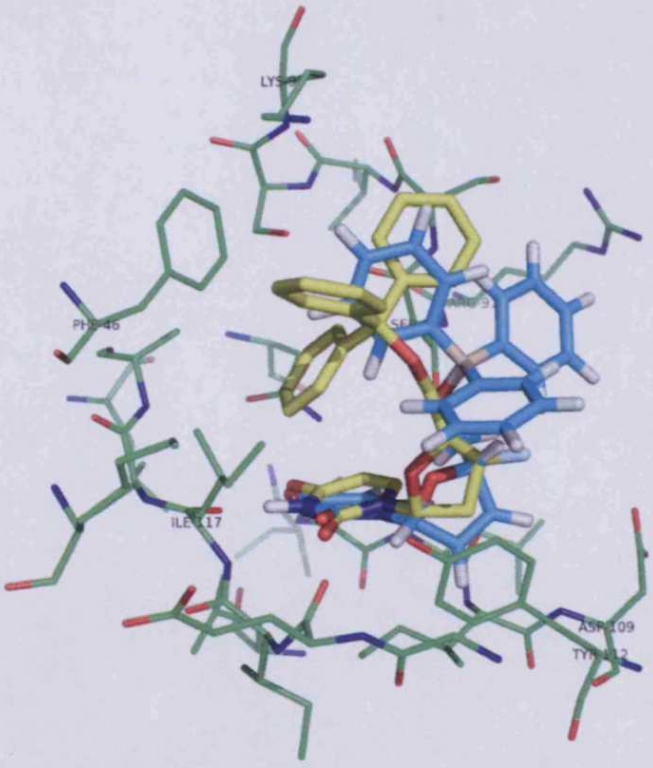


Figure 6.23 Docking result of the compound WSP1073 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

The docking of WSP867 (*Ki* 1.16 μM) was particularly interesting. This has in the 5'-position a *tert*-butyl diphenyl silyl substituent. The docking program found for this compound the same binding mode as the crystal structure of WSP869, showing the two phenyl rings involved in stacking and T-shape interactions with the Phe 46 (Figure 6.24), while the tert butyl seems make no interaction with the protein. This suggested that the presence of all three phenyl rings is not strictly necessary for the inhibition of the *P. falciparum* enzyme.

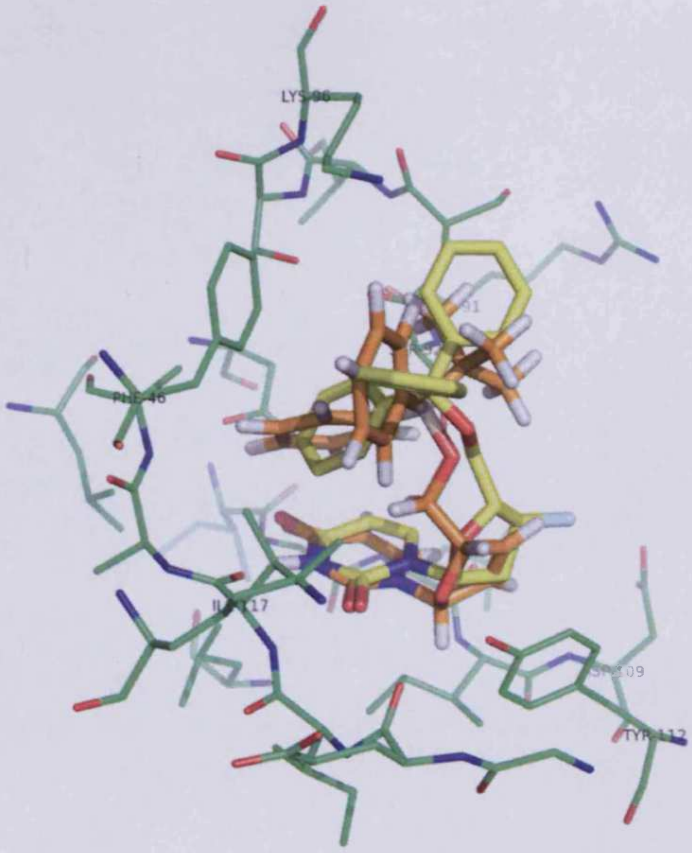
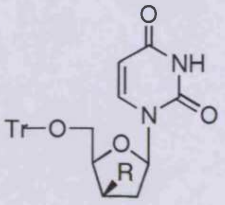


Figure 6.24 Docking result of the compound WSP867 (orange) superimposed with WSP869 obtained crystallographically (yellow).

Group 6. This group of inhibitors is characterised by the different stereochemistry of the 3'-position. This different design of the compounds was made in order to form additional interactions with the protein. These substituents in the 3'-position might point toward the inside of the active site, which should increase the possibility of making extra interactions, possibly with the Ser 92.

This series is composed of three compounds (WSP1361, WSP1362 and WSP1363, Figure 6.25); the inhibitor WSP1361 is one of the most active compounds of this group (0.3 μM), which also showed the most interesting docking result. This inhibitor was docked in the same location as the experimental structure of WSP869 (Figure 6.26). The hydroxyl group in the 3'-position points towards Ser 92; the hydroxyl might form a hydrogen bond with the Ser 92 side chain. The other ligands of the group were docked in a conformation that resembled

the X-ray structure, but the substituent in 3'-position did not show any important interactions with the protein. This is in agreement with experimental data.



WSP Group 6

1361 R = OH
1362 R = OAc
1362 R = N_3

Figure 6.25. Structures of compounds assayed of the group 6

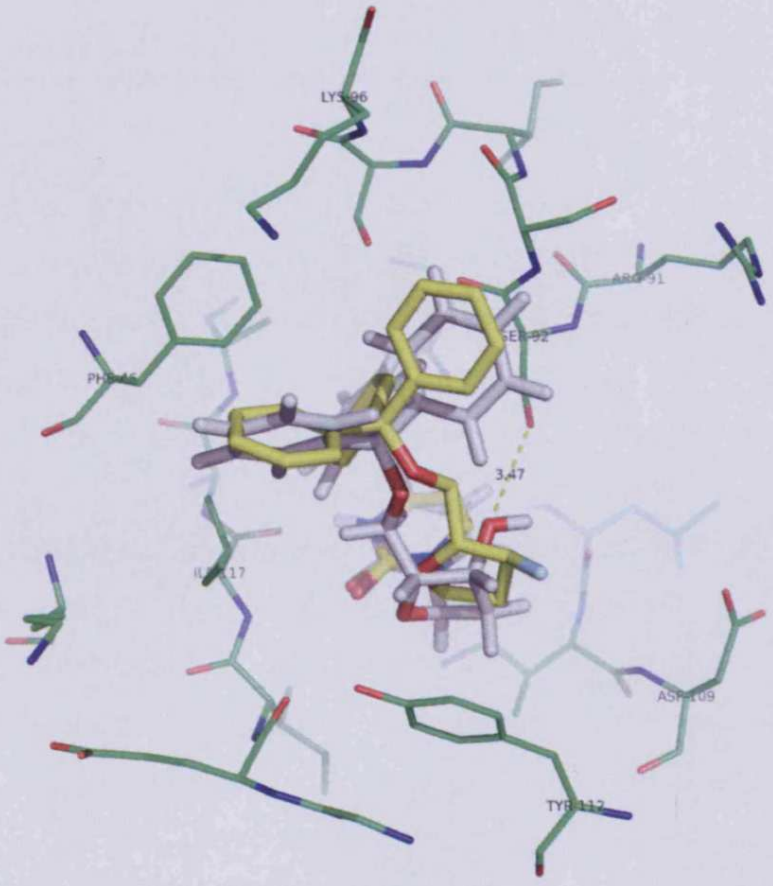


Figure 6.26. Docking result of the compound WSP1361 (cyan) superimposed with WSP869 obtained crystallographically (yellow).

6.1.1 Comparison between the binding score and the experimentally determined enzyme inhibition (*Ki*).

The estimated free energies of binding were plotted against the experimental inhibitory activities for all the compounds with an activity less than 500 μM (Table 6.3). The automated docking was carried out by using two different setting of the *P. falciparum* enzyme, one with all the water removed and one with the presence of 5 water molecules (Figure 6.27), which seemed to be conserved in other crystal structures[38]. The docking calculations revealed the same binding mode in both the protein settings. There was little affect on the binding score, whether the conserved water molecules were included or not (Table 6.3-4), except for the compounds of group III, which did not dock when conserved waters were included. This may be due to the bulky hydrophobic group in the 3'-position, which clash with the conserved water molecules.

In absence of the water molecules, considering all the 59 compounds, the linear correlation between the calculated binding free energy and the inhibition activity (*Ki*) was R^2 0.24 (Figure 6.28), which is relatively low. A particular reason for this low R^2, is the compounds of group 3. There was no correlation of the predicted binding score with the experimental binding. FlexX overestimated the energy binding of the compounds of group 3. Hence, they were considered outliers (they had a low scoring energy and poor activity).

WSP	*Ki* (μM)	Scoring energy (kJ/mol)
871	0.2	-64.92
1149	0.23	-63.38
1361	0.3	-59.91
1289	0.4	-59.02
1068	0.72	-58.85
1134	0.76	-64.83
1288	1.14	-58.48
867	1.16	-59.88
1206	1.27	-61.81
1073	1.3	-58.6
1207	1.43	-61.07
483	1.8	-61.12
864	1.9	-64.78
1324	2.11	-63.09
1243	2.17	-63.01
1244	2.56	-61.24
949	2.8	-60.12
1247	3.11	-49.91
1290	3.36	-62.88
1292	4.06	-54.08
950	4.2	-53.19
1229	4.22	-60.04
869	4.98	-61.83
1059	5.2	-65.7
828	10	-47.29
946	12.42	-64.02
1203	25.5	-64.9
665	26	-58
1291	27.6	-61.77
1245	31.55	-51.08
786	40	-61.53
1205	48	-61.26
1363	49	-63.05
865	58	-52
704	67	-45.24
790	88	-43
948	89	-37
1329	103	-53.63
952	110.7	-51.3
706	115	-62.04
788	123	-51.21
666	142	-49.72
661	178	-46.12
668	189	-49.16
660	206	-50.15
784	217	-54.89
669	228	-49.03
787	244	-44
826	252	-33.49
830	254	-34.6
866	255	-70

933	268	-30.28
662	294	-51.33
829	298	-30.68
953	313	-68.29
659	316	-45.49
785	324	-54.16
1362	350	-60.21
958	426	-51.00
789	468	-46.89

Table 6.3. Energy scores obtained from FlexX program with no water molecules and the experimentally determined inhibition constant.

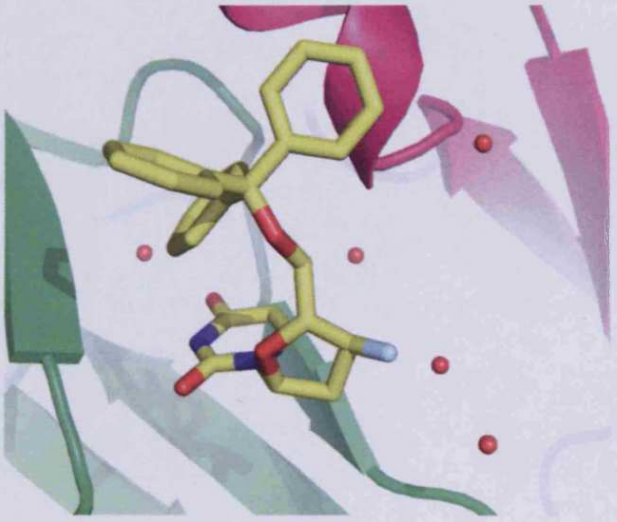


Figure 6.27 *P. falciparum* dUTPase active site, with the co-crystallized inhibitor WSP869 (yellow) and 5 conserved water molecules.

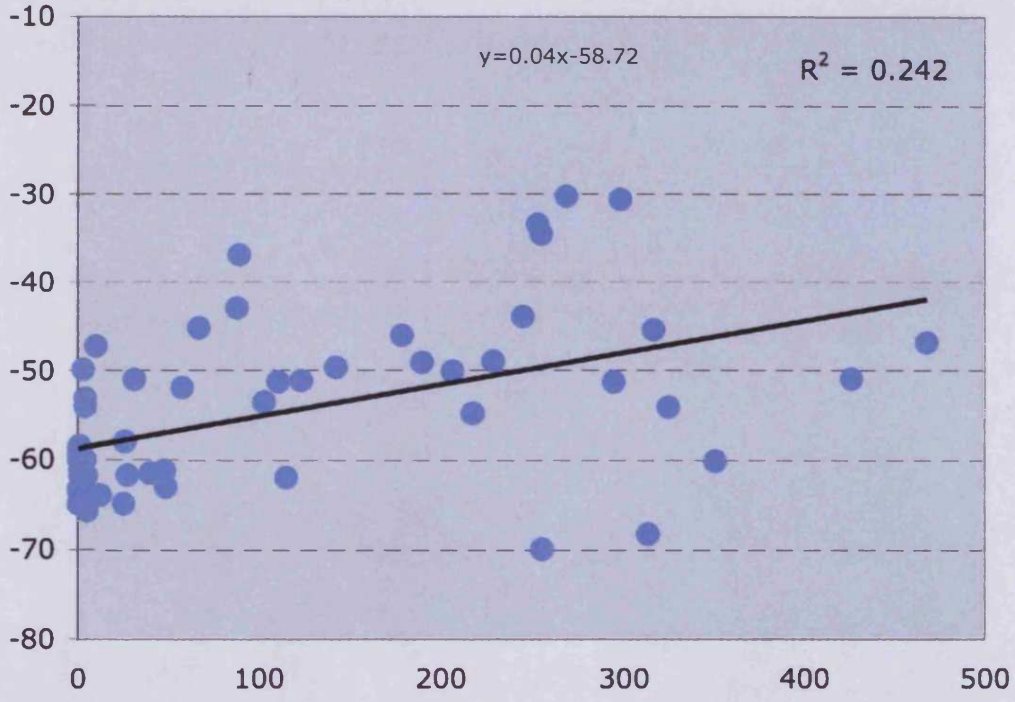


Figure 6.28 Correlation between *Ki* (*x*-axis) and FlexX score (*y*-axis) with no water molecules, for compounds with a *Ki* < 500 μM. For this graph were used the full set of compounds.

WSP	*Ki* (μM)	Scoring energy (kJ/mol)
871	0.2	-65.86
1149	0.23	-64.32
1361	0.3	-60.69
1289	0.4	-59.74
1068	0.72	-54.72
1134	0.76	-63.63
1288	1.14	-59.07
867	1.16	-59.88
1206	1.27	-61.71
1073	1.3	-58.6
1207	1.43	-61.07
483	1.8	-61.55
864	1.9	-64.78
1324	2.11	-63.09
1243	2.17	-63.46
1244	2.56	-62.42
949	2.8	-59.06
1247	3.11	-48.33

1290	3.36	-61.85
1292	4.06	-53.85
950	4.2	-53.52
1229	4.22	-60.04
869	4.98	-61.83
1059	5.2	-66.43
828	10	-47.29
946	12.42	-64.52
1203	25.5	-60.87
665	26	-58.08
1291	27.6	-60.8
1245	31.55	-50.92
786	40	-61.53
1205	48	-62.16
1363	49	-63.82
865	58	-52.85
948	89	-37.92
1329	103.2	-56.35
952	110.7	-51.31
706	115	-62.04
788	123	-50.85
666	142	-51.18
661	178	-46.12
668	189	-49.16
660	206	-51.46
669	228	-50.07
826	252	-34.15
830	254	-39.7
866	255	-72.49
662	294	-52.28
829	298	-34.88
785	324	-54.16
1362	350	-40.83
958	426	-52.51

Table 6.4 Energy scores obtained from FlexX program with 5 conserved water molecules and the experimentally determined inhibition constant.

A better correlation was observed between experimental binding energy and FlexX score, when docking was carried out with the presence of the molecules of water. This had revealed an R^2 value of 0.32 (Figure 6.29, number of compounds 54). This improvement of the linear correlation is probably due to the exclusion from the calculation of the compounds of the Group 3 (WSP659, WSP784, WSP789, WSP790 and WSP704, see also Figure 6.15). In absence of the conserved water molecules, the docking could have sufficient space to accommodate the substituent in 3'-position in the active site

pocket. These compounds did not have any significant activity against the *P. falciparum* dUTPase (see appendix 1). Thus, an additional docking calculation was performed using the protein (number of compounds 54) in the absence of the conserve waters and excluding the compounds of the Group III. This gave a slightly improvement R^2 of 0.27 (Figure 6.30).

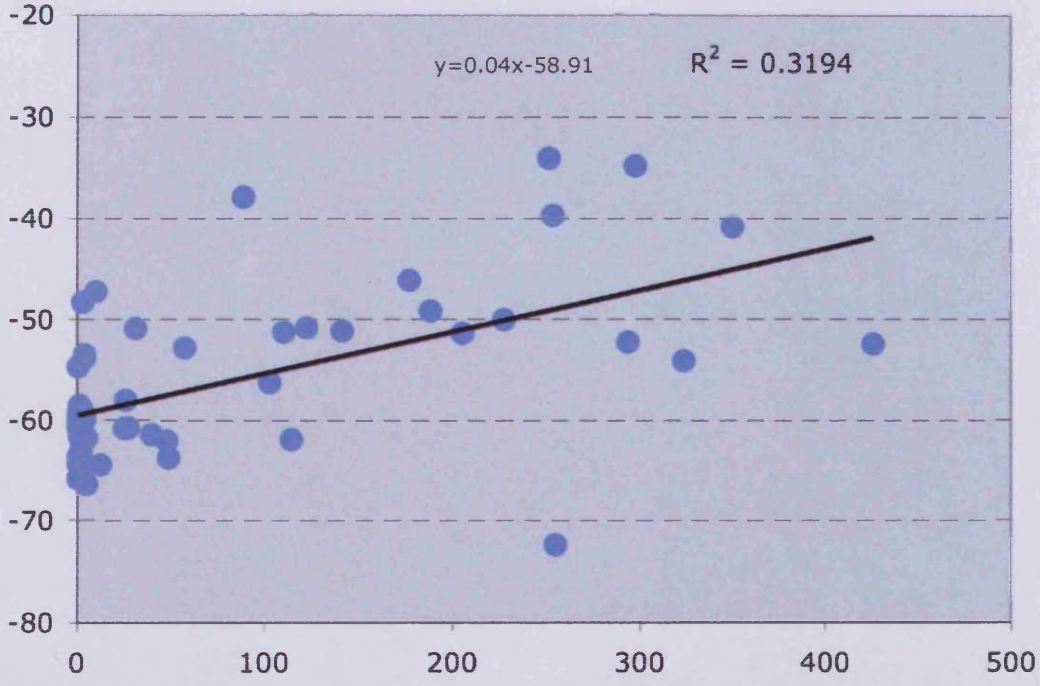


Figure 6.29 Correlation between *Ki* (*x*-axis) and FlexX score (*y*-axis), for compounds with a *Ki* < 500 μM. The calculations of the scoring energies were carried out with the protein in presence of 5 conserved water molecules, using 54 compounds.

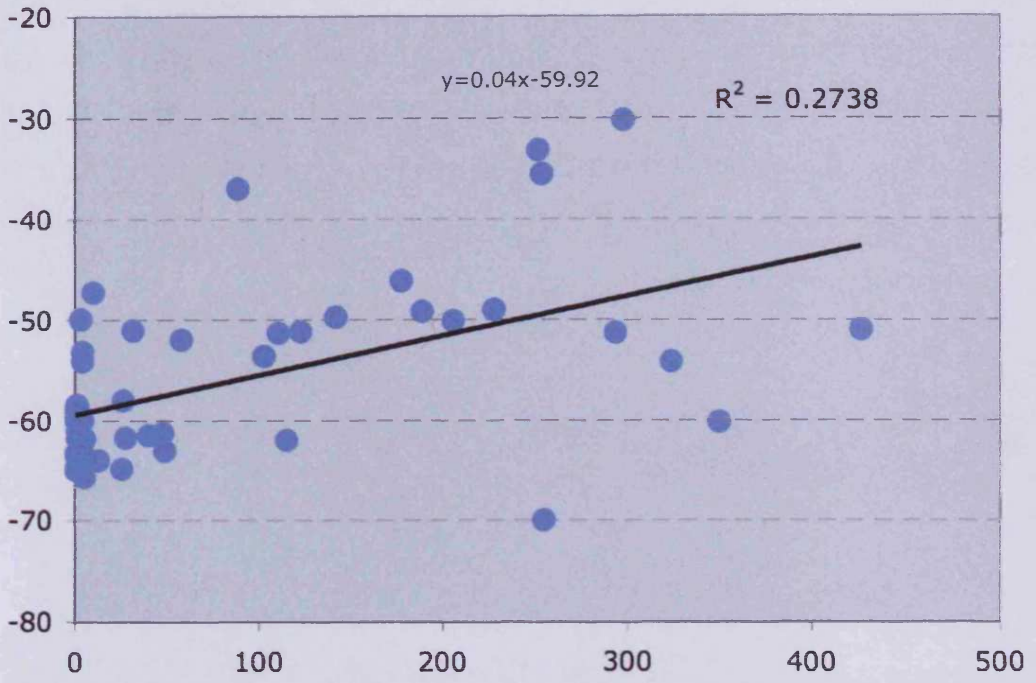


Figure 6.30. Correlation between *Ki* (*x*-axis) and FlexX score (*y*-axis), for compounds with a *Ki* < 500 μM. The calculations of the scoring energies were carried out with the protein in absence of water molecules using 54 compounds.

The calculated free energies of binding in both models (in absence of the water and in presence of the water molecules) were plotted against the experimental inhibitory activities (*Ki*, see appendix 1) for the set of compounds with an activity less than 10 μM, which included 23 inhibitors (Table 6.5). The liner correlation analysis was in contrast with the tendency of the previous linear correlation, which considered a bigger set of compounds. The docking model in presence of the water gave a R^2 of 0.26; the model with no water molecules gave a R^2 of 0.30. The main difference between the two docking models is in the computed energy score of the compound WSP1068 (Figure 6.31). This inhibitor contains a methoxy group in *para* position on one of the trityl phenyl ring. The automated docking placed this substituent in the bottom of the active site, where a molecule of water is normally conserved. Since the methoxy group has a hydrophobic character, it is not surprising that in presence of the water FlexX program under estimated the binding energy of this

compound. In fact, in absence of the water molecules, the energy score showed more consistency compared with the rest of the results.

In general, the models did not show highly significant linear correlation, however the computed energy score obtained from FlexX gave a certain consistence in the prediction of the binding affinity of the majority of the ligands.

WSP	*Ki* (μM)	Scoring energy[a] (kJ/mol)	Scoring energy[b] (kJ/mol)
871	0.2	-64.92	-65.86
1149	0.23	-63.38	-64.32
1361	0.3	-59.91	-60.69
1289	0.4	-59.02	-59.74
1068	0.72	-58.85	-54.72
1134	0.76	-64.83	-63.63
1288	1.14	-58.48	-59.07
1206	1.27	-61.81	-61.71
1207	1.43	-61.07	-61.07
483	1.8	-61.12	-61.55
864	1.9	-64.78	-64.78
1324	2.11	-63.09	-63.09
1243	2.17	-63.01	-63.46
1244	2.56	-61.24	-62.42
949	2.8	-60.12	-59.06
1247	3.11	-49.91	-48.33
1290	3.36	-62.88	-61.85
1292	4.06	-54.08	-53.85
950	4.2	-53.19	-53.52
1229	4.22	-60.04	-60.04
869	4.98	-61.83	-61.83
1059	5.2	-65.7	-66.43
828	10	-47.29	-47.29

Table 6.5 Correlation between *Ki* (*x*-axis) and FlexX score (*y*-axis), for compounds with a *Ki* < 10 μM. [a] scoring energy for the protein with deleted water. [b] scoring energy for the protein with 5 conserved water molecules.

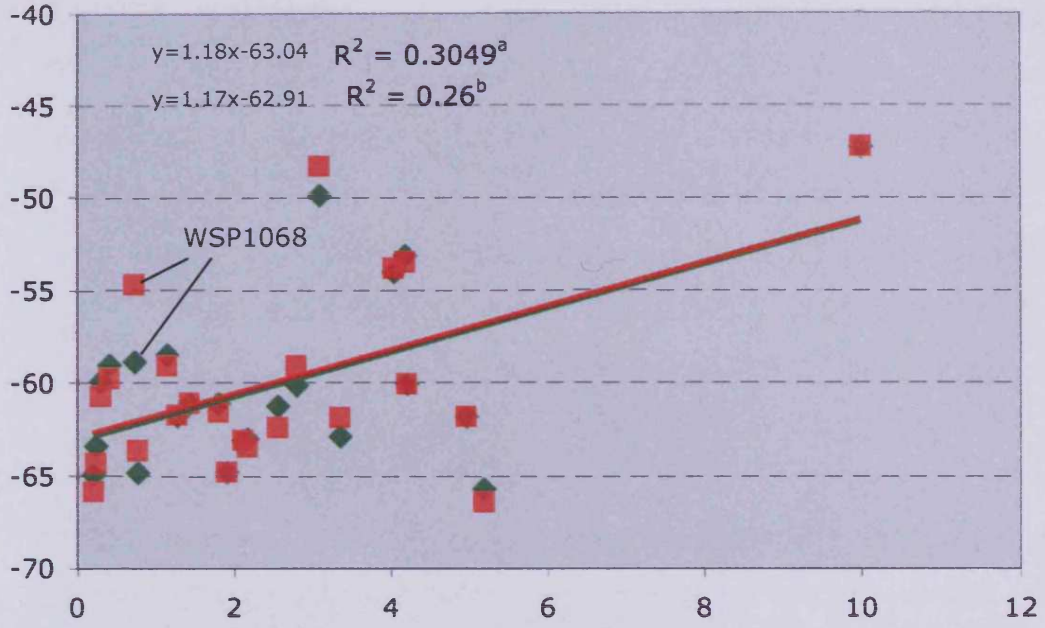


Figure 6.31 The series in green represents the correlation between *Ki* (*x*-axis) and FlexX score (*y*-axis), using the protein with deleted water, for compounds with a *Ki* < 10 μM. The series in red represents the correlation between *Ki* (*x*-axis) and FlexX score (*y*-axis), using the protein with the 5 conserved water molecules, for compounds with a *Ki* < 10 μM. R^2 = [a] linear correlation of the series in green, [b] linear correlation of the series in green.

6.2 Conclusion

FlexX was successfully able to dock the majority of the compounds into the active site of the *P. falciparum* dUTPase in a conformation similar to that of the experimentally determined co-crystal structure of WSP869. That is the uracil was bound deep in the active site in the uracil binding pocket and the trityl group in a hydrophobic pocket, principally bound by Phe 46, Ile 117 and Lys 96.

In general compounds, which showed a favourable binding score, had a trityl group or 2 or 3 aromatic rings in the 5'-position. Whilst, FlexX was able to dock other compounds in a similar binding conformation to WSP869. These compounds had an unfavourable binding score. This is probably due to the lack of interaction in the

hydrophobic pocket. These latter binding interactions are very unlikely to have any physiological relevance.

The presence of conserved water molecules may also pay a part in the binding of some of the inhibitors. The conserved water molecules appear to restrict the size of substituent at the 3'-position. In order for these to bind would require displacement of the conserved water molecules, which presumably is not energetically favourable.

In summary the docking appears to predict a reasonable trend conformation for the inhibitors. There also appears to be a correlation between binding score and experimentally determined binding energy. This is not perfect; scoring function of docking programs to be approximate, aiming to the approximations that have to be made in order to have a scoring function that it is feasible to calculate in a computationally rapid manner.

However the correlation that has been obtained is useful and validates the docking calculations. This may have some utilities in prediction of the binding of novel compounds, structurally related to the inhibitors described here.

In this thesis a more sophisticated methods to calculating free energy changes on binding has been carried out (see chapter 7). This yield a good correlation

An alternate binding mode for the inhibitors has also been discovered using FlexX. In this binding mode, the uracil still binds in the binding pocket, but now the trityl group binds in the triphosphate binding pocket. This also allows a hydrogen bonding interaction between the heteroatom at the 5'-position and the Ser 92. Presumably the interaction between the trityl and the triphosphate

binding pocket is not very favourable as the triphosphate binding pocket is very polar.

In general this alternate binding conformation has a much less favourable binding energy. It is probably not physiologically relevant, in at least the majority of the ligands. There is of course no experimental evidence for this second binding mode.

7 Chapter: Calculation of the binding affinity using linear interaction energy method

The calculation of free energies of binding can play a critical role in rational drug design. Such calculations have traditionally been carried out using the free energy perturbation (FEP) or thermodynamic integration (TI) methods[75, 76]. These procedures are theoretically exact and, if implemented appropriately, can yield very precise estimates of free energies of binding. However, to obtain high precision, the extensive simulation of nonphysical intermediate states is often required. Although accurate, these methods are highly computationally intensive making them slow and impractical for application to many industrial pharmaceutical problems.

To provide a possible solution Åqvist and co-workers recently introduced a linear interaction energy (LIE) method, which compromise between accuracy and computational speed. This is a semi-empirical technique for the calculation of free energy changes based on the simulation of only two states. These two states are (i) the solvated ligand and (ii) the ligand bound to the solvated protein. The theoretical foundation for this method[77, 78], was based on the following equation proposed by Åqvist and co-workers:

$$\Delta G = 0.5\langle \Delta U_{elec} \rangle + \alpha \langle \Delta U_{vdw} \rangle; \qquad \beta = 0.5 \qquad (1)$$

ΔU_{elec} and ΔU_{vdw} are the differences in the averaged inhibitor-environment electrostatic and van der Waals energies, respectively, between the two simulations, and the angled brackets denote ensemble averages. The original work fitted the equation to the experimental free energies of binding for a series of four endothiapepsin inhibitors, generating a value of 0.161 for α.

The proposed method has since been successfully applied to a variety of different systems[77], including HIV-1 protease[79],[80], trypsin[81] and thrombin[82]. While Åqvist has reported good transferability of the parameters to a limited number of other systems, other workers[82, 83] have not been so successful. Moreover, the data sets to which the method has been applied have mostly been relatively small. Consequently, the issue of the "transferability" of LIE parameters has not been fully addressed. Recently, however, Wang et al.[84] investigated the transferability of the van der Waals coefficient α in more detail. Simulations of trypsin-benzamidine[81] and camphor-P450cam[83] systems had previously been carried out using the GROMOS and CVFF force fields, respectively[75]. These simulations were repeated using the force field developed by Cornell et al.[85] with a similar simulation setup resulting in similar values of α. Hence it was concluded that the value of α was not strongly dependent on the force field. Additional simulations on five other systems, giving a total of seven ligands interacting with five proteins, showed that one fixed α could not give results in agreement with experiment in all cases. This work was carried out with the value of β fixed to 0.5, and hence a similar analysis is not available for this parameter. However, it has been noted by Hansson et al. that the value of β becomes smaller than the theoretical 0.5 as the number of ligand–OH groups increases[86]. This might be due to an overestimation by the force field of the electrostatic interactions between groups and their surroundings possibly because polarization effects are not properly described by the available force fields. Therefore, the values of α and β depend on the system under investigation, the force field used, and the computational methods that are applied. A proper value of these two parameters must first be determined by comparing calculated and experimentally estimated binding affinities.

In conclusion, to set up an LIE model for the prediction of ligand

binding affinities using a particular force field, it is necessary to perform a double regression (α and β) analysis for a set of compounds with known experimental ΔG values[87].

The current study discussed in this chapter analysed two sets of nucleoside derivatives (acyclic and cyclic nucleosides derivatives described in the chapters 5 and 6) dUTPase inhibitors. The LIE method was used to build predictive binding affinity models, in order to forecast ligand-protein affinity of new potential inhibitors of *P. falciparum* dUTPase.

7.1 Materials and Methods

7.1.1 LIE Methodology.

The LIE method employs experimental data on binding free energy values for a set of ligands (referred as training set) to estimate the binding affinities for a set of novel compounds. The method is based on the linear response approximation (LRA), which dictates that binding free energy of a protein-ligand system is a function of polar and non-polar energy components that scale linearly with the electrostatic and van der Waals interactions between a ligand and its environment. The free energy of binding for the complex is derived from considering only two states (Figure 7.1), (i) free ligand in the solvent, (ii) ligand bound to the solvated protein[88].

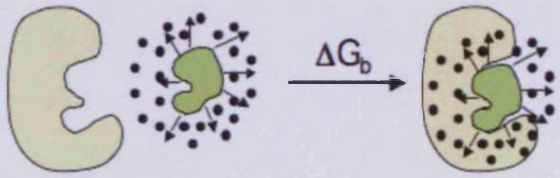


Figure 7.1 Schematic representation of a binding event showing different environments for dUTPase inhibitors. Small arrows depict potential interactions of drug with water (unbound state) or water and protein (bound state).

The conformational changes and entropic effects pertaining to unbound receptor are taken into account implicitly and only interactions between the ligand and either the protein or solvent are computed during the molecular mechanic calculations. Among the various formulations of the LIE methodology developed in the past, the SGB-LIE method implementing a surface generalized Born (SGB) model for the solvation has been shown to be 1 order of magnitude faster than the methods based on explicit solvent[88] with the same

order of accuracy. The SGB-LIE method also offers better accuracy in treating the long-range electrostatic interactions. The SBG-LIE method implements the original formulation proposed by Jorgensen[89] for the case of continuum solvent replacing the solvent accessible surface area term by a cavity term as follows:

$$\Delta G = \alpha(\langle U_{vdw}^{b}\rangle - \langle U_{vdw}^{f}\rangle) + \beta(\langle U_{elec}^{b}\rangle - \langle U_{elec}^{f}\rangle) + \gamma(\langle U_{cav}^{b}\rangle - \langle U_{cav}^{f}\rangle) \quad (2)$$

Where bracketed terms represents the ensemble average of the energy terms, such as van der Waals (U_{vdw}), electrostatic (U_{elec}), or cavity (U_{cav}) energy. The energy terms involved can be computed using energy minimization, molecular dynamics, or Monte Carlo calculations. All the terms are evaluated for interaction between ligand, both in the free (f) and bound (b) state, and its environment. The α, β, and γ are LIE fitting parameters. The transferability and dependence of LIE parameters on force fields and protein-ligand system are still the subject of discussion. In the Jorgensen formulation[89], LIE parameters are free coefficients that need to be determined by fitting the experimental data on the training set compounds. In the SGB model of solvation, there is no explicit van der Waals or electrostatic interaction between the solute and solvent. The contribution for net free energy of solvation comes from two energy terms, namely, reaction field energy (U_{rxn}) and cavity energy (U_{cav}):

$$U_{SGB} = U_{rxn} + U_{cav} \quad (3)$$

The cavity and reaction field energy terms implicitly take into account the van der Waals and the electrostatic interactions, respectively, between the ligand and solvent. The application of the SGB-LIE method for a given protein-ligand system essentially involves computing four energy components, i.e., the van der Waals

and Coulombic energy between the ligand and protein and the reaction field and cavity energy between the ligand and continuum solvent. The total electrostatic energy in the SGB-LIE method is the sum of Coulombic and reaction field energy terms.

7.1.2 Computational details.

Preparation of receptor and ligands was undertaken using the SYBYL package. All the calculations for the SGB-LIE method were performed in the LIAISON package from Schrodinger Inc.[90]. The LIAISON module performs LIE calculations in the OPLS (Optimized Potential for Liquid Simulations) force field with a residue-based cutoff of 15 Å. The OPLS force field was also used for charge assignment and all energy calculations performed in the SYBYL and Liaison packages. The simulations were performed on a set of 26 acyclic and 57 cyclic nucleoside derivatives.

7.1.3 Protein Preparation.

Initial coordinates for the *P. falciparum* dUTPase bound with a 5'-tritylated nucleoside (compound WSP869) were taken from the RSCB Protein Data Bank (code: 1VYQ), which describes the structure of protein in its trimeric form. All residues were kept at their default protonation states at physiological pH. The net charge on the protein was 0. After the addition of hydrogen atoms, the complex structure was energy minimized in the SYBYL package while all atoms except hydrogen were kept fixed. The motivation behind keeping protein atoms fixed was to preserve the structure of the binding site during energy calculations. All water molecules were removed, although, in the case of the cyclic derivatives binding affinity calculation, five

conserved water molecules were left in the active site to generate an additional model.

7.1.4 Ligand preparation.

All the ligands used in this study were prepared previously in the docking study (see chapter 5-6). The amino moiety present in the analogues was protonated as in the docking study. The partial charges on the atoms were assigned using the OPLS force field. The binding modes used for all the compounds were the same as predicted by the docking study. These putative binding modes resembled the conformation of the *P. falciparum* dUTPase crystal structure. There are in total 26 acyclic (the ligand WSP869 was included in the acyclic training set) and 57 cyclic derivatives. The acyclic ligand WSP961 was left out from the original training set.

7.1.5 LIE Calculation for Acyclics.

Two methods were employed for obtaining the average LIE energy. In the first procedure, a conjugate gradient minimization was performed, with a maximum of 1000 steps. The root mean square (rms) gradient value of 0.001 was the criterial for the convergence of minimization, followed by collection of the LIE energies. This procedure was carried out using a 15 Å residue-based cutoff. In the second procedure, the conjugate gradient minimization was followed by a hybrid Monte Carlo (HMC) step that employed 20 *ps* of heating (before collecting the LIE energies), a sample target temperature of 300 K, and 30 ps of sampling for the LIE energies (total simulation time = 50 *ps*). For this set of HMC simulations, the time step was 0.002 *ps*, there were five molecular dynamics (MD) steps per HMC cycle, and energies were sampled every 10 steps. A shorter

simulation was applaied (total time 20 *ps*) showing that 50 ps were adeguate to obtain the convergence. For all the simulations, protein residues beyond 10 Å from the binding pocket (as defined by the ligand WSP869) were frozen. Equation 2 was used for fitting of the values.

Knowing from the literature the importance of the charge state of the system, careful attention was paid to the charge neutralization used in the protein preparation protocol. For comparison, three different variants of the charge neutralization were employed (using the protein preparation utility of the LIAISON program). In method I, all the residues within the 8.3 Å of compound WSP869, were left charged. Residues outside this range were then adjusted to make the overall charge of the protein neutral. This resulted in a total charge of +2 within the 8.3 Å shell around the ligand WSP869.

For the method II all the residues within 12 Å of the compound WSP869 were left charged and the residues outside this range were then adjusted to give an overall neutral charge for the protein.

In order to study the contribution of the positive charge of the lysine in the binding affinity in the presence of trityl derivatives (see chapter 2), in method III all the residues within the 8.3 Å of compound WSP869, were left charged, as in the first method, except the Lys 96 was left neutral. In the methods I-II this residue was left charged (+1).

7.2 Results and Discussion Acyclic Nucleoside Derivatives

The SGB-LIE method was applied to a training set of 26 acyclic inhibitors of *P. falciparum* dUTPase to build a binding affinity model, which could be used to predict the free energy of binding of new nucleotide analogues.

For all the 26 compounds the experimental *Ki* was available. The 26 ligands were potentially ideal to be considered as a training set, with four orders of magnitude difference between the *Ki* and a relatively good diversity in the structures. Also, the training set contains enough data points not to suffer from over-parametrization by the LIE model.

Training set compounds were docked into the *P. falciparum* binding site, as described in the previous chapter (see chapter 5), and the SGB-LIE calculations were performed using the LIAISON module.

Two major factors that influence the energy terms in the LIE equation are the treatment of the protein charge state and the sampling method used to obtain the average energies. In order to choose the best method for the model, four LIE calculations were performed to compare the different effects, which influence the predicting model. The best fit to experimental values, based on the root mean square deviation (RMSD) calculation between experimental and predicted free energy of binding values, was obtained using with the method I. A summary of the results using the three methods (I-II-III) can be found in Table 7.1.

	Method	RMSD (kcal/mol)
A	Protein prep. = method I sampling = minimization	0.506
B	Protein prep. = method I sampling = HMC	0.539
C	Protein prep. = method II sampling = minimization	0.732
D	Protein prep. = method III sampling = minimization	0.763

Table 7.1 Summary of the fits to Equation 2 using different protein preparation methods.

For the protein preparation method I was used with two different sampling procedures for determining the LIE average energy (Table 7.2). It was found, based on the RMSD values, that for this set of ligands, the conformation space search provided by simple minimization sampling produced a better fit of the experimental binding energies than HMC method (Figure 7.2).

Theoretically, Monte Carlo (MC) methods should be more efficient in sampling the ensemble average values for energy terms; however sampling using MC could be computationally expensive and in this case did not improve the results accuracy, confirming previous studies[91, 92] which revealed that the conformational search provided by Molecular Dynamics (MD) or MC methods did not produced significantly better fit than the energy-minimization sampling. Method I with minimization sampling was chosen to build our definitive binding affinity model as generated the best prediction of the free energy of binding (see Table 7.1, this method was chosen for both the acyclic and cyclic derivatives).

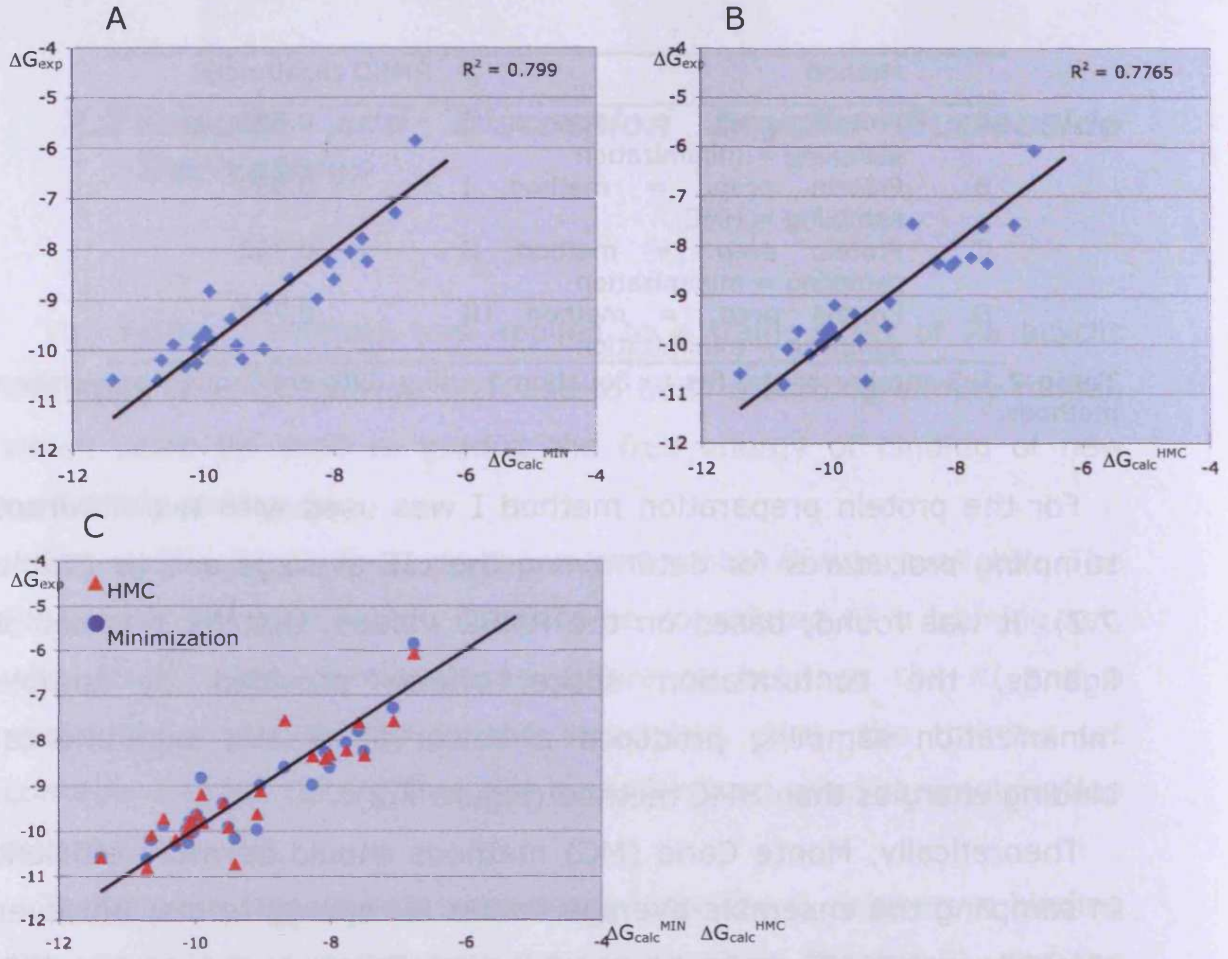


Figure 7.2 Comparison of the experimental versus calculated ΔG_{bind} for two different sampling methods: A) Minimization sampling method; B) HMC sampling method; C) comparison between the two sampling methods

WSP	ΔG_{exp}	ΔG_{calc}^{MIN}	ΔG_{JK}^{MIN}	ΔG_{calc}^{HMC}	ΔG_{JK}^{HMC}
870	-10.652	-10.201	-10.085	-10.088	-9.979
872	-7.581	-7.790	-7.807	-7.609	-7.611
873	-8.078	-8.249	-8.255	-8.406	-8.424
874	-8.006	-8.591	-8.619	-8.319	-8.331
962	-7.761	-8.051	-8.091	-8.222	-8.261
1000	-10.108	-10.268	-10.285	-10.027	-10.041
1001	-9.908	-8.824	-8.750	-9.194	-9.137
1002	-7.073	-7.282	-7.293	-7.575	-7.593
1004	-9.892	-9.806	-9.802	-9.812	-9.805
1005	-9.089	-9.976	-10.044	-9.624	-9.705
1060	-10.470	-9.886	-9.814	-9.722	-9.623
1063	-8.266	-8.981	-9.052	-8.351	-8.361
1064	-9.038	-8.985	-8.982	-9.122	-9.128
1065	-8.678	-8.586	-8.575	-7.561	-7.483
1066	-7.499	-8.24	-8.285	-8.334	-8.394
1067	-10.082	-9.782	-9.758	-9.773	-9.742
1069	-6.766	-5.849	-5.731	-6.071	-6.012
1072	-10.131	-9.965	-9.924	-9.965	-9.949
1142	-9.978	-9.634	-9.498	-9.634	-9.583
1224	-9.508	-9.907	-10.059	-9.907	-9.492
1228	-9.589	-9.382	-9.330	-9.382	-10.394
1230	-11.396	-10.593	-10.544	-10.593	-10.622
1241	-10.030	-10.034	-10.034	-9.869	-9.844
1315	-10.293	-10.327	-10.331	-10.189	-10.177
1317	-10.709	-10.610	-10.582	-10.811	-10.827
rmsd		0.506	0.560		

Table 7.2 Binding affinity model calculation for the two different sampling methods. ΔG_{exp} refers to the experimental binding free energy of the dUTPase inhibitors; ΔG_{calc}^{MIN} and ΔG_{calc}^{HMC} refer to the calculated binding free energy using the Minimization and HMC sampling methods respectively; ΔG_{JK}^{MIN} and ΔG_{JK}^{HMC} refer to energy values obtained using jackknife test using Minimization and HMC sampling methods respectively. rmsd (root mean square deviation)

Method II-III did not improve the RMSD values. For Method III, the high RMSD values might be due to the loss of the interaction between the positive charge of the Lysine residue and the trityl moiety (see chapter 2). The result of the minimization sampling in method III revealed a movement of the Lys 96 (uncharged) that could suggest the loss of a cation-π interaction with the aromatic network of the trityl moiety (Figure 7.3).

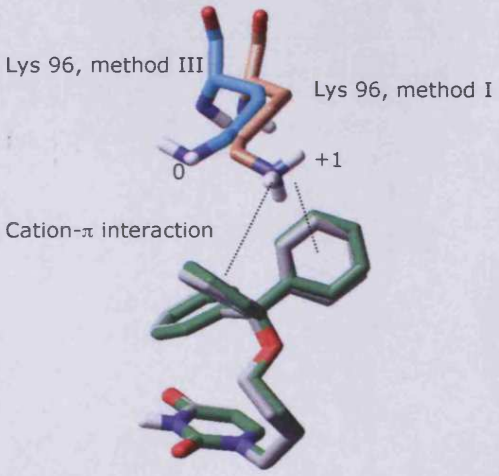


Figure 7.3 Superimposition of the result of the minimization sampling of the methods I and III in presence of the ligand WSP1000 (method I in green, method III in grey). In the method III the Lys moves, losing the cation-π interaction with the ligand.

The simulations were performed both for the ligand-free and ligand-bound state. The various interaction energy terms described in the methods were collected by the LIAISON package and are presented here in Table 7.3 (Method I, minimization). The largest contribution for the binding energy comes from the van der Waals (VDW) interactions, which is expected since that the compounds are mostly lipophilic molecules that interact favorably with a binding cavity lined with hydrophobic residues (see chapter 2). The cavity energy term in the bound state is smaller than in the free state for all compounds, as there is less energy penalty for creating a cavity in solvent when part of the ligand is buried into the hydrophobic binding site. The reaction field energy term in the free state lies in a very narrow range for all compounds and the value of ΔU_{rxn} (Equation 3) is normally positive, which is expected since the ligands are very lipophilic. Some of the compounds present negative ΔU_{rxn} values, which may be due to the positive charges of these compounds. This might suggest that the solvation of these compounds would be favored under physiological conditions[88].

WSP	U_{vdw}^{f}	U_{ele}^{f}	U_{rxn}^{f}	U_{cav}^{f}	U_{vdw}^{b}	U_{elec}^{b}	U_{rxn}^{b}	U_{cav}^{b}
1230	0.00	0.00	-83.31	4.33	-46.771	-15.9	-89.31	-2.492
1317	0.00	0.00	-22.01	4.53	-41.123	-26.24	-4.24	-2.472
870	0.00	0.00	-20.28	4.19	-48.949	-21.98	-4.32	-2.329
1060	0.00	0.00	-80.29	4.28	-48.663	-34.06	-97.89	-2.369
1315	0.00	0.00	-20.42	4.27	-49.181	-31.08	-2.34	-2.413
1072	0.00	0.00	-22.84	4.26	-48.897	-24.44	-6.30	-2.283
1000	0.00	0.00	-20.31	4.29	-49.063	-30.69	-2.16	-2.393
1067	0.00	0.00	-23.35	4.76	-44.370	-39.24	-4.23	-2.358
1241	0.00	0.00	-76.52	4.65	-43.863	-38.71	-91.19	-2.313
1142	0.00	0.00	-20.72	4.68	-40.431	-41.2	-3.19	-2.391
1001	0.00	0.00	-24.08	4.57	-39.367	-39.58	-4.76	-2.130
1004	0.00	0.00	-20.74	4.63	-42.228	-30.12	-4.00	-2.325
1228	0.00	0.00	-19.24	4.34	-49.107	-33.57	-3.66	-2.252
1224	0.00	0.00	-78.21	4.37	-50.139	-28.11	-95.96	-2.437
869	0.00	0.00	-23.42	4.38	-52.013	-30.47	-6.81	-2.389
1005	0.00	0.00	-21.81	4.64	-42.069	-24.42	-7.15	-2.353
1064	0.00	0.00	-19.49	3.86	-37.794	-31.86	-1.81	-2.124
1065	0.00	0.00	-21.67	3.86	-37.183	-36.57	-4.48	-2.091
1063	0.00	0.00	-19.47	3.87	-37.788	-31.35	-1.79	-2.123
873	0.00	0.00	-23.60	3.55	-36.199	-27.41	-8.31	-1.950
874	0.00	0.00	-18.47	3.80	-37.009	-23.58	-4.31	-2.017
962	0.00	0.00	-18.76	3.62	-33.109	-26.31	-6.51	-1.965
872	0.00	0.00	-20.57	3.50	-31.416	-25.88	-4.92	-1.825
1066	0.00	0.00	-19.92	3.44	-34.242	-28.85	-3.85	-1.954
1002	0.00	0.00	-18.45	3.39	-34.843	-32.74	-2.59	-1.752
1069	0.00	0.00	-22.66	2.78	-21.804	-23.87	6.59	-1.34

Table 3 Average van der Waals (vdw), Electrostatic (ele), Reaction Field (rxn), and Cavity (cav) Energy Terms for the Free (f) and Bound State (b) Obtained from SGB-LIE Calculations. The van der Waals and electrostatic energy terms for the free state are zero for all ligands, since there is no explicit interaction between ligand and solvent in continuum solvent.

The energy values in Table 7.3 were used to fit Equation 2 using the least-square error-fitting method. The values obtained for the three fitting parameters, α, β, and γ, are 0.015, -0.042, and 3.881, respectively. The large value of the cavity energy term signifies that binding is largely driven by the ligand's ability to bury itself in the binding cavity, which is understandable given that most of the ligands are highly hydrophobic in nature. Even when the observed VDW values are low, the total VDW interaction can contribute significantly toward the total free energy of binding of the final complex.

In Table 7.2, the experimental free energy values obtained from the *Ki* values and the free energy values estimated using fitting parameters are presented. The rmsd error between the experimental values and the values obtained by the fit was 0.506 kcal/mol, which

is an indicator of the robustness of the fitting. The quality of the fit can also be judged by the value of the squared correlation coefficient (R^2), which was 0.799 for this training set (Figure 7.2A, see also Table 7.2).

The jackknife cross-validation test was performed in order to estimate the bias and prediction power of the binding affinity model developed using the training set. The jackknife test involves leaving one compound out of the training set and then using fitting parameters obtained from the rest of the training set to predict the binding free energy of the ligand left out. The RMSD error between the experimental and the predicted free energy values for the jackknife test is 0.56 kcal/mol, which suggests a reasonable prediction capability of the SGB-LIE method for the estimation of free energy values. The biggest error contribution came from the ligand WSP1001 (Figure 7.4).

The jackknife test produces an underestimated (higher value than the experimental value, see Table 7.2) value for the free energy binding of this compound. Interestingly, WSP1000, which was well predicted in the calculation, is very similar to WSP1001 (Table 7.2).

This difference between the predicted and the experimental value for ligand WSP1001 might be due to the presence of the silicon atom in the structure, which is not well parameterized in the OPLS force field[93]. By including this compound in the training set, we train the LIE to make a close prediction for similar analogues having a triphenylsilyl moiety with the possibility of a different orientation and size of this motif group. When the compound WSP1001 were left out from the training set, the values of the LIE parameters did not differ much (α=0.020, β=-0.050, γ=3.766), but the quality of fit improved (RMSD =0.474 kcal/mol, R^2 =0.83) as would be expected when an outlier is omitted.

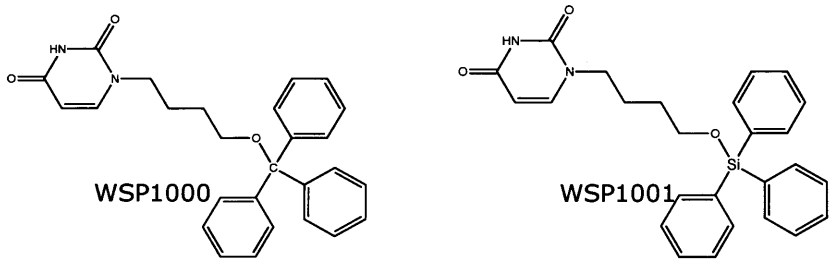


Figure 7.4 Compounds WSP1000 and WSP1001.

Based on these data, we have been able to model a set of 26 dUTPase inhibitors. And have managed to generate robust models, as indicated by the low rmsd values obtained by the jackknife cross-validation studies. However, a closer investigation of the trends of the individual terms of the LIE model versus the binding energy illustrated some areas where the method merits further investigation.

Plots of the correlations between the experimental ΔG_{bind} and each term (VDW, electrostatic and cavity terms) in the LIE equation are shown in Figure 7.5. One of the features highlighted by these plots there is relative good correlation between the experimental ΔG_{bind} and the calculated ΔG_{VDW} and ΔG_{elec}. A different situation was observed for the correlation between the LIE electrostatic energy *versus* experimental binding (Figure 7.5B), whre there was very little correlation. However, this lack of correlation is compensated by the VDW energies and the β. Factors that may explain this result include limitations in the underlying methods for computing the electrostatic energy terms (U_{ele}^{f} and U_{ele}^{b}, see Table 7.3). The literature indicates that the balancing of the subtle interactions between electrostatic and VDW terms poses a serious challenge[92]. It might be possible that a more theoretically rigorous treatment of the electrostatic energy would improve this situation.

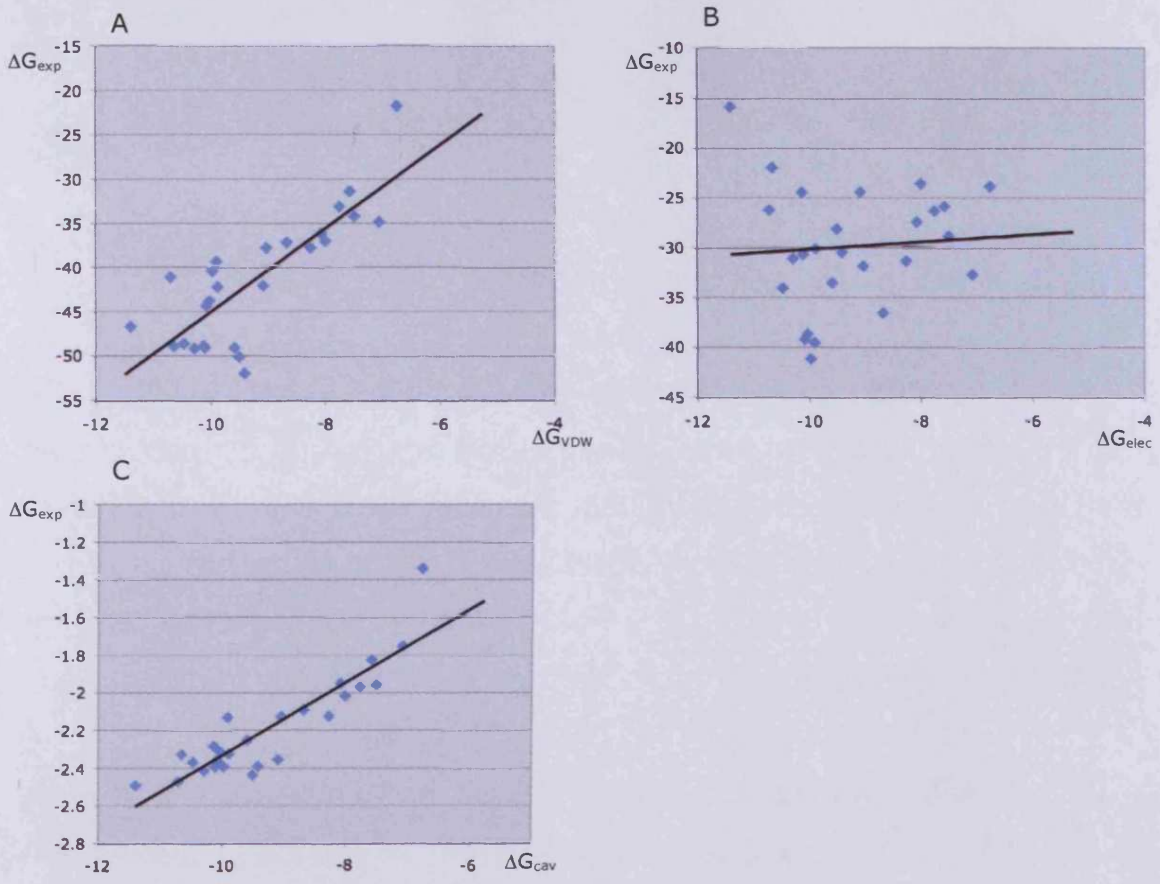


Figure 5 Individual trends for the LIE energies: A) van der Waals, B) electrostatic, C) cavity, versus the experimental ΔG_{exp}.

7.3 LIE Calculation for Cyclic Derivatives.

To compute the interaction terms required in the LIE model, energy minimization sampling was performed in the LIAISON package, and the various energy terms were collected at the end of the minimization. Energy minimization sampling was used since the previous LIE study of acyclic derivatives revealed that the conformational search provide by HMC method did not produced any improvement in the calculations compared with the minimization sampling. For this set of cyclic compounds a conjugate gradient minimization was performed, with a maximum of 1000 steps. The root mean square (rms) gradient value of 0.001 was the criteria for the convergence of minimization, followed by collection of the LIE energies. This procedure was carried out using a 15 Å residue-based cut-off. The simulations were performed removing all the water molecules and a second energy minimization was employed with the presence of 5 water molecules, which seem to be conserved in the active site of the dUTPase enzymes.

7.3.1 Result and Discussion of Cyclic Derivatives.

The SGB-LIE method was applied to a training set of 57 acyclic inhibitors of *P. falciparum* dUTPase to build a binding affinity model.

As stated previously, two models were generated using the protein in absence of water (model 1) and with 5 water molecules in the active site (model 2). For each model differences in VDW, electrostatic, reaction field, and cavity energies between the inhibitor and its surroundings in continuum-solvent *P. falciparum* dUTPase and in continuum-solveted state were calculated after 1000 energy minimization steps. Correlations between the experimental data and

the calculated energy terms were derived using the least-square fit of Equation 2.

The overall root mean square deviation (RMSD) of the LIE prediction for this training set in the model 1 (no water) was 1.322 kcal/mol, which is quite high but encouraging considering this large number of compounds (57 ligands). Model 2 revealed an RSMD of 1.339 kcal/mol, therefore the presence of the water molecules did not improve the prediction model. The three LIE parameters found for the two models are summarized in Table 7.4.

	α	β	γ
Model 1	0.0406	-0.015	2.814
Model 2	0.0381	-0.003	2.811

Table 7.4 Summary of the fitting coefficient values for the equation 2 using two different protein models: model 1 without water molecules; model 2 with 5 water molecules.

Both the binding affinity models failed to reproduce the experimental activities yielding a jackknife correlation coefficient R^2 and RMSD of 0.175 and 1.407 kcal/mol respectively for the model 1 (no water); model 2 (5 water molecules) revealed an R^2 and RMSD of 0.155 and 1.407 kcal/mol respectively. These results might be an indication of the role of the water molecules in the binding affinity of this class of compounds. Initially the water molecules were added to the model in order to investigate the possibility of the formation of interactions between the substituent in 3'-position and the protein. A closer look to the simulation results suggested that one water molecule (W4), could play a major role in the inhibition process of these ligands (Figure 7.6).

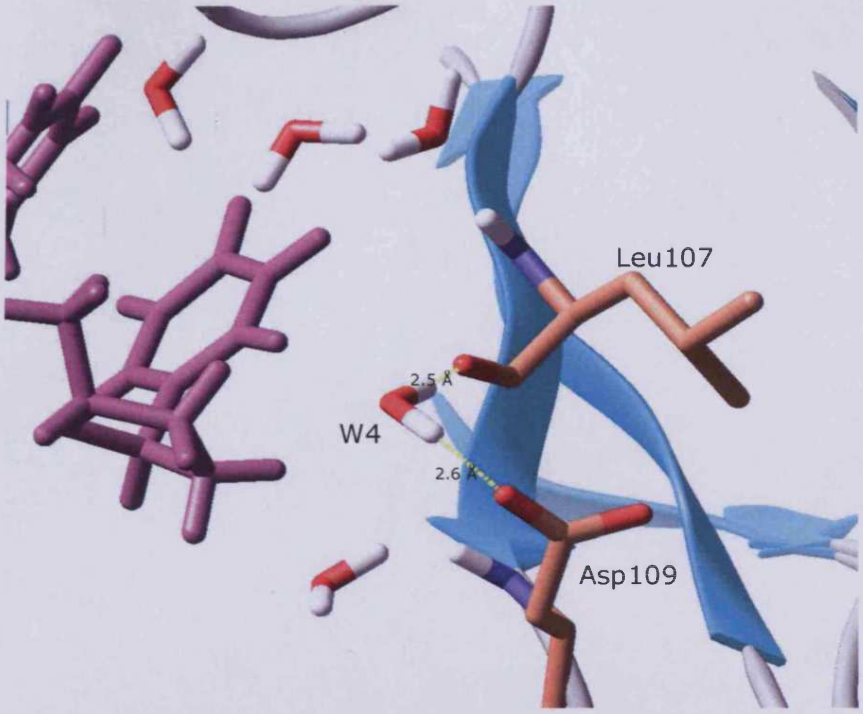


Figure 7.6 Over view of the *P. falciparum* dUTPase active site after minimization sampling in presence of the ligand WSP869 (violet) and 5 water molecules (red). Water molecule W4 form hydrogen bond with the Asp 109 side chain and the Leu 107 backbone.

The results of the minimization sampling on the protein showed this water molecule (W4) forming two hydrogen bonds with the Leu 107 backbone and the Asp 109 side chain. Hence, this water can only be an acceptor of hydrogen bond for the ligand.

From the structural analysis of the simulations result, it seemed that only one compound could interact, by a hydrogen bond with this water molecule, compound WSP1361. This ligand was designed with the 3'-substituent "up" (xylo) conformation (see chapter 3). In fact, this type of structure might allow polar groups in 3'-position to better interact with the protein or possibly with the conserved water molecules. The more potent inhibition of the compound WSP1361 (*Ki* 0.3 μM) compared to WSP483 (Ki 1.8 μM), suggests that the OH group in the "up" conformation might be the reason for the increased potency against the enzyme. The result of the minimization sampling on this ligand (Figure 7.7) confirmed the formation of a hydrogen bond between the OH in 3'-position a water molecule (1.7 Å), and

also an additional hydrogen bond between the 3'-OH and the Ser 92 (1.9 Å).

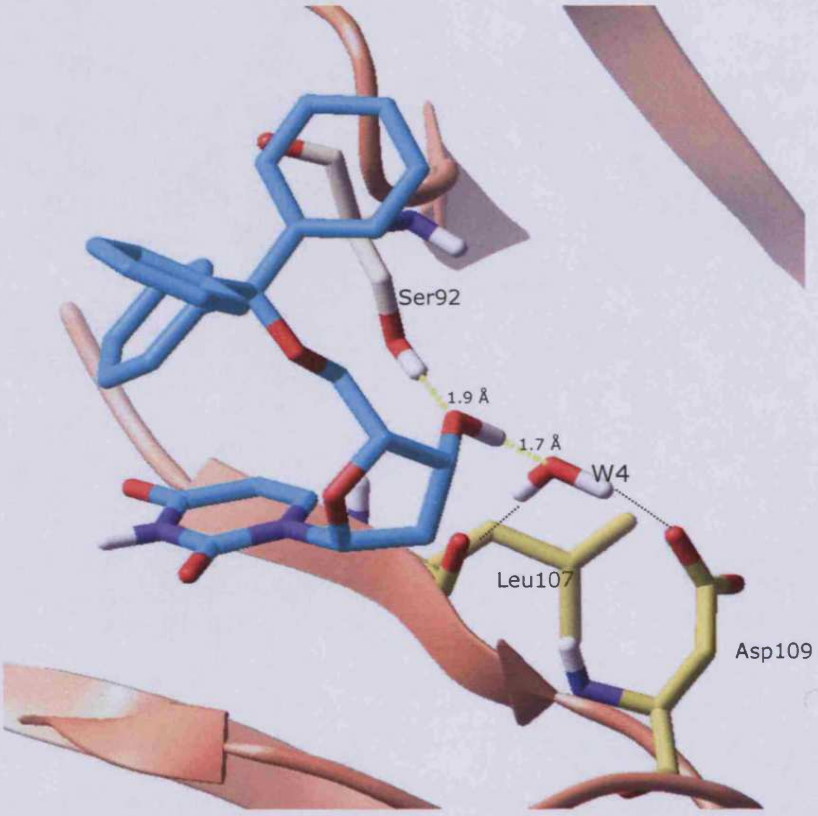


Figure 7.7 Over view of the *P. falciparum* dUTPase active site after minimization sampling in presence of the ligand WSP1361 (cyan) and 5 water molecules. The OH group in 3' position forms a direct hydrogen bond with a Ser 92 and by the water W4 with the Asp 109 and Leu 107.

The LIE calculation under estimated the free energy of binding of this inhibitor (see Table 7.5). On the other hand, the result from the model 1 of the ligand WSP1361 revealed an estimated binding affinity very similar to the one in model 2, even though, the interactions between the ligand an W4 and Ser 92 (Figure 7.8).

WSP	ΔG_{exep}	ΔG_{calc} no water	ΔG_{calc} 5 water
871	-11.4	-9.76	-9.4
1149	-11.31	-8.62	-8.86
1361	-11.15	-8.91	-8.87
1289	-10.97	-9.20	-9.28
1068	-10.61	-9.71	-8.86
1134	-10.57	-9.30	-9.35
1288	-10.32	-8.98	-8.98
867	-10.31	-8.61	-8.81
1206	-10.26	-9.63	-9.59
1073	-10.24	-8.06	-7.85
483	-10.04	-9.15	-9.00
864	-10.01	-9.38	-9.40
1324	-9.95	-9.20	-9.24
1243	-9.86	-9.80	-9.75
1207	-9.83	-9.24	-9.27
1244	-9.83	-9.03	-8.93
949	-9.77	-9.00	-9.09
1247	-9.71	-8.63	-8.87
1290	-9.66	-9.19	-9.14
1292	-9.54	-8.21	-8.19
1229	-9.52	-9.63	-9.38
950	-9.52	-9.15	-8.96
869	-9.42	-8.95	-9.02
1059	-9.39	-9.44	-9.53
828	-8.99	-7.75	-7.94
946	-8.85	-9.16	-9.00
1203	-8.41	-9.51	-9.47
665	-8.40	-9.21	-9.27
1291	-8.36	-8.90	-8.88
1245	-8.28	-7.69	-7.73
786	-8.13	-8.82	-9.07
1205	-8.02	-8.98	-8.92
1363	-8.01	-9.09	-8.97
865	-7.91	-7.59	-7.88
704	-7.82	-10.06	-9.91
948	-7.64	-7.68	-7.85
952	-7.51	-7.07	-7.04
1329	-7.50	-9.30	-9.46
706	-7.48	-9.23	-9.62
788	-7.44	-8.53	-8.21
666	-7.35	-7.86	-8.01
661	-7.21	-6.73	-6.91
680	-7.21	-6.69	-6.59
668	-7.18	-7.81	-7.83
669	-7.06	-7.58	-7.69
826	-7.00	-5.93	-5.98
830	-6.99	-6.13	-6.24
866	-6.99	-9.48	-9.31
660	-6.96	-7.44	-7.59
662	-6.9	-6.37	-6.49

829	-6.9	-5.07	-5.14
953	-6.87	-9.84	-9.85
659	-6.86	-8.15	-8.14
785	-6.84	-9.46	-9.25
1362	-6.8	-9.83	-10.16
958	-6.68	-7.91	-8.07
789	-6.61	-7.82	-7.45

Table 7.5 Binding affinity model calculation for the minimization sampling method. ΔG_{exp} refers to the experimental binding free energy of the dUTPase inhibitors; ΔG_{calc} refers to the calculated binding free energy using the Minimization sampling methods respectively in absence of the water and in presence of the 5 water molecules in the active site.

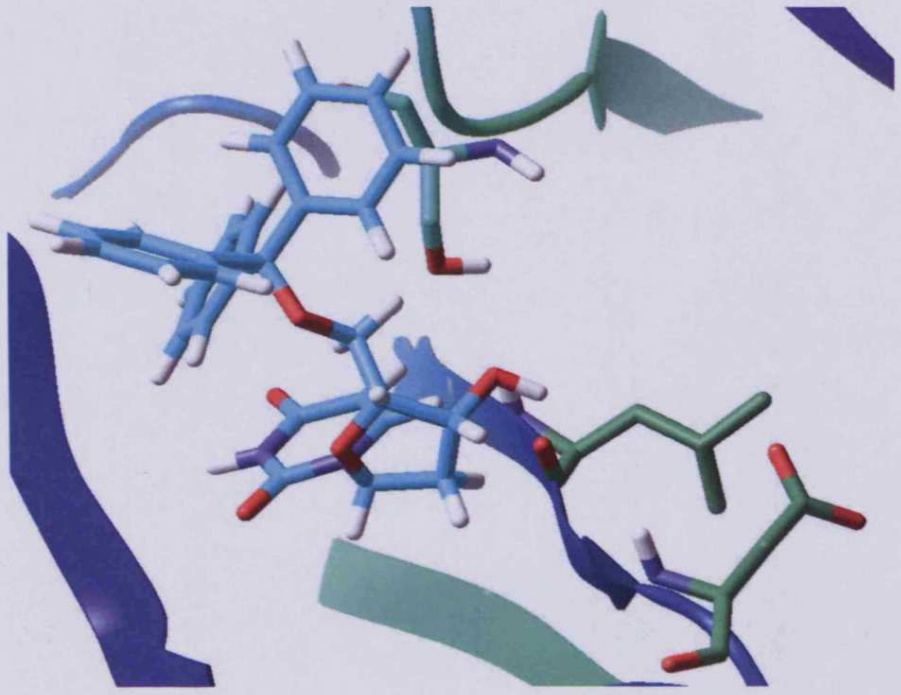


Figure 7.8 Overview of the *P. falciparum* active site in presence of the ligand WSP1361 (cyan) after minimization sampling in absence of the water molecules. The OH group in 3' position does not form any interaction with the protein.

WSP1362, which gave the biggest error in both the models, also has 3'-substituent in the "up" (xylo). This compound has low activity agaist the *P. falciparum* enzyme (*Ki* 350μM). The SAR study suggested that it might simply be due to the high lipophilicity of the substituent in 3'-position. However, the LIE calculation overestimated the free energy of binding for this ligand (Table 7.5). The result of the minimization sampling showed a hydrogen bond between the oxygen of the acetyl and the NH of the Asp 109 backbone. It can be noticed that the water molecule W4 has been pushed away from the pocket

(Figure 7.9). It might suggest that this water plays an important role in the inhibition mechanism.

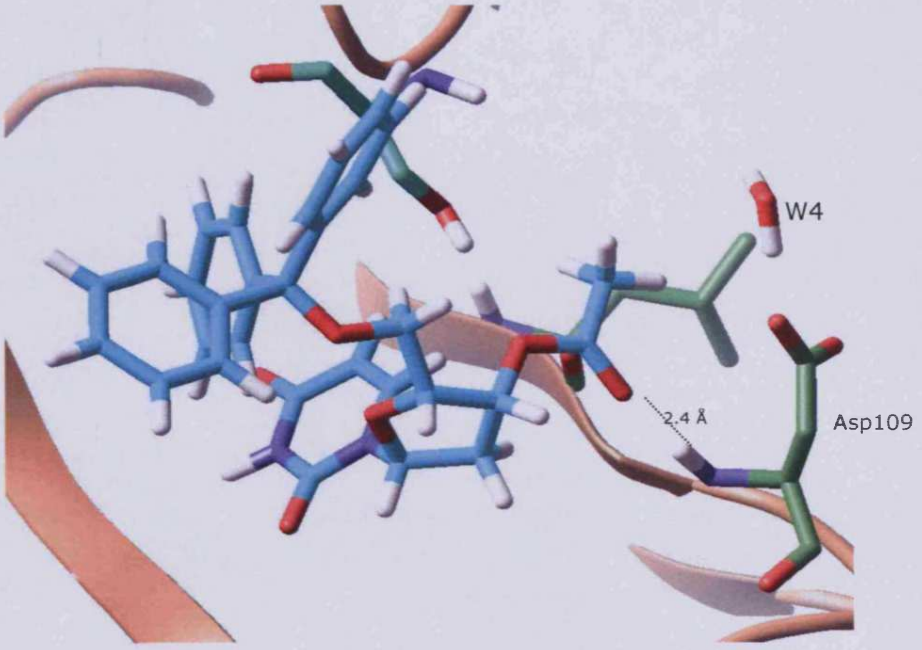


Figure 7.9 Overview of the *P. falciparum* active site in presence of the ligand WSP1362 (cyan) after minimization sampling in presence of water molecules. The water molecule W4 has been pushed away from the pocket losing interaction with the protein.

In general, the LIE did not show a significant difference in free energy of binding between the model 1 and the model 2 (Figure 7.10). Particularly significant is the case of ligand WSP1361, where both methods estimated similar binding energy, although there are some key H-bonding interactions absent in model 2 (model 1 -8.907 kcal/mol, model 2 -8.874 kcal/mol). This result could indicate that the interactions with the conserved water are not particularly significant. Additionally, this could suggest that the LIAISON did not distinguish the two models.

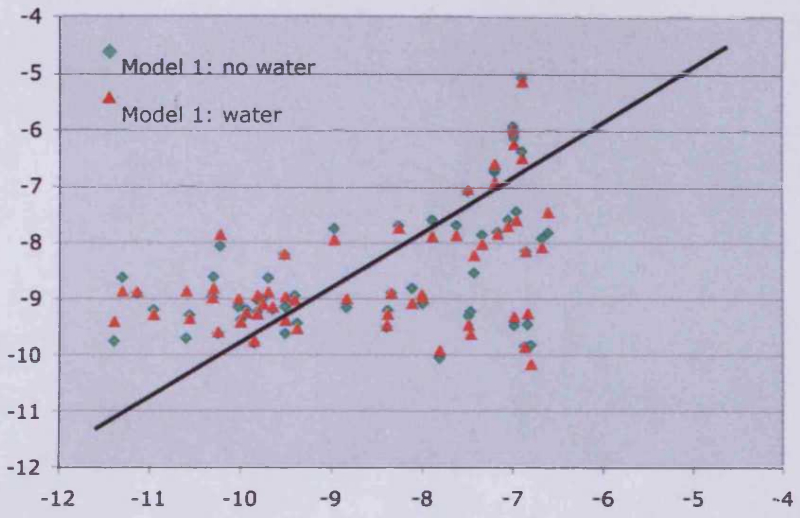


Figure 7.10 Comparison of the experimental versus calculated free energy of binding for two different models: model 1 (red) in absence of water; model 2 in presence of water (green). The plot shows similar energy for both the models.

7.4 Conclusion

In this study it has been shown that the SGB-LIE method could be applied to estimate the free energy of binding with a relatively good level of accuracy for a range of compounds with varying inhibition potencies. The energy minimization method has reproduced experimental data with reasonably small error for the majority of acyclic nucleotides derivatives analogues, indicating that it should be valuable in assessing new *P. falciparum* inhibitors.

Unfortunately, the same results could not be reproduced for the cyclic derivatives. The model for this set of compounds showed poor rates of prediction. It may be due to a variety of reasons:

- The putative binding modes of ligands used to generate the models could be inaccurate due to unpredictable docking failures.
- The LIE method fails to predict the free energy of binding for the cyclic derivatives

- Relatively small errors in accuracy of the assay results could have an impact on correlation, particularly for compounds with low affinity or lack of stability in assay medium.

Further investigation into how all these parameters affect the final models is required before coming to any conclusion.

8 Chapter: Comparative molecular field analysis (CoMFA) of cyclic and acyclic nucleoside derivatives.

CoMFA methodology describes 3D structure activity relationships in a quantitative manner. In this method a model is generated which can be used to predict the activities of related compounds[94]. For this purpose, a set of molecules is first selected which will be included in the analysis. An important precondition is that all molecules in the set must interact with the same receptor (or enzyme, ion channel, transporter) in the same manner, i.e., with identical binding sites in the same relative geometry. This set of molecules is used to construct a "training set" to derive the CoMFA model. In order to prove the validity of the derived model(s), a subset of molecules are selected to be the "test set", to test the CoMFA model to ensure that reasonable activities are predicted by the CoMFA model.

Once a suitable set of molecules has been defined, atomic partial charges are calculated and low energy conformations are generated. A pharmacophore hypothesis is derived to orient the superposition of all individual molecules and afford a rational and consistent alignment.

The molecules are positioned in a 3D-lattice which grid distance is defined. Different atomic probes are used, e.g., a carbon atom, a positively or negatively charged atom, a hydrogen bond donor or acceptor, or a lipophilic probe, interact with the molecules. The energy values, which the probes would experience in the corresponding position of the regular 3D lattice, are used to calculate field values at each grid point.

These "fields" which correspond to hundreds thousands of values, must be correlated with the binding affinities or with other biological activity values. Partial list-squares (PLS) statistical analysis is an appropriate method for this purpose. Normally the cross-validation "leave one out" (LOO) method is used to estimate the "robustness" of the derived model in predicting activities.

In a partial least-squares (PLS) analysis, two factors are important: the number of components used in the regression equation and the correlation coefficient (R) usually squared. This number (R^2), which is also used in (multiple) linear regression, is between zero and one and expresses the quality of the PLS analysis. It indicates the proportion of the variation in the dependent variable (here the activity) that is explained by the regression equation and its value should be as close to one as possible for a optimum correlation. However, R^2 expresses the quality of the data fit rather than the quality of prediction. To express the predictive power of the analysis, the cross-validated R^2, usually indicated by q^2, is used. In cross-validation, one value is left out, a model is derived using the remaining data, and the model is used to predict the value originally left out (LOO). This procedure is repeated for all values, yielding q^2. The final q^2 value is normally (much) lower than R^2, q^2 values greater than 0.5 indicate significant predictive power.

After a cross-validated PLS analysis, it will possible to determine the optimal number of components. These components are linear combinations of the variables, ordered in such a way that the first component will describe most of the variation in the activity, and the second most of the remaining variation, etc. The cross-validated PLS analysis will be carried out for different numbers of components. The number of components should not exceed one-third of the number of molecules. More components would lead to a model that is "over-

trained", a concept also known as over-fitting; it has a better fit to the training data but its predictive power is diminished.

The result of the PLS analysis are usually presented as a set of contour maps. These contour maps show favorable and unfavorable steric regions around the molecules as well as favorable and unfavorable regions for electropositive or electronegative substituents in certain positions. Predictions for the test set (the compounds not included in the analysis) and for other new compounds can be made, either in a quantitative manner, by calculating the molecular fields of these molecules and by inserting the grid values into the PLS model, or by a qualitative inspection of these contour maps.

The current study discussed in this chapter showed new insights into the relationship between the structural information of the series of 26 acyclic and 57 cyclic compounds and the inhibitory potency. The 3D-QSAR (CoMFA) was used to build a model in order to identify structural features in *P. falciparum* dUTPase that could be used to identify new inhibitors.

8.1 Materials and methods

The 3-D structures of the series of 26 acyclic and 57 cyclic nucleoside analogues were constructed using standard geometric parameters of molecular modelling software package SYBYL 7.2. Atomic charges were calculated using the Gasteiger–Hückel protocol. With full minimization by the Tripos force field[95], the optimized structures of these analogues were obtained.

FlexX, a molecular docking program, was used (see chapter 5 and 6) to identify appropriate binding orientations and conformations of these nucleoside analogues in the *P. falciparum* dUTPase. The docking results were optimized by the Schrödinger package using the default parameters of the Liaison module by minimization method (see chapter 7).

8.1.1 Structural alignment.

The putative conformation with the strongest predicted binding affinity to the protein, obtained by the docking calculation, was extracted from the optimized inhibitor-*P. falciparum* dUTPase complex. Considering that 3D-QSAR relies on the assumption that drug-receptor interactions are responsible for the biological activity, then the binding conformations found in the *P. falciparum* enzyme by the docking (Figure 8.1) are probably more suitable for CoMFA analysis than randomly generated low energy conformations. The alignment of these conformations was derived from the docking results and used directly for CoMFA to explore the specific contributions of electrostatic and steric effects to the bioactivities[96].

8.1.2 CoMFA

Steric and electrostatic interactions were calculated using the Tripos force field. The distance-dependent dielectric constant at all intersections in a (2 Å) regular grid was used taking an sp^3 carbon atom as the steric probe and a +1 charge as the electrostatic probe. The cutoff was set to 30 kcal/mol. A regression analysis was carried out using the full cross-validation partial least-squares (PLS) methods of LOO (leave-one-out), with the standard options for scaling variables. For the final model, a non-cross-validation was performed using the optimum number of components to yield to a non-cross-validated R^2 value.

8.2 Results and discussion

In order to determine a putative binding conformation of the two series of acyclic and cyclic analogues to build CoMFA models, the FlexX program was used to dock all the compounds into the active site of *P. falciparum* dUTPase (Figure 8.1). The docking results resemble the conformation of the known X-ray structure of *P. falciparum* dUTPase complexed with cyclic nucleoside derivative WSP869 (see chapter 5 and 6).

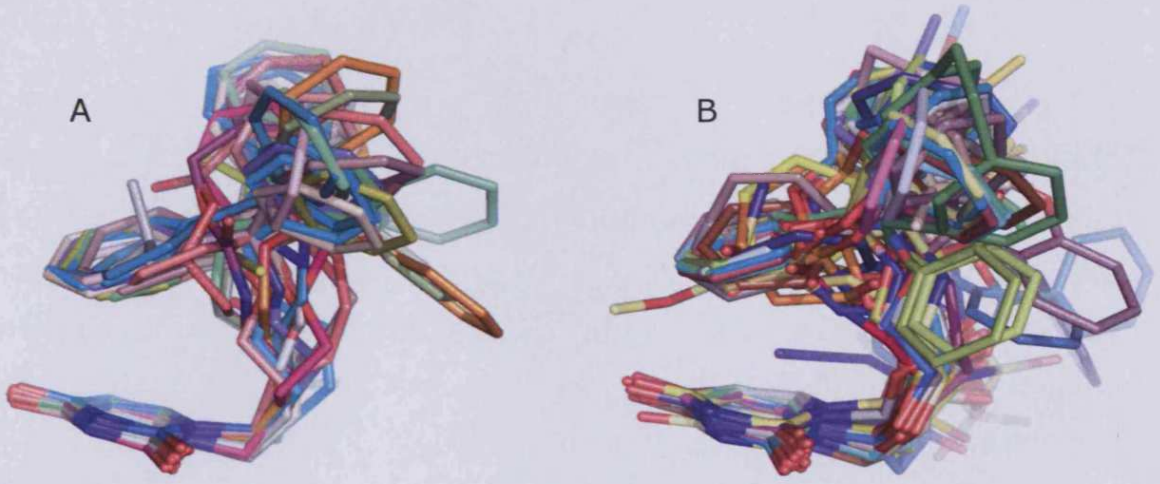


Figure 8.1 CoMFA (and CoMSIA) structure alignment: A) superimposition of 26 acyclic compounds; B) superimposition of 57 cyclic compounds

8.2.1 CoMFA model

The activities of the compounds investigated (see also appendix) are shown in Table 1 as *Ki* (μM), IC_{50} (μM) and pIC_{50} (=-log IC_{50} (M)) values, where the IC_{50} values represent the concentrations of the nucleoside derivatives required to inhibit the enzyme by 50%. The IC_{50} values were determined from the *Ki* values by the relationship[97]:

$$IC_{50} = K_i\,(1 + K_m/[\text{dUTP}]) \qquad (1)$$

Where [dUTP]= 50μM and K_m = 2.42μM.
CoMFA 3D-QSAR methodology was applied using the pIC_{50} values.

A CoMFA analysis was performed separately on two series of compounds (26 acyclic and 57 cyclic). On both the series, a partial least square (PLS) analysis was performed, first in order to find the optimum number of components by determining the cross validated R^2 value (q^2) obtained from the leave-one-out cross validation technique, and then standard error prediction values.

Acyclic compounds

The CoMFA model (26 compounds) containing five components obtained a high correlation coefficient (R^2) of 0.99 and a cross-validation q^2 of 0.74. These statistical parameters are given in Table 8.1, along with experimental and predicted pIC_{50} (-log IC_{50}(M)) values. The training set results for the 26 acyclic compounds are shown graphically in Figure 8.2.

To further validate this CoMFA analysis, a series of calculations were carried out to evaluate the degree to which the CoMFA model is predictive. To do this, three compounds were randomly deleted form the training set, then a new CoMFA model was computed and used to predict the activities of the excluded compounds.

This process was repeated twice, giving a total of 9 predicted values. The results of the three sets of calculations (9 points) are given in Table 8.1, with the predicted values showed as bold entries, and Figure 8.2 shows the results graphically. In all three cases, the CoMFA models so obtained showed high q^2 values indicating statistical robustness (Table 8.1). The predicted test values were generally in good agreement with those of the experimental values of the training set, indicating that the training set model is robust with respect to the training set composition. The average pIC_{50} error for

the 9 predicted values from the three sets of calculations was 0.29, this corresponds to a factor of 1.94 (~2) in the IC_{50} predictions.

WSP ligand	*Ki* (μM)	*IC50* (μM)	*pIC50*	*CoMFA prediction*	*CoMFA Test set 1*	*CoMFA Test set 2*	*CoMFA Test set 3*
869	4.98	0.23	6.64	6.66	6.72	6.67	6.62
870	0.67	0.03	7.51	7.46	7.41	7.47	7.43
872	98	4.52	5.34	5.26	5.57	5.27	5.28
873	44	2.03	5.69	5.63	**5.37**	5.67	5.67
874	49	2.26	5.65	5.59	5.60	5.58	5.69
962	73	3.37	5.47	5.55	5.53	**5.77**	5.45
1000	1.62	0.07	7.13	7.22	7.23	7.20	7.20
1001	2.24	0.10	6.99	7.03	6.99	7.02	7.05
1002	224	10.34	4.99	5.07	5.00	5.04	**5.75**
1004	2.3	0.11	6.97	6.94	**6.76**	6.96	6.95
1005	8.47	0.39	6.41	6.45	6.45	6.46	6.43
1060	0.9	0.04	7.38	7.38	7.39	**7.17**	7.45
1063	32	1.48	5.83	6.15	6.16	6.14	6.16
1064	8	0.37	6.43	6.15	6.16	6.14	6.16
1065	17	0.78	6.11	6.01	6.04	**5.75**	6.03
1066	112	5.17	5.29	5.37	5.34	5.35	5.33
1067	1.69	0.08	7.11	7.11	7.08	7.12	7.07
1069	368	16.99	4.77	4.76	4.79	4.77	4.80
1072	1.56	0.07	7.14	7.17	**7.22**	7.18	7.18
1142	2	0.09	7.03	7.00	7.04	7.02	7.02
1224	4.29	0.20	6.70	6.66	6.69	6.65	**7.16**
1228	3.76	0.17	6.76	6.67	6.70	6.71	6.69
1230	0.2	0.01	8.03	8.08	8.08	8.05	8.04
1241	1.84	0.08	7.07	7.06	7.01	7.02	7.05
1315	1.2	0.06	7.26	7.23	7.26	7.21	**7.07**
1317	0.61	0.03	7.55	7.54	7.52	7.53	7.53
R^2				0.99	0.99	0.99	0.99
q^2				0.74	0.71	0.68	0.70
N^e				5	5	5	5
N^f				26	23	23	23
Field contribiution							
Steric[g]				0.83	0.82	0.83	0.84
Electrostatic[h]				0.17	0.18	0.17	0.16

Table 8.1 Experimental (*Ki*, IC_{50} and pIC_{50}) and CoMFA predicted (pIC_{50}) values for acyclic nucleoside derivatives against *P. falciparum* dUTPase and statistical results of 3D-QSAR CoMFA models. The bold values represent predicted activities of compounds that were not included in the training set. R^2 Correlation coefficient. q^2 cross-validated correlation coefficient after leave-one-out procedure. N^e optimal number of principal components. N^f number of compounds. [g] denotes the steric contribution to the model. [h] denotes the electrostatic contribution to the model.

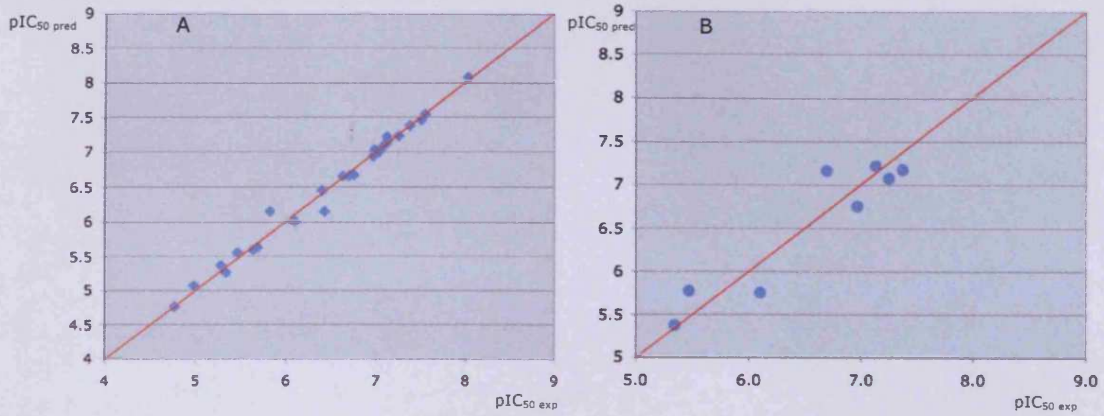


Figure 8.2 plots of the experimental pIC_{50} ($pIC_{50\ exp}$) versus predicted pIC50 ($pIC_{50\ pred}$): A) 26 compounds training set; B) results of the nine compounds from prediction studies (test sets).

The CoMFA result can be represented as 3D "coefficient contour" (Figure 8.3). It shows regions where variations of steric and electrostatic, nature in the structural features of the different molecules contained in the training set, lead to increases or decreases in the activity. The CoMFA fields are presented as contour plots in Figure 8.3. To aid in visualization, compound WSP1230 (the most active inhibitor of the series) is displayed with magenta in the map as reference structure. Green region indicate regions where increased steric bulk is associated with the enhanced *P. falciparum* dUTPase inhibitory activity, and yellow regions suggest where increased steric bulk is unfavorable to activity. Blue regions show areas where more positively charged groups are favorable and while red areas where increased negatively charged groups are favorable for activity.

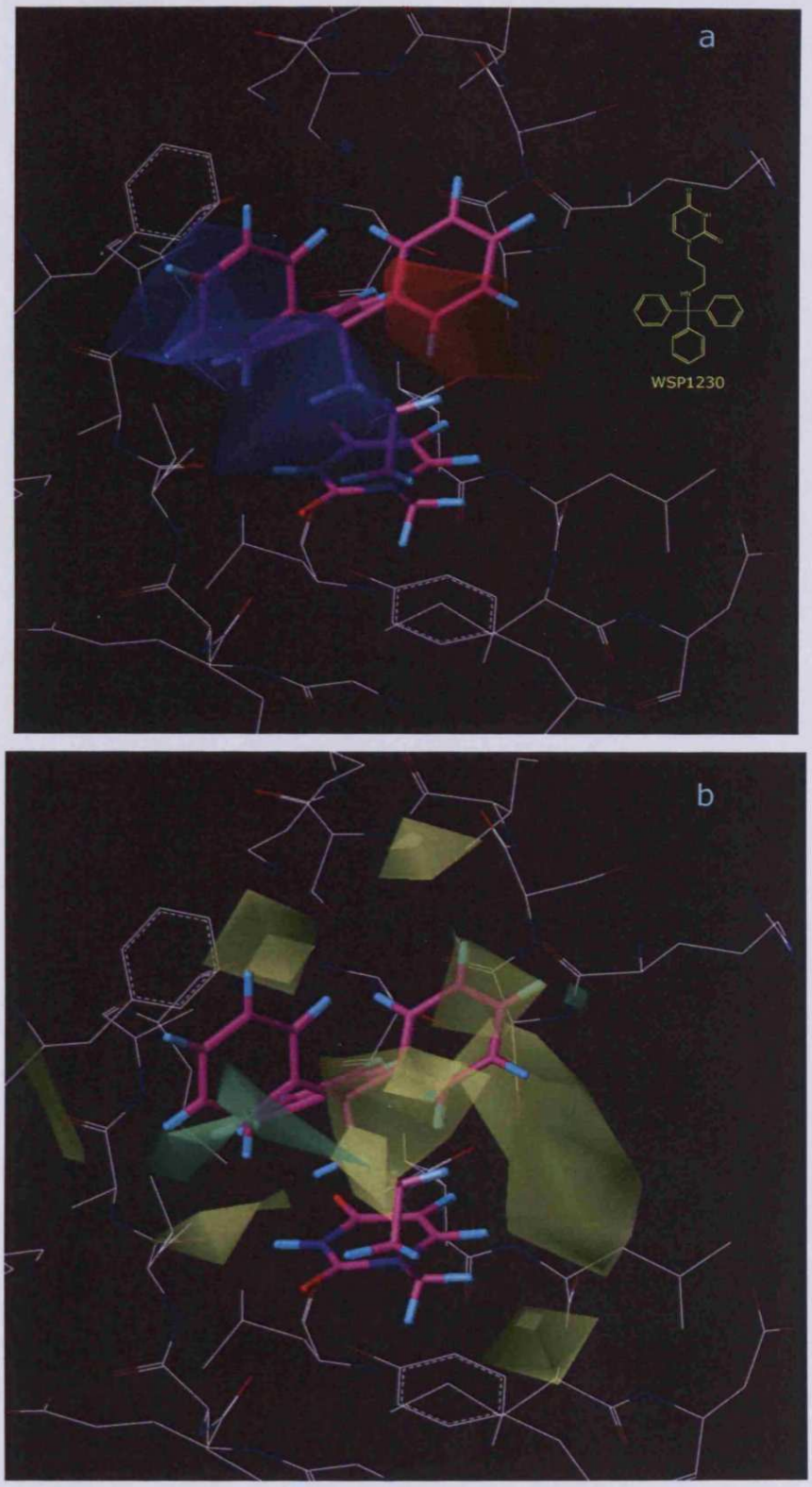


Figure 8.3 CoMFA contour maps in combination with inhibitor WSP1230: (a) the electrostatic field distribution; (b) the steric field distribution. The inhibitor is shown in magenta stick. Positive potential favored areas in blue; positive potential unfavored areas in red. Sterically favored areas in green; sterically unfavored areas in yellow.

The electrostatic contribution to the whole molecular field is only 17.5% (Table 8.1), which suggests this active site favours hydrophobic compounds. The electrostatic contour map is dominated by the region favorable to the positive charges. Such large blue regions are mainly distributed near the linker group, which connects the trityl moiety with the aliphatic spacer. This is consistent with the observation that compounds containing the amino group linker are more active than the ether linked (see chapter 3). In these compounds, the nitrogen is predominantly protonated under physiological conditions and this argument can also be extended to the bound ligands. The red contour region is distributed in an area mainly exposed to the solvent and thus may not possess any significant interactions to our protein.

The steric contour map is dominated by yellow regions, which indicate the unfavorable steric regions. A high density of yellow region can be identified near Ser 92. This might be in agreement with some experimental data of the cyclic derivatives. The cyclic compounds WSP1362 and WSP1363 (see appendix 1), which have relatively large substituents in 3'-position that point toward the Ser 92, have very poor activity compared with WSP1361. This compound has a small substituent in 3'-position than that in WSP1362 or WSP1363. Hence, the different activities of these compounds might be due to unfavorable steric interactions (see chaper 6).

And finally, the favorable green contour map is present only in a region of the active site exposed to the solvent.

8.2.2 Cyclic compounds

A CoMFA model was developed for the cyclic nucleoside analogues (57 compound). The model contained two components (Table 8.2), showed a correlation coefficient of 0.61 and a cross-validation q^2 of 0.36. These values do not indicate a good statistical correlation, and suggested poor predictive ability for the model. The cyclic compound CoMFA model is shown graphically in Figure 8.4, it can be seen that the biggest contribution to error is from compound WSP1073 (Figure 8.5). This might be explained by analysing the conformation of this compound in the active site. Its trityl moiety is placed out the hydrophobic pocket (see chapter 6). This difference in the conformation compared with the rest of the compounds might be sufficient to classified WSP1073 as an outlier. In an attempt to improve the model, WSP1073 was excluded.

A new CoMFA model (based on) was derived from the remaining compounds. It contained five components, gave an R^2 value of 0.95 and a q^2 of 0.60 after cross-validation. These statistical parameters are given in Table 8.2, along with the experimental and predicted pIC_{50} values. These values seemed to indicate improved statistical correlation and predictive ability compared to the initial model (Figure 8.6).

To validate this CoMFA analysis, a series of calculations were carried out to evaluating its statistical robustness. To do this, six compounds randomly selected were removed from the training set, then a new CoMFA model was calculated and used to predict the activities of the excluded compounds. As already described for the acyclic compounds the process was repeated two more times, giving a total of 18 predicted values.

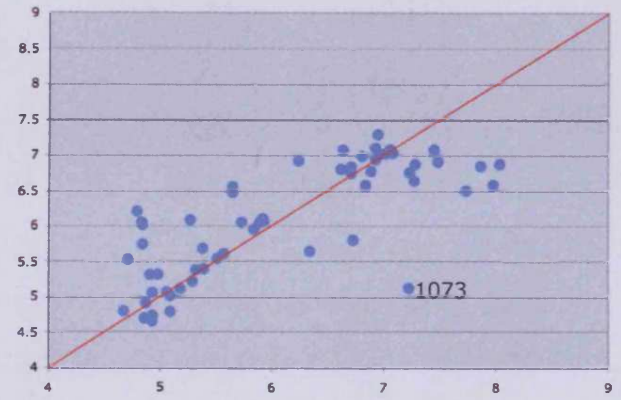


Figure 8.4 Plots of the experimental pIC_{50} ($pIC_{50\ exp}$) versus predicted pIC50 ($pIC_{50\ pred}$) with 57 compounds training set. The compound WSP1073 is highlighted as an outlier.

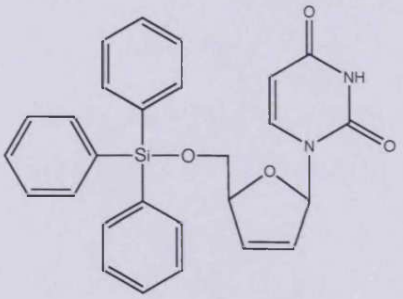


Figure 8.5 Compound WSP1073.

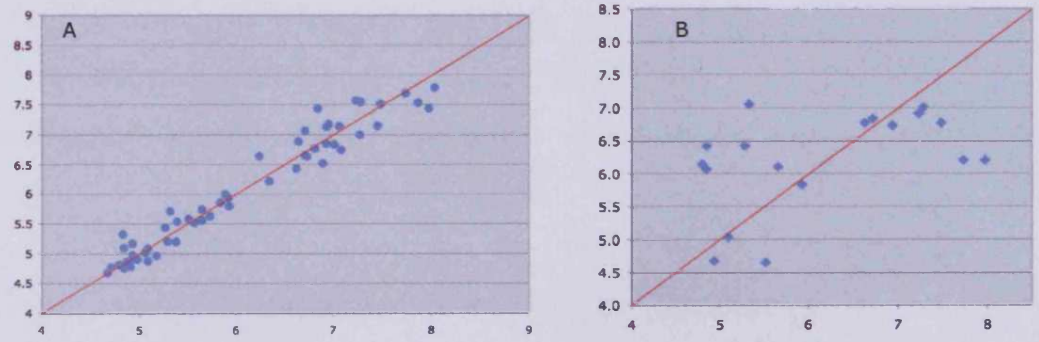


Figure 8.6 Plots of the experimental pIC_{50} ($pIC_{50\ exp}$) versus predicted pIC_{50} ($pIC_{50\ pred}$) for the model excluding WSP1073: A) 56 compounds training set; B) results of the 18 compounds from prediction studies.

The results of the three sets of calculations (18 points) are given in Table 8.2, and the predicted values are shown as red and bold entries. Figure 8.6 shows the results of the 18 predictions, graphically. In more than half of the cases, the CoMFA models so obtained were statistically significant (Table 8.2) and the predicted test values were in generally good agreement with those of the training set. However in a few cases (38% of the predicted compounds) the model showed poor prediction, suggesting that the training set model may not be sufficiently robust with respect to the training set composition.

wsp ligand	*Ki (μM)*	*IC_{50} (μM)*	*pIC_{50}*	*CoMFA Prediction*	*CoMFA Test set 1*	*CoMFA Test set 2*	*CoMFA Test set 3*
483	1.8	0.08	7.08	6.74	5.31	6.82	7.06
659	316	14.59	4.84	5.09	4.93	**6.06**	5.53
660	206	9.51	4.91	4.78	6.80	4.95	5.03
661	178	8.22	5.09	5.09	7.07	5.05	**5.03**
662	294	13.57	4.87	4.86	7.18	4.91	4.65
665	26	1.20	5.92	5.93	**5.83**	6.35	6.00
666	142	6.56	5.18	4.96	6.57	4.94	5.06
668	189	8.73	5.06	5.02	5.65	4.92	5.02
669	228	10.53	4.98	4.90	6.78	5.06	5.05
680	178	8.22	5.09	4.87	7.11	4.97	4.85
704	67	3.09	5.51	5.58	**4.65**	5.06	5.53
706	115	5.31	5.27	5.43	5.41	**6.42**	6.04
785	324	14.96	4.83	5.32	4.78	5.91	5.92
786	40	1.85	5.73	5.62	4.61	5.59	6.24
789	468	21.61	4.67	4.67	6.62	4.83	4.98
790	88	4.06	5.39	5.53	7.45	5.57	5.51
826	252	11.63	4.93	4.97	7.97	4.99	**4.67**
828	10	0.46	6.34	6.21	5.93	5.78	5.47
829	298	13.76	4.86	4.81	5.65	4.95	4.42
830	254	11.73	4.93	4.92	6.93	4.83	4.74
864	1.9	0.09	7.06	7.14	6.71	7.05	7.01
865	58	2.68	5.57	5.51	6.95	5.59	5.30
866	255	11.77	4.93	5.17	6.93	4.52	4.87
867	1.16	0.05	7.27	7.00	5.84	6.89	6.58
869	4.98	0.23	6.64	6.88	6.84	6.84	7.07
871	0.2	0.01	8.03	7.78	7.28	7.30	6.99
946	12.42	0.57	6.24	6.63	6.81	6.84	7.10
948	89	4.11	5.38	5.19	5.89	6.16	5.40
949	2.8	0.13	6.89	6.52	6.73	6.54	6.83
950	4.2	0.19	6.71	6.64	7.01	6.88	6.70
952	110.7	5.11	5.29	5.20	5.32	5.03	5.28

953	312.8	14.44	4.84	4.75	7.86	4.99	**6.42**
958	426	19.67	4.71	4.77	5.65	5.38	5.44
1059	5.2	0.24	6.62	6.43	6.80	6.54	**6.77**
1068	0.72	0.03	7.48	7.51	**6.77**	7.47	7.03
1134	0.76	0.04	7.45	7.15	5.33	7.05	7.01
1149	0.23	0.01	7.97	7.44	4.79	7.79	**6.20**
1203	25.5	1.18	5.93	5.79	5.30	5.78	6.21
1205	48	2.22	5.65	5.74	5.06	5.68	6.02
1206	1.27	0.06	7.23	7.57	**6.91**	7.33	6.79
1207	1.43	0.07	6.93	7.13	4.58	**6.74**	7.17
1229	4.22	0.19	6.71	7.06	5.19	**6.83**	7.20
1243	2.17	0.10	6.95	7.17	5.04	7.21	7.26
1244	2.56	0.12	6.93	6.85	4.90	6.89	7.07
1245	31.55	1.46	5.84	5.85	5.13	6.04	5.74
1247	3.11	0.14	6.84	7.44	5.29	7.79	6.20
1288	1.14	0.05	7.28	7.54	5.50	**7.01**	6.87
1289	0.4	0.02	7.73	7.69	**6.20**	7.44	6.60
1290	3.36	0.16	6.81	6.76	4.85	7.01	7.12
1291	27.6	1.27	5.89	5.99	5.54	5.95	6.07
1292	4.06	0.19	6.73	6.63	5.01	6.08	5.85
1324	2.11	0.10	7.01	6.83	5.87	6.74	**7.05**
1329	103.2	4.76	5.32	5.71	4.91	5.24	5.30
1361	0.3	0.01	7.86	7.53	4.89	6.99	6.79
1362	350	16.16	4.79	4.81	**6.14**	5.47	6.19
1363	49	2.26	5.65	5.55	7.08	**6.10**	6.03
R^2				0.95	0.91	0.85	0.71
q^2				0.60	0.59	0.55	0.57
N^e				5	4	3	2
N^f				56	50	50	50
Field contribiution							
Steric[g]				0.73	0.74	0.79	0.81
Electrostatic[h]				0.27	0.26	0.21	0.19

Table 8.2 Experimental (*Ki*, IC_{50} and pIC_{50}) and CoMFA predicted (pC_{50}) values for cyclic nucleoside derivatives against *P. falciparum* dUTPase and statistical results of 3D-QSAR CoMFA models. The bold values represent predicted activities of compounds that were not included in the training set. R^2 Correlation coefficient. q^2 cross-validated correlation coefficient after leave-one-out procedure. N^e optimal number of principal components. N^f number of compounds. [g] denotes the steric contribution to the model. [h] denotes the electrostatic contribution to the model.

The average pIC_{50} error for the 18 predicted values from the three sets of calculations was 0.77, corresponding a factor of 5.86 uncertainty in the IC_{50} predictions.

The electrostatic contribution to the whole molecular field is 27.2%. An important feature of this CoMFA model is that the electrostatic contour map is more dominated by the red regions

rather than blue (Figure 8.7a). A large red region is displayed in close proximity to the nitrogen group, which links the trityl moiety with the rest of the molecule. This is in contrast with the experimental data, which indicates that the presence of protonated nitrogen (see charges 3-6), usually increases the activity of these compounds. Although the models are statistically robust, this lack of agreement may highlight some ambiguity in this CoMFA analysis. In addition, a blue region is present at the bottom of the active site, near the uracil moiety. This is consistent with the fact that, in this region, the compounds are stabilized by the hydrogen bonds formed between the two residues (Asn 103 and Ile 117) and the two carbonyl oxygens of the uracil group (see chapter 2). The largest blue region is in close proximity to the substituent in the 3'-position of the ligand. This is in agreement with the requirement for a polar group in this part of the binding site, due to the presence of an aspartate residue (Asp 109, see also chapter 4). A large blue region is also observed near the Ser 95 and the trityl moiety, which is partially in agreement with the requirements of the binding site.

The rest of the red regions seem to be consistent with the requirement of the protein structure. One red region is distributed near the Lys 96, which this is consistent with the fact that the lysine has a protonated nitrogen, and is potentially able to interact with a negative ligand substituent. In addition, a red region is adjacent to Arg 91, it may interact with the negative functionality on the ligand. The remaining red regions are mainly distributed in areas of the active site exposed to the solvent.

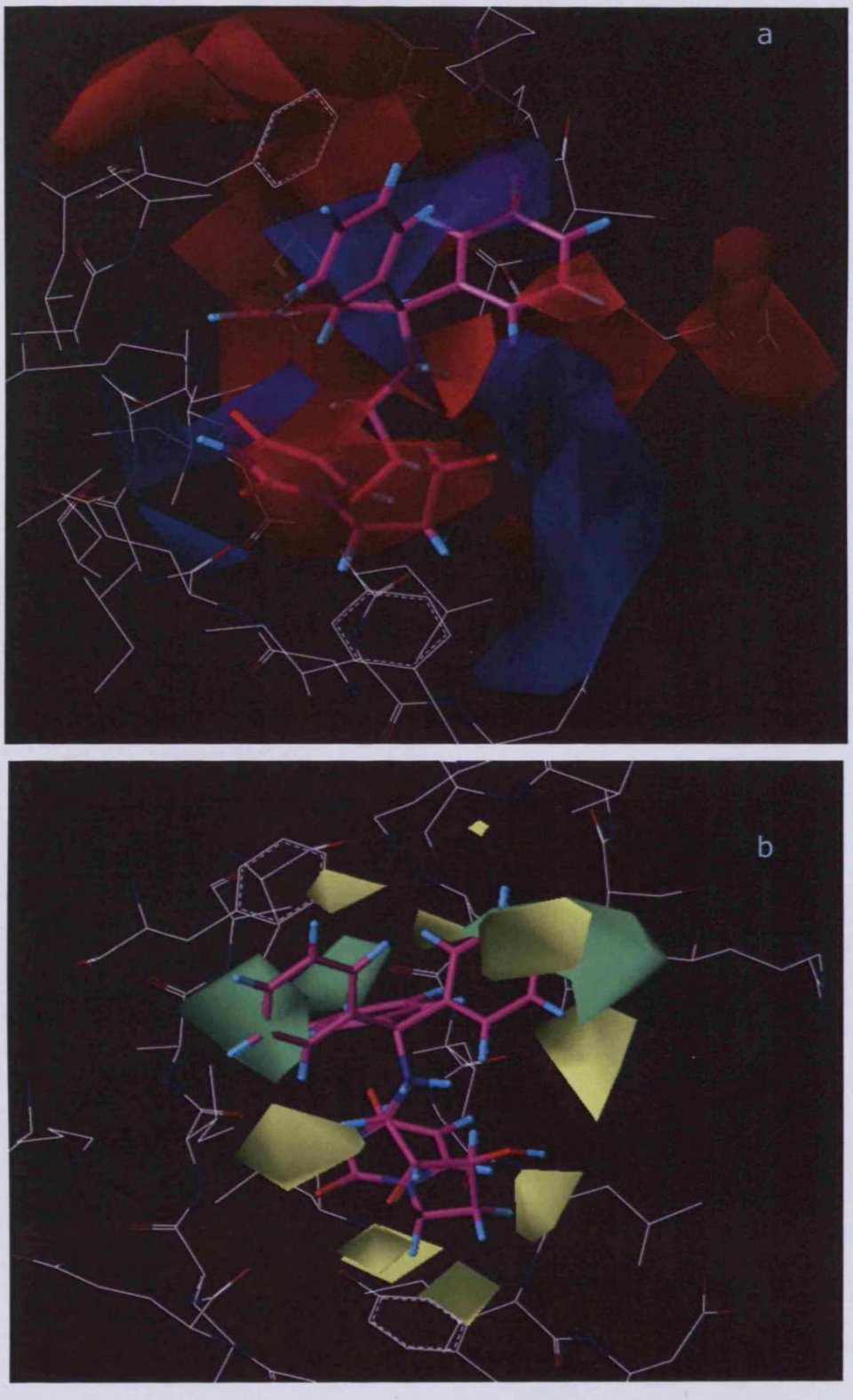


Figure 8.7 CoMFA contour maps in combination with inhibitor WSP1230: (a) the electrostatic field distribution; (b) the steric field distribution. The inhibitor is shown in magenta stick. Positive potential favored areas in blue; positive potential unfavored areas in red. Sterically favored areas in green; sterically unfavored areas in yellow.

The steric field CoMFA contour maps (Figure 8.7b) showed that

there are favorable steric regions near each phenyl ring of the trityl group. Two of the phenyl rings are mainly exposed to the solvent; therefore the presence of a green region in these areas is not particularly surprising and does not suggest any interaction with the protein. One of the phenyl rings points toward the bottom of the active site; the green region near this aromatic ring is consistent with the presence of a relatively shallow pocket in this area.

Consistently, most yellow regions are in agreement with the requirement of the active site. In fact, some of the sterically unfavorable regions are present near the Tyr 112, which suggest that a substituent in 2'-position on the sugar moiety could lead to a loss of activity[34]. A yellow region is also present in proximity to the Ser 92. This might explain the low activity of the two cyclic compounds WSP1362 and WSP1363, which have relatively large substituents in the 3'-position and that probably point toward the Ser 92. The rest of the yellow regions are found in areas that are mainly exposed to the solvent.

8.2.3 CoMFA model of the completed series cyclic and acyclic compounds.

In an attempt to design a CoMFA model to overcome some of the problems encountered in the case of the cyclic compound series. A suitable training set was created using both cyclic and acyclic derivatives.

This new CoMFA model was composed of 81 compounds (all compounds from the acyclic and cyclic series' were used with the exception of WSP1073). This problems surrounding this compound have been described in great detail in the previous section. The CoMFA analysis was performed using the same procedures employed for the previous models (as described in the Materials and Methods Section). In addition, the structural alignment and the grid parameters were kept same (see section above). For the test set, eight points were deleted, and this procedure was repeated three times, giving a total number of 24 values for the model to predict. The training set and the results of the three test sets are given in the Table 8.3, in which the predicted values are given in bold entries. Additionally, the Figure 8.8 shows graphically the training set of the 81 compounds and the results of the 24 predictions.

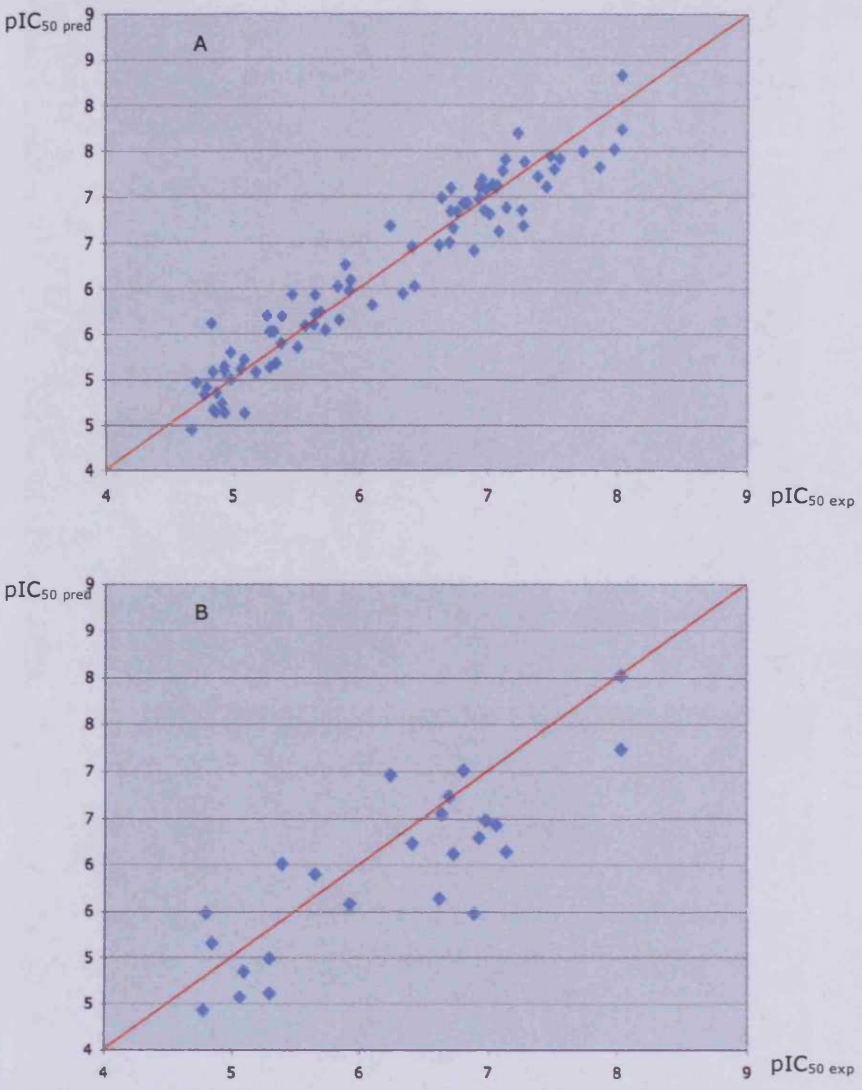


Figure 8.8 Plots of the experimental pIC_{50} ($pIC_{50\ exp}$) versus predicted pIC_{50} ($pIC_{50\ pred}$): A) 81 compounds training set; B) results of the 24 compounds prediction studies.

The CoMFA models obtained were statistically significant and their predictions were generally in good agreement with the experimental values (R^2 0.93, q^2 0.64, see Table 8.3). The average pIC_{50} error for the 24 predicted values from the three sets of calculations was 0.5, corresponding to a factor of 3.2 uncertainty in the IC_{50} [98]. This model showed better potency prediction compared with the cyclic CoMFA model.

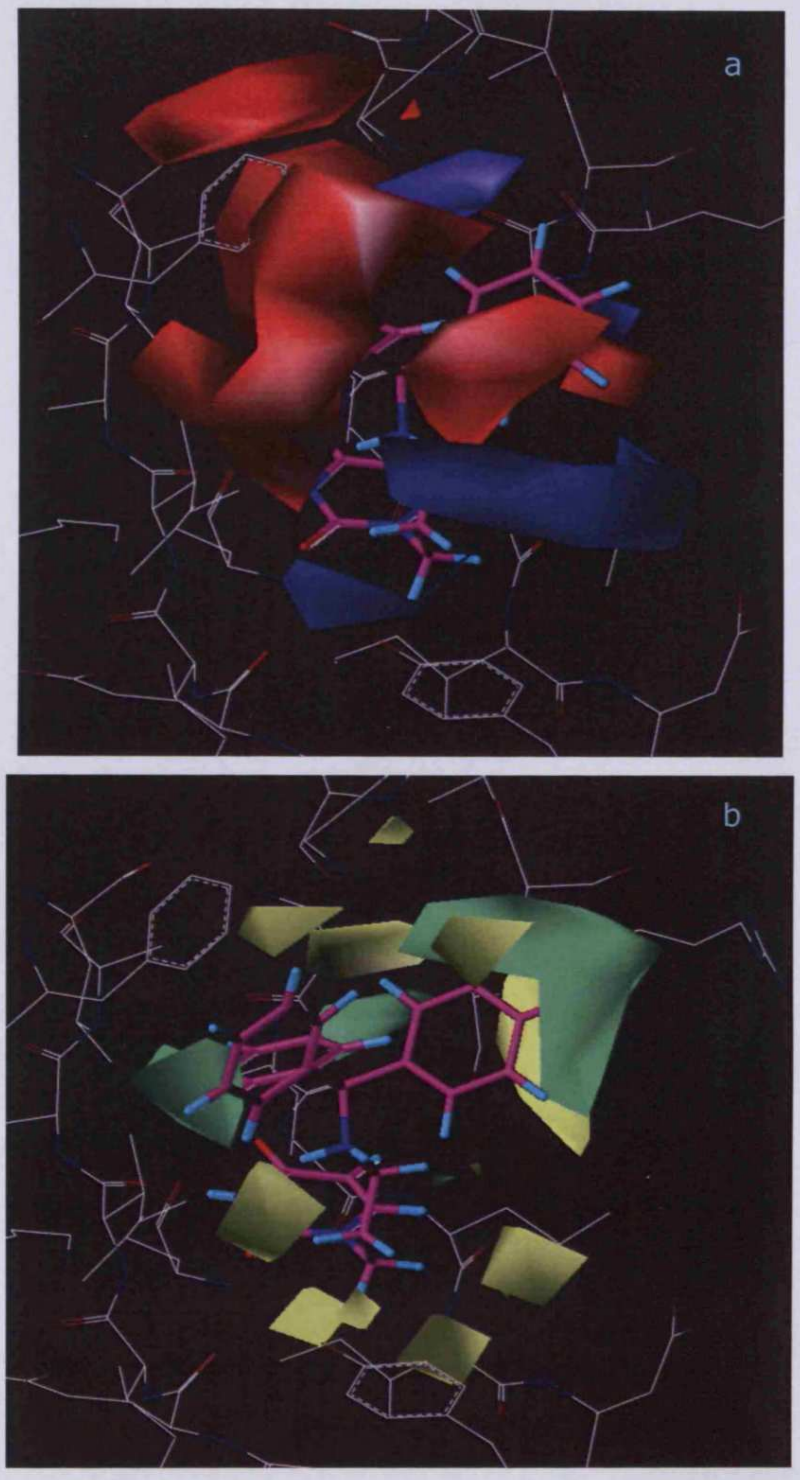


Figure 8.9 CoMFA contour maps in combination with inhibitor WSP1230: (a) the electrostatic field distribution; (b) the steric field distribution. The inhibitor is shown in magenta stick. Positive potential favored areas in blue; positive potential unfavored areas in red. Sterically favored areas in green; sterically unfavored areas in yellow.

The contour map with the compound WSP1230 displayed are showed in Figure 9a-b. The electrostatic contribution to the whole

molecular field is 25.3%. Figure 8.9a displays a favorable region for positive charge near the amino group, which links the trityl moiety with the aliphatic linker.

This is an important feature of the CoMFA model, because it is in good agreement with requirements of the binding site and the experimental values. The nitrogen group points to the solvent, with apparently no direct interaction with the protein. However, the X-ray structure of *P. falciparum* dUTPase in complex with WSP871[73] revealed the presence of a water molecule which was at hydrogen bonding distance from the nitrogen group in 5'-position and the OH of the conserved Tyr 112 (see chapter 5). Near this nitrogen group, the CoMFA model showed a distinct blue region sandwiched between two large solvent exposed red regions. From the experimental values it seemed that the orientation of the nitrogen group might affect the activity of these compounds. In fact, a careful analysis of the putative conformation of the ligand WSP1224 compared to WS1230 (Figure 8.10a and b respectively) showed the nitrogen group pointing to the inside of the pocket, near one of the red regions (Figure 8.11).

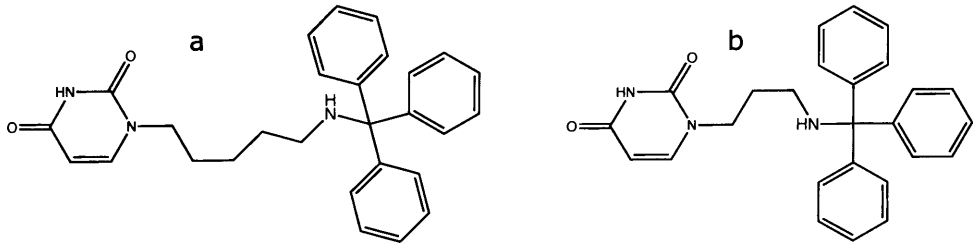


Figure 8.10 schematic representation of a) WSP1224; b) WSP1230

The change in the conformation affecting the position of the "5-amino" group is probably due to the different length of the spacer compared with that of WSP1230. This might affect the interaction with the water molecule, hence influencing the activity of the inhibitors (WSP1230 *Ki* of 0.2 μM and 1224 *Ki* of 4.29 μM).

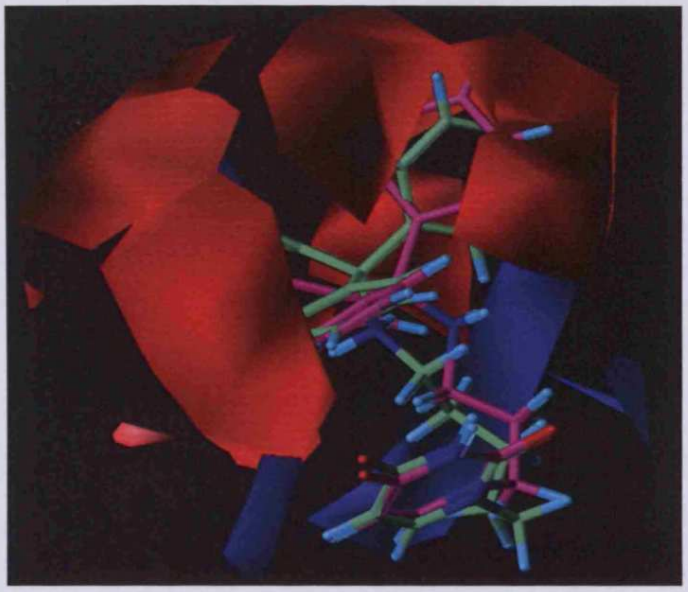


Figure 8.11 CoMFA contour maps in combination with inhibitors WSP1230 (magenta) and WSP1224 (green) superimposed, showing the electrostatic field distribution. Positive potential favored areas in blue; positive potential unfavored areas in red.

The presence of large red regions (favourable to a negative charge) in the other models, near the nitrogen group might be explained by analyzing the composition of the training set. The majority of the compounds, in both the cyclic and acyclic series, have the oxygen-linker in place of the more favoured amino-linker group. This group has a relatively more negative character. To further confound issues some compounds did not show significant changes in activity with respect to the linkers i.e. compounds, which are expected to be more active because of the presence of the amino-linker group, had activity values relatively close in magnitude to their ethoxy-linker analogues. The small activity differences may not be well detected by CoMFA analysis. It maybe due to these reasons that the CoMFA model, of the cyclic series failed to generate the appropriate blue contour (a region favorable to positive charge) near the linker. A more explicit way of describing the problem is to say that the training set of the cyclic series did not contain sufficient numbers of compounds with a nitrogen in the 5'-position having sufficient activity to affect the CoMFA model.

A Combination of the two nucleoside derivatives series yields a new training set that has more compounds with the nitrogen group in the linker. In this way we were able to generate a CoMFA model, which is in good agreement with the experimental data.

The other areas of the electrostatic contour map favorable to positive charge are found in 3 additional defined regions. One blue region is near the Tyr 112 and the main chain of Glu 115. This indicates that compounds with a low electron density are preferred in this part of the active site. This is consistent with the fact that the tyrosine residue can be an acceptor of hydrogen bond (see chapter 2-5), and additionally form π-cation interaction. The remaining two blue regions are displayed near the Ser 92 and Ser 90 respectively, which is in agreement with the requirements of this active site.

Further analysis of the new model shows that a red contour map, which represents the regions favorable to negative charge, dominates the top section of the binding pocket. This is consistent with fact that all the compounds with trityl moiety, which interacts principally with the superior section of the active site, show a good activity. This increase stability is not due only to a hydrophobic interaction, but the red region indicates that this part of the active site prefers a high electron density. In fact the trityl group has an elevated electron density due to the presence of the aromatic rings. It is also explaining that the replacement of the trityl group with a purely hydrophobic group such as the tri-isopropylsilyl group (compound WSP951, see appendix 1) leads to a reduced activity. The tert-butyldiphenylsilyl group has more activity than the tri-isopropyl derivative due to the presence of two aromatic rings.

The steric contour map showed in Figure 8.9b displays favorable regions for steric interactions (green color) mainly in the top section of the binding pocket, near each phenyl ring of the trityl moiety. Two

of the largest regions are mainly exposed to the solvent; therefore they do not seem to be importanti in increasing potency. Another large green region, less exposed to solvent, indicates favorable steric interaction between the residues Ser 92 and Asn 103, which might likely form a relatively broad and shallow pocket.

Some of the yellow regions are found near the Tyr 112, in agreement with the requirements of the active site (see chapter 2). A large yellow region is present near the Ser 92. It may explain the low activity of the compound WSP1205, which has an azide group in 3'-position. This group is not extremely bulky; however it is very rigid, and the poor flexibility forces this group to occupy an unfavorable steric region.

WSP ligand	*Ki (μM)*	*IC50 (μM)*	*pIC50 (M)*	*CoMFA prediction*	*CoMFA set test 1*	*CoMFA set test 2*	*CoMFA set test 3*
483	1.8	0.08	7.08	6.63	6.67	**5.90**	6.72
659	316	14.59	4.84	5.08	**5.16**	4.60	5.05
660	206	9.51	4.91	4.74	4.72	4.70	4.70
661	178	8.22	5.09	5.22	**4.84**	5.14	4.73
662	294	13.57	4.87	4.85	5.03	4.84	4.82
665	26	1.20	5.92	5.98	6.06	5.87	**5.58**
666	142	6.56	5.18	5.08	5.16	5.09	5.18
668	189	8.73	5.06	5.11	5.25	5.06	**4.57**
669	228	10.53	4.98	5.29	5.09	4.97	4.91
680	178	8.22	5.09	4.63	4.82	4.71	5.15
704	67	3.09	5.51	5.35	5.40	5.25	5.47
706	115	5.31	5.27	5.70	5.48	5.70	5.63
785	324	14.96	4.83	5.61	5.58	5.53	5.53
786	40	1.85	5.73	5.54	5.65	5.54	5.47
789	468	21.61	4.67	4.45	4.40	4.65	4.46
790	88	4.06	5.39	5.70	5.51	**6.01**	5.71
826	252	11.63	4.93	5.10	4.98	5.15	5.17
828	10	0.46	6.34	5.94	5.96	5.97	5.98
829	298	13.76	4.86	4.64	4.83	4.65	4.65
830	254	11.73	4.93	4.63	5.20	5.16	4.72
864	1.9	0.09	7.06	7.10	7.16	**6.43**	7.16
865	58	2.68	5.57	5.59	5.48	5.63	5.51
866	255	11.77	4.93	5.14	4.75	4.79	5.13
867	1.16	0.05	7.27	6.68	6.88	6.67	6.75
869	4.98	0.23	6.64	6.99	**6.55**	6.98	6.95
870	0.67	0.03	7.51	7.30	7.45	7.26	7.23
871	0.2	0.01	8.03	8.33	**7.24**	7.79	7.70
872	98	4.52	5.34	5.18	5.07	5.17	5.16
873	44	2.03	5.69	5.74	5.75	5.71	5.72

874	49	2.26	5.65	5.60	5.57	5.66	5.51
946	12.42	0.57	6.24	6.68	6.72	6.64	**6.96**
948	89	4.11	5.38	5.40	5.15	5.51	5.30
949	2.8	0.13	6.89	6.41	6.49	6.35	**5.48**
950	4.2	0.19	6.71	6.84	6.87	6.82	6.95
952	110.7	5.11	5.29	5.52	5.25	5.49	**4.61**
953	312.8	14.44	4.84	4.66	4.75	5.03	4.61
958	426	19.67	4.71	4.96	4.83	4.97	4.95
962	73	3.37	5.47	5.94	6.06	5.90	5.92
1000	1.62	0.07	7.13	7.41	7.38	7.41	7.46
1001	2.24	0.10	6.99	7.08	**6.48**	7.12	6.99
1002	224	10.34	4.99	5.00	5.23	5.31	5.20
1004	2.3	0.11	6.97	6.85	6.85	6.91	6.84
1005	8.47	0.39	6.41	6.45	6.39	**6.23**	6.41
1059	5.2	0.24	6.62	6.47	6.52	**5.43**	6.46
1060	0.9	0.04	7.38	7.22	7.21	7.20	7.38
1063	32	1.48	5.83	6.02	6.11	6.03	6.06
1064	8	0.37	6.43	6.02	6.11	6.03	6.06
1065	17	0.78	6.11	5.82	5.91	5.80	5.91
1066	112	5.17	5.29	5.13	**5.00**	5.08	5.54
1067	1.69	0.08	7.11	7.29	7.20	7.27	7.24
1068	0.72	0.03	7.48	7.45	7.50	7.43	7.61
1069	368	16.99	4.77	4.83	**4.43**	4.85	4.78
1072	1.56	0.07	7.14	6.88	6.85	**6.14**	6.83
1134	0.76	0.04	7.45	7.11	7.16	7.14	7.17
1142	2	0.09	7.03	7.13	7.12	7.10	7.14
1149	0.23	0.01	7.97	7.52	7.80	7.47	7.79
1203	25.5	1.18	5.93	6.09	5.98	6.17	5.98
1205	48	2.22	5.65	5.93	5.75	5.71	5.78
1206	1.27	0.06	7.23	7.70	7.49	7.68	7.76
1207	1.43	0.07	6.93	7.11	7.05	7.00	7.07
1224	4.29	0.20	6.70	6.50	6.60	6.42	**6.74**
1228	3.76	0.17	6.76	6.84	6.82	6.78	6.86
1229	4.22	0.19	6.71	7.10	7.04	7.04	6.80
1230	0.2	0.01	8.03	7.74	8.26	8.35	**8.02**
1241	1.84	0.08	7.07	7.12	7.04	7.21	7.17
1243	2.17	0.10	6.95	7.19	7.11	7.17	7.17
1244	2.56	0.12	6.93	6.98	7.06	7.10	**6.29**
1245	31.55	1.46	5.84	5.65	5.74	5.64	5.63
1247	3.11	0.14	6.84	6.93	6.86	6.93	6.84
1288	1.14	0.05	7.28	7.38	7.25	7.32	7.28
1289	0.4	0.02	7.73	7.50	7.56	7.56	7.54
1290	3.36	0.16	6.81	6.93	6.86	**7.01**	7.03
1291	27.6	1.27	5.89	6.26	6.17	6.21	6.26
1292	4.06	0.19	6.73	6.66	6.83	**6.12**	6.74
1315	1.2	0.06	7.26	6.86	6.88	6.83	6.84
1317	0.61	0.03	7.55	7.42	7.46	7.41	7.53
1324	2.11	0.10	7.01	6.81	6.93	6.78	6.73
1329	103.2	4.76	5.32	5.53	5.31	5.55	5.58
1361	0.3	0.01	7.86	7.33	7.43	7.27	7.30
1362	350	16.16	4.79	4.91	**5.48**	4.88	4.80
1363	49	2.26	5.65	5.72	5.94	5.91	5.97
R^2				0.93	0.94	0.93	0.94

q^2		0.64	0.61	0.63	0.61
N^e		5	5	5	5
N^f		81	73	73	73
Field contribiution					
Steric[g]		0.75		0.77	0.75
Electrostatic[h]		0.25		0.23	0.25

Table 8.3 Experimental (*Ki*, IC_{50} and pIC_{50}) and CoMFA predicted (pC_{50}) values for all the nucleoside derivatives against *P. falciparum* dUTPase and statistical results of 3D-QSAR CoMFA models. The bold values represent predicted activities of compounds that were not included in the training set. R^2 Correlation coefficient. q^2 cross-validated correlation coefficient after leave-one-out procedure. N^e optimal number of principal components. N^f number of compounds. [g] denotes the steric contribution to the model. [h] denotes the electrostatic contribution to the model.

8.2.4 Comparative molecular similarity index analysis (CoMSIA).

In this section the comparative molecular similarity index analysis approach was used to investigate structure-activity relactionships with the nucleoside derivatives (cyclic and acyclic). In contrast to CoMFA, which uses standard Lennard-Jones and Coulombic terms, CoMSIA indices are calculated using Gaussian-type functions[99]. The CoMSIA indices vary less rapidly at short gridpoint-molecule distances than do the CoMFA energy. In addition, the lattice itself is used to elucidate similarity, and CoMSIA maps are thought by some workers to be more useful since they focus on the areas that are actually occupied by ligand atoms[98]. CoMSIA analyses are also typically less sensitive to small alignment errors than are CoMFA analyses, since there are no singularities at ligand atom sites. The CoMSIA analysis was performed following the procedure outlined previously for the CoMFA analysis. The CoMSIA alignment and grid were the same as those used in the CoMFA analysis discussed previously (see Figure 8.1). The result for the training set, are given in Table 8.4 and graphically in Figure 8.12. The resulting CoMSIA field maps are shown in Figure 8.13. We also carried out an additional set of calculations in which the activities of these species were predicted were predicted using a reduced training set, basically as discussed previously. These results are shown in Table 8.4 and are generally similar to the CoMFA results.

For the training set was obtained an R^2 of 0.95 and a q^2 of 0.69 for 81 compounds using six components (Table 8.4), using Gasteiger-Huckel charges. As with the CoMFA, a series of calculations were carried out to validate the CoMSIA model and evaluate to what extent the model had predictive value. To do this, we again used the 73 compounds training-set to predict the activities of three sets of eight compounds. The data for a typical 24 compounds test set is shown in

Figure 8.12b. On average, the error in prediction is about a factor of 3.6, slightly higher than that obtained with the CoMFA analysis. Over all, the CoMSIA model showed similar results compared to the CoMFA analysis.

The steric field map results from the training set (81 compounds) are shown in Figure 8.13a. The CoMSIA steric map showed (in yellow) a rather large unfavorable steric region along the side of the aliphatic spacer (for the cyclic compounds the yellow region is near the 3'-position). This unfavorable steric region is in agreement with requirements of the protein binding site. It might explain the low activities of the compounds WSP1362 and WSP1363 (cyclic compounds, see appendix 1), which have a substituent in 3'-position; therefore the loss of activity might be related to unfavorable steric interactions in this part of the active pocket (see chapter 4). Favorable steric regions are found near the two phenyl rings of the trityl moity, which are the most exposed to the solvent.

CoMSIA electrostatic contour maps are shown in Figure 8.13b. A blue contour region is in the bottom of the active site, which overlaps the uracil group. This is consistent with the fact that residues Glu and Leu for H-bonds with the uracil group. In addition, there are another two blue regions near the linker group exposed to the solvent, which is consistent with the fact that generally the secondary amine in this position enhances the activity. There is one small red contour region near the linker suggesting that increased electron density may favour activity. Also this suggests that different orientations of the nitrogen linker may affect the activity. The activity is expected to improve when the nitrogen group is exposed to the solvent. Additionally, the red region indicated that a more electronergative group, as the oxygen, in an area less exposed to the solvent might also increase the activity.

CoMSIA hydrophobic contour maps are shown in Figure 8.13c. It shows two unfavorable hydrogen regions in orange. One of the regions overlaps the centre of the trityl group. This may seem contradict the experimental data; however this result is consistent with the fact that a purely hydrophobic group does not improve the activity, such as compounds WSP1066, WSP873, WSP874, WSP962, WSP865, WSP951, WSP958. On the other hand, the presence of a relatively high electron density in this position, such as the trityl group, increases the activity.

The second orange region is displayed not far from the Asp 109. This seems consistent with the requirements of the active site.

The contour maps of hydrogen bond donor (Figure 8.13d) and acceptor fields (Figure 8.13e) describe the partial arrangement of favorable and disfavored hydrogen bond interactions to acceptor or donator groups of the target molecule. One favorable hydrogen bond donor (cyan) is near to the nitrogen linker, which is consistent with the experimental data. Also, a large favorable contour H-bond acceptor (green) surrounds the group linker. This is consistent consitent with the fact that the majority of the most active compounds have an H-bond acceptor, such as the oxygen, in this position.

In general, it appears that improvement in binding affinity can be achieved by having a polar substituent in the linker position. Conformational restriction of this substituent may affect the protein binding of this linker group.

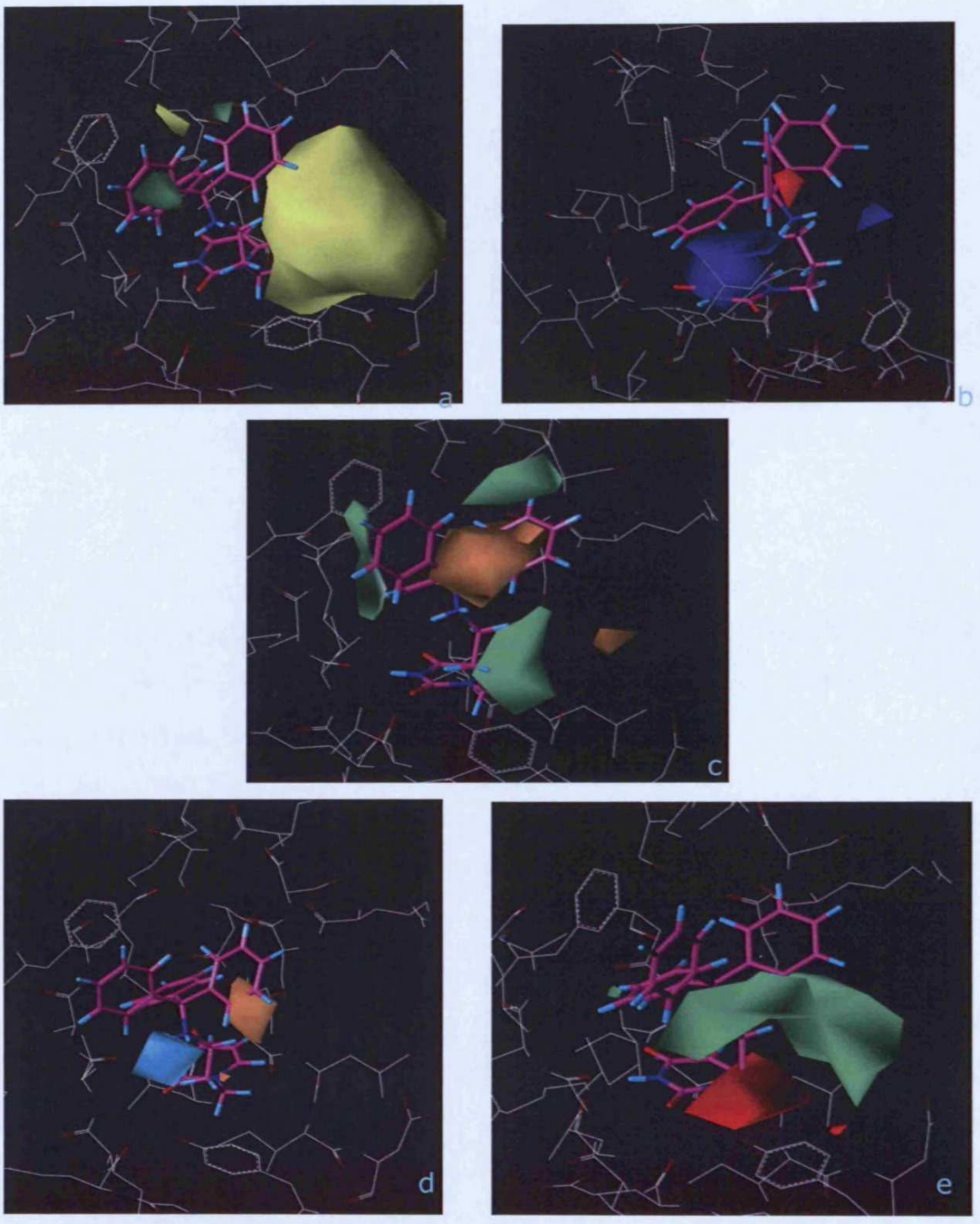


Figure 8.13 CoMSIA contour maps in combination with inhibitor WSP1230: (a) the steric field (green favorable, yellow disfavored); (b) the electrostatic field (blue positive potential favorable, red disfavored);(c) hydrophobic field (green favorable, orange disfavored); (d) donor field (cyan favorable, orange disfavored); (e) acceptor field (green favorable, red disfavored).

WSP ligand	*Ki (μM)*	*IC50 (μM)*	*pIC50 (M)*	*CoMSIA prediction*	*CoMSIA set test 1*	*CoMSIA set test 2*	*CoMSIA set test 3*
483	1.8	0.08	7.08	7.00	6.87	**6.46**	6.94
659	316	14.59	4.84	4.96	**5.15**	4.90	4.90
660	206	9.51	4.91	4.77	4.86	4.79	4.82
661	178	8.22	5.09	4.81	**5.25**	4.94	4.85
662	294	13.57	4.87	4.69	4.81	4.67	4.75
665	26	1.20	5.92	6.04	5.93	5.95	**5.59**
666	142	6.56	5.18	4.98	4.93	4.93	4.96
668	189	8.73	5.06	5.46	5.52	5.58	**5.85**
669	228	10.53	4.98	5.02	5.08	5.09	5.12
680	178	8.22	5.09	4.78	4.64	4.82	4.98
704	67	3.09	5.51	5.61	5.13	5.22	5.55
706	115	5.31	5.27	5.49	5.42	5.38	5.37
785	324	14.96	4.83	5.20	5.24	5.18	5.16
786	40	1.85	5.73	5.77	5.68	5.64	5.85
789	468	21.61	4.67	4.72	4.60	4.73	4.59
790	88	4.06	5.39	5.39	5.43	**5.63**	5.35
826	252	11.63	4.93	4.99	5.09	4.97	5.11
828	10	0.46	6.34	6.03	5.83	5.93	6.11
829	298	13.76	4.86	4.78	4.73	4.70	4.88
830	254	11.73	4.93	4.87	5.05	5.04	5.04
864	1.9	0.09	7.06	7.28	7.10	**6.76**	7.33
865	58	2.68	5.57	5.53	5.58	5.51	5.59
866	255	11.77	4.93	5.16	4.95	4.92	4.83
867	1.16	0.05	7.27	6.67	6.64	6.65	6.70
869	4.98	0.23	6.64	6.74	**6.90**	6.78	6.83
870	0.67	0.03	7.51	7.43	7.21	7.23	7.40
871	0.2	0.01	8.03	8.14	**7.20**	7.66	7.89
872	98	4.52	5.34	5.23	5.32	5.29	5.27
873	44	2.03	5.69	5.64	5.76	5.69	5.68
874	49	2.26	5.64	5.55	5.56	5.60	5.60
946	12.42	0.57	6.24	6.50	6.56	6.60	**7.19**
948	89	4.11	5.38	4.89	5.24	5.25	4.89
949	2.8	0.13	6.89	6.86	6.98	6.96	**5.55**
950	4.2	0.19	6.71	6.87	6.67	6.62	7.00
952	110.7	5.11	5.29	5.35	5.45	5.46	**5.39**
953	312.8	14.44	4.84	4.73	5.02	5.05	4.87
958	426	19.67	4.71	5.05	5.16	5.14	5.02
962	73	3.37	5.47	5.75	5.93	5.93	5.75
1000	1.62	0.07	7.13	7.62	7.46	7.40	7.55
1001	2.24	0.10	6.98	6.80	**6.52**	6.95	6.82
1002	224	10.34	4.98	5.49	5.54	5.46	5.54
1004	2.3	0.11	6.97	6.89	7.00	7.05	6.85
1005	8.47	0.39	6.41	6.53	6.65	**6.60**	6.57
1059	5.2	0.24	6.62	6.70	6.74	**6.49**	6.70
1060	0.9	0.04	7.38	7.28	7.08	7.12	7.25
1063	32	1.48	5.83	5.99	5.94	5.85	6.07
1064	8	0.37	6.43	5.99	5.94	5.85	6.07
1065	17	0.78	6.1	5.85	5.65	5.64	5.88
1066	112	5.17	5.29	5.51	**5.76**	5.28	5.55
1067	1.69	0.08	7.11	7.05	6.86	7.15	7.07

1068	0.72	0.03	7.48	7.67	7.89	7.90	7.82
1069	368	16.99	4.77	4.77	**4.74**	4.77	4.75
1072	1.56	0.07	7.14	6.86	6.91	**6.60**	6.82
1134	0.76	0.04	7.45	7.28	7.10	7.07	7.33
1142	2	0.09	7.03	7.12	7.02	7.12	7.11
1149	0.23	0.01	7.97	7.62	7.70	7.77	7.56
1203	25.5	1.18	5.93	5.85	5.83	5.74	5.77
1205	48	2.22	5.65	5.94	5.88	5.78	5.64
1206	1.27	0.06	7.23	7.52	7.23	7.50	7.63
1207	1.43	0.07	6.93	7.40	6.90	7.54	6.72
1224	4.29	0.20	6.7	6.73	6.46	6.26	**6.28**
1228	3.76	0.17	6.76	7.03	7.02	7.01	7.03
1229	4.22	0.19	6.71	6.94	7.02	6.95	6.84
1230	0.2	0.01	8.03	7.84	8.21	8.31	**7.22**
1241	1.84	0.08	7.07	6.90	6.92	7.14	6.96
1243	2.17	0.10	6.95	6.62	6.46	6.44	6.66
1244	2.56	0.12	6.93	6.68	7.56	6.71	**7.20**
1245	31.55	1.46	5.84	5.75	5.83	5.88	5.70
1247	3.11	0.14	6.84	6.81	6.94	6.72	6.89
1288	1.14	0.05	7.28	7.26	7.47	7.34	7.44
1289	0.4	0.02	7.73	7.41	7.63	7.65	7.55
1290	3.36	0.16	6.81	6.74	6.63	**6.47**	6.80
1291	27.6	1.27	5.89	6.21	5.86	5.75	6.06
1292	4.06	0.19	6.73	6.50	6.48	**5.32**	6.70
1315	1.2	0.06	7.26	7.51	7.30	7.28	7.49
1317	0.61	0.03	7.55	7.45	7.47	7.37	7.37
1324	2.11	0.10	7.01	7.02	7.13	7.04	7.03
1329	103.2	4.76	5.32	5.36	5.70	5.78	5.37
1361	0.3	0.01	7.86	7.34	7.37	7.32	7.51
1362	350	16.16	4.79	4.96	**6.69**	5.38	4.91
1363	49	2.26	5.65	5.73	6.05	5.91	5.82
R^2				0.95	0.91	0.92	0.95
q^2				0.69	0.68	0.70	0.66
N^e				6	5	5	6
N^f				81	73	73	73
Steric[g]				0.14	0.15	0.15	0.137
Electrostatic[h]				0.08	0.08	0.07	0.08
Hydrophobic[i]				0.34	0.32	0.31	0.33
Acceptor[l]				0.19	0.18	0.19	0.19
Donator[m]				0.25	0.27	0.28	0.27

Table 8.4 Experimental (*Ki*, IC_{50} and pIC_{50}) and CoMSIA predicted (pC_{50}) values for all the nucleoside derivatives against *P. falciparum* dUTPase and statistical results of 3D-QSAR CoMSIA models. The bold values represent predicted activities of compounds that were not included in the training set. R^2 Correlation coefficient. q^2 cross-validated correlation coefficient after leave-one-out procedure. N^e optimal number of principal components. N^f number of compounds. [g] denotes the steric contribution to the model. [h] denotes the electrostatic contribution to the model. [i] denotes the hydrophobic contribution to the model. [l,m] denotes the hydrogen bond contribution to the model.

8.3 Conclusions

In this chapter have been reported a 3D-QSAR/CoMFA analysis of 26 acyclic nucleoside derivatives, yielding a theoretical-versus-experimental pIC_{50} correlation of $R^2 \sim 0.99$ and a $q^2 \sim 0.74$. This is the report of a QSAR/CoMFA correlation for these compounds. The predictive capability of this methodology was tested with activity predictions for a series of test compounds, with an average pIC_{50} error of 0.29. Additionally, an alternative CoMFA and CoMSIA models were designed using a training set of 81 compounds, with a theoretical-versus-experimental R^2 correlations of ~0.93 and ~0.95, respectively, with q^2 values ~0.64 (CoMFA) and ~0.69 (CoMSIA). The models were validated using a series of test compounds (24 compounds in total). This showed an average pIC_{50} error of 0.50 (CoMFA) and 0.48 (CoMSIA).

In general the models appear to have reasonable power perditivity, which could be use for the design of other *P. falciparum* dUTPase inhibitors.

9 Chapter: Conclusion and future work

9.1 Conclusion

In this work molecular and structural studies have provided some insight into the structure and function of the dUTPase enzyme from a variety of sources including: human, bacterial and viral. In particular six homotrimeric dUTPase structures were analysed using different molecular modelling methodologies, with the aim of improving the activity and selectivity of the series of cyclic and acyclic nucleoside derivatives for the *P. falciparum* dUTPase:

- The broad SAR study has been developed highlighting key features that are important for activity and selectivity of two compound series for *P. falciparum* dUTPase. It suggests that there is a requirement for a trityl-type moiety at the 5'-position for cyclic series, and at the end of the aliphatic spacer for the acyclic series. In addition, selectivity at the enzyme level is generally enhanced either by the O/NH switch at the 5'-position in the cyclic, and at the linker position for the acyclic.

- Crystallographic studies together with GRID analysis on the *P. falciparum* dUTPase, human, *E. coli* and *M. tuberculosis* dUTPases have shown some selective binding features, and highlighted the possibility of improvements in activity against the *P. falciparum* enzyme.

- The docking of the *P. falciparum* dUTPase appears to predict a reasonable trend in conformation for the inhibitors studied. There also appears to be a reasonable correlation between

docking predicted binding score and experimentally determined binding energy for the majority of compounds in both the series. This also may have some utilities for prediction of a binding affinity of novel compounds, structurally related to the inhibitors described here.

In addition, an alternate binding mode ("unfolded" conformation) for the inhibitors has also been discovered using FlexX. However there is no experimental evidence for this second binding mode.

This study also highlighted the possibility of reducing the number of phenyl rings from three to two, at the 5'-position for cyclic, and at the end of the aliphatic spacer for acyclic compounds. This change is believed not affect the activity and selectivity but could improve the general physico- chemical properties of the inhibitors.

- More sophisticated methods (LIE) to calculate free energy changes of binding have been carried out (see chapter 7). This study yielded good correlation between the experimental and predicted biological data. The models generated for this study showed relatively good robustness in free energy binding prediction. They should be valuable to assess new potential *P. falciparum* dUTPase inhibitors.

- Part of the thesis has been focused on developing and validating predictive 3D-QSAR (CoMFA and CoMSIA) models for the *P. falciparum* dUTPase. In general the models appear to have reasonable power predictivity, which could be used to development of other *P. falciparum* dUTPase inhibitors.

To date these methods have allowed us to discover a new set of novel inhibitors for the *P. falciparum* dUTPase, which may lead to new

treatment against malaria[60,100, 101].

9.2 Future Work

9.2.1 Prediction of Blood-Brain Barrier permeation from 3D-molecular field

A major challenge for treatment of most brain disorders is overcoming the difficulty of delivering therapeutic agents to specific regions of the brain. The ability of a central nervous system (CNS) drug to penetrate the blood-brain barrier (BBB) is a fundamental requirement for the drug to be active. Conversely, peripherally acting drugs ideally possess negligible BBB penetration in order to minimize undesired CNS-related side-effects. The uptake of a compound into the brain is a complex process due to the unique properties of the endothelial cells of the brain capillaries that act as both a physical barrier and a biochemical interface[102].

It is known that relatively lipophilic drugs can cross the BBB by passive diffusion influenced by their hydrogen-bonding capacity. Polar molecules normally do not cross the BBB, but sometimes a process of active transport facilitates their penetration, which is very difficult to predict. Local hydrophobicity, ionization profile, molecular size, lipophilicity, and flexibility are other important parameters which play a role in BBB permeation[102].

Using a computational method that is able to convert 3D fields into new descriptors well suited for structure–pharmacokinetic relationships (Volfsurf)[103], we could predict the BBB permeation of our inhibitors. The experimental determination of BBB penetration is laborious, expensive, time-consuming and requires a sufficient

quantity of pure compound, often in radio-labeled form, to obtain good experimental data. Therefore, a fast and reliable computational method to predict BBB permeation at an early stage of discovery could help decrease the attrition curve since it would allow the virtual screening of many compounds prior to synthesis[104].

A training set of 67 compounds was compiled from literature to build a predicting model of the BBB permeation. The final model, showed an R^2 of 0.70 and a q^2 (cross validated R^2 value) of 0.63.

However, further investigation is necessary to generate improved models. This could be achieved by selecting a larger and more diverse training set of compounds where uptake is purely passive.

Bibliography

[1] WHO; Malaria, available at:
http://www.who.int/topics/malaria/en/

[2] Snow, R. W.; Guerra, C. A.; Noor, A. M.; Myint, H. Y.; Hay, S. I. The global distribution of clinical episodes of Plasmodium falciparum malaria *Nature* **2005,** 434, (7030), 214-217.

[3] Encarta; online encyclopedia:
http://encarta.msn.com/media_461541582/Life_Cycle_of_the_Malaria_Parasite.html

[4] Dr. B.S. Kakkilaya's Malaria Web Site; Malaria site:
http://www.malariasite.com/malaria/Pathology.htm#BLOOD

[5] Markell and Voge's Medical Parasitology, 9th Edition, by David T. John, MSPH, PhD and William A. Petri, Jr., MD, PhD **2005.**

[6] Leech, J. H.; Barnwell, J. W.; Aikawa, M.; Miller, L. H.; Howard, R. J. Plasmodium-Falciparum Malaria - Association Of Knobs On The Surface Of Infected Erythrocytes With A Histidine-Rich Protein And The Erythrocyte Skeleton *Journal Of Cell Biology* **1984,** 98, (4), 1256-1264.

[7] Wiser, M. F.; Lanners, H. N.; Bafford, R. A. Export of Plasmodium proteins via a novel secretory pathway *Parasitology Today* **1999,** 15, (5), 194-198.

[8] Deitsch, K. W.; Wellems, T. E. Membrane modifications in erythrocytes parasitized by Plasmodium falciparum *Molecular And Biochemical Parasitology* **1996,** 76, (1-2), 1-10.

[9] Francis, S. E.; Sullivan, D. J.; Goldberg, D. E. Hemoglobin metabolism in the malaria parasite Plasmodium falciparum *Annual Review Of Microbiology* **1997,** 51, 97-123.

[10] Slater, A. F. G.; Swiggard, W. J.; Orton, B. R.; Flitter, W. D.; Goldberg, D. E.; Cerami, A.; Henderson, G. B. An Iron Carboxylate Bond Links The Heme Units Of Malaria Pigment *Proceedings Of The National Academy Of Sciences Of The United States Of America* **1991,** 88, (2), 325-329.

[11] Casteel, D. A. Antimalarial Agents. Burger's Medicinal Chemistry and Drug Discovery:Chemotherapeutic Agents Vol 5; Sixth ed.; John Wiley & Sons, Inc., **2003**; pp 919-1032.

[12] Macomber, P. B., O'brien, R. L., Hahn, E. F. Chloroquine: phsysiological basis of drug resistance in Plasmodium berghei. *Science* **1966**, 152, 1374-1375.

[13] Eggleson, K. K.; Duffin, K. L.; Goldberg, D. E. Identification and Characterization of Falcilysin, a Metallopeptidase Involved in Hemoglobin Catabolism within the Malaria Parasite Plasmodium falciparum. *J. Biol. Chem.* **1999**, 274, 32411-32417.

[14] Egan, T. J.; Marques, H. M. The role of heme in the activity of chloroquine and related antimalarial drugs. *Coordin. Chem. Rev.* **1999**, 190-192, 493-517.

[15] Foley, M.; Tilley, L. Quinoline antimalarials: Mechanisms of action and resistance. *Int. J. Parasitol.* **1997**, 27, 231-240.

[16] Chou, A. C.; Chevli, R.; Fitch, C. D. Ferriprotoporphyrin IX fulfills the criteria for identification as the chloroquine receptor of malaria parasites. *Biochemistry* **1980**, 19, 1543-1549.

[17] Kumar, A.; Katiyar, S. B.; Agarwal, A.; Chauhan, P. M. S. Perspective in Antimalarial Chemotherapy. *Current Medicinal Chemistry* **2003**, 14, 1137-1150.

[18] Meshnick, S. R. Artemisinin: mechanisms of action, resistance and toxicity *International Journal For Parasitology* **2002,** 32, (13), 1655-1660.

[19] Wei, N.; Sadrzadeh, S. M. H. Artemisinin-Induced Membrane Damage - Requirement For Iron Or Hemin *Faseb Journal* **1994,** 8, (5), A673-A673.

[20] Berman, P. A.; Adams, P. A. Artemisinin enhances heme-catalysed oxidation of lipid membranes *Free Radical Biology And Medicine* **1997,** 22, (7), 1283-1288.

[21] Rosenthal, P. J.; Meshnick, S. R. Hemoglobin catabolism and iron utilization by malaria parasites *Molecular And Biochemical Parasitology* **1996,** 83, (2), 131-139.

[22] Nontprasert, A.; Pukrittayakamee, S.; Nosten-Bertrand, M.; Vanijanonta, S.; White, N. J. Studies of the neurotoxicity of oral artemisinin derivatives in mice *American Journal Of Tropical Medicine And Hygiene* **2000,** 62, (3), 409-412.

[23] Management of severe and complicated malaria: a practical handbook. WHO, Geneva. **1991**.

[24] Management of severe and complicated malaria: Drug resistance in malaria. WHO, Geneva. **2001**.

[25] Biagini, G. A.; O'Neill, P. M.; Nzila, A.; Ward, S. A.; Bray, P. G. Antimalarial chemotherapy: young guns or back to the future? *Trends Parasitol.* **2003**, 19, 479-487.

[26] Galperin, M. Y.; Moroz, O. V.; Wilson K. S.; Murzin, A. G. House cleaning,

a part of good housekeeping *Mol. Microbiol.* **2005**, 59, 5–19.

[27] McIntosh, E. M.; Haynes, R. H. dUTP pyrophosphatase as a potential target for chemotherapeutic drug development *Acta Biochimica Polonica* **1997,** 44, (2), 159-171.

[28] Warner, H. R.; Duncan, B. K.; Garrett, C.; Neuhard, J. Synthesis and Metabolism of Uracil-Containing Deoxyribonucleic- Acid in Escherichia-Coli *Journal of Bacteriology* **1981,** 145, (2), 687-695.

[29] Barclay, B. J.; Kunz, B. A.; Little, J. G.; Haynes, R. H. Genetic and Biochemical Consequences of Thymidylate Stress *Canadian Journal of Biochemistry* **1982,** 60, (3), 172-194.

[30] Goulian, M.; Bleile, B.; Tseng, B. Y. The Effect Of Methotrexate On Levels Of Dutp In Animal-Cells *Journal Of Biological Chemistry* **1980,** 255, (22), 630-637.

[31] Gadsden, M. H.; McIntosh, E. M.; Game, J. C.; Wilson, P. J.; Haynes, R. H. dUTP Pyrophosphatase Is an Essential Enzyme in Saccharomyces-Cerevisiae *EMBO Journal* **1993,** 12, (11), 4425-4431.

[32] Larsson, G.; Svensson, L. A.; Nyman, P. O. Crystal structure of the Escherichia coli dUTPase in complex with a substrate analogue (dUDP) *Nature Structural Biology* **1996,** 3, (6), 532-538.

[33] Barabas, O.; Pongracz, V.; Kovari, J.; Wilmanns, M.; Vertessy, B. G. Structural insights into the catalytic mechanism of phosphate ester hydrolysis by dUTPase *Journal of Biological Chemistry* **2004,** 279, (41), 42907-42915.

[34] Mol, C. D.; Harris, J. M.; McIntosh, E. M.; Tainer, J. A. Human dUTP pyrophosphatase: Uracil recognition by a beta hairpin and active sites formed by three separate subunits *Structure* **1996,** 4, (9), 1077-1092.

[35] Dauter, Z.; Persson, R.; Rosengren, A. M.; Nyman, P. O.; Wilson, K. S.; Cedergren-Zeppezauer, E. S. Crystal structure of dUTPase from equine infectious anaemia virus; active site metal binding in a substrate analogue complex *Journal of Molecular Biology* **1999,** 285, (2), 655-673.

[36] Prasad, G. S.; Stura, E. A.; McRee, D. E.; Laco, G. S.; HasselkusLight, C.; Elder, J. H.; Stout, C. D. Crystal structure of dUTP pyrophosphatase from feline immunodeficiency virus *Protein Science* **1996,** 5, (12), 2429-2437.

[37] Moroz, O. V.; Harkiolaki, M.; Galperin, M. Y.; Vagin, A. A.; Gonzalez-Pacanowska, D.; Wilson, K. S. The crystal structure of a complex of Campylobacter jejuni dUTPase with substrate analogue sheds light on the mechanism and suggests the "basic module" for dimeric d(C/U)TPases *Journal Of Molecular Biology* **2004,** 342, (5), 1583-1597.

[38] Chan, S.; Segelke, B.; Lekin, T.; Krupka, H.; Cho, U. S.; Kim, M.; So, M. Y.; Kim, C. Y.; Naranjo, C. M.; Rogers, Y. C.; Park, M. S.; Wald, G. S.; Pashkov, I.; Cascio, D.; Perry, J. L.; Sawaya, M. R. Crystal structure of the Mycobacterium tuberculosis dUTPase: Insights into the catalytic mechanism *Journal of Molecular Biology* **2004,** 341, (2), 503-517.

[39] Whittingham, J. L.; Leal, I.; Nguyen, C.; Kasinathan, G.; Bell, E.; Jones, A. F.; Berry, C.; Benito, A.; Turkenburg, J. P.; Dodson, E. J.; Perez, L. M. R.; Wilkinson, A. J.; Johansson, N. G.; Brun, R.; Gilbert, I. H.; Pacanowska, D. G.; Wilson, K. S. dUTPase as a platform for antimalarial drug design: Structural basis for the selectivity of a class of nucleoside inhibitors *Structure* **2005,** 13, (2), 329-338.

[40] Tarbouriech, N.; Buisson, M.; Seigneurin, J. M.; Cusack, S.; Burmeister, W. P. The monomeric dUTPase from Epstein-Barr virus mimics trimeric dUTPases *Structure* **2005,** 13, (9), 1299-1310.

[41] Persson, R.; Cedergren-Zeppezauer, E. S.; Wilson, K. S. Homotrimeric dUTPases; Structural solutions for specific recognition and hydrolysis of dUTP *Current Protein & Peptide Science* **2001,** 2, (4), 287-300.

[42] Gouet, P.; Courcelle, E.; Stuart, D. I.; Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript *Bioinformatics* **1999,** 15, (4), 305-308.

[43] Kovari, J.; Barabas, O.; Takacs, E.; Bekesi, A.; Dubrovay, F.; Pongracz, V.; Zagyva, I.; Imre, T.; Szabo, P.; Vertessy, B. G. Altered active site flexibility and a structural metal-binding site in eukaryotic dUTPase - Kinetic characterization, folding, and crystallographic studies of the homotrimeric Drosophila enzyme *Journal Of Biological Chemistry* **2004,** 279, (17), 17932-17944.

[44] Hidalgo-Zarco, F.; Camacho, A. G.; Bernier-Villamor, V.; Nord, J.; Ruiz-Perez, L. M.; Gonzalez-Pacanowska, D. Kinetic properties and inhibition of the dimeric dUTPase-dUDPase from Leishmania major *Protein Science* **2001,** 10, (7), 1426-1433.

[45] Sawaya, M. R.; Prasad, R.; Wilson, S. H.; Kraut, J.; Pelletier, H. Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: Evidence for an induced fit mechanism *Biochemistry* **1997,** 36, (37), 11205-11215.

[46] Mustafi, D.; Bekesi, A.; Vertessy, B. G.; Makinen, M. W. Catalytic and structural role of the metal ion in dUTP pyrophosphatase *Proceedings Of The National Academy Of Sciences Of The United States Of America* **2003,** 100, (10), 5670-5675.

[47] Harkiolaki, M.; Dodson, E. J.; Bernier-Villamor, V.; Turkenburg, J. P.; Gonzalez-Pacanowska, D.; Wilson, K. S. The crystal structure of Trypanosoma cruzi dUTPase reveals a novel dUTP/dUDP binding fold *Structure* **2004,** 12, (1), 41-53.

[48] Gonzalez-Pacanowska, D. Personal Communication; **2004**; *Instituto de Parasitologia y Biomedicina "Lopez-Neyra"*, Granada (Spain).

[49] Hidalgo-Zarco, F.; Gonzalez-Pacanowska, D. Trypanosomal dUTPases as potential targets for drug design *Current Protein & Peptide Science* **2001,** 2, (4), 389-397.

[50] Mc Carthy, O. K.; Schipani, A.; Buendia, A. M.; Ruiz-Perez, L. M.; Kaiser, M.; Brun, R.; Pacanowska, D. G.; Gilbert, I. H. Design, synthesis and evaluation of novel uracil amino acid conjugates for the inhibition of Trypanosoma cruzi dUTPase *Bioorganic & Medicinal Chemistry Letters* **2006,** 16, (14), 3809-3812.

[51] Buisson, M.; Hernandez, J. F.; Lascoux, D.; Schoehn, G.; Forest, E.; Arlaud, G.; Seigneurin, J. M.; Ruigrok, R. W. H.; Burmeister, W. P. The crystal structure of the Epstein-Barr Virus protease shows rearrangement of the processed C terminus *Journal Of Molecular Biology* **2002,** 324, (1), 89-103.

[52] Brocchieri, L. Low-complexity regions in Plasmodium proteins: In search of a function *Genome Research* **2001,** 11, (2), 195-197.

[53] Ladner, R. D.; Caradonna, S. J. The human dUTPase gene encodes both nuclear and mitochondrial isoforms - Differential expression of the isoforms and characterization of a cDNA encoding the mitochondrial species *Journal Of Biological Chemistry* **1997,** 272, (30), 19072-19080.

[54] Ma, J. C.; Dougherty, D. A. The cation-pi interaction *Chemical Reviews* **1997,** 97, (5), 1303-1324.

[55] Vertessy, B. G.; Larsson, G.; Persson, T.; Bergman, A. C.; Persson, F.; Nyman, P. O. The complete triphosphate moiety of non-hydrolyzable substrate analogues is required for a conformational shift of the flexible C-terminus in E-coli dUTP pyrophosphatase *Febs Letters* **1998,** 421, (1), 83-88.

[56] Williams, M. V. Effects Of Mercury (II) Compounds On The Activity Of dUTPase From Various Sources *Molecular Pharmacology* **1986,** 29, (3), 288-292.

[57] Zalud, P.; Wachs, W. O.; Nyman, P. O.; Zeppezauer, M. Inhibition of the Proliferation of Human Cancer Cells In-vitro by Substrate-Analogous Inhibitors of dUTPase. Purine and Pyrimidine Metabolism in Man VIII; Plenum Press: New York, **1995**; pp 135-138.

[58] Bernier-Villamor, V.; Camacho, A.; Hidalgo-Zarco, F.; Perez, J.; Ruiz-Perez, L. M.; Gonzalez-Pacanowska, D. Characterization of deoxyuridine 5 '-triphosphate nucleotidohydrolase from Trypanosoma cruzi *Febs Letters* **2002,** 526, (1-3), 147-150.

[59] Nguyen, C.; Kasinathan, G.; Leal-Cortijo, I.; Musso-Buendia, A.; Kaiser, M.; Brun, R.; Ruiz-Perez, L. M.; Johansson, N. G.; Gonzalez-Pacanowska,

D.; Gilbert, I. H. Deoxyuridine triphosphate nucleotidohydrolase as a potential antiparasitic drug target *Journal Of Medicinal Chemistry* **2005,** 48, (19), 5942-5954.

[60] Nguyen, C.; Ruda, G. F.; Schipani, A.; Kasinathan, G.; Leal, I.; Musso-Buendia, A.; Kaiser, M.; Brun, R.; Ruiz-Perez, L. M.; Sahlberg, B. L.; Johansson, N. G.; Gonzalez-Pacanowska, D.; Gilbert, I. H. Acyclic nucleoside analogues as inhibitors of Plasmodium falciparum dUTPase *Journal Of Medicinal Chemistry* **2006,** 49, (14), 4183-4195.

[61] GRID version 22; Molecular Discovery Ltd., West Way House, Elms Parade, Oxford, U.K.; **1998**.

[62] Sotriffer, C; Klebe, G; Stahl, M; Böhm, H. J.; Docking, Scoring Functions, and Virtual Screening *Manuscript to appear in "Burger's Medicinal Chemistry Handbook", 6th ed., Vol. 1*

[63] Babine, R. E.; Bender, S. L. Molecular recognition of protein-ligand complexes: Applications to drug design *Chemical Reviews* **1997,** 97, (5), 1359-1472.

[64] Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application Of The 3-Dimensional Structures Of Protein Target Molecules In Structure-Based Drug Design *Journal Of Medicinal Chemistry* **1994,** 37, (8), 1035-1054.

[65] Vonitzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. V.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. Rational Design Of Potent Sialidase-Based Inhibitors Of Influenza-Virus Replication *Nature* **1993,** 363, (6428), 418-423.

[66] Klebe, G. Recent developments in structure-based drug design *Journal Of Molecular Medicine* **2000,** 78, (5), 269-281.

[67] SYBYL 6.9, TRIPOS Inc, 1699 South Hanley Road, St Louis, Missouri 63144, USA

[68] Klebe, G.; Mietzner, T. A Fast and Efficient Method to Generate Biologically Relevant Conformations *Journal of Computer-Aided Molecular Design* **1994,** 8, (5), 583-606.

[69] Bohm, H. J. The Computer-Program Ludi - a New Method for the Denovo Design of Enzyme-Inhibitors *Journal of Computer-Aided Molecular Design* **1992,** 6, (1), 61-78.

[70] Klebe, G. The Use of Composite Crystal-Field Environments in Molecular Recognition and the De-Novo Design of Protein Ligands *Journal of Molecular Biology* **1994,** 237, (2), 212-235.

[71] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking

method using an incremental construction algorithm *Journal of Molecular Biology* **1996,** 261, (3), 470-489.

[72] Bohm, H. J. The Development Of A Simple Empirical Scoring Function To Estimate The Binding Constant For A Protein Ligand Complex Of Known 3-Dimensional Structure *Journal Of Computer-Aided Molecular Design* **1994,** 8, (3), 243-256.

[73] Whittingham J. and Wilson K.S., personal communication **2006**; *University of York* (UK).

[74] ACD/Lab 7.00 Release; Product Version 7.09, Built: 03 Nov **2003** Toronto, Canada.

[75] Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding constants of neuraminidase inhibitors: An investigation of the linear interaction energy method *Journal Of Medicinal Chemistry* **1999,** 42, (25), 5142-5152.

[76] Kollman, P. A. Free-Energy Calculations - Applications To Chemical And Biochemical Phenomena *Chemical Reviews* **1993,** 93, (7), 2395-2417.

[77] Aqvist, J.; Medina, C.; Samuelsson, J. E. New Method For Predicting Binding-Affinity In Computer-Aided Drug Design *Protein Engineering* **1994,** 7, (3), 385-391.

[78] Aqvist, J.; Hansson, T. On the validity of electrostatic linear response in polar solvents *Journal Of Physical Chemistry* **1996,** 100, (22), 9512-9521.

[79] Hansson, T.; Aqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations *Protein Engineering* **1995,** 8, (11), 1137-1144.

[80] Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Aqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 protease inhibitors derived from mannitol: Synthesis, inhibitory potencies, and computational predictions of binding affinities *Journal Of Medicinal Chemistry* **1997,** 40, (6), 885-897.

[81] Aqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions *Journal Of Computational Chemistry* **1996,** 17, (14), 1587-1597.

[82] JonesHertzog, D. K.; Jorgensen, W. L. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method *Journal Of Medicinal Chemistry* **1997,** 40, (10), 1539-1549.

[83] Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes *Protein Engineering* **1996,** 9, (7), 567-571.

[84] Wang, J.; Dixon, R.; Kollman, P. A. Ranking ligand binding affinities with avidin: A molecular dynamics-based interaction energy study *Proteins-Structure Function And Genetics* **1999,** 34, (1), 69-81.

[85] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules *Journal Of The American Chemical Society* **1996,** 117, 5179-5197.

[86] Hansson, T.; Marelius, J.; Aqvist, J. Ligand binding affinity prediction by linear interaction energy methods *Journal Of Computer-Aided Molecular Design* **1998,** 12, (1), 27-35.

[87] Van Lipzig, M. M. H.; ter Laak, A. M.; Jongejan, A.; Vermeulen, N. P. E.; Wamelink, M.; Geerke, D.; Meerman, J. H. N. Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method *Journal Of Medicinal Chemistry* **2004,** 47, (4), 1018-1030.

[88] Singh, P.; Mhaka, A. M.; Christensen, S. B.; Gray, J. J.; Denmeade, S. R.; Isaacs, J. T. Applying linear interaction energy method for rational design of noncompetitive allosteric inhibitors of the sarco- and endoplasmic reticulum calcium-ATPase *Journal Of Medicinal Chemistry* **2005,** 48, (8), 3005-3014.

[89] Carlson, H. A.; Jorgensen, W. L. An Extended Linear-Response Method For Determining Free-Energies Of Hydration *Journal Of Physical Chemistry* **1995,** 99, (26), 10667-10673.

[90] *FirstDiscovery2.7,* 2.7 ed.; Schrodinger Inc.: Portland, 2004.

[91] Svab, I.; Alexandru, D.; Vitos, G.; Flonta, M. L. Binding affinities for sulfonamide inhibitors with matrix metalloproteinase-2 using a linear response method *Journal Of Cellular And Molecular Medicine* **2004,** 8, (4), 551-562.

[92] Tounge, B. A.; Reynolds, C. H. Calculation of the binding affinity of beta-secretase inhibitors using the linear interaction energy method *Abstracts Of Papers Of The American Chemical Society* **2003,** 226, U444-U444.

[93] *Liaison* (Schrodinger) documentation, manual **2005.**

[94] Kubinyi, H. QSAR and 3D QSAR in drug design.2. Applications and problems *Drug Discovery Today* **1997,** 2, (12), 538-546.

[95] Clark, M.; Cramer, R. D.; Vanopdenbosch, N. Validation Of The General-Purpose Tripos 5.2 Force-Field *Journal Of Computational Chemistry* **1989,** 10, (8), 982-1012.

[96] Pan, X. L.; Tan, N. H.; Zeng, G. Z.; Han, H. J.; Huang, H. Q. 3D-QSAR and docking studies of aldehyde inhibitors of human cathepsin K *Bioorganic*

& *Medicinal Chemistry* **2006,** 14, (8), 2771-2778.

[97] Cheng, Y.; Prusoff, W. H. Relationship Between Inhibition Constant (K1) And Concentration Of Inhibitor Which Causes 50 Per Cent Inhibition (I50) Of An Enzymatic-Reaction *Biochemical Pharmacology* **1973,** 22, (23), 3099-3108.

[98] Kotsikorou, E.; Song, Y. C.; Chan, J. M. W.; Faelens, S.; Tovian, Z.; Broderick, E.; Bakalara, N.; Docampo, R.; Oldfield, E. Bisphosphonate inhibition of the exopolyphosphatase activity of the Trypanosoma brucei soluble vacuolar pyrophosphatase *Journal Of Medicinal Chemistry* **2005,** 48, (19), 6128-6139.

[99] Klebe, G., Abraham, U. Comparative Molecular Similarity Index Analysis (CoMSIA) to Study Hydrogen Bonding Properties and to Score Combinatorial Libraries *Journal of Computer-Aided Molecular Design* **1999**, 13, 1-10

[100] Gilbert, Ian; Nguyen, Corinne; Ruda, Gian Filippo; Schipani, Alessandro; Kasinathan, Ganasan; Johansson, Nils-Gunnar; Pacanowska, Dolores Gonzales. Preparation of deoxyuridine nucleosides as dUTPase inhibitors. *PCT Int. Appl.* **2005**, 73 pp. CODEN: PIXXD2 WO 2005066160 A1 20050721 CAN 143:133640 AN 2005:638870 CAPLUS

[101] Gilbert, Ian; Nguyen, Corinne; Ruda, Gian Filippo; Schipani, Alessandro; Kasinathan, Ganasan; Johansson, Nils-Gunnar; Pacanowska, Dolores Gonzales. Preparation of pyrimidinediones as deoxyuridine triphosphate nucleotidohydrolase (dUTPase) inhibitors for treatment of parasitic infections. *PCT Int. Appl.* **2005**, 76 pp. CODEN: PIXXD2 WO 2005065689 A1 20050721 CAN 143:153391 AN 2005:638741 CAPLUS

[102] de Boer, A. G.; van der Sandt, I. C. J.; Gaillard, P. J. The role of drug transporters at the blood-brain barrier *Annual Review Of Pharmacology And Toxicology* **2003,** 43, 629-656.

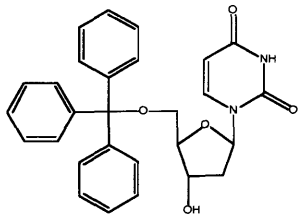
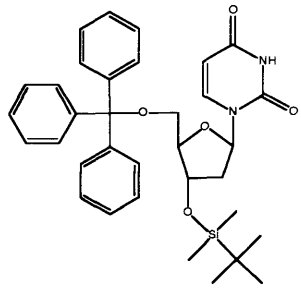
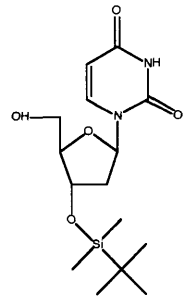
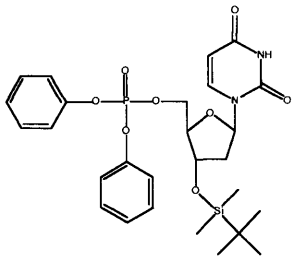
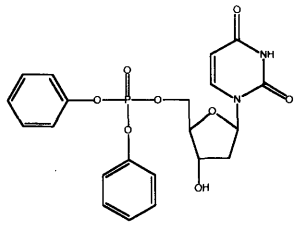
[103] Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds *European Journal Of Pharmaceutical Sciences* **2000,** 11, S29-S39.

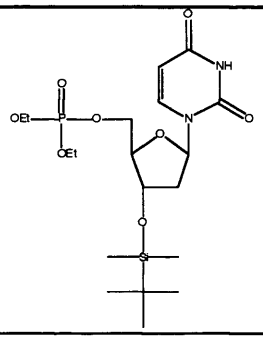
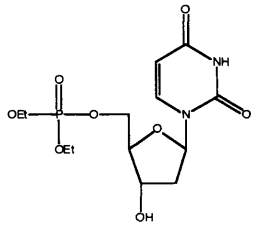
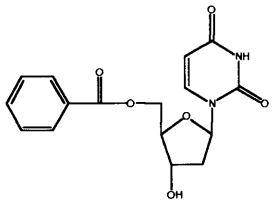
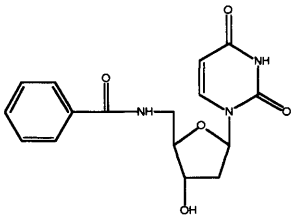
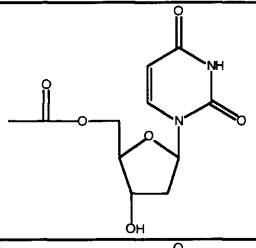
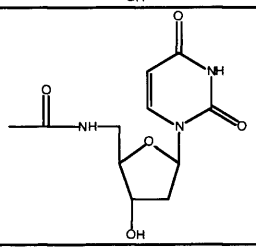
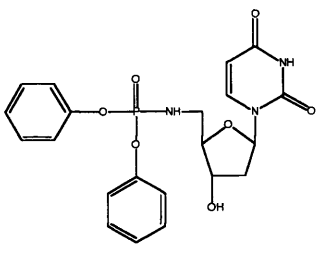
[104] Ooms, F.; Weber, P.; Carrupt, P. A.; Testa, B. A simple model to predict blood-brain barrier permeation from 3D molecular fields *Biochimica Et Biophysica Acta-Molecular Basis Of Disease* **2002,** 1587, (2-3), 118-125.

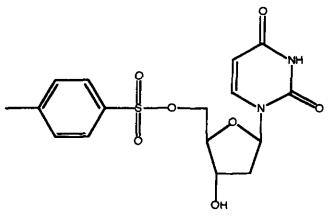
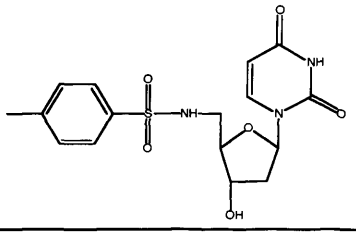
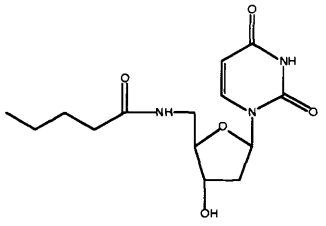
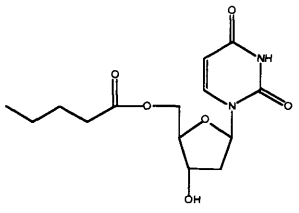
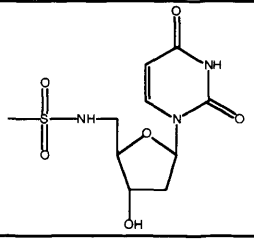
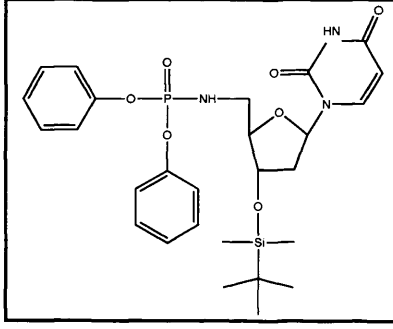
Appendix 1

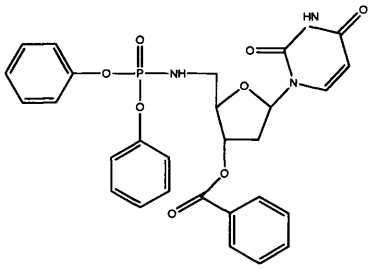
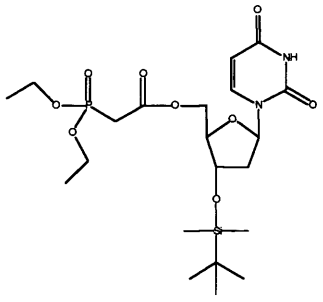
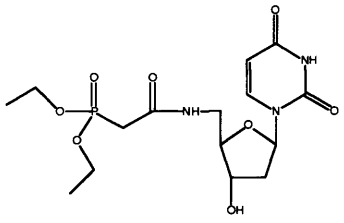
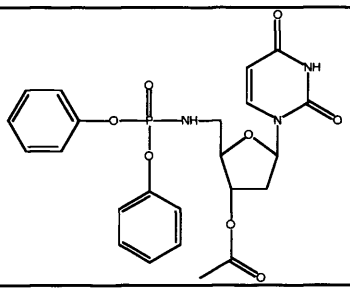
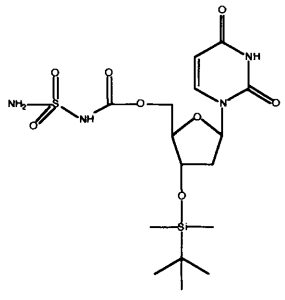
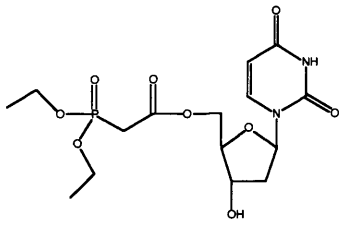
Complete biological data for the compounds discusse in this thesis[48, 100, 101].

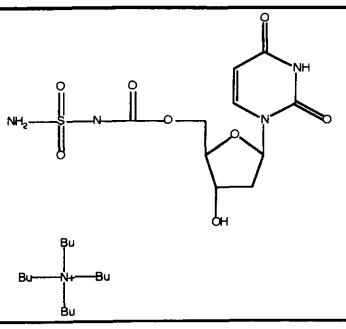
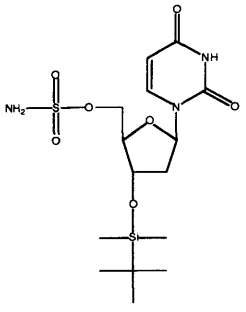
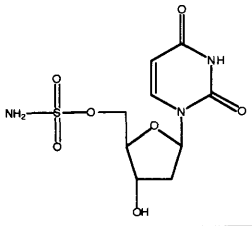
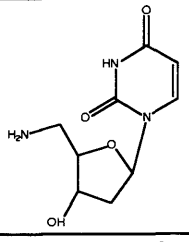
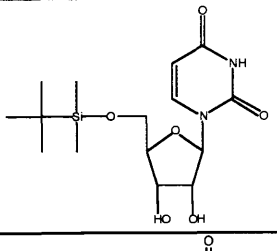
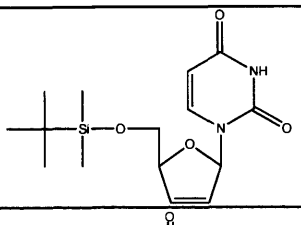
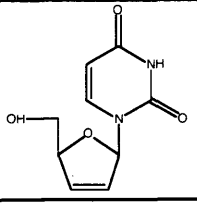
P. f.: Plasmodium falciparum; Tox.: toxicity; MW: molecular weight

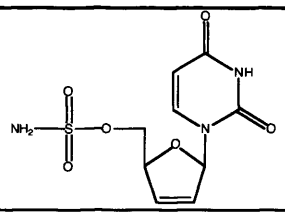
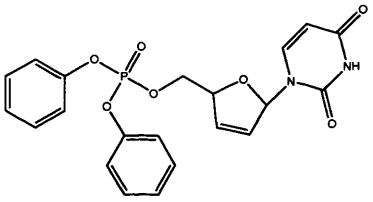
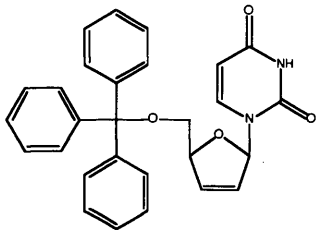
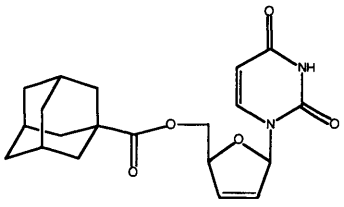
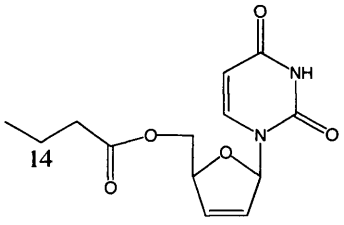
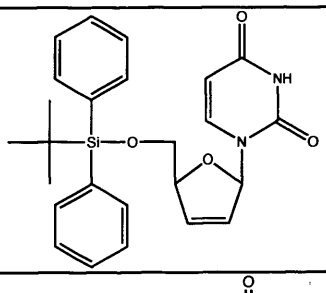
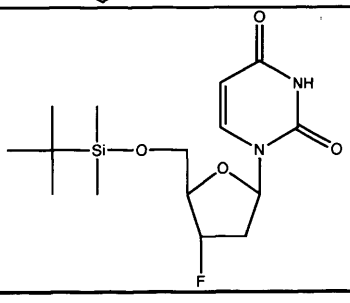
		Enzyme assay *Ki* (μM); (SI)		*In vitro* assay IC_{50} (μM)		
Structure	**WSP**	***P. f.***	***human***	***P. f.***	***T. x***	***MW***
	483	1.8 (9.8)	17.7	6	192	470.528
	484	515.2 (>1.9)	>1mM	1	>154	584.791
	485	648.0 (0.2)	119	>15	>263	342.469
	486	596.9 (>1.6)	>1mM	6	>157	574.646
	487	238.3 (0.6)	135.1	>11	>196	460.382

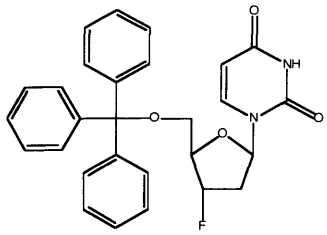
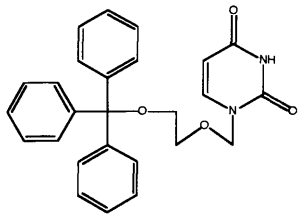
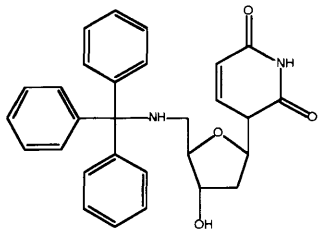
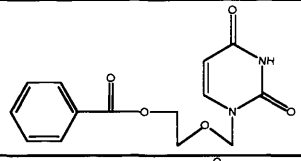
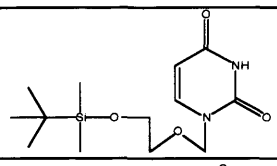
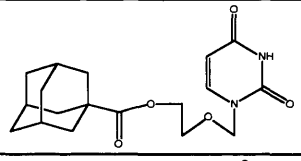
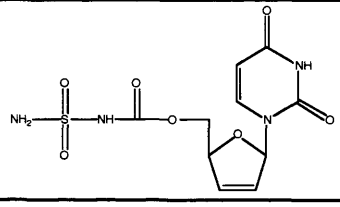
	658	340	nd	>10	108	478.557
	659	316	nd	>14	198	364.293
	660	268	nd	>15	>271	332.315
	661	178	nd	>15	>272	331.33
	662	294	nd	>19	64	270.243
	663	660	nd	>19	>335	269.258
	665	26	nd	>11	>196	459.398

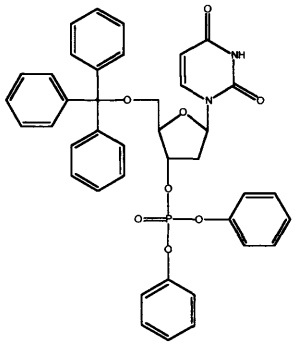
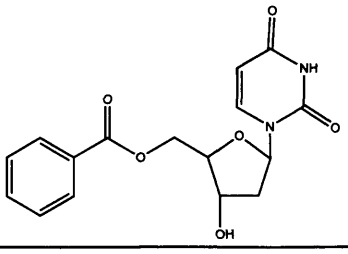
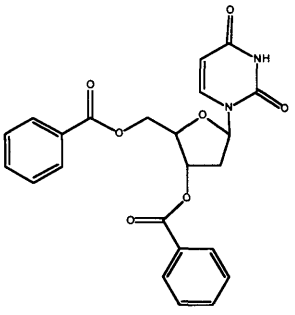
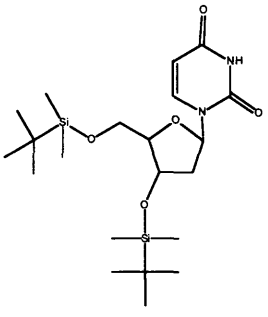
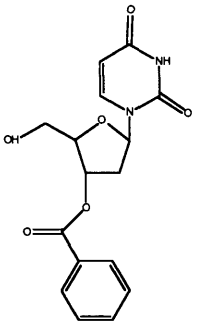
	666	142	nd	>13	>236	382.396
	667	163	nd	>13	>236	381.412
	668	189	nd	>16	>289	311.339
	669	228	nd	>16	>289	312.324
	680	108	nd	>16	>295	305.313
	704	67	nd	4	18	573.661

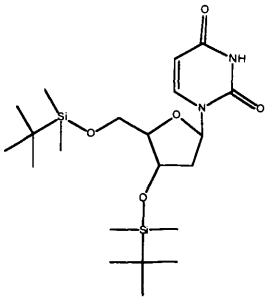
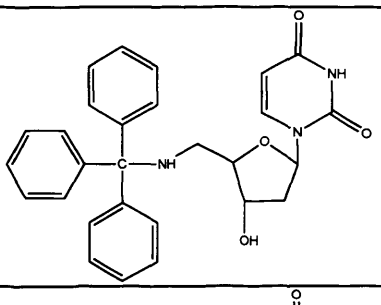
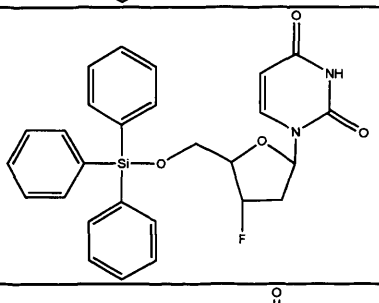
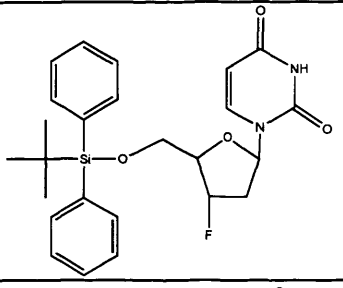
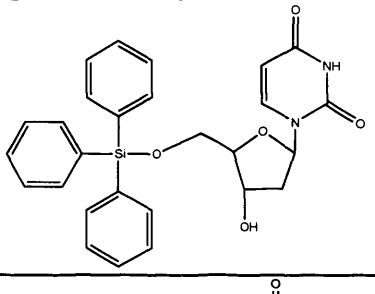
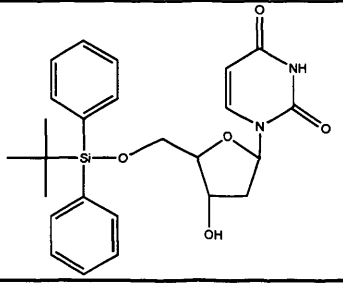
	706	115	nd		34	563.5068
	784	217	nd	18	-	520.594
	785	86	nd	78	-	405.346
	786	40	nd	4	34	501.435
	787	244	nd	37	-	464.574
	788	123	nd	98	168	406.331

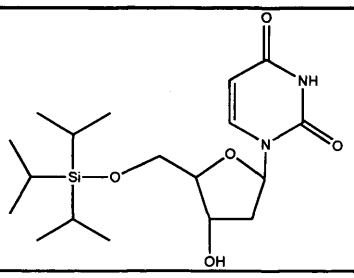
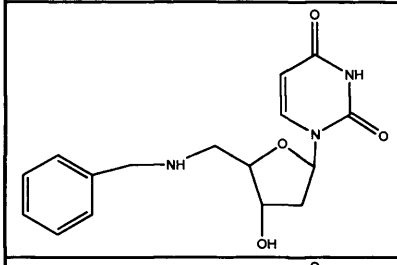
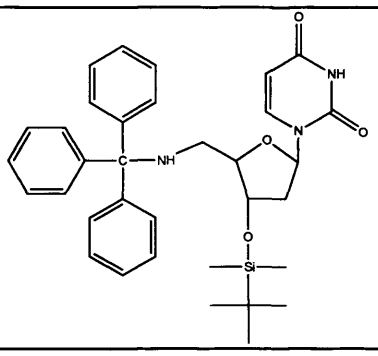
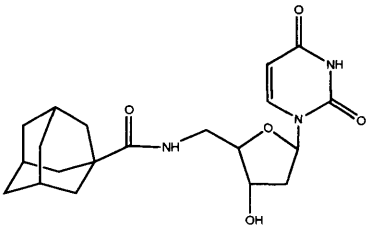
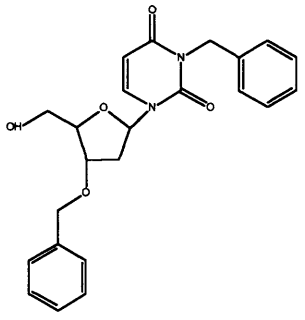
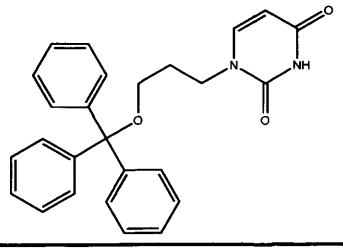
	789	468	nd	20	-	591.773
	790	88	nd	40	-	421.549
	791	400	nd	>163	-	307.286
	804	655	nd	-	-	227.221
	826	252	nd	25	-	358.468
	828	10	nd	51	-	324.454
	829	298	nd	214	-	210.19

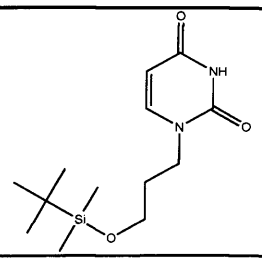
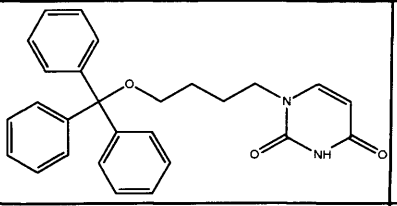
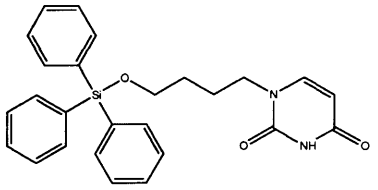
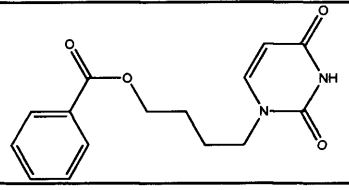
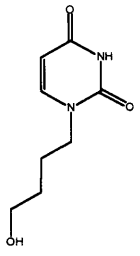
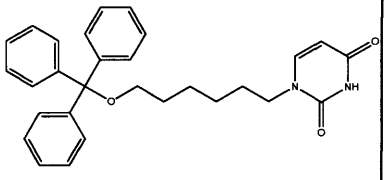
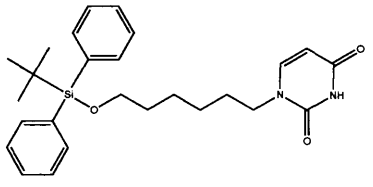
	830	254	nd	163	-	289.27
	831	99	nd	54	-	442.367
	864	1.9 (82.4)	156.5	>11	-	452.512
	865	57.7 (5.2)	298	>13	-	372.423
	866	255.7 (1.4)	350	>11	-	448.606
	867	1.16 (124.13)	144	3	67	448.596
	868	628 (0.46)	288.4	>14.5	-	344.46

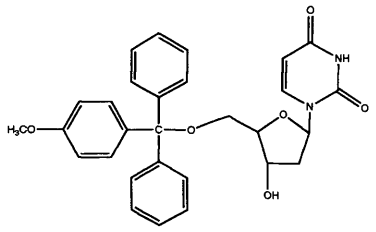
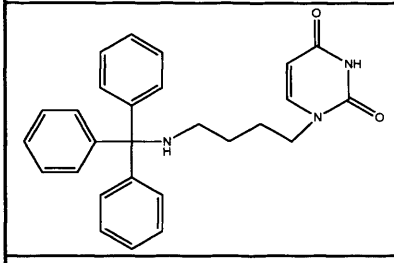
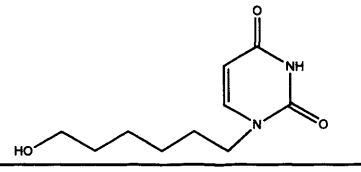
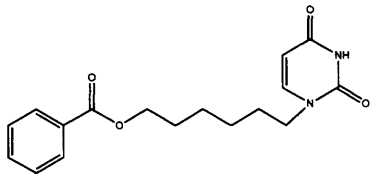
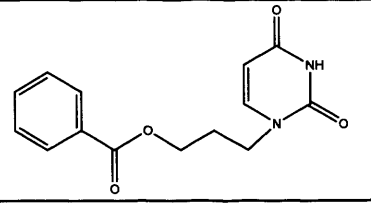
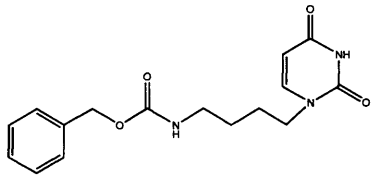
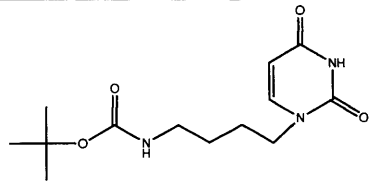
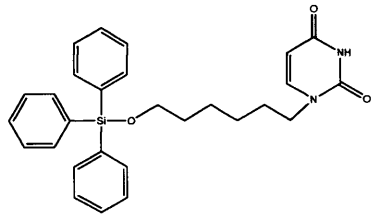
	869	4.98 (91.2)	457	2		472.519
	870	0.6 (25.3)	17.01	0.9	70.0 (MIC)	428.49
	871	0.2 (231.5)	46.3	4.5		468.555
	872	98 (3.2)	311.06	127.1	201.5	290.277
	873	43.7 (1.43)	62.7	62.2	>299.5	300.431
	874	49.1 (6.8)	334.8	36.4	175	348.401
	875	82.3 (9.6)	798.9	>150.4	261.5	332.296

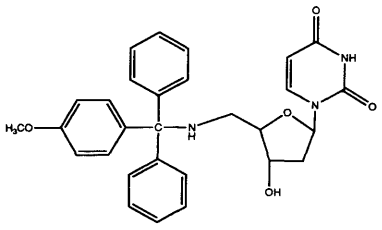
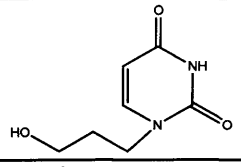
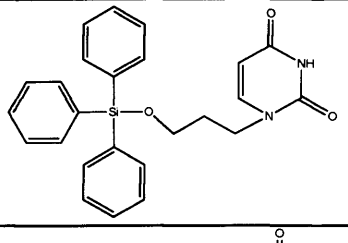
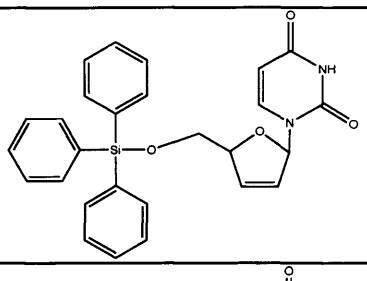
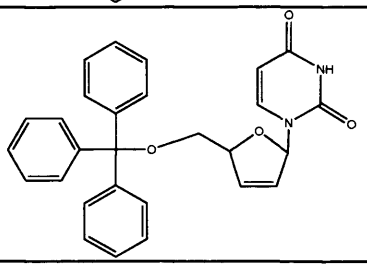
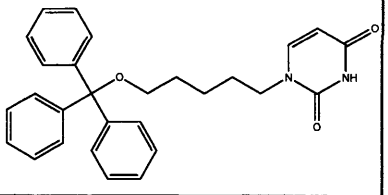
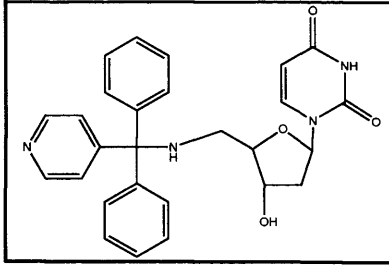
	876	115.5	nd	0.8	18	702.704
	929	206.3 (>4.8)	>1mM	9.8	-	332.315
	930	>1mM (>0.9)	952.3	>11.5	95.7	436.423
	931	>1mM (>1)	>1mM	-	-	468.509
	932	>1mM (>1)	>1mM	>15.0	34	332.315

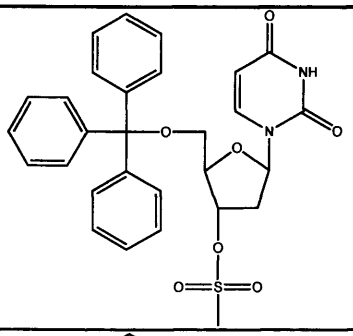
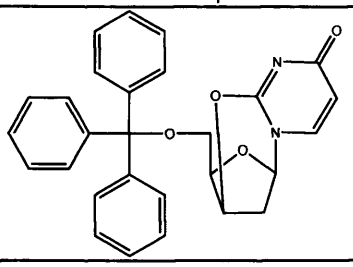
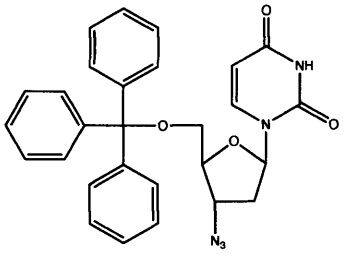
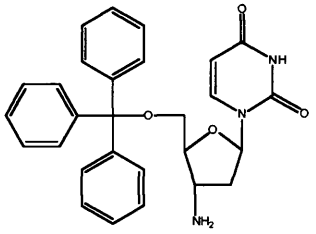
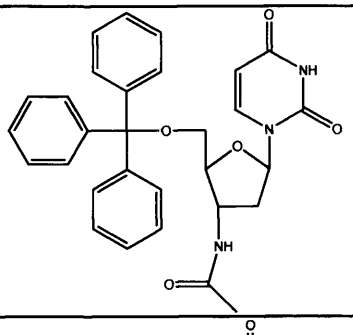
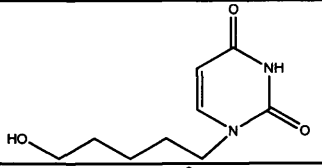
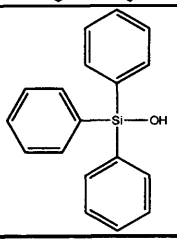
	933	268.5 (>3.7)	>1mM	3	10.7	456.732
	946	12.42 (>80)	>1mM	5.3	30.3	471.534
	947	975 (>1) >1mM (>1)	>1mM	1	85.5	488.593
	948	89.39 (9.3)	807.86	8.8	16.4	468.603
	949	2.8 (324.6)	908.9	1.1	-	486.602
	950	4.2 (191.8)	805.8	6.6	22.9	466.612

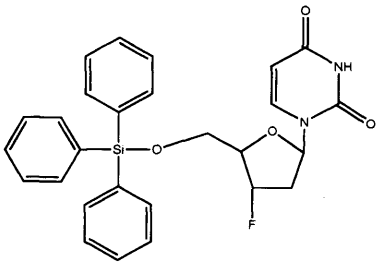
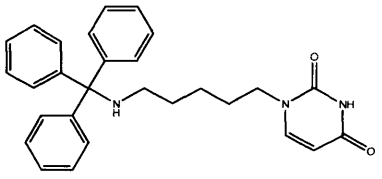
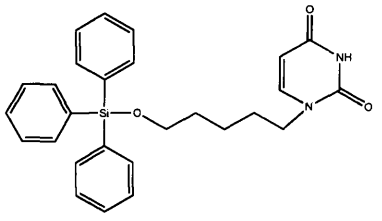
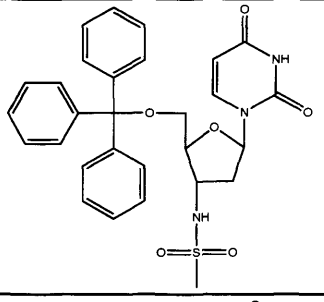
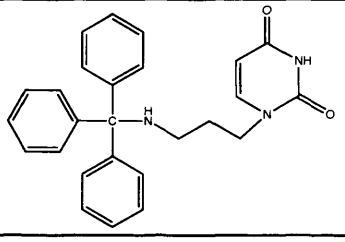
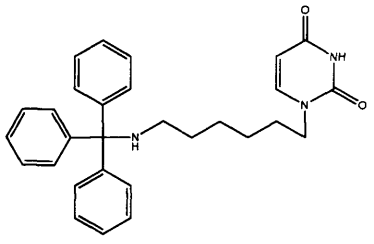
	951	227.1 (>4.4)	>1mM	>13	25.2	384.55
	952	110.7 (4.23)	469.5	>16	-	317.346
	953	312.8 (>3)	>1mM	1.8	7.8	583.806
	958	>1mM (<0.65)	658.75	>13	-	389.454
	959	>1mM (>1)	>1mM	>12	123.8	408.456
	961	86.55 (3.62)	313.46	7.5	34.4	412.491

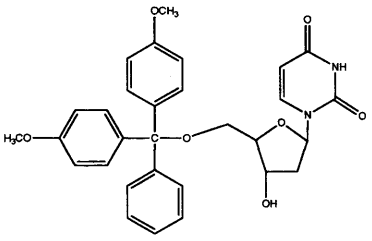
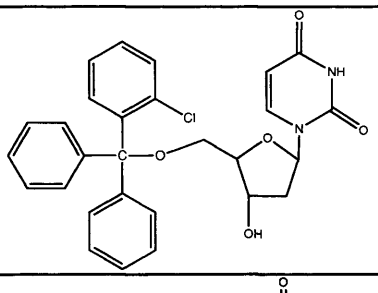
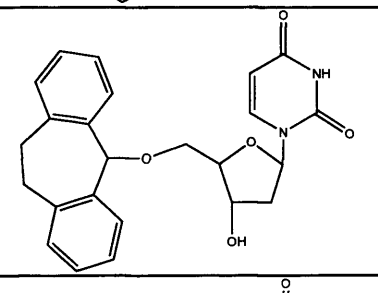
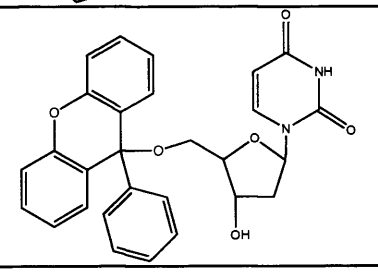
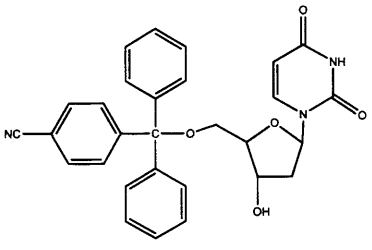
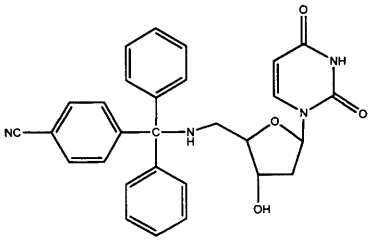
	962	73.1 (1.01)	73.46	>18	191.9	284.432
	1000	1.62 (>617)	>1mM	4.9	42.4	426.518
	1001	2.24 (>446)	>1mM	0.36	na	442.592
	1002	223.5 (2.21)	495.9	37.8	na	288.305
	1003	>1mM	>1mM	235	na	184.196
	1004	9.58 (49.7)	476.29	2.3	43.8	454.572
	1005	8.47 (<118)	>1mM	3.9	18	450.656

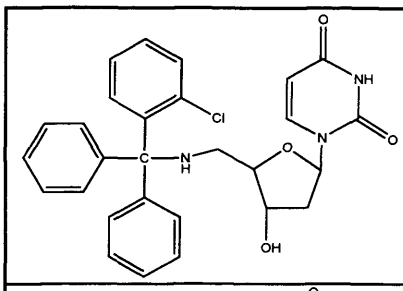
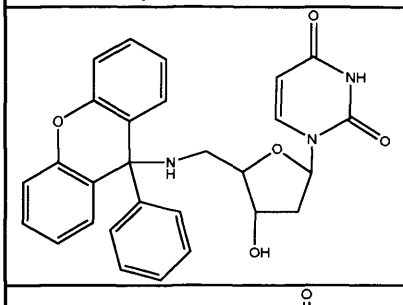
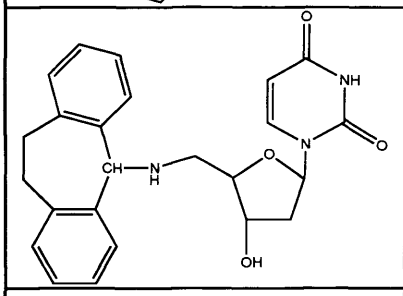
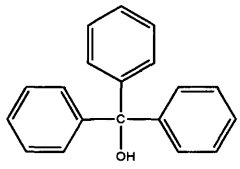
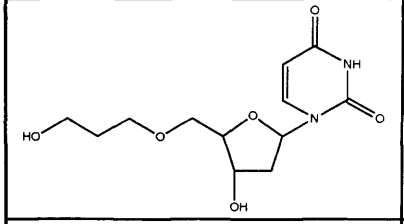
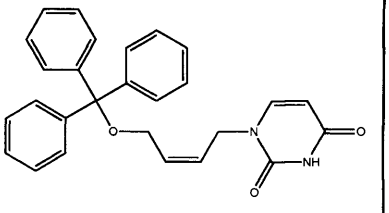
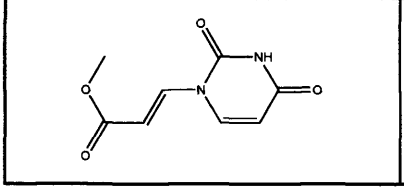
	1059	5.2 (>192)	>1mM	2.2		500.54
	1060	0.9 (>1111)	>1mM	3.8	33.4	425.52
	1062	>1m (<0.6)	635.5	19.1		212.25
	1063	32.2 (>31)	>1mM	3.4		316.35
	1064	9.2 (>98)	898	>18.2		274.27
	1065	16.5 (>60)	>1mM	14.5		317.34
	1066	111.8 (>9)	>1mM	16.9		283.32
	1067	1.69 (226)	383.2	0.4	50	470.63

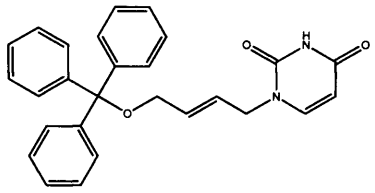
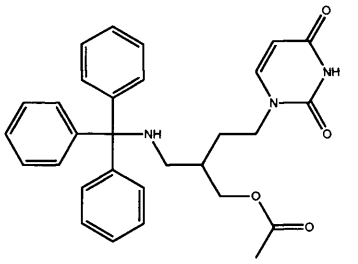
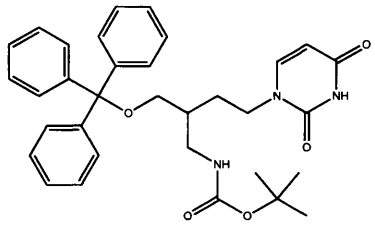
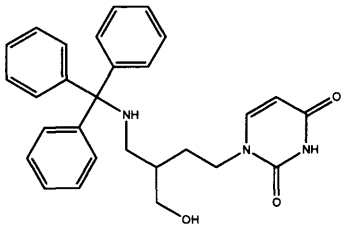
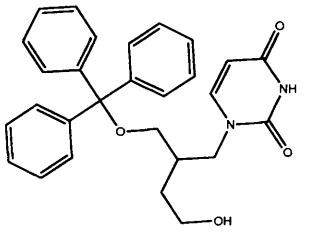
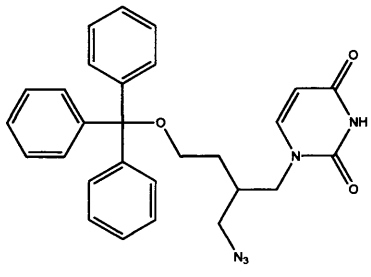
	1068	0.72 (520)	374.3	6.6		499.56
	1069	367.5 (>2)	>1mM	>29.4		170.17
	1072	1.56 (16)	25.1	0.58	70.2	428.56
	1073	1.3 (>769)	>1mM	1	85.4	468.58
	1134	0.76 (>1315)	>1mM	>11	-	452.5
	1142	2	>1mM	1.1	23.4	440.53
	1149	0.23 (4345)	>1mM	>10		470.52

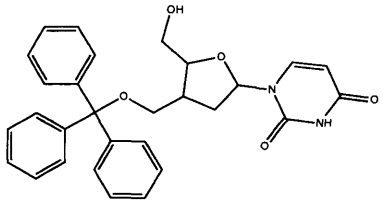
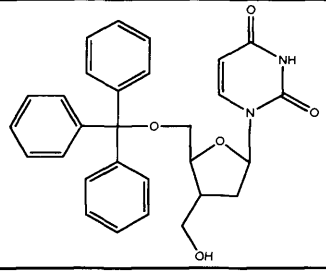
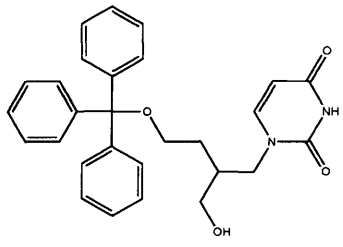
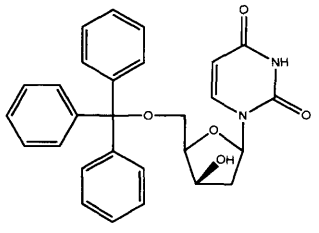
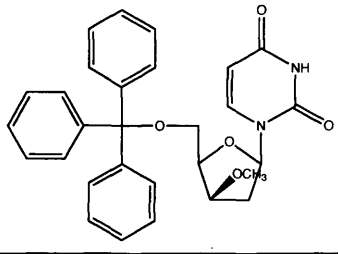
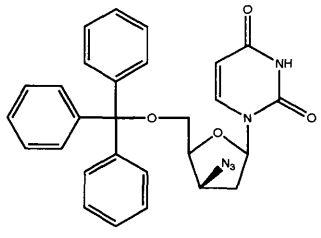
	1203	25.5 (>39)	>1mM	0.7	37.7	548.62
	1204	>1mM (>1)	>1mM	0.6	54.6	452.512
	1205	48.11 (>20)	>1mM	1.3	25.8	495.541
	1206	1.27 (>787)	>1mM	1	3.2	469.543
	1207	1.43 (>699)	>1mM	2.2	44.2	511.58
	1208	>1mM	>1mM	19.5	>454	198.22
	1211	>1mM	>1mM	0.6	137.7	276.4

	1223	>1mM (>1)	>1mM			
	1224	4.29 (>233)	>1mM	2.2	39.1	439.56
	1228	3.76 (>265)	>1mM	0.4	46.6	456.61
	1229	4.22 (>236)	>1mM	1.3	18.4	547.635
	1230	0.2 (6.75)	1.35	4.4	107.4	411.506
	1241	1.84 (>543)	>1mM	1.1	40.3	453.587

	1243	2.17 (>460)	>1mM	3.2	27.3	530.58
	1244	2.56 (>390)	>1mM	2.2	39.8	504.972
	1245	31.55 (>31)	>1mM	4.6	92.8	420.467
	1247	3.11 (>321)	>1mM	2.8	38.2	484.511
	1288	1.14 (715)	816	1.6	51.3	495.53
	1289	0.40 (579)	231.5	3.3	30.5	494.54

	1290	3.36 (70.8)	238	2	29.8	503.98
	1291	27.6 (>36)	>1mM	3.4	47.4	483.52
	1292	4.06 (173.6)	705	5.7	203.6	419.47
	1294	>1mM	>1mM	12.3	-	260.33
	1314	>1mM	>1mM	>17.5	>340	286.28
	1315	1.2 (4.3)	5.16	3.46	38.7	424.5
	1316	>1mM	>1mM	23.14	206.7	196.2

	1317	0.61 (>1639)	>1mM	3.65	33.9	424.5
	1318	1.33 (55.5)	73.9	2.84	37.3	497.6
	1319	2.48 (>403)	>1mM	0.89	14.7	555.7
	1320	0.24 (23.9)	5.74	3.17	55.4	455.6
	1321	0.9 (23.7)	21.4	1.85	57.1	456.5
	1322	23.1 (>43.2)	>1mM	0.64	18.4	481.6

	1323	>1mM	>1mM	0.48	20.8	484.5
	1324	2.11 (55.9)	118	3.5	28.6	484.5
	1328	2.15 (120.9)	260	1.15	32.6	456.5
	1361	0.3	>1000	1.9	26.8	470.52
	1362	350	>1000	2.1	24.6	512.55
	1363	49	>1000	4.5	25.6	495.53

Appendix 2

Computational details

- *GRID program (version 22)*

Probes used: DRY, C3, C1=, N2+, N1:, O, O::, O1, F
GRID box dimentions: chosen to encompass all the important parts of the active site.
GRID spacing 1 Å
MOVE option

- *Minimization*

Conjugate gradient method
Gradient: 0.001 kcal/mol
TRIPOS force field
Dielectric constant: 80.00
Dielectric funtion distant dependent
Interaction: 10000

- *FlexX program (multiple ligands)*

Active site dimention 6.5 Å (the ligand WSP869 was considered the centre of the active site)
Formal charges were defined by FlexX program
Verbosity=5
Num. answer=30
Output format=Mol2
Create answer table=on
Core SLN=uracil structure

- *LIAISON program*

1. Sampling method: Minimization
   Minimization algoritm: Conjugate gradient
   Residue-based cutoff distance: 15.00 Å
   Force field: OPLS_2001
   Maximum minimization steps: 1000
   RMS gradient for convergence: 0.01
2. Sampling method: Hybrid Monte Carlo
   Minimization algoritm: Conjugate gradient
   Residue-based cutoff distance: 15.00 Å
   Heating time: 20ps
   Simulation temperature: 300K
   Simulation time: 30ps
   Time semple between data collection: 10ps

- *3D-QSAR*

CoMFA Field Class: TRIPOS standar
Dielectric: Distance
Smoothing: none
Drop Electrostatic: Within Steric Cutoff for Each Row

Steric cutoff: 30.0
Electrostatic cutoff: 30.0
Transition: Smooth
Validation method: Leave-One-Out
Colum Filtering: 2.0

Appendix 3

Publications

- Gilbert, Ian; Nguyen, Corinne; Ruda, Gian Filippo; Schipani, Alessandro; Kasinathan, Ganasan; Johansson, Nils-Gunnar; Pacanowska, Dolores Gonzales. Preparation of deoxyuridine nucleosides as dUTPase inhibitors. *PCT Int. Appl.* **2005**, 73 pp. CODEN: PIXXD2 WO 2005066160 A1 20050721 CAN 143:133640 AN 2005:638870 CAPLUS

- Gilbert, Ian; Nguyen, Corinne; Ruda, Gian Filippo; Schipani, Alessandro; Kasinathan, Ganasan; Johansson, Nils-Gunnar; Pacanowska, Dolores Gonzales. Preparation of pyrimidinediones as deoxyuridine triphosphate nucleotidohydrolase (dUTPase) inhibitors for treatment of parasitic infections. *PCT Int. Appl.* **2005**, 76 pp. CODEN: PIXXD2 WO 2005065689 A1 20050721 CAN 143:153391 AN 2005:638741 CAPLUS

- Mc Carthy, O. K.; Schipani, A.; Buendia, A. M.; Ruiz-Perez, L. M.; Kaiser, M.; Brun, R.; Pacanowska, D. G.; Gilbert, I. H. Design, synthesis and evaluation of novel uracil amino acid conjugates for the inhibition of Trypanosoma cruzi dUTPase *Bioorganic & Medicinal Chemistry Letters* **2006,** 16, (14), 3809-3812.

- Nguyen, C.; Ruda, G. F.; Schipani, A.; Kasinathan, G.; Leal, I.; Musso-Buendia, A.; Kaiser, M.; Brun, R.; Ruiz-Perez, L. M.; Sahlberg, B. L.; Johansson, N. G.; Gonzalez-Pacanowska, D.; Gilbert, I. H. Acyclic nucleoside analogues as inhibitors of Plasmodium falciparum dUTPase *Journal Of Medicinal Chemistry* **2006,** 49, (14), 4183-4195.

Acyclic Nucleoside Analogues as Inhibitors of *Plasmodium falciparum* dUTPase

Corinne Nguyen,[†] Gian Filippo Ruda,[†] Alessandro Schipani,[†] Ganasan Kasinathan,[†] Isabel Leal,[‡] Alexander Musso-Buendia,[‡] Marcel Kaiser,[§] Reto Brun,[§] Luis M. Ruiz-Pérez,[‡] Britt-Louise Sahlberg,[∥] Nils Gunnar Johansson,[∥] Dolores González-Pacanowska,[‡] and Ian H. Gilbert*,[†]

Welsh School of Pharmacy, Cardiff University, Redwood Building, King Edward VII Avenue, Cardiff, CF10 3XF, UK, Instituto de Parasitología y Biomedicina, Consejo Superior de Investigaciones Científicas, Parque Tecnológico de Ciencias de la Salud, Avenida del Conocimiento, 18100 Armilla (Granada), Spain, and Swiss Tropical Institute, Socinstrasse 57, CH-4002 Basel, Switzerland, Medivir AB, Lunastigen 7, S-141 Huddinge, Sweden

Received February 4, 2006

We report the discovery of novel uracil-based acyclic compounds as inhibitors of deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase), an enzyme involved in nucleotide metabolism that has been identified as a promising target for the development of antimalarial drugs. Compounds were assayed against both *P. falciparum* dUTPase and intact parasites. A good correlation was observed between enzyme inhibition and cellular assays. Acyclic uracil derivatives were identified that showed greater or similar potency and in general increased selectivity compared to previously reported inhibitors. The most active compound reported here against the *P. falciparum* enzyme had a K_i of 0.2 μM. Molecular modeling studies provided a good rationale for the observed activities. Preliminary ADME studies indicated that some of the lead compounds are drug-like molecules. These compounds are useful tools for further investigating *P. falciparum* dUTPase for the development of much-needed novel antimalarial drugs.

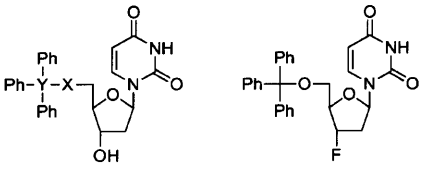
Introduction

Malaria is a major health problem, especially in the tropical regions of the world. Annually there are 300 million cases of malaria, and at least a million people, mainly children in sub-Saharan Africa, die from the disease.[1] The parasite (*Plasmodium falciparum*) that causes the most acute form of malaria is quick to become resistant to chemotherapeutic agents. For example, in many places there is a major problem of resistance to chloroquine, a 4-aminoquinoline that was for many years the mainstay of treatment for malaria. Hence, there is urgent need for the development of new drugs against new molecular targets, where cross-resistance to existing agents is not likely to be a problem. MMV[a] estimates a new drug is required for combating malaria every 5 years.[2]

We have recently discovered that some 5′-tritylated nucleosides (**1a**−**d**, Table 1) are selective inhibitors of the *P. falciparum* enzyme deoxyuridine nucleotidohydrolyase (dUTPase).[3] This enzyme, which has not been exploited as a drug target, is involved in nucleotide metabolism. It is responsible for the hydrolysis of dUTP to dUMP and in doing so fulfils two roles. First, it provides a source of dUMP, the substrate for thymidylate synthase involved in the only route that is available for biosynthesis of dTMP (which is subsequently converted into dTTP). Second, it helps maintain a low dUTP/dTTP ratio, reducing erroneous incorporation of uracil into DNA. dUTPase is found in almost all organisms, and it has been shown to be essential for viability where this has been investigated.[4−6] The *Plasmodium* and human dUTPase enzymes both belong to the trimeric subclass of dUTPases and share many kinetic and structural characteristics. They are very selective for dUTP over other deoxyribose nucleoside triphosphates and also ribose nucleoside triphosphates (e.g. UTP).[7] Both enzymes share a common overall topology and show the five conserved motifs characteristic of trimeric dUTPases.[7,8] Interestingly there is an insertion of approximately 23 amino acids in the *Plasmodium* dUTPase.[9] Despite the overall similarities, there are subtle structural differences between the *Plasmodium* and human dUTPases, which allow for selective inhibition of the *P. falciparum* enzyme.[3,9]

Previously reported inhibitors of dUTPases include non-hydrolyzable isosteres of nucleoside triphosphates, where the

Table 1. Previously Reported Data for Inhibition of *P. falciparum* DUTPase and Growth Inhibition of *P. falciparum* Cultured in Red Blood Cells by 5′-Tritylated Nucleosides[3]



1a, X = O, Y = C
1b, X = NH, Y = C
1c, X = O, Y = Si

1d

	dUTPase		in vitro	
compound	*P. falciparum* K_i (μM)	human K_i (μM)	*P. falciparum* IC_{50} (μM)	L6-cells IC_{50} (μM)
1a	1.8	18	6	192
1b	0.2	46	4.5	44
1c	2.8	909	1.1	nd
1d	5.0	457	2.0	nd

[a] nd: not determined. The IC_{50} values are the means of two independent assays performed in duplicate.

* To whom correspondence should be addressed. Tel: +44 (0) 1382 386240. Fax: +44 (0) 1382 386373. E-mail: i.h.gilbert@dundee.ac.uk. Current address: School of Life Sciences, University of Dundee, MSI/WTB/CIR Complex, Dow Street, Dundee, DD1 5EH, Scotland, UK.
[†] Cardiff University.
[‡] Instituto de Parasitología y Biomedicina.
[§] Swiss Tropical Institute.
[∥] Medivir.

[a] Abbreviations: dUTPase, deoxyuridine nucleotidohydrolase; MMV, Medicine for Malaria Venture, d4U, 2′,3′-deoxy-2′,3′-didehydrouridine, ADME, adsorption, distribution, metabolism, excretion; PS-PPh_3, polymer-supported triphenylphosphine; TPSO, triphenylsilyloxy; DTT, dithiothreitol; IPTG, isopropyl β-D-thiogalactopyranoside; PMSF, α-toluenesulfonyl fluoride, FBS: fetal bovine serum; BSA,bovine serum albumin.

10.1021/jm060126s CCC: $33.50
Published on Web 06/13/2006

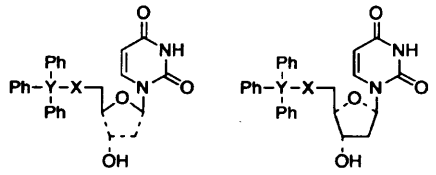


Figure 1. Analogy between acyclic and deoxyuridine derivatives.

α,β bridging oxygen atom is replaced by a methylene or an amino group.[10,11] These have been excellent tools for investigating the kinetics of the dUTPase enzyme and for generating crystal structures of the enzyme from various species. However, such compounds are unlikely to be drug candidates owing to chemical and enzymatic instability, poor bioavailability, and membrane permeability. On the contrary, the 5′-tritylated deoxyuridine analogues that we have reported and shown to possess selective antiplasmodial activity are more drug-like, having improved stability, reasonable logP values and low molecular weight.[3]

In this paper we discuss further modifications of the 5′-tritylated deoxyuridine derivatives, in particular generation of acyclic analogues (Figures 1 and 2). These may offer advantages in terms of optimizing the distance between the base (uracil) and trityl group and allow for further alterations and optimization of drug structure. Acyclic nucleoside derivatives are successful antiviral agents; e.g. aciclovir for the treatment of herpes infections. Initially we prepared an acyclic derivative in which there was an oxygen atom in the chain (Figure 2, compound **2a**). However, this particular analogue proved to be metabolically unstable, poorly selective, and fairly cumbersome to purify. Therefore we prepared a series of compounds with no oxygen in the chain, using simple and versatile methodologies (Figure 2, Series **3** to **6**). In addition, to investigate the effect of adding rigidity to the side-chain, compounds in which there was both an *E* and *Z* double bond in the acyclic chain (compounds **7a** and **8a**) were synthesized. Previous studies have shown that d4U analogues[3] were potent inhibitors of the *Plasmodium* dUTPase, possibly due to favorable π interactions with a tyrosine residue (Tyr112) within the enzyme active site. Finally, some branched acyclic derivatives were prepared: γ-branched analogues of the penciclovir type form Series **9**, while Series **10** encompasses the β-branched analogues. Such branched structures were designed to partially mimic the backbone of 2′-deoxyuridine and possibly increase affinity for the enzyme active site.

Synthesis

Series 2. 1-[2-(Benzoyloxyethoxy)methyl]uracil **2g** was synthesized by condensation of bis-silylated uracil (prepared in situ by reaction of uracil with *N,O*-bis[trimethylsilyl]-acetamide) with the commercially available 2-(chloromethoxy)-ethyl benzoate, in the presence of the Lewis acid catalyst trimethylsilyl triflate. Purification proved cumbersome (possibly due to impurities present in chloromethoxyethyl benzoate), and **2g** was obtained in 30% yield, although with small traces of impurities. Removal of the benzoate ester by treatment with potassium carbonate in methanol afforded the alcohol 1-[2-(hydroxyethoxy)methyl]uracil (**11**, 49%), which was further derivatized with trityl chloride and *tert*-butyldimethylsilyl chloride (TDMSCl), to produce the trityl ether (**2a**, 20%) and TBDMS ether (**2f**, 56%), respectively.

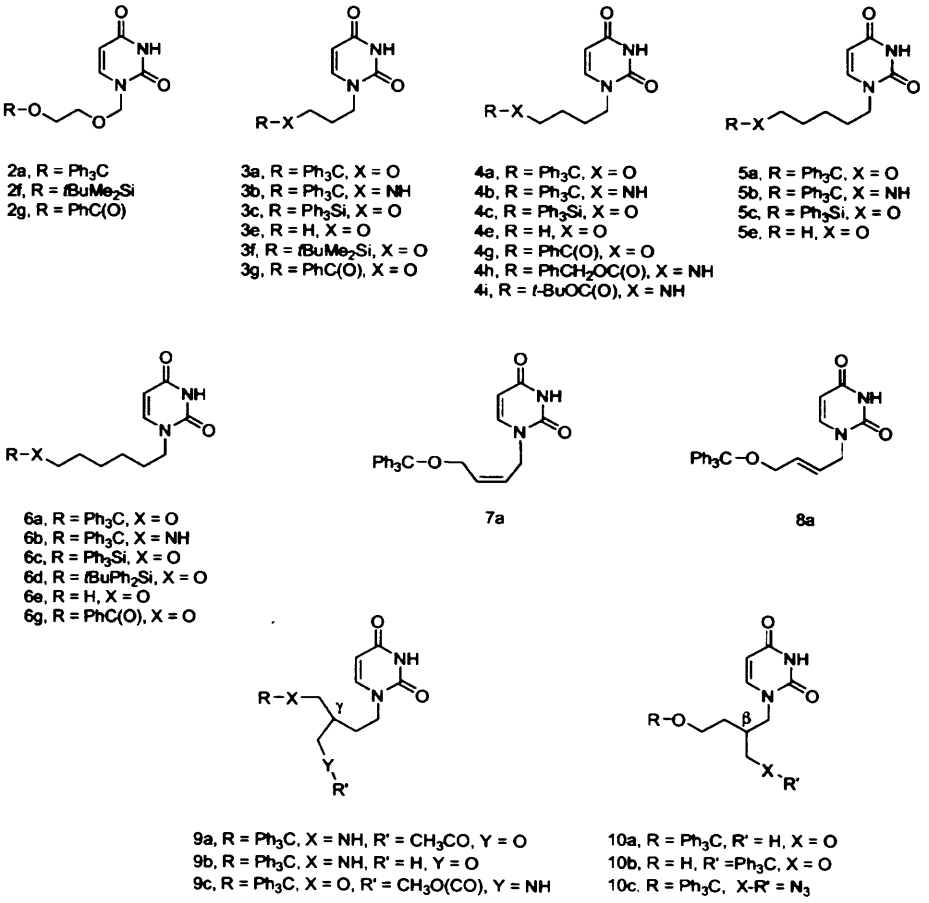
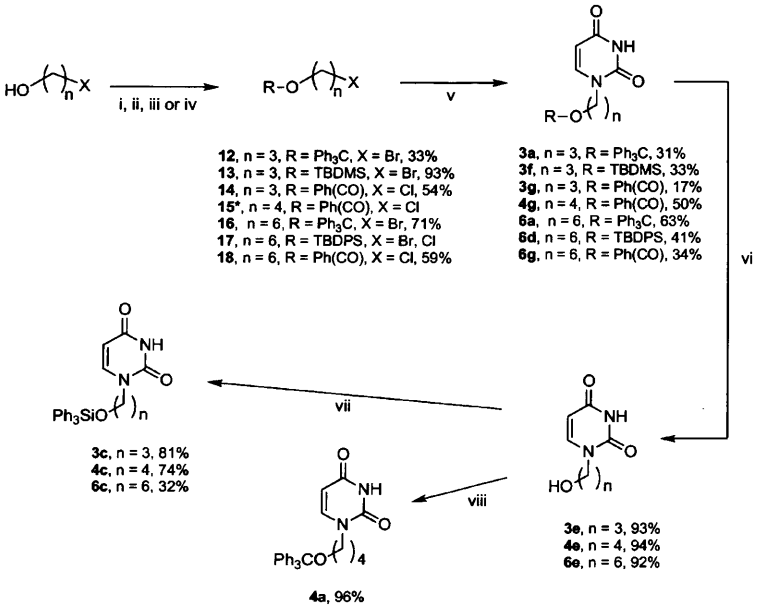


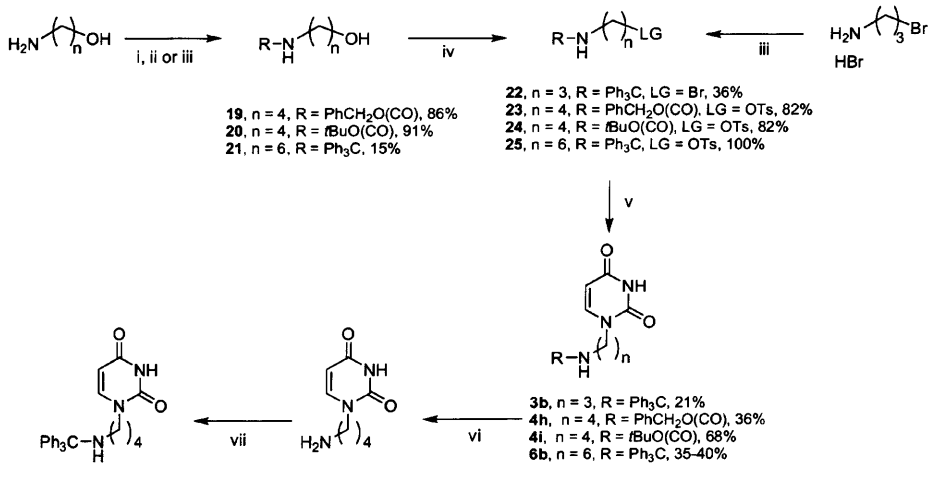
Figure 2. Structures of compounds assayed.

Scheme 1[a]



[a] (i) TrCl, Et_3N, DMAP, DCM, rt; (ii) TBDMSCl, imidazole, DMF, rt; (iii) PhCOCl, pyridine, 0 °C−rt; (iv) TBDPSCl, imidazole, DMF, 0 °C−rt; (v) uracil, Cs_2CO_3, DMF, 40−90 °C; (vi) From **3g**, **4g**, **6g**; MeONa, MeOH, rt; (vii) Ph_3SiCl, pyridine, 0 °C; (viii) TrCl, pyridine, 50 °C. *Compound **15** was purchased from Acros.

Scheme 2[a]



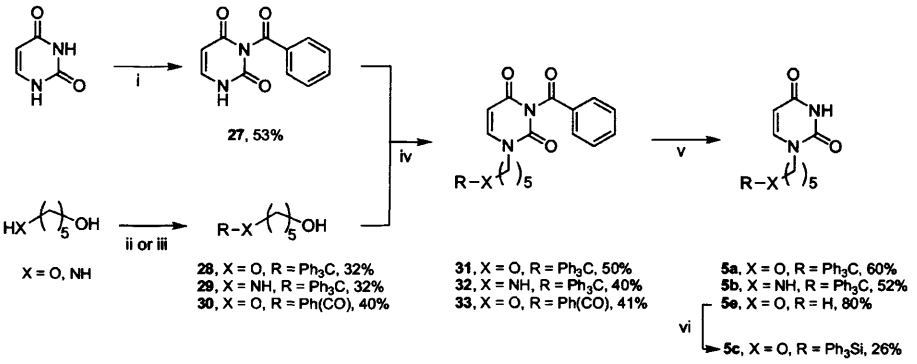
[a] (i) Boc_2O, Et_3N, DCM, rt; (ii) $PhCH_2OCOCl$, Na_2CO_3, H_2O, 0 °C-rt; (iii) TrCl, Et_3N, DCM, 0 °C; (iv) TsCl, pyridine, 0 °C or TsCl, Et_3N, DMAP, 0 °C; (v) uracil, Cs_2CO_3, DMF, 40 °C; (vi) From **4h**, H_2, 5% Pd/C, MeOH, rt; (vii) TrCl, pyridine, 50 °C.

Series 3, 4, and 6. The preparation of these compounds is summarized in Schemes 1 and 2. The key step was the direct coupling of uracil with the relevant halogenated aliphatic side-chain, in the presence of cesium carbonate in DMF under moderate heating (40 °C). The chemoselectivity of the alkylation was verified by two-dimensional NOESY NMR (showing the coupling through space between 1′-H of the side-chain and 6-H of the uracil, see Supporting Information).

As depicted in Scheme 1, the alcohol functionality of the linear side-chain fragment was protected as either a benzoate ester, silyl ether, or trityl ether. In the case of the C4 analogues, an ether protection was not suitable due to the ease of intramolecular cyclization of the side-chain moiety (formation of a five-membered ring). Therefore, the trityl ether and triphenylsilyl ether analogues were not prepared directly but via etherification of 1-(4-hydroxybutyl)uracil (**4e**). Hydroxy uracil analogues **3e**, **4e**, and **6e** were prepared very neatly in high yields by treatment of the benzoate ester derivatives **3g**, **4g**, and **6g** with sodium methoxide in methanol. Furthermore, as the triphenylsilyl (TPS) group was expected not to be completely stable under the reaction conditions required for the coupling step, its introduction was left until last. Analogues **3c**, **4c**, and **6c** were obtained by reaction of the corresponding alcohols with TPSCl in pyridine at 0 °C, in 81, 74, and 32% yields, respectively.

Derivatives bearing a terminal amino group in their side-chain (Scheme 2) were prepared following a methodology similar to their oxygen analogues. Tritylaminopropyl and tritylaminohexyl

Scheme 3[a]



[a] (i) 1. PhCOCl, pyridine, CH_3CN, rt; 2. K_2CO_3, dioxane; 30 min; (ii) TrCl, Et_3N, DCM, 0 °C−rt; (iii) PhCOCl, Me_2SnCl_2, K_2CO_3, THF, rt; (iv) DIAD, PS-PPh_3, THF, rt; (v) NaOH, dioxane, rt, or MeONa, MeOH, rt; (vi) Ph_3SiCl, pyridine, rt.

analogues **3b** and **6b** could be obtained by direct coupling of uracil with the suitably activated aliphatic fragment (**22** and **25**, respectively). For the C4 analogue, the side-chain was N-protected as a carbamate (**23** or **24**) prior to coupling to the uracil to avoid intramolecular cyclization of the side-chain. Removal of the benzyloxycarbonyl group by hydrogenation in the presence of 5% Pd/C yielded the amino analogue **26**, which was finally tritylated to give the target **4b**.

Series 5, Compounds 7a and 8a. The synthesis of compounds from Series 5 is described in Scheme 3. As with their C4 analogues, the C5 side-chain precursors can be prone to intramolecular cyclization, in this case due to the ease of formation of six-membered rings. A way to overcome this problem is to use the Mitsunobu reaction in the key coupling step: in the presence of diisopropyl azadicarboxylate (DIAD) and triphenylphosphine (PPh_3), uracil can react with a variety of aliphatic chains bearing an alcohol without need for it to be activated. A convenient procedure developed by Hernández et al.[12] employs polymer-supported PPh_3 (PS-PPh_3) which facilitates the otherwise potentially tricky removal of the triphenylphosphine oxide byproduct and excess PPh_3 reagent. To avoid N^3 alkylation, uracil was protected prior to coupling. This was achieved in a two-step procedure consisting of N^1,N^3 double protection with benzoyl chloride in pyridine, followed by selective N^1 deprotection carried out using potassium carbonate in acetonitrile.[13] Pentane-1,5-diol and 5-aminopentan-1-ol were tritylated using trityl chloride and triethylamine in DCM, with an excess of diol or amino alcohol to minimize ditritylation. The diol was also monobenzoylated using benzoyl chloride in the presence of dimethyl tin dichloride as the catalyst inducing monoprotection.[14] Alcohols **28**−**30** thus obtained were successfully coupled to N^3-benzoyluracil using the Mitsunobu methodology to give the N^3-benzoyl precursors **31**−**33** in 40−50% yield (Scheme 3). Cleavage of the benzoyl group was performed either with sodium methoxide in methanol or with NaOH in dioxane to yield the target analogues **5a**, **5b**, and **5e** in 52−80% yield. Finally, the C5 triphenylsilyl ether **5c** was obtained from the hydroxyuracil derivative **5e**, as previously described for the corresponding C3, C4, and C6 analogues.

The same general strategy was applied to the synthesis of unsaturated derivatives **7a** and **8a**, using (*Z*)-4-trityloxybut-2-en-1-ol (**34**) and the *E* isomer (**36**), as the respective side-chain units in the Mitsunobu coupling step. Intermediates **34** and **36** were prepared as described by Hernandez et al.[12] and McDonald et al.[15] (see Supporting Information).

Series 9 and 10. The branched acyclic analogues (Series 9 and 10) were synthesized through a combination of standard protections/deprotections of orthogonal protecting groups, from the branched precursors **39**, **42** (Schemes 4 and 5), and **47** (1-[2-azidomethyl-4-*tert*-butyldimethylsilyloxybutyl]uracil). For a first screen, no attempts were made at producing enantiomerically pure samples, and the compounds were tested as enantiomeric mixtures. Noteworthy is the successful use of microwave irradiation to produce the tritylamino analogue **9c**, where various attempts using conventional heating failed. Finally, compound **10c** was prepared in two straightforward steps from the TBDMSO branched analogue **47**: removal of the TBDMS group with TBAF in THF to free the hydroxyl group (intermediate **48**) followed by tritylation.

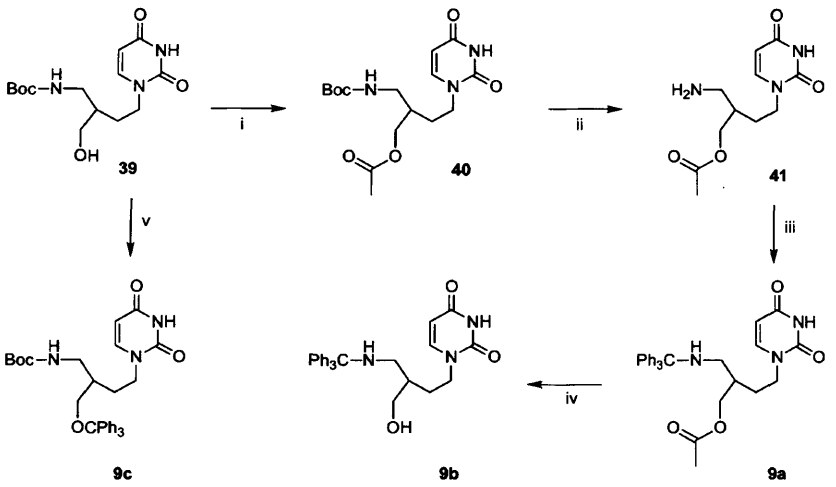
Biology

Enzyme Assays. Compounds were screened against the recombinant *P. falciparum* dUTPase and recombinant human dUTPase, to evaluate inhibition and selectivity (SI, selectivity index). The results from these assays are shown in Tables 2 and 3.

Compound **2a** was the closest single chain acyclic analogue of 5′-trityloxy-2′-deoxyuridine **1a**; it retained similar or even slightly improved (2-fold) enzyme activity and selectivity (**2a**, K_i = 0.7 μM, SI = 25; **1a**, K_i = 1.8 μM, SI = 10). Replacing the trityl (Tr) ether by a *tert*-butyldimethylsilyl (TBDMS) ether, a benzoate ester or an adamantyl amide group resulted in loss of activity, suggesting that lipophilicity and steric bulk alone did not account for enzyme inhibition. This is in agreement with conclusions previously reported for deoxyuridine derivatives[3] and is discussed later in this paper for other series of compounds assayed.

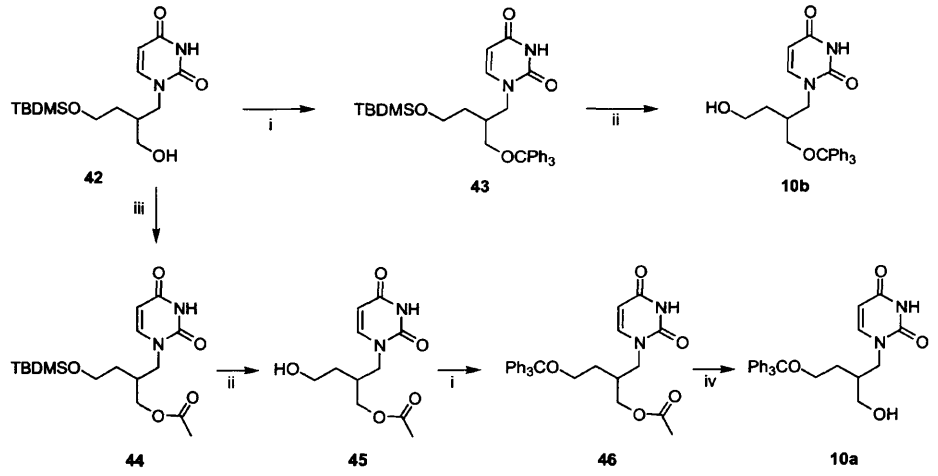
Compound **4a** gave rise to a similar inhibition of the *P. falciparum* enzyme compared to compound **2a** (**2a**, K_i = 0.7 μM; **4a**, K_i = 1.6 μM), which demonstrated that while having an oxygen in the side-chain probably is a closer mimic of the nucleoside, this is not a key feature for activity. Furthermore, **4a** showed a much lower affinity for the human enzyme than **2a** (Table 2; **4a**, K_i > 1000 μM; **2a**, K_i = 17 μM) and consequently was at least 25 times more selective (selectivity index = [K_i human]/[K_i *P. falciparum*)]; Table 3). Hence, it can be postulated that while the oxygen does not play a significant role in *P. falciparum* dUTPase inhibition, it undergoes interactions that are more important within the human enzyme active site. In addition, the 2-(trityloxyethoxy)methyl acyclic derivative **2a** was found (i) to be prone to rapid degradation by human liver microsomes (see Discussion, ADME tests) and (ii) to require a cumbersome purification (see Synthesis). Therefore, Series **2** was not progressed further.

Scheme 4[a]



[a] (i) Ac_2O, pyridine, rt, 88%; (ii) TFA, DCM, rt; (iii) TrCl, Et_3N, DCM, rt, 70%; (iv) MeONa, MeOH, rt, 99%; (v) TrCl, pyridine, microwave irradiation (three times 160 °C for 5 min), 79%.

Scheme 5[a]



[a] (i) TrCl, pyridine, 70 °C; **42**, 76%; **45**, 90%; (ii) TBAF, THF, rt; **10b**, 47%; **44**, 78%; (iii) Ac_2O, pyridine, rt, 67%; (iv) MeONa, MeOH, rt, 57%.

Instead, using the simple and versatile synthetic methodology used to prepare compound **4a**, more analogues with varying side-chain length (n = 3 to 6) in combination with a selection of functional groups at the end of their aliphatic chain were assayed (Table 3). Functional moieties considered (Figure 2, Table 3) included trityl ether, tritylamine, various silyl ethers, alcohol, benzoate ester, and carbamates (Cbz and Boc) for varying steric bulk, lipophilicity and electronic characteristics.

Effect of Side-Chain Length. No general correlation between activity and side-chain length was revealed. Within the trityloxy Series **a**, the lowest K_i values were obtained for **4a** and **5a**. Within the tritylamino Series **b**, compound **3b** was the most active, although the least selective as well. Within the triphenylsilyloxy Series **c**, activity was retained with no significant change in K_i when the chain length was varied from three to six carbons. However, the chain length of three carbon atoms correlated with a lower selectivity for the parasite over the human enzyme. Possibly there is also a slight reduction in selectivity for the chain length of six carbon atoms. Compound **4b** (acyclic analogue of lead **1b**) appeared as the best overall inhibitor (K_i = 0.9 μM, SI > 1111).

In addition, when focusing on the effect of the O/NH switch at the end of the side-chain within a series (series **3** to **6** considered individually), few variations are observed as regards activity, with the clear exception of the C3 analogues; indeed while the trityl ether **3a** did not show significant inhibition of *P. falciparum* dUTPase (K_i = 87 μM), the corresponding tritylamine **3b** showed a greatly improved activity (over 400-fold) with a K_i = 0.2 μM. A 5-fold increase in activity along with a 10-fold improvement in selectivity can also be noted when replacing the trityloxy by a tritylamino group in the C6 series (Table 2; **6a**, K_i = 9.6 μM, SI = 50; and **6b**, K_i = 1.8 μM, SI > 543). Very similar observations can be made when considering a C/Si switch within the trityl group itself (that is when replacing the trityloxy by a triphenylsilyloxy group). Indeed, (i) the silyl ether **3c** was much more active than its analogue **3a** (K_i = 1.6 and 87 μM, respectively); (ii) within Series **6**, the silyl ether **6c** showed both increased activity and greater selectivity compared to **6a** (Table 2; **6c**, K_i = 1.7 μM,

Table 2. Inhibition of the *P. falciparum* and Human dUTPases, and Growth Inhibition of *P. falciparum* and Mammalian Cell Line (rat L6-cells)

	dUTPase		in vitro	
compound	*P. falciparum* K_i (μM)	Human K_i (μM)	*P. falciparum*[a] IC_{50} (μM)	L6-cells[b] IC_{50} (μM)
2a	0.7	17	0.9	39
2f	44	63	62	>299
2g	98	311	127	202
3a	87	313	7.5	34
3b	0.2	1.4	4.4	107
3c	1.6	25	0.58	70
3e	368	>1000	>29	nd
3f	73	73	>18	192
3g	9.2	898	>18	nd
4a	1.6	>1000	4.9	42
4b	0.9	>1000	3.8	33
4c	2.2	>1000	0.36	nd
4e	>1000	>1000	235	nd
4g	224	496	38	nd
4h	17	>1000	15	nd
4i	112	>1000	17	nd
5a	2.0	>1000	1.1	24
5b	4.3	>1000	2.2	39
5c	3.8	>1000	0.38	47
5e	>1000	>1000	20	>454
6a	2.3	476	2.3	44
6b	1.8	>1000	1.1	40
6c	1.7	383	0.46	50
6d	8.5	>1000	3.9	18
6e	>1000	636	19	nd
6g	32	>1000	3.4	nd
7a	1.2	5.2	3.5	39
8a	0.6	>1000	3.7	34
9a	1.3	74	2.8	37
9b	0.24	5.7	3.2	55
9c	2.5	>1000	0.9	15
10a	2.2	260	1.2	33
10b	0.9	21	1.9	57
10c	23	>1000	0.64	18
Ph_3SiOH	>1000	>1000	0.65	138
Ph_3COH	>1000	>1000	12	nd
TDBMSOH	nd	nd	>37	>680

[a] *Plasmodium falciparum*; [b] Measure of cytotoxicity; nd: not determined. Controls: for *P. falciparum*, chloroquine, IC_{50} = 0.1 μM; for cytotoxicity, podophyllotoxin, IC_{50} = 0.012 μM. The IC_{50} values are the means of four values of two independent assays done in duplicate.

SI = 226; and **6a**, K_i = 9.6 μM, SI = 50); and (iii) little variation was noticeable within Series **4** and **5**.

Unsaturated Acyclic Analogues 7a and 8a. The two isomers **7a** and **8a** showed similar levels of inhibition of the *Plasmodium* dUTPase, also comparable to that of their saturated analogue **4a**, with the *E* isomer **8a** being possibly slightly more active (Table 2; **4a**, K_i = 1.6 μM; **7a**, K_i = 1.2 μM; **8a**, K_i = 0.6 μM). On the other hand, a great difference in activity against the human dUTPase was observed: **7a** gave rise to some inhibition of the enzyme (K_i = 5.2 μM) whereas **8a** was totally inactive (K_i > 1 mM). Consequently, while **7a** showed very poor selectivity (SI = 4), its *E* isomer **8a** was very selective toward the *P. falciparum* enzyme (SI > 1666). This striking observation suggests that introduction of rigidity through a double bond in position 2,3 of the side-chain can be beneficial (Table 2, compare **4a** [K_i = 1.6 μM; SI > 617] and **8a** [K_i = 0.6 μM; SI > 1666]); however, the choice of conformation of the double bond is crucial if selectivity is to be retained.

Branched Acyclic Analogues (Series 9 and 10). One of the key SAR features established so far for both acyclic and cyclic[3] analogues is the requirement for a group bearing two or even three phenyl rings, trityl and triphenylsilyl being the groups of choice. Therefore, the branched acyclic derivatives prepared to further our investigations all contained one such group (either TrO or TrNH), as well as another functional group in the branched chain. Furthermore, it was decided that one chain would be four carbons long as a mimic of the nucleoside analogues.

The γ-branched derivative **9b** represented the closest analogue to lead **1b** and was found to be as potent an inhibitor of *P. falciparum* dUTPase, with a K_i value of 0.2 μM. However this result was partially contrasted by the 10-fold loss in selectivity observed for **9b** (Table 2). This selectivity loss was a direct consequence of an increased inhibition of the human dUTPase (Table 2 and Figure 1; **1b**, K_i = 46 μM; **9b**, K_i = 5.7 μM). Thus, increasing the flexibility of the inhibitor seemed to allow enhanced interactions unfortunately only within the human dUTPase.

Blocking the free hydroxyl in **9b** as an acetate ester in **9a** resulted in a noticeable 7-fold loss of activity against the *P. falciparum* dUTPase. However, a slight increase in selectivity was also observed. This suggests that the loss of the OH hydrogen bond donor has a greater effect on interactions within the human dUTPase than with the *P. falciparum* enzyme.

Compound **9c**, that possesses a trityl ether at the end of one C4 chain and a Boc carbamate at the end of the other C4 chain, gave very similar results to the unbranched analogue **4a**.

In series **10**, there is branching at the β-position. Similar inhibition of the *Plasmodium* dUTPase was seen for the analogue with a β-hydroxymethyl compared to the unbranched analogue **4a**, although with a slight loss of selectivity. Replacing the hydroxyl group with an azide group (**10c**) led to a 10-fold loss in activity.

Compound **10b** has a three carbon chain with a β-hydroxyethyl branch. Comparison of **10b** with its unbranched C3 trityloxy analogue **3a**, shows that the introduction of a hydroxyethyl β-branch enhanced antiplasmodial activity (40-fold) as well as selectivity (6-fold). However, the latter remains in the low range (SI = 23).

In Vitro Assays

Compounds were screened against *P. falciparum* cultured in red blood cells to evaluate antiplasmodial activity and against a mammalian cell line (L-6 rat cells) as a test for cytotoxicity (Tables 2 and 4).

In general, compounds with a trityloxy (Series **a**), tritylamino (Series **b**), or triphenylsilyloxy (Series **c**) substituent inhibited the growth of *P. falciparum*, with IC_{50} values in the order of 0.4 to 8 μM (Table 2). The most potent inhibitors were the TPSO analogues **4c**, **5c**, and **6c** (IC_{50} = 0.36, 0.38, and 0.43 μM, respectively), with IC_{50} values up to 10-fold lower than compounds from the TrO and TrNH series. Compound **6d**, in which there is a diphenyl substituent, also gave good inhibition of parasite growth (IC_{50} = 3.9 μM). Furthermore, these compounds were selective for the parasite over mammalian (L-6) cells, the degree of selectivity varying from about 5-fold to over 100-fold.

Compounds lacking a triphenyl or diphenyl substituent showed much weaker inhibition of *P. falciparum* growth, with only compound **6g** giving an IC_{50} value under 10 μM (Table 2). This mirrors the trends observed for the nucleoside analogues.[3] Furthermore, acyclic inhibitors show similar or more potent parasite growth inhibition than their deoxyuridine counterparts (e.g. **4a**, IC_{50} = 4.9 μM and **1a**, IC_{50} = 6 μM; **4b**, IC_{50} = 3.8 μM and **1b**, IC_{50} = 4.5 μM; **4c**, IC_{50} = 0.36 μM and **1c**, IC_{50} = 1.1 μM). The side-chain length did not appear

Table 3. Overview of Enzyme Inhibition for Compounds from Series **2** to **6**[a]

no.	R	**2**, $R(CH_2)_2OCH_2U$	**3**, $R(CH_2)_3U$	**4**, $R(CH_2)_4U$	**5**, $R(CH_2)_5U$	**6**, $R(CH_2)_6U$
a	Ph_3CO	0.7 (25)	87 (4)	1.6 (>617)	2.0 (>500)	9.6 (50)
b	Ph_3CNH		0.2 (7)	0.9 (>1111)	4.3 (>233)	1.8 (>543)
c	Ph_3SiO		1.6 (16)	2.2 (>446)	3.8 (>265)	1.7 (227)
d	*t*-$BuPh_2SiO$					8.5 (>118)
e	HO		368 (>2)	>1 mM	>1 mM	>1 mM
f	*t*-$BuMe_2SiO$	44 (1.4)	73 (1)			
g	PhC(O)	98 (3)	9.2 (>97)	224 (2)		32 (>31)
h	$PhCH_2OC(O)NH$			17 (60)		
i	*t*-BuOC(O)NH			112 (>8)		

[a] K_i values against *P. falciparum* dUTPase are given in μM. The selectivity index defined as (K_i human/K_i *P. falciparum*) is given in parentheses. U = uracil.

Table 4. Overview of *P. falciparum* Growth Inhibition for Compounds from Series **2** to **6**[a]

no.	R	**2**, $R(CH_2)_2OCH_2U$	**3**, $R(CH_2)_3U$	**4**, $R(CH_2)_4U$	**5**, $R(CH_2)_5U$	**6**, $R(CH_2)_6U$
a	Ph_3CO	0.9 (78)	7.5 (5)	4.9 (8)	1.1 (21)	2.3 (19)
b	Ph_3CNH		4.4 (24)	3.8 (8.8)	2.2 (18)	1.1 (37)
c	Ph_3SiO		0.58 (121)	0.36 (na)	0.38 (123)	0.46 (109)
d	*t*-$BuPh_2SiO$					3.9 (5)
e	HO		>29 (na)	235 (na)	20 (>23)	19 (na)
f	*t*-$BuMe_2SiO$	62 (>4)	>18 (<11)			
g	PhC(O)	127 (1.6)	>18 (na)	38 (na)		3.4 (na)
h	$PhCH_2OC(O)NH$			14.5 (na)		
i	*t*-BuOC(O)NH			17 (na)		

[a] IC_{50} values are given in μM. The selectivity index defined as (IC_{50} *P. falciparum*/IC_{50} L6-cells) is given in parentheses. U = uracil. na: not available.

to have a clear influence on antiplasmodial activity in vitro, as seen in the case of enzyme inhibition.

Evidence from the literature suggests that bonds between silicon and oxygen could be labile in cellular conditions.[16,17] Therefore, the commercially available triphenylsilanol (TPSOH) and *tert*-butyldimethylsilanol (TBDMSOH) were also investigated for their antiparasitic activity (Table 2). TBDMSOH was inactive against *P. falciparum* (IC_{50} > 37 μM). On the other hand, TPSOH was found to be a potent inhibitor of *P. falciparum* growth, with a level of activity (IC_{50} = 0.3 μM) similar to that of the TPSO analogues (Series **c**). It may be possible that compounds from Series **c** owe some or all of their antiplasmodial activity to TPSOH potentially released as a result of hydrolysis under cellular assay conditions. However, it was shown previously[3] that the *P. falciparum* dUTPase was inhibited by the TPSO derivatives but not by TPSOH (K_i > 1000 μM) nor by the corresponding alcohols (Series **e**), which demonstrates that the TPSO analogues, as such, were responsible for enzyme inhibition. Therefore, if TPSOH is inhibiting parasite growth then it is acting on a molecular target other than dUTPase (and not yet identified).

Both unsaturated analogues **7a** and **8a** were good inhibitors of *P. falciparum* growth with virtually identical IC_{50} values (3.5 and 3.7 μM, respectively), but with relatively low selectivity compared to mammalian cells growth (IC_{50} = 39 and 34 μM, respectively).

Branched analogues from Series **9** and **10** all showed strong inhibition of *P. falciparum* growth in vitro, with IC_{50} values ranging from 0.6 to 3.2 μM. None of the compounds were found to be very cytotoxic (Table 2). Furthermore, compared to their nucleoside and single chain analogues, the branched derivatives exhibited similar or slightly better antiplasmodial activity (Table 2). In particular, compare the IC_{50} values for γ-branched **9b** (3.2 μM) and its nucleoside analogue **1b** (4.5 μM); for the γ-branched **9a** (2.8 μM, hydroxymethyl branch) and **9b** (3.2 μM, acetoxymethyl branch) and their single chain analogue **4b** (3.8 μM); for **9c** (0.9 μM) and the straight chain **4a** (4.9 μM). A hydroxymethyl (**10a**) or azidomethyl (**10c**) β-branch added onto **4a** increased antiplasmodial activity (four times in the case of **10a** and a more significant eight times in the case of **10c**). Similarly, compared to acyclic **3a** (IC_{50} = 7.5 μM), its hydroxyethyl β-branched analogue **10b** (IC_{50} = 1.9 μM) was slightly more active in vitro.

Molecular Modeling

In an attempt to rationalize the results of the enzyme inhibition, the inhibitors were docked into the active site of the *P. falciparum* enzyme. Previously we have reported the crystal structure of *P. falciparum* dUTPase complexed with **1d**.[9] This was used as a starting point for the modeling. FlexX[18] as implemented in the Sybyl software[19] was used for the docking. To validate the modeling, compound **1d** was docked in to the P. *falciparum* dUTPase active site, in an attempt to reproduce the binding conformation found in the crystal structure. There was an RMSD of 1.14 Å between the docked and crystallographically determined conformations of the **1d** (Figure 3a), suggesting that the docking was reliable.

Hence, all the inhibitors were docked into the active site. In general, it may be observed that the inhibitors bind in a manner similar to which **1d** was found to bind crystallographically (Figure 3b and 3c).[9] All the inhibitors have the uracil moiety positioned in the bottom of the active site forming hydrogen bonds with the side chain of the residue Asn103 and the chain of residue Ile117. Also the inhibitors have the trityl moiety positioned in a hydrophobic part of the active site, interacting with the side chain of residues Phe46 (forming a $\pi-\pi$ interaction) and Ile117 (Figure 3d). Furthermore, it is possible that Lys96 might also interact with the triphenyl group forming a cation$-\pi$ interaction. Interestingly, Phe46 is not present in the human structure, but is replaced by Val42, which may partially account for the selectivity of the compounds for the parasite.

The aliphatic chain that links the uracil group and the trityl moiety interacts with Tyr112 by van der Waals interaction. The flexibility of these acyclic nucleoside derivatives allows them to adopt a folded conformation which still retains the uracil and trityl compounds in their binding pockets.

A complete table of docking scores obtained with FlexX is reported and included in the Supporting Information of this

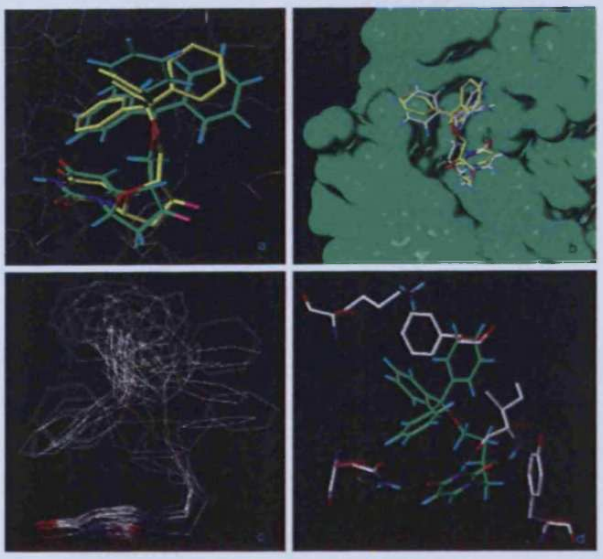


Figure 3. (a) Superimposition of the binding conformation of compound **1d** obtained crystallographically (yellow) and by modeling (green). (b) As Figure 3a, with the protein shown as a surface. (c) Superimposition of the docked conformations of compounds of Series **2**–**6** (except **4h**). (d) Interactions of **4a** with key residues in the active site of the *P. falciparum* dUTPase.

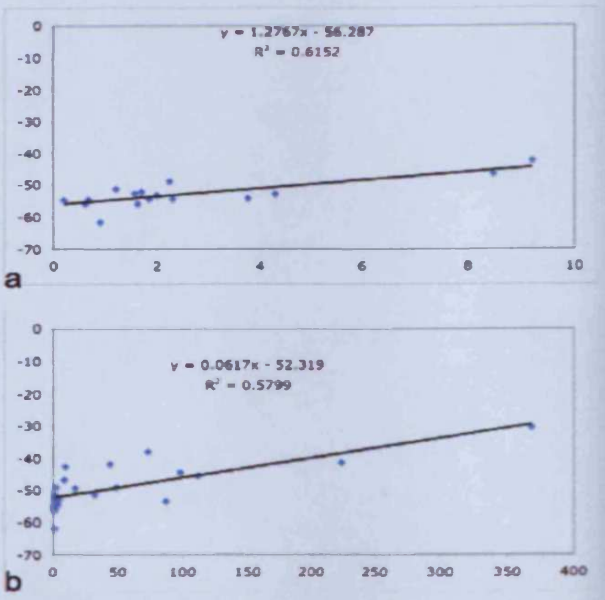


Figure 4. (a) Correlation between K_i (*x*-axis) and FlexX score (*y*-axis) for compounds with a K_i < 10 μM against *P. falciparum* dUTPase. (b) Correlation between K_i (*x*-axis) and FlexX score (*y*-axis) for compounds with a K_i < 400 μM.

article. The computed energy scores obtained from FlexX suggest a favorable interaction between *P. falciparum* dUTPase and most of the ligands with low constant of inhibition; this is evident from the high negative binding score values. Compounds with very poor experimental binding affinity (K_i > 500 μM) also showed poor scores with FlexX, which is consistent with the model possibly showing predictive potency. There appears to be a correlation between observed K_i and docking score, which gave a linear correlation with an R^2 of 0.61 (Figure 4) {this correlation excludes the branched chain compounds **9** and **10**}.

Discussion

Correlation Enzyme/in Vitro Assays. In general, a good correlation between enzyme and cellular assays results was observed. Potent inhibitors of the enzyme contained a triphenyl substituent (or in one case a diphenyl substituent), and all of these gave good inhibition of parasite growth. Compounds lacking the triphenyl moiety failed to give inhibition of the enzyme and also gave much weaker inhibition of parasite growth. There was not an obvious correlation between chain length and enzyme inhibition or growth inhibition. The docking calculations suggest that the chains can fold to accommodate the molecules so that the uracil and triphenyl groups bind in their respective pockets. Slight differences were observed regarding the extent of inhibition in the enzyme and cellular assays. This may be due to differences in cellular permeability of compounds. Additionally, some compounds may be metabolized; for example an acetoxy group may be hydrolyzed in cellular conditions to produce a more active analogue. Thus, whereas the hydroxyl derivative **9a** was 5–10 times more active against *P. falciparum* dUTPase than analogues **9b** (acetoxy derivative) and **9c** (carbamate derivative), it showed similar or even possibly slightly less activity in the cellular assay.

However, in four instances, assay results at the enzyme and parasite levels did not correlate. The acyclic derivative **3g** (with a benzoyloxypropyl side-chain) inhibited the enzyme with a K_i of 9.2 μM, but showed no *P. falciparum* growth inhibition (IC_{50} > 18 μM). This may be explained by hydrolysis of the benzoate occurring in cell culture, with **3g** being converted into the inactive hydroxyl analogue **3e**. On the other hand, the trityl ether **3a**, the benzoate acyclic **6g** and the γ-branched derivative **10c** inhibited *P. falciparum* growth but not the target dUTPase. Here, the lack of correlation could be due to the compounds acting on a molecular target other than *P. falciparum* dUTPase.

In the case of the unsaturated analogues **7a** and **8a**, although there was a good correlation between inhibition of the target dUTPase and inhibition of *P. falciparum* growth, the striking difference of affinity for the human enzyme was not reflected in the cytotoxicity cellular assay. Indeed, although compound **7a** inhibited the human dUTPase (K_i = 5.4 μM) whereas its geometrical isomer **8a** did not (K_i > 1 mM), both analogues showed similar cytotoxicity on mammalian cells (IC_{50} around 35 μM), although only one mammalian cell type has been investigated.

Preliminary ADME Studies. Preliminary ADME tests showed that compounds **1b**, **2a**, and **1d** have reasonable drug-like properties,[20] with the acyclic analogue **2a**, however, lacking metabolic stability as evidenced by the high clearance rate by human liver microsomes (Table 5). Tests on additional acyclic analogues focused on stability. These included (i) incubation with human liver microsomes (HLM), which assesses the overall response to all liver metabolising enzymes and therefore provides information regarding metabolic stability and also gives early guidance relating to drug half-life; and (ii) stability in simulated gastric, which is particularly relevant, as oral administration is a key requirement for the antimalarial drugs sought. Key results are shown in Table 5.

HLM intrinsic clearance for **2a** and **4a** indicated that having an oxygen in the side-chain facilitated degradation by liver enzymes. The Si/C replacement (Ph_3Si in place of Ph_3C) did not seem to have any effect on metabolic stability (compare HLM intrinsic clearance for **4a** and **4c**). On the other hand,

Table 5. Preliminary ADME tests carried out

compound	HLM intrinsic clearance[a] μL/min/mg	HLM intrinsic clearance relative to control	incubation in simulated gastric fluid amount remaining/incubation time	log D^b
2a	96 (55)	1.7	Na	4.93
4a	68 (29)	2.3	2%/5 min	5.48
4c	69 (29)	2.4	<2%/0.2 min	6.28
6c	35 (29)	1.2	<2%/0.2 min	7.04
4b	64 (52)	1.2	94%/150 min	4.42

[a] Value for control (MV026048) is given in parentheses. [b] Estimated using ACDLabs 7.0; na: not available.

introducing a tritylamino group in place of a trityloxy group (NH/O replacement) appeared to enhance stability (compare **4a** and **4b**). Finally, results suggested that side-chain length may be important for metabolic stability. Indeed, compound **6c** (side-chain of six carbons) was found to be much more stable toward liver enzymes than its analogue with a side-chain of four carbons (**4c**).

Looking at stability in simulated gastric fluid, results obtained for compounds **4a**, **4b**, and **4c** showed that the tritylamino group provided stability whereas the trityl ether and triphenylsilyl ether were predicted to undergo significant degradation in the stomach, especially so in the case of the silyl ether **4c**, almost instantly hydrolyzed.

That inhibition of *P. falciparum* dUTPase was found to be closely associated with the presence of trityl or triphenylsilyl groups raises the issue of high lipophilicity. According to Lipinski investigations, good bioavailability or permeability is more likely for drugs with a logP below 5.[21] LogP values can only be predicted for uncharged substances, and for compounds that can exist in different ionic forms (particularly relevant for tritylamines at physiological pH), logD values represent better predictions. Clearly, predicted logD values were high due to the three phenyl groups (Table 5). However, the presence of the ionizable amine moiety lowered lipophilicity (by 1 logD unit; compare **4b** and **4a**), while the introduction of a silicon atom had the opposite effect (compare **4a** and **4c**).

Overall, these preliminary results suggest that for further lead development (especially in view of an orally administered drug), a tritylamino unit is a better option than trityloxy and triphenylsilyloxy groups from a lipophilicity (logD) point of view, as well as for stability, both in simulated gastric fluid simulation and toward liver enzymes.

Conclusion

We have described a cost-effective synthesis of a selection of acyclic analogues of deoxyuridine. Potent and selective nontoxic inhibitors were identified. The compounds prepared showed equal or better antiplasmodial properties (inhibition of *P. falciparum* dUTPase and parasite growth, nontoxicity) compared to cyclic analogues previously reported,[3] especially with regards to selectivity. Structure−activity relationships were found to be in good agreement with previous observations, emphasizing the importance of a trityl or triphenylsilyl group for activity. Furthermore, potential issues related to stability and lipophilicity associated with such groups were considered for some of the most promising inhibitors. Finally, we have tested a simple docking model which reproduced well experimental data and may have some potential as a qualitative screening tool. Globally, the once neglected *P. falciparum* dUTPase is now much better understood and stands as a very promising target for the development of novel antimalarial drugs.

Experimental Section

Chemistry. Chemicals were purchased from Aldrich, Sigma, Lancaster, Acros, Avocado, or Fluka and were used without further purification. Compounds **39**, **42**, and **47** were provided by Medivir AB. Dry solvents were generally purchased from Fluka in sure-seal bottles and stored over molecular sieves. Reactions were performed in predried apparatus under an atmosphere of nitrogen unless otherwise stated. Reactions using microwave irradiation were carried in a Biotage Initiator microwave. Thin-layer chromatography (TLC) was performed on Merck silica gel $60F_{254}$ plates. Column chromatography was carried out using either Fisons matrix silica 60 (35−70 micron) or prepacked ISOLUTE SI columns purchased from Argonaut. ^1H, ^{13}C, and ^{19}F NMR were recorded on a Bruker Avance DPX 300 spectrometer or on a Bruker Avance DPX 500 with tetramethylsilane as the internal standard. Deuterated solvents were purchased from Goss unless stated otherwise. Infrared (IR) spectra were recorded on a Perkin-Elmer 1600 FT-IR spectrometer. Melting points (mp) were measured on a Gallenkamp melting point apparatus and were uncorrected. Low-resolution mass spectra were recorded on a Fisons VG Platform II spectrometer or on a Brucker MicroTof spectrometer. High-resolution mass spectra were determined by the EPSRC Mass Spectrometry Centre, Swansea, UK. Microanalyses were carried out by MEDAC Ltd.

General Procedure A. Coupling of Uracil with Protected Side-Chain Activated via a Bromo, Chloro. or Tosylate. Uracil (1−1.5 equiv) and cesium carbonate (1 equiv) in dry DMF were stirred at room temperature under nitrogen for 30 min. The suitably protected and activated side-chain fragment (1 equiv) in dry DMF was added dropwise via a syringe. The reaction mixture was stirred at 30−60 °C for 4−64 h. Water was added and the solution was extracted with EtOAc. The organic layers were washed with water, dried over Na_2SO_4, and concentrated in vacuo. The desired compound was obtained after purification by flash column chromatography.

General Procedure B. Ester Hydrolysis with Sodium Methoxide. The benzoate or acetate ester (1 equiv) was dissolved in a 0.2 M solution of sodium methoxide (10 equiv) in CH_3OH. The reaction was stirred at room temperature until the disappearance of the starting esters was observed by TLC. The solution was neutralized with Dowex 50WX8−200 ion-exchange resin (H^+ form). The resin was filtered and washed with CH_3OH. The filtrate was concentrated in vacuo, and the residue was coevaporated with ethanol several times to give the desired alcohol. Alternatively (when specified) the residue was chromatographed.

General Procedure C. Tritylation. The alcohol or amine precursor (1 equiv), trityl chloride (1.2 equiv), and, when stated, DMAP (catalytic amount) were stirred in dry pyridine at 50 °C for up to 64 h. The reaction mixture was partitioned between cold water and CH_2Cl_2. The organic layers were washed with brine, dried over $MgSO_4$, and concentrated in vacuo. The crude residue was purified by flash column chromatography to give the desired product.

General Procedure D. Silylation of Alcohols with Triphenylsilyl Chloride. A solution of triphenylsilyl chloride (TPSCl, 1.4 equiv) in dry pyridine was added dropwise to a solution of the alcohol precursor (1 equiv) in dry pyridine previously cooled in an ice-salt bath. The reaction mixture was stirred at 0 °C under nitrogen for 2−4 h. If necessary (presence of unreacted alcohol detected by TLC), more TPSCl in dry pyridine was added. Once the reaction had reached completion, it was quenched with CH_3OH. The solvent was then removed in vacuo, and the crude residue was purified by flash column chromatography (ISOLUTE SI column eluted with 0 → 5% CH_3OH in $CHCl_3$ unless otherwise stated) to give the desired triphenylsilyl ether.

1-[2-(Trityloxy)ethoxymethyl]uracil (2a). Compound **2a** was prepared following general procedure C, from the alcohol **11** (0.50 g, 2.69 mmol) and trityl chloride (0.82 g, 2.96 mmol) in dry pyridine (20 mL). White solid (0.23 g, 20%). R_f = 0.71 (10% CH_3OH/$CHCl_3$); mp 145−148 °C; ^1H NMR (300 MHz, $CDCl_3$) δ 3.32 (2H, m, 2′-H), 3.78 (2H, m, 3′-H), 5.29 (2H, s, 1′-H), 5.82 (1H, d, J = 7.9 Hz, 5-H), 7.29−7.51 (16H, m, Ph-H and 6-H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 63.3 (3′-CH_2), 69.7 (2′-CH_2), 77.4 (1′-CH_2), 87.2 (Ph_3C), 103.6 (5-CH), 127.5 (Ph-CH), 127.7 (Ph-CH), 128.3 (Ph-CH), 129.1 (Ph-CH), 143.5 (6-CH), 144.2 (Ph-C), 151.4 (2-C), 163.8 (4-C); IR (KBr) 3021, 1702, 1673, 760, 1452, 1087, 703 cm^{-1}; LRMS (ES^+) m/z 451 ([M + Na]$^+$, 100%); 473 (6%), 452 (13%), 243 (7%); HRMS (ES^+) calcd for $C_{26}H_{24}N_2NaO_4$ (M + Na)$^+$ 451.1634, found 451.1626.

1-(3-Trityloxypropyl)uracil (3a). Compound **3a** was prepared following general procedure A, from 3-bromo-1-*O*-tritylpropanol **12** (0.40 g, 1.05 mmol), Cs_2CO_3 (0.38 g, 1.05 mmol), and uracil (0.13 g, 1.15 mmol) in DMF (10 mL). White solid (134 mg, 31%). Flash column chromatography was performed eluting the column (ISOLUTE SI) with 0 → 30% EtOAc/hexane. Mp 161−164 °C; ^1H NMR (300 MHz, $CDCl_3$) δ 2.14−2.02 (2H, m, 2′-H), 3.22 (2H, t, J = 5.67 Hz, 3′-H), 4.02 (2H, t, J = 6.58 Hz, 1′-H), 5.47 (1H, dd, J = 2.4, 7.9 Hz, 5-H), 6.95 (1H, d, J = 7.9 Hz, 6-H), 7.50−7.22 (15H, m, Ph-H), 8.10 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 26.5 (2′-CH_2), 48.0 (1′-CH_2), 67.0 (3′-CH_2), 97.2 (Ph_3C C-10), 101.7 (5-CH), 128.9 (Ph-CH), 128.4 (Ph-CH), 127.7 (Ph-CH), 137.6 (Ph-C), 148.7 (6-CH), 150.7 (2-C), 160.0 (4-C); Anal. ($C_{26}H_{24}N_2O_3$) C, H, N.

1-(3-Tritylaminopropyl)uracil (3b). Compound **3b** was prepared following general procedure A from uracil (0.23 g, 2.05 mmol), and Cs_2CO_3 (0.46 g, 1.42 mmol) in DMF (8 mL), reacted with 3-bromo-1-tritylaminopropane **22** (0.51 g, 1.33 mmol) in DMF (3 mL). Chromatography was carried out using 20 → 50% EtOAc in hexane; a second column (0 → 10% CH_3OH in $CHCl_3$) was necessary to remove traces of DMF. Compound **3b** was isolated as a white solid (0.116 g, 21%). Mp 228−230 °C (dec); ^1H NMR (300 MHz, $CDCl_3$ + CD_3OD) δ 1.87 (2H, m, 2′-H), 2.20 (2H, t, J = 6.5 Hz, 3′-H), 3.85 (2H, t, J = 7.2 Hz, 1′-H), 5.61 (1H, d, J = 7.9 Hz, 5-H), 7.09 (1H, d, J = 7.9 Hz, 6-H), 7.15−7.35 (9H, m, Ph-H), 7.47 (6H, m, Ph-H); ^{13}C NMR (75 MHz, $CDCl_3$ + CD_3OD) δ 30.3 (2′-CH_2), 40.4 (3′-CH_2), 47.2 (1′-CH_2), 71.2 (Ph_3C), 102.4 (5-CH), 126.8 (Ph-CH), 128.3 (Ph-CH), 129.0 (Ph-CH), 144.9 (6-CH), 146.1 (Ph-C), 151.4 (2-C), 164.8 (4-C); LRMS (ES^+) m/z 845 ([2M + Na]$^+$, 3%), 434 ([M + Na]$^+$, 22%), 412 ([M + H]$^+$, 10%), 243 (Ph_3C^+, 42%), 87 (100%); HRMS (ES^+) calcd for $C_{26}H_{26}N_3O_2$ (M + H)$^+$ 412.2020, found 412.2018; Anal. ($C_{26}H_{25}N_3O_2$, 0.19 HBr) C, H, N; Calcd Br, 3.56; found Br, 4.91.

1-(3-Triphenylsilyloxypropyl)uracil (3c). Compound **3c** was prepared following general procedure D, from a solution of 1-(3-hydroxypropyl)uracil **3e** (0.19 g, 1.13 mmol) in dry pyridine (4 mL) stirred at 0 °C with TPSCl (0.43 g, 1.46 mmol) in dry pyridine (3 mL) for 4 h 30 min, and for 15 min after further addition of TPSCl (0.204 g, 0.692 mmol) in dry pyridine (1 mL). White solid (0.392 g, 81%). R_f = 0.52 (10% CH_3OH/$CHCl_3$); mp 152−153 °C; ^1H NMR (300 MHz, $CDCl_3$) δ 2.01 (2H, m, 2′-H), 3.94 (4H, m, 1′-H and 3′-H), 5.54 (1H, d, J = 7.9 Hz, 5-H), 7.04 (1H, d, J = 7.9 Hz, 6-H), 7.51 (9H, m, Ph-H), 7.68 (6H, m, Ph-H), 9.49 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 31.5 (2′-CH_2), 46.4 (1′-CH_2), 60.3 (3′-CH_2), 102.1 (5-CH), 128.3 (Ph-CH), 130.8 (Ph-CH), 134.0 (Ph-C), 135.8 (Ph-CH), 145.6 (6-CH), 151.3 (2-C), 164.3 (4-C); LRMS (ES^+) m/z 879 ([2M + Na]$^+$, 8%), 451 ([M + Na]$^+$, 38%), HRMS (ES^+) calcd for $C_{25}H_{28}N_3O_3Si$ (M + NH_4)$^+$ 446.1894, found 446.1887; Anal. ($C_{25}H_{24}N_2O_3Si$) C, H, N.

1-(3-*tert*-Butyldimethylsilyloxypropyl)uracil (3f). Compound **3f** was prepared following general procedure A, from 3-bromo-1-*tert*-butyldimethylsilylpropanol **13** (0.64 g, 2.68 mmol), uracil (0.30 g, 2.68 mmol), and Cs_2CO_3 (0.87 g, 2.68 mmol) in DMF (10 mL). Colorless oil (250 mg, 33%). Flash column chromatography was performed eluting the column (ISOLUTE SI) with 0 → 30% EtOAc/hexane. Mp 120−122 °C; ^1H NMR (300 MHz, $CDCl_3$) δ 0.00 (6H, s, [CH_3]$_2$Si), 0.95 (9H, s, C[CH_3]$_3$), 1.94 (2H, quintet, J = 6.6 Hz, 2′-H), 3.70 (2H, t, J = 6.6 Hz, 1′-H), 3.91 (2H, t, J = 6.6 Hz, 3′-H), 5.72 (1H, d, J = 7.9 Hz, 5-H), 7.30 (1H, d, J = 7.9 Hz, 6-H), 9.15 (1H, bs, 3-N); ^{13}C NMR (75 MHz, $CDCl_3$) δ −5.0 ([CH_3]$_2$Si), 18.6 (C[CH_3]$_3$), 26.3 (C[CH_3]$_3$), 31.6 (2′-CH_2), 46.6 (1′-CH_2), 59.4 (3′-CH_2), 102.0 (5-CH), 145.8 (6-CH), 151.2 (2-C), 164.3 (4-C); LRMS (ES^+) m/z 307 ([M + Na]$^+$, 100%).

1-(5-Trityloxypentyl)uracil (5a). The N^3-benzoyluracil **31** (231 mg; 0.42 mmol) was stirred for 48 h at room temperature with a mixture of 1 M NaOH and dioxane (v/v 1:1, 8 mL). After 48 h, brine (8 mL) was added to the solution which was then extracted with EtOAc (3 × 8 mL). The organic phase was dried over $MgSO_4$ and concentrated in vacuo. The crude residue was purified by flash column chromatography eluting the silica gel column with 10% CH_3OH in CH_2Cl_2. The colorless oil isolated was treated with a mixture of water/ethanol (v/v, 1:1) to give compound **5a** as a white solid (112 mg; 60%). R_f = 0.64 (10% CH_3OH/CH_2Cl_2); ^1H NMR (300 MHz, $CDCl_3$) δ 1.50−1.60 (2H, m, 3′-H), 1.74−1.84 (4H, m, 2′-H and 4′-H), 3.22 (2H, t, J = 6.2 Hz, 5′-H), 3.83 (2H, t, J = 7.3 Hz, 1′-H), 7.34−7.46 (9H, m, Ph-H), 5.80 (1H, d, J = 7.9 Hz, 5-H), 7.21 (1H, d, J = 7.9 Hz, H-6), 7.56 (6H, m, Ph-H), 9.13 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 23.6 (3′-CH_2), 29.2 (2′-CH_2), 29.9 (4′-CH_2), 49.2 (1′-CH_2), 63.5 (5′-CH_2), 86.9 (Ph_3C), 102.5 (5-CH), 127.4 (Ph-CH), 128.4 (Ph-CH), 129.1 (Ph-CH), 144.8 (Ph-C), 144.9 (6-CH), 151.4 (2-C), 164.6 (4-C); LRMS (ES^+) m/z 463 ([M + Na]$^+$, 100%); HRMS (ES^+) calcd for $C_{28}H_{32}N_3O_3$ (M + NH_4)$^+$ 458.2438, found 458.2439.

1-(5-Tritylaminopentyl)uracil (5b). Compound **5b** was prepared following general procedure B, from the N^3-benzoyluracil derivative **32** (0.19 g; 0.35 mmol) treated with 0.2 M CH_3ONa in CH_3OH (17 mL, 3.4 mmol). After removal of the Dowex resin and concentration of the filtrate in vacuo, the residue was purified by flash column chromatography, eluting the silica gel column with 10% CH_3OH in CH_2Cl_2. White foam (240 mg; 52%). R_f = 0.50 (10% CH_3OH/CH_2Cl_2); ^1H NMR (300 MHz, $CDCl_3$) δ 1.36−1.49 (2H, m, 3′-H), 1.63−1.75 (4H, m, 2′-H and 4′-H), 2.18 (2H, t, J = 7.2 Hz, 5′-H), 3.74 (2H, t, J = 7.3 Hz, 1′-H), 5.73 (1H, d, J = 7.9 Hz, 5-H), 7.14 (1H, d, J = 7.9 Hz, 6-H), 7.22−7.36 (9H, m, Ph-H), 7.52 (6H, d, J = 9.9 Hz, Ph-H), 8.70 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 24.6 (3′-CH_2), 29.5 (4′-CH_2), 30.8 (2′-CH_2), 43.7 (5′-CH_2), 49.2 (1′-CH_2), 71.3 (Ph_3C), 102.5 (5-CH), 126.7 (Ph-CH), 128.2 (Ph-CH), 129.0 (Ph-CH), 144.8 (6-CH), 146.6 (Ph-C), 151.0 (2-C),163.8 (4-C); LRMS (ES^+) m/z 440 [M$^+$, 50%], 243 (Ph_3C^+, 100%); HRMS (ES^+) calcd for $C_{28}H_{30}N_3O_2$ (M + H)$^+$ 440.2341, found 440.2333.

1-(5-Triphenylsilyloxypentyl)uracil (5c). Compound **5c** was prepared following general procedure D, from a solution of 1-(5-hydroxypentyl)uracil **5e** (0.05 g, 0.25 mmol) in pyridine (1 mL) treated with TPSCl (0.08 g, 0.27 mmol) in pyridine (1 mL) at room temperature for 24 h. Flash column chromatography was performed on a silica gel column eluted with 1% CH_3OH in CH_2Cl_2. Light orange syrup (30 mg, 26%). R_f = 0.69 (5% CH_3OH/CH_2Cl_2); ^1H NMR (300 MHz, $CDCl_3$) δ 1.43−1.51 (2H, m, 3′-H), 1.66−1.76 (4H, m, 2′-H and 4′-H), 3.72 (2H, t, J = 7.3 Hz, 1′-H), 3.87 (2H, t, J = 6.1 Hz, 5′-H), 5.69 (1H, d, J = 7.9 Hz, 5-H), 7.08 (1H, d, J = 7.9 Hz, 6-H), 7.43−7.54 (9H, m, Ph-H), 7.67 (6H, m, Ph-H), 9.26 (1H, s, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 23.2 (3′-CH_2), 29.0 (2′-CH_2), 32.2 (4′-CH_2) 49.2 (1′-CH_2), 63.8 (5′-CH_2), 102.4 (5-CH), 128.7 (Ph-C), 130.5 (Ph-CH), 134.4 (Ph-CH) 136.4 (Ph-CH), 144.9 (6-CH), 151.2 (2-C), 164.2 (4-C); LRMS (ES^+): m/z 479 ([M + Na]$^+$, 100%); HRMS (ES^+) calcd for $C_{27}H_{28}N_2NaSiO_3$ (M + Na)$^+$ 479.1761, found 479.1780; Anal. ($C_{27}H_{28}N_2SiO_3$, 0.55 H_2O) C, H; Calcd N, 6.00; found N, 5.45.

1-[(*Z*)-4-Trityloxy-2-butenyl]uracil (7a). Compound **7a** was prepared following general procedure B, from the N^3-benzoyluracil **37** (0.05 g, 0.09 mmol) treated with 0.2 M CH_3ONa in CH_3OH (5 mL, 1.0 mmol). After removal of the Dowex resin and concentration of the filtrate in vacuo, the residue was purified by flash column chromatography, eluting the silica gel column with 10% CH_3OH in CH_2Cl_2. White solid (38 mg, 90%). R_f = 0.66 (10% CH_3OH/CH_2Cl_2); mp 65−68 °C; ^1H NMR (300 MHz, $CDCl_3$) δ 3.79 (2H,

d, J = 6.2 Hz, 4′-H), 4.28 (2H, d, J = 7.4 Hz, 1′-H), 5.69−5.98 (2H, m, 5-H and 3′-H), 6.01−6.09 (1H, m, 2′-H), 7.16 (1H, d, J = 8.0 Hz, 6-H), 7.29−7.41 (9H, m, Ph-H), 7.49−7.52 (6H,m, Ph-H), 8.21 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 44.9 (1′-CH_2), 59.9 (4′-CH_2), 87.8 (Ph_3C), 102.1 (5-CH), 126.1 (3′-CH), 127.7 (Ph-CH), 128.4 (Ph-CH), 129.0 (Ph-CH), 132.5 (2′-CH), 144.1 (Ph-C), 144.2 (6-CH), 151.2 (2-C), 164.0 (4-C); LRMS (ES^+) m/z 447 ([M + Na]$^+$, 100%); HRMS (ES^+) calcd for $C_{27}H_{24}N_2O_3Na$ (M + Na)$^+$ 447.1679, found 447.1680; Anal. ($C_{27}H_{24}N_2O_3$, 0.4 H_2O) C, H, N.

1-[(*E*)-4-trityloxy-2-butenyl]uracil (8a). Compound **8a** was prepared following general procedure B, from the N^3-benzoyluracil **38** (0.05 mg, 0.09 mmol) treated with 0.2 M CH_3ONa in CH_3OH (5 mL, 1.0 mmol). After removal of the Dowex resin and concentration of the filtrate in vacuo, the residue was purified by flash column chromatography, eluting the silica gel column with 0→5% CH_3OH in $CHCl_3$. White solid (38 mg, 90%). R_f = 0.51 (10% $CH_3OH/CHCl_3$); mp 70−72 °C; ^{1}H NMR (300 MHz, $CDCl_3$) δ 3.74 (2H, d, J = 3.9 Hz, 1′-H), 4.42 (2H, d, J = 5.6 Hz, 4′-H), 5.77 (1H, d, J = 7.9 Hz, 5-H), 5.83−5.98- (2H, m, 2′-H and 3′-H), 7.19 (1H, d, J = 7.9 Hz, 6-H), 7.27−7.39 (9H, m, Ph-H), 7.49 (6H, d, J = 7.8 Hz, Ph-H), 9.04 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 49.6 (1′-CH_2), 64.0 (4′-CH_2),87.5 (Ph_3C), 102.9 (5-CH), 124.1 (3′-CH), 124.1 (Ph-CH), 128.3 (Ph-CH), 129.0 (Ph-CH), 133.3 (2′-CH), 144.1 (6-CH), 144.3 (Ph-C), 151.1 (2-C), 163.9 (4-C); LRMS (ES^+) m/z 447 ([M + Na]$^+$, 100%); HRMS (ES^+) calcd for $C_{27}H_{28}N_3O_3$ (M + NH_4)$^+$ 442.2125, found 442.2128; Anal. ($C_{27}H_{24}N_2O_3$, 0.2 H_2O) C, H, N.

1-[4-Acetyloxy-3-(tritylaminomethyl)butyl]uracil (9a). Trityl chloride (0.13 g, 0.46 mmol) was added to a solution of compound **41** (0.10 g, 0.43 mmol) and triethylamine (0.64 mL, 0.46 mmol) in dry CH_2Cl_2 (5 mL). The reaction mixture was stirred at room temperature for 6 h. Further trityl chloride (0.06 g, 0.23 mmol) and triethylamine (0.32 mL, 0.23 mmol.) were added, and the reaction was left stirring overnight. Water (5 mL) was added. The organic layer was extracted and washed with 0.1M HCl (5 mL), dried over $MgSO_4$, concentrated in vacuo, and purified by flash column chromatography eluting with 0 → 5% CH_3OH in CH_2Cl_2. The title compound was obtained as a white foam (150 mg, 70%). R_f = 0.60 (10% $CH_3OH/CHCl_3$); mp 68−72 °C; ^{1}H NMR (300 MHz, $CDCl_3$) δ 1.65−1.83 (3H, m, 2′-H and 3′-H), 2.03 (3H, s, CH_3CO), 2.17 (2H, d, J = 6.0 Hz, CH_2NHTr), 3.68 (2H, m, 1′-H), 4.27 (2H, d, J = 5.0 Hz, 4′-H), 5.66 (1H, d, J = 7.9 Hz, 5-H), 7.02 (1H, d, J = 7.9 Hz, 6-H), 7.19−7.30 (9H, m, Ph-H), 7.46 (6H, d, J = 7.3 Hz, Ph-H), 9.07 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 21.3 (CH_3CO), 29.7 (2′-CH_2), 37.0 (3′-CH), 44.6 (CH_2-NHTr), 47.4 (1′-CH_2), 65.3 ($CH_3CO_2CH_2$ or 4′-CH_2), 71.3 (Ph_3C), 102.7 (5-CH), 126.8 (Ph-CH), 128.3 (Ph-CH), 129.0 (Ph-CH), 144.7 (6-CH), 146.2 (Ph-C), 151.1 (2-C), 164.0 (4-C), 171.5 (CH_3CO); LRMS (CI) m/z 498.3 ([M + H]$^+$, 100%); HRMS (ES^+) calcd for $C_{30}H_{32}N_3O_4$ (M + H)$^+$ 498.2387, found 498.2384; Anal. ($C_{30}H_{31}$-N_3O_4,1.5 H_2O) C, H, N.

1-[4-Hydroxy-3-(tritylaminomethyl)butyl]uracil (9b). The intermediate **9a** (0.11 g, 0.22 mmol) was deacetylated according to general procedure B using 0.2 M CH_3ONa in CH_3OH (5 mL, 1.0 mmol). Compound **9b** was obtained as a white foam (100 mg, 99%). R_f = 0.66 (10% $CH_3OH/CHCl_3$); mp 75−80 °C; ^{1}H NMR (300 MHz, $CDCl_3$) δ 1.47−1.68 (2H, m, 2′-H), 1.84 (1H, bs, 3′-H), 2.40−2.48 (2H, m, CH_2NHTr), 3.57−3.82 (3H, m, 1′-H and 4′-C*H*H), 3.90−3.95 (1H, m, 4′-CH*H*), 5.26 (2H, bs, CH_2OH and Ph$_3$CN*H*), 5.65 (1H, d, J = 7.9 Hz, 5-H), 7.03 (1H, d, J = 7.9 Hz, 6-H), 7.35−7.23 (9H, m, Ph-H), 7.41−7.48 (6H, m, Ph-H), 8.16 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 29.4 (2′-CH_2), 38.3 (3′-CH), 47.5 (1′-CH_2), 48.1 (CH_2NHTr), 67.0 (4′-CH_2), 97.8 (Ph_3C), 102.7 (5-CH), 127.2 (Ph-CH), 128.5 (Ph-CH), 128.9 (Ph-CH), 144.7 (6-CH), 145.1 (Ph-C), 150.9 (2-C), 163.7 (4-C); LRMS (ES^+) m/z 456 ([M + H]$^+$, 100%); HRMS (ES^+) calcd for $C_{28}H_{30}N_3O_3$ (M + H)$^+$ 456.2282, found 456.2283; Anal. ($C_{28}H_{29}$-N_3O_3, 0.5 H_2O, 1.0 HCl) C, H, N.

1-[4-(*tert*-Butoxycarbonylamino)-3-(trityloxymethyl)butyl]-uracil (9c). 1-[4-(*tert*-Butoxycarbonylamino)-3-(hydroxymethyl)-butyl]uracil **39** (0.05 g, 0.15 mmol), trityl chloride (0.05 mg, 0.19 mmol), and dry pyridine (2 mL) were irradiated with microwaves to reach the temperature of 160 °C for 5 min. The irradiating cycle was repeated three times. The solvent was removed in vacuo, and the crude product was purified by flash column chromatography using 40% hexane in EtOAc. Compound **9c** was obtained as a white foam (70 mg, 79%). R_f = 0.27 (40% hexane/EtOAc); mp 86−88 °C; ^{1}H NMR (300 MHz, $CDCl_3$) δ 1.47 (9H, s, C[CH_3]$_3$), 1.53−1.62 (1H, m, 3′-H), 1.67−1.85 (2H, m, 2′-H), 3.05−3.10 (1H, m, 4′-C*H*H), 3.23−3.31 (3H, m, 1′-H and 4′-CH*H*), 3.76−3.84 (2H, m, 3′-H), 4.58 (1H, t, J = 6.1 Hz, N*H*Boc), 5.67 (1H, dd, J = 2.4, 7.9 Hz, 5-H), 7.27−7.40 (10H, m, 6-H and Ph-H), 7.48 (6H, d, J = 7.0 Hz, Ph-H), 8.08 (1H, bs, 3-NH); ^{13}C NMR (75 MHz, $CDCl_3$) δ 28.4 (C[CH_3]$_3$), 29.7 (2′-CH_2), 36.9 (3′-CH_2), 40.8 (4′-CH_2), 46.8 (1′-CH_2), 64.3 (CH_2OTr), 79.2 (C[CH_3]$_3$), 86.8 (Ph_3C), 101.9 (5-CH), 126.9 (Ph-CH), 127.8 (Ph-CH), 128.6 (Ph-CH), 143.7 (Ph-C), 145.0 (6-CH), 150.8 (2-C), 156.5 (NHCO_2), 163.8 (4-C); HRMS (ES^+) calcd for $C_{33}H_{38}N_3O_5$ (M + H)$^+$ 498.2387, found 498.2384; Anal. ($C_{33}H_{37}N_3O_5$, 0.30 H_2O) C, H; Calcd N, 7.49; found N, 6.98.

1-[2-(Hydroxymethyl)-4-(trityloxy)butyl]uracil (10a). The intermediate **46** (0.14 g, 0.28 mmol) was deacetylated according to general procedure B using 0.2 M CH_3ONa in CH_3OH (5 mL, 1.0 mmol). Compound **10a** was obtained as a white solid (73 mg, 57%). R_f = 0.51 (10% CH_3OH/CH_2Cl_2); mp 83−85 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ 1.64−1.77 (2H, m, 3′-H), 3.21−3.28 (2H, m, 4′-H), 2.03−2.10 (1H, m, 2′-H), 3.44 (1H, dd, J = 4.4, 12.0 Hz, C*H*HOH), 3.49 (1H, dd, J = 3.4 Hz, 12.0 Hz, CH*H*OH), 3.63 (1H, dd, J = 4.9, 14.0 Hz, 1′-C*H*H), 3.78 (1H, dd, J = 9.0, 14.0 Hz, 1′-CH*H*), 5.69 (1H, d, J = 7.9 Hz, 5-H), 7.12 (1H, d, J = 7.9 Hz, 6- H), 7.27 (3H, m, Ph-H), 7.33 (6H, d, J = 7.5 Hz, Ph-H), 7.45 (6H, d, J = 7.4 Hz, Ph-H), 9.67 (1H, bs, 3-NH); ^{13}C NMR (125 MHz, $CDCl_3$) δ 29.1 (3′-CH_2), 38.3 (2′-CH), 49.0 (1′-CH_2), 61.0 (CH_2OH), 61.1 (4′-CH_2), 87.0 (Ph_3C), 102.4 (5-CH), 127.3 (Ph-CH), 127.9 (Ph-CH), 128.7 (Ph-CH), 143.9 (Ph-C), 145.6 (6-CH), 151.9 (2-C), 163.8 (4-C); LRMS (ES^+) m/z 479 ([M + Na]$^+$, 100%); HRMS (ES^+) calcd for $C_{28}H_{28}N_2NaO_4$ (M + Na)$^+$ 479.1941, found 479.1946; Anal. ($C_{28}H_{28}N_2O_4$, 1.20 H_2O) C, H, N.

1-[4-Hydroxy-2-(trityloxymethyl)butyl]uracil (10b). The TBDMS protected intermediate **43** (0.13 g, 0.23 mmol) was dissolved in THF (3 mL). Tetrabutylammonium fluoride (0.07 g, 0.27 mmol) was added, and the reaction mixture was stirred at room temperature until complete disappearance of the starting material was observed by TLC. The solvent was removed in vacuo, and the crude product was purified by flash column chromatography eluting with 5 → 10% CH_3OH in $CHCl_3$. The title compound was obtained as a white solid (49 mg, 47%). R_f = 0.57 (10% $CH_3OH/CHCl_3$); mp 85−88 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ 1.69−1.77 (2H, m, 3′-H), 2.14 (1H, bs, 4′-OH), 2.21 (1H, m, 2′-H), 3.14−3.19 (2H, m, 4′-H), 3.71−3.76 (2H, m, 1′-C*H*H and CH*H*OTr), 3.79−3.84 (1H, m, C*H*HOTr), 3.99 (1H, dd, J = 4.1, 13.8 Hz, 1′-CH*H*), 5.32 (1H, d, J = 7.9 Hz, 5-H), 6.69 (1H, d, J = 7.9 Hz, 6-H), 7.27−7.30 (6H, m, Ph-H), 7.34 (3H, t, J = 7.5 Hz, Ph-H), 7.43 (6H, d, J = 7.6 Hz, Ph-H), 8.23 (1H, bs, 3-NH); ^{13}C NMR (125 MHz, $CDCl_3$) δ 32.3 (3′-CH_2), 36.1 (2′-CH), 49.4 (1′-CH_2), 60.2 (CH_2OTr), 62.5 (4′-CH_2), 86.7 (Ph_3C), 101.5 (5-CH), 127.3 (Ph-CH), 128.0 (Ph-CH), 128.6 (Ph-CH), 143.4 (Ph-C), 145.3 (6-CH), 151.1 (2-C), 163.7 (4-C); LRMS (ES^+) m/z 479 ([M + Na]$^+$, 100%); HRMS (ES^+) calcd for $C_{28}H_{28}N_2O_4Na$ (M + Na)$^+$ 479.1941, found 479.1935; Anal. ($C_{28}H_{28}N_2O_4$, 1.40 H_2O) C, H, N.

1-[2-(Azidomethyl)-4-(trityloxy)butyl]uracil (10c). Compound **10c** was prepared by tritylation of 1-[2-(azidomethyl)-4-(hydroxy)-butyl]uracil **48** (0.10 g, 0.42 mmol) with trityl chloride (0.16 g, 0.58 mmol) in dry pyridine (5 mL), according to general procedure C. White solid (163 mg, 82%). R_f = 0.57 (10% $CH_3OH/CHCl_3$); mp 75−80 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ 1.42−1.51 (2H, m, 3′-H), 2.03−2.11 (1H, m, 2′-H), 2.96−3.14 (4H, m, 1′-H and 4′-H), 3.42−3.46 (1H, m, C*H*HN_3), 3.50 (1H, dd, J =6.1, 13.9 Hz, CH*H*N_3), 5.49 (1H, dd, J = 2.0, 7.9 Hz, 5-H), 6.90 (1H, d, J = 7.9 Hz, 6-H), 7.06−7.09 (3H, m, Ph-H), 7.14 (6H, t, J = 7.6 Hz, Ph-H), 7.25 (6H, d, J = 7.3 Hz, Ph-H), 8.52 (1H, bs, 3-NH); ^{13}C NMR

(125 MHz, $CDCl_3$) δ 29.6 (3′-CH_2), 36.5 (2′-CH), 50.5 (1′-CH_2), 51.9 (CH_2N_3), 60.5 (4′-CH_2), 102.5 (5-CH), 127.1 (Ph-CH), 127.9 (Ph-CH), 128.6 (Ph-CH), 143.9 (Ph-C), 144.8 (6-CH), 150.8 (2-C), 163.3 (4-C); IR (KBr) 3061, 2102, 1593, 1449 cm^{-1}; LRMS (ES^+) *m/z* 504 ($[M + Na]^+$, 100%); LRMS (ES^-) *m/z* 482 (M^-, 100%); HRMS (ES^+) calcd for $C_{28}H_{24}N_5O_2Na$ $(M + Na)^+$ 504.2006, found 504.1997; Anal. ($C_{28}H_{24}N_5O_2$) C, H; Calcd N, 14.54; found N, 13.79.

Biological Assays. Enzyme Purification and Inhibition Assays. Both recombinant *P. falciparum* and human dUTPases were expressed in *E. coli* BL21 (DE3) cells which had been transformed with the pET11*Pfdut*[9] and pET3*Hudut* (kindly provided by P. O. Nyman, Lund University, Sweden) expression vectors, respectively. For dUTPase purification, the same procedure was used for both the human and the *Plasmodium* enzymes. Cell pellets from a 2-L IPTG-induced culture were resuspended in 40 mL of buffer A (20 mM sodium acetate (pH 5.5), 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT) and 20 μM PMSF. The cells were lysed by sonication, and the cell extract was cleared by centrifugation at 15 000 rpm for 45 min. The supernatant was loaded onto a 50 mL phosphocellulose (Whatman P-11) column at 4 °C and was eluted with a 50 mM to 2 M NaCl gradient in buffer A. The enzyme was then dialyzed against buffer A, prior to gel filtration chromatography on a Superdex 200 HA 10/30 column at 4 °C. Purified fractions contained dUTPase of approximately 96% purity.

Nucleotide hydrolysis was monitored by mixing enzyme and substrate with a rapid kinetic accessory (Hi-Tech Scientific) attached to a spectrophotometer (Cary 50) and connected to a computer for data acquisition and storage as described previously.[22] Protons, released through the hydrolysis of nucleotides, were neutralized by a pH indicator in weak buffered medium with similar pK_a and monitored spectrophotometrically at the absorbance peak of the basic form of the indicator. The ratio between the indicator and the buffer concentration was 50:2000 (μM), and the absorbance changes were kept within 0.1 units. The indicator/buffer pair used was red cresol/bicine (pH 7.5–8.5, 573 nm). Assay mixes contained 30 nM of *Pf*dUTPase, 50 μM dUTP, 5 mM $MgCl_2$, 2.5 mM DTT, 1.25 mg/mL BSA, and 100 mM KCl. V_{max} and K_{Mapp} were calculated by fitting the resulting data to the integrated Michaelis–Menten equation. The apparent K_M values were plotted against inhibitor concentration, and K_i values (Table 2) were obtained according to eq 1.

$$K_{Mapp} = \frac{K_M}{K_i}[I] + K_M \qquad (I)$$

In Vitro Assays. All in vitro assays were carried out twice independently in duplicate.

Plasmodium falciparum. In vitro activity against the erythrocytic stages of *P. falciparum* was determined by using a ^3H-hypoxanthine incorporation assay[23,24] using the chloroquine and pyrimethamine resistant K1 strain and the standard drugs chloroquine (Sigma C6628) and artemisinin (Arteannuin, Qinghaosu; Sigma 36,159-3). Compounds were dissolved in DMSO at 10 mg/mL and added to parasite cultures incubated in RPMI 1640 medium without hypoxanthine, supplemented with HEPES (5.94 g/L), $NaHCO_3$ (2.1 g/L), neomycin (100U/mL), Albumax (5 g/L), and washed human red cells A^+ at 2.5% haematocrit (0.3% parasitaemia). Serial doubling dilutions of each drug were prepared in 96-well microtiter plates and incubated in a humidified atmosphere at 37 °C; 4% CO_2, 3% O_2, 93% N_2.

After 48 h, 50 μL of ^3H-hypoxanthine (= 0.5 μCi) was added to each well of the plate. The plates were incubated for an additional 24 h under the same conditions. The plates were then harvested with a Betaplate cell harvester (Wallac, Zurich, Switzerland), and the red blood cells were transferred onto a glass fiber filter and then washed with distilled water. The dried filters were inserted into a plastic foil with 10 mL of scintillation fluid and were counted in a Betaplate liquid scintillation counter (Wallac, Zurich, Switzerland). IC_{50} values were calculated from sigmoidal inhibition curves using Microsoft Excel.

Cytotoxicity. Rat skeletal myoblasts (L-6 cells) were seeded in 96-well microtiter plates in RPMI 1640 medium with 10% FBS and 2 mM L-glutamine at a density of 4×10^4 cells/mL. After 24 h, the medium was removed and replaced by fresh medium containing a serial drug dilution, and the plate was incubated at 37 °C under a 5% CO_2 atmosphere for 72 h. Alamar Blue (10 μL) (Trinova, Giessen, Germany) was then added to each well and incubation continued for a further 2–4 h. The plates were read in a microplate fluorescence scanner (Spectramax Gemini XS by Molecular Devices) using an excitation wavelength of 536 nm and an emission wavelength of 588 nm. From the sigmoidal inhibition curve, IC_{50} values were calculated.

Molecular Modeling. Molecular Docking Studies. The ligands under study were built employing the SKETCH module of SYBYL6.9 and minimized using Tripos Force Field. Docking studies were carried out using the FlexX program[19] interfaced with SYBYL 6.9. FlexX is a fast-automated program based on incremental construction procedure. In this method the flexibility of the ligands is considered by including several conformations of ligands while maintaining a rigid structure for the biomolecule. All the molecules were docked into the active sites of the *P. falciparum* dUTPase. The 3D coordinates of the active sites were taken from the X-ray crystal structures of the *P. falciparum* dUTPase reported as complexes with the inhibitor **1d**, deposited in the Brookhaven Protein Databank (PDB codes: 1VYQ). Molecules of water were deleted. While creating the receptor description file (rdf), the active site was defined as the area within 6.5 Å around the cocrystallized ligand. Formal charges were assigned to all the molecules and FlexX run was submitted.

For most compounds the top rank-score was selected as a result of the docking. For a few compounds (**2g**, **3a**, **4a**,**g**,**i**, **5b**, **6b**, and **7a**) a lower rank-score was chosen, as this represented a better mimic the conformation of the crystal structure in these cases.

Preliminary ADME Studies. Metabolic Stability in Human Liver Microsomes. Materials. NADPH was purchased from Sigma Aldrich, Sweden AB. Pooled human liver microsomes were purchased from In Vitro Technologies, Inc..

Metabolic Incubation. Incubation mixture consisted of liver microsomes (0.5 mg microsomal protein/ml) substrate, (2 μM), and NADPH (1 mM) in a total volume of 0.3 mL of phosphate buffer (100 mM, pH 7.4). The reaction was initiated by the addition of NADPH, and the sample was incubated at 37 °C. At time points 0, 10, 20, and 30 min, aliquots (50 μL) were removed and added to 150 μL of acetonitrile in order terminate the reaction. Following mixing and centrifugation the supernatant was analyzed by LC-MS.

Calculation of in Intrinsic Clearance, Clint. The natural logarithm (residual concentration) versus time (min) is plotted. The slope of the line gives the elimination rate constant = k. The elimination half-life $T_{1/2} = -0.693/k$. Clint = $(0.693/T_{1/2}) \times$ mL incubation/mg microsomes.

Stability in Simulated Gastric Fluid. Materials. The Simulated Gastric Fluid consisted of 0.2% sodium chloride, 0.32% pepsin, and 0.7% hydrochloric acid. The pH of the final solution was 1.2 (USP 23).

Incubation. The compound was incubated in 0.5 mL of the gastric fluid at a final concentration of 10 μM. At time points; 0, 0.2, 5, 15, 30, 60, 150 min, aliquots (50 μl) were removed and mixed with 50 μL of phosphate buffer, (pH 7.4) and 0.2 mL of acetonitrile. The sample was centrifuged and the supernatant analyzed by LC–UV (254 nm).

Acknowledgment. We would like to thank the European Union (FP5 Cell Factory QLRT-2001-00305), the FIS Network RICET/C03, and the State Secretariat for Education and Research of Switzerland (R.B. and M.K.) for funding. The Royal Pharmaceutical Society of Great Britain is acknowledged for a studentship (G.K.) and the British Government for an ORS award (G.K.). The EPSRC National Mass Spectrometry Service

Centre in Swansea is acknowledged for carrying out accurate mass spectrometry.

Supporting Information Available: Experimental procedures and analytical data for all the compounds not included in the Experimental Section, a table of elemental analysis data obtained for selected compounds, a table with the docking scores of FlexX and a NOESY spectrum of compound **4b**. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) http://www.who.int/topics/malaria/en/.
(2) http://www.who.int/tdr/publications/tdrnews/news60/mmv.htm.
(3) Nguyen, C.; Kanisathan, G.; Leal-Cortijo, I.; Musso-Buendia, A.; Kaiser, M.; Brun, R.; Ruiz-Pérez, M.; Johansson, N.-G.; González-Pacanowska, D.; Gilbert, I. H. Deoxyuridine Triphosphate Nucleotidohydrolase as a Potential Antiparasitic Drug Target. *J. Med. Chem.* **2005**, *48*, 5942−5954.
(4) El-Hajj, H. H.; Zhang, H.; Weis, B. Lethality of a dut (deoxyuridine triphosphatase) Mutation in *Escherichia Coli*. *J. Bacteriol.* **1988**, *170*, 1069−1075.
(5) Gadsden, M. H.; McIntosh, E. M.; Game, J. C.; Wilson, P. J.; Haynes, R. H. dUTP Pyrophosphatase is an Essential Enzyme in *Saccharomyces cerevisiae*. *EMBO J.* **1993**, *12*, 4425−4431.
(6) Hidalgo-Zarco, F.; González-Pacanowska, D. Trypanosomatid dUTPases as Potential Targets for Drug Design. *Curr. Protein Pept. Sci.* **2001**, *2*, 389−397.
(7) Mol, C. D.; Harris, J. M.; McIntosh, E. M.; Tainer, J. A. Humand dUTP pyrophosphatase: uracil recognition by a β hairpin and active sites formed by three separate subunits. *Structure* **1996**, *4*, 1077−1092.
(8) Persson, R.; Cedergren-Zeppezauer, E. S.; Wilson, K. S. Homotrimeric dUTPases; Structural solutions for specific recognition and hydrolysis of dUTP. *Curr. Protein Pept. Sci.* **2001**, *2*, 287−300.
(9) Whittingham, J. L.; Leal, I.; Kasinathan, G.; Nguyen, C.; Bell, E.; Berry, C.; Benito, A.; Turkenburg, J.; Dodson, E. J.; Ruiz-Pérez, L. M.; Wilkinson, A. J.; Johansson, N. G.; Brun, R.; Gilbert, I. H.; González-Pacanowska, D.; Wilson, K. Novel Inhibitors of *Plasmodium falciparum* dUTPase Provide a Platform for Anti-Malarial Drug Design. *Structure* **2005**, *13*, 329−338.
(10) Zalud, P.; Wachs, W. O.; Nyman, P. O.; Zeppezauer, M. Inhibition of the Proliferation of Human Cancer Cells in vitro by Substrate Analogues Inhibitors of dUTPase. *Adv. Exp. Med. Biol.* **1995**, *370*, 135−138.
(11) Persson, T.; Larsson, G.; Nyman, P. O. Synthesis of 2′-Deoxyuridine 5′(α,β-imido)triphosphate, a Substrate Analogue and Potent Inhibitor of dUTPase. *Bioorg. Med. Biochem.* **1996**, *4*, 553−556.
(12) Hernández, A.-I.; Balzarini, J.; Karlsson, A.; Camarasa, M.-J.; Pérez-Pérez, M.-J. Acyclic Nucleoside Analogues as Novel Inhibitors of Human Mitochondrial Thymidine Kinase. *J. Med. Chem.* **2002**, *45*, 4254−4263.
(13) Frieden, M.; Giraud, M.; Reese, C. B.; Song, Q. Synthesis of 1-[cis-3-(Hydroxymethyl)cyclobutyl]-uracil, -Thymine and -Cytosine. *J. Chem. Soc., Perkin Trans. 1* **1998**, 2827−2832.
(14) Maki, T.; Iwasaki, F.; Matsumura, Y. A new convenient method for selective monobenzoylation of diols. *Tetrahedron Lett.* **1998**, *39*, 5601−5604.
(15) McDonald, W. S.; Verbicky, C. A.; Zercher, C. K. Two-Directional Synthesis of Polycyclopropanes. An Approach to the Quinquecyclopropane Fragment of U-106305. *J. Org. Chem.* **1997**, *12*, 11215−1222.
(16) Showell, G. A.; Mills, J. S. Chemistry Challenges in Lead Optimization: Silicon Isosteres in Drug Discovery. *Drug Discov. Today.* **2003**, *8*, 551−556.
(17) Bains, W.; Tacke, R. Silicon Chemistry as a Novel Source of Chemical Diversity in Drug Design. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 526−543.
(18) Rarey, M.; Krame, B.; Lengauer, Y.; Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(19) Sybyl 6.9, Tripos Inc., 1699 South Hanley Road, St Louis, MO 63144.
(20) Properties considered included cellular permeability (Caco-2 cells test), metabolic stability (incubation with human liver microsomes), and possible effect of CYP3A4 (substrate properties and inhibition of this important cytochrome P450 enzyme gives an early warning of metabolism, toxicity, and potential interactions with coadministered drugs). N. G. Johansson (Medivir AB), personal communication.
(21) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(22) Hidalgo-Zarco, F.; Camacho, A. G.; Bernier-Villamor, V.; Nord, J.; Ruiz-Perez, L. M.; Gonzalez-Pacanowska, D. Kinetic properties and inhibition of the dimeric dUTPase-dUDPase from Leishmania major. *Protein Sci.* **2001**, *10*, 1426−1433.
(23) Desjardins, R. E.; Canfield, C. J.; Haynes, D.; Chulay, J. D. Quantitative Assessment of Antimalarial Activity in Vitro by a Semiautomated Microdilution Technique. *Antimicrob. Agents Chemother.* **1979**, *16*, 710−718.
(24) Matile, H. Pink, J. R. L. *Plasmodium falciparum* Malaria Parasite Cultures and their Use in Immunology. In *Immunological Methods*; Academic Press: San Diego, 1990; pp 221−234.

JM060126S

Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 3809–3812

Design, synthesis and evaluation of novel uracil amino acid conjugates for the inhibition of *Trypanosoma cruzi* dUTPase

Orla K. Mc Carthy,[a,b] Alessandro Schipani,[a,b] Alex Musso Buendía,[c] Luis M. Ruiz-Perez,[c] Marcel Kaiser,[d] Reto Brun,[d] Dolores González Pacanowska[c] and Ian H. Gilbert[a,b,*]

[a]*Welsh School of Pharmacy, Cardiff University, King Edward VII Avenue, Cardiff CF10 3XF, UK*
[b]*School of Life Sciences, University of Dundee, MSI/WTB/CIR complex, Dow Street, Dundee DD1 5EH, UK*
[c]*Instituto de Parasitologia y Biomedicina "Lopez-Neyra", Consejo Superior de Investigaciones Científicas, Avda. del Conocimiento s/n Parque Tecnológico de Ciencias de la Salud, 18100-Armilla, Granada, Spain*
[d]*Swiss Tropical Institute, Socinstrassee 57, CH-4002 Basel, Switzerland*

Received 21 March 2006; revised 11 April 2006; accepted 11 April 2006
Available online 3 May 2006

Abstract—Potential inhibitors of the *Trypanosoma cruzi* dUTP nucleotidohydrolase were docked into the enzyme using the program FlexX. Compounds that docked selectively were then selected and synthesized using solid phase methodology, giving rise to a novel library of amino acid uracil acetamide compounds which were evaluated for enzyme inhibition and anti-parasitic activity.

The protozoan parasite *Trypanosoma cruzi* is the causative agent of Chagas' disease, also called American trypanosomiasis. The parasite is transmitted by blood sucking triatomine insects and is one of three species of the genus Trypanosoma that are pathogenic to humans. Chagas' disease occurs mainly in South and Central America. It is estimated that 16–18 million people are infected with the disease, of which approximately 50,000 will die each year.[1] Although the number of incidences of the disease has declined over the past 20 years due to vector control initiatives,[2] there is still no satisfactory cure for American trypanosomiasis and current drug treatment is only effective against the early stages of the disease. The need for new drugs to treat this disease is therefore urgent.

Recently, deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase) has been identified as a novel and valid target against trypanosomatidae.[3] The enzyme is essential in both eukaryotes and prokaryotes, where investigated.[4,5] dUTPase catalyses the hydrolysis of dUTP to dUMP in the presence of Mg^{2+} and plays a critical role in maintaining the level of dUTP in the cell 10^{-5} times lower than that of dTTP. In doing so, incorporation of uracil into DNA is minimized, excessive incorporation of which would normally lead to DNA fragmentation and cell death. dUTPases are known to exist in several oligomeric forms. Monomeric enzymes are encoded by herpes virus and Epstein–Barr virus, while the homotrimeric forms are present in mammals, various bacteria and viruses and also in the *Plasmodium falciparum* parasite, the causative agent for malaria. The homodimeric enzymes have been found in *Leishmania major*, *T. cruzi* and *Campylobacter jejuni*. The dimeric enzymes show no similarity in sequence or structure to the monomeric or trimeric forms and are thought to have reached their catalytic potential through a different evolutionary route. Crystal structures of seven trimeric (*Escherichia coli*,[6] human,[7] equine infectious anemia virus,[8] feline immunodeficiency virus,[9] *Methanococcus jannaschii*,[10] *Mycobacterium tuberculosis*,[11] and *P. falciparum*[12]) and two dimeric (*T. cruzi*[13] and *C. jejuni*[14]) dUTPases have been published to date.

It has been shown that relative to other eukaryotic dUTPase enzymes, α,β-imido-dUTP is a strong inhibitor of the dimeric enzymes. It has been proposed therefore that the triphosphate moiety is necessary for inhibition of the dimeric dUTPases.[15] To emphasise the difference, tritylated nucleoside derivatives prepared by our group were good inhibitors of the trimeric enzymes[12] but are inactive against the dimeric parasitic

Keywords: dUTPase; Structure-based drug design.
* Corresponding author. Tel.: +44 0 1382 386 240; fax: +44 0 1382 386 373; e-mail: ih.gilbert@dundee.ac.uk

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.04.027

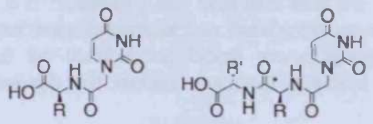
dUTPases. In contrast to the trimeric dUTPases where a glycine-rich site for phosphate binding is common, the phosphates in the dimeric enzymes are held in place with hydrogen bonds to charged side chains.[13] Taking advantage of the structural differences between these enzymes, it was thought that selective inhibitors for the dimeric *T. cruzi* dUTPase (TcdUTPase) could be designed and synthesized as lead compounds for the treatment of Chagas' disease.

Considering the hydrophilic nature of the TcdUTPase active site, compounds in which chains with good hydrogen bond donating or hydrogen bond accepting capability attached to N1 of the uracil ring were considered as potential inhibitors. Ease of synthesis of uracil acetamide derivatives, to which were attached one or two amino acids (Fig. 1), by the Fmoc solid phase strategy allowed for a wide range of structural diversity to be implemented and were therefore chosen for further study. Compounds such as those shown in Figure 1 were drawn, minimized and docked into the active site of TcdUTPase using the program FlexX. The ligand-free (native) TcdUTPase exists in an open form (pdb 1OGL). Binding of the substrate induces substantial structural changes so that the enzyme closes over the ligand (pdb 1OGK).[13] Docking studies were carried out on the closed (complexed) form of the enzyme. The compounds which docked with the best superimposition over the endogenous dUDP ligand were synthesized using solid phase chemistry and tested for biological activity.

FlexX is a docking program that takes into account the flexibility of the ligand but not that of the receptor. The docking method it uses is based on an incremental construction algorithm which consists of three phases: (1) base selection, (2) base placement, and (3) complex construction.[16] The interaction types and scoring functions in FlexX are based on work by Böhm and Klebe.[17]

Docking studies showed that: (1) the R side chain should be H or CH_3. Structures where the side chains, R, were bigger than those of Ala or Gly were too bulky to fit into the active site. (2) The optimum distance between the carbonyl marked with an * (Fig. 1) and the terminal atom was 2 or 3 but no more than 5 atoms. (3) The best scoring compounds were those where R′ was a charged residue. A large degree of variation of the R′ side chain without effecting the superimposition was, however, allowed. Compounds **1–11** (Table 1) were among those which docked best into the TcdUTPase active site (Fig. 2).

Compounds **1–11** were synthesized by solid phase Fmoc chemistry as shown in Scheme 1. Amino acids **12–19**



8 **1-7, 9-11**

Figure 1. Structure of potential TcdUTPase inhibitors.

Table 1. Compounds docked into the TcdUTPase active site using FlexX and subsequently synthesized

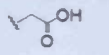
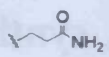
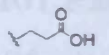
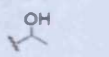
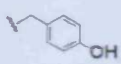
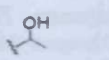
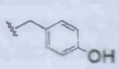
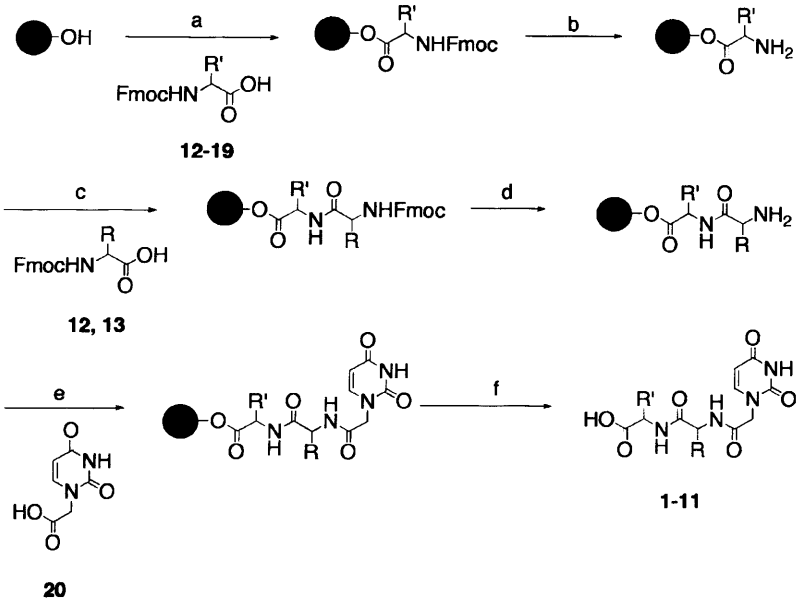
Compound	R	R′	Colour
1	CH_3		Orange
2	CH_3		Red
3	CH_3		Green
4	CH_3	H	Blue
5	CH_3		Purple
6	CH_3		Magenta
7	CH_3		Violet
8	H	–	Greenblue
9	H		Redorange
10	H		By atom type
11	H		White
dUDP			Yellow

Figure 2. Docking of compounds **1–11** in TcdUTPase active site. Superimposition with dUDP.



Scheme 1. General synthesis of compounds **1–11**: (a) DIC, **12–19**, DMAP, DMF (**12**, Ala; **13**, Gly; **14**, Asp; **15**, Gln; **16**, Glu; **17**, Lys; **18**, Thr; **19**, Tyr); (b) 20% piperidine, DMF; (c) TBTU, HOBt, **12–13**, DIPEA, DMF; (d) 20% piperidine, DMF; (e) TBTU, HOBt, **20**, DIPEA, DMF (f); 50% TFA, DCM, TES.

were first converted to the appropriate anhydrides using DIC. The anhydrides and DMAP were then added to Wang resin[18] and the percentage loading was calculated by UV absorption.[19] Subsequent deprotection of the Fmoc group with piperidine was then followed by coupling to the second amino acid using TBTU and HOBt.[20,21] A final deprotection and coupling to 1-carboxymethyl uracil, **20**,[22] followed by cleavage from the resin beads using 50% TFA released the desired compounds **1–11**. All final compounds were analysed and characterized by ^{1}H NMR and MS.

Compounds **1–11** were tested for both TcdUTPase enzyme inhibition[23] and in vitro effects on parasite growth.[24] As can be seen from **Table 2**, none of the compounds inhibited the TcdUTPase enzyme at a concentration of 1 mM. Neither did the compounds inhibit parasite growth at the maximum concentrations shown.

Table 2. Biological activity of compounds **1–11**

Compound	Enzyme inhibition K_i	In vitro IC_{50} (μM)	Cytotoxicity IC_{50}[a] (μM)
1	>1 mM	>84	>252
2	>1 mM	>81	>243
3	>1 mM	>81	210
4	>1 mM	>100	>301
5	>1 mM	>81	>243
6	>1 mM	>87	>262
7	>1 mM	>74	214
8	>1 mM	>132	>396
9	>1 mM	>84	>253
10	>1 mM	>91	>274
11	>1 mM	>76	>230
Standard	18.40[b] μM	1.04[c]	0.15[d]

[a] Cytotoxicity tests were carried out on rat L6 cells.
[b] dUMP.
[c] Benznidazole.
[d] Podophyllotoxin.

Thus, the search so far for lead inhibitors of the dimeric dUTPase enzymes remains unsuccessful. There are a number of reasons why this design paradigm was not successful in this case. (i) Recently, the crystal structure of the dimeric *C. jejuni* dUTPase revealed the coordination of Mg^{2+} ions within the active site of the enzyme.[14] These ions are absent in the TcdUTPase crystal structure. It is unclear if the Mg^{2+} ions are bound to the unliganded enzyme, but it is possible that the presence of these ions must be taken into account when designing inhibitors in silico.

(ii) As previously mentioned, FlexX does not take into account the flexibility of the protein in question. Although this approximation reduces run time significantly, in this case where binding of the ligand induces a large conformational change in protein structure it is possible that a docking program in which protein flexibility is taken into account should be used.

(iii) There must be a mechanism by which the dUTPase undergoes the large conformational change on binding dUTP. Two mechanisms of binding could be envisaged. First, the uracil and deoxyribose moieties, which bind to motifs from the rigid domain of protein, would be recognized. The α, β, and subsequently γ phosphates, which are recognized by mobile domain motifs, would then bond to the Mg^{2+} ions and respective amino acid side chains sequentially and induce 'active site closure'. Alternatively, the phosphate moieties would be recognized by the mobile domain first and the protein would then undergo such a conformational change so as to

encapsulate the uracil and sugar moieties. In the latter case, the phosphate moieties would be more essential for ligand recognition and therefore replacement of them would be detrimental to ligand recognition. This would explain the lack of activity of compounds **1–11**. In either case, the experimental biological data indicate that compounds **1–11** do not possess the required structure to bind to the open form of the enzyme sufficiently enough to induce the conformational change required for inhibition.

Although compounds **1–11** docked with good superimposition and favourable binding energies into the TcdUTPase active site with FlexX, it is apparent that other factors such as protein flexibility and the presence of Mg^{2+} ions should be taken into account when designing competitive inhibitors for these enzymes in silico.

Acknowledgments

The authors thank the European Union (QLRT-2001-00305), Cardiff University and the FIS Network RICET/C03 for funding of this project. The National EPSRC Mass Spectrometry service centre (Swansea) is acknowledged for accurate mass spectrometry.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2006.04.027.

References and notes

1. http://www.cdc.gov/ncidod/dpd/parasites/chagasdisease.
2. http://www.who.int/ctd/chagas/.
3. Hidalgo-Zarco, F.; González-Pacanowska, D. *Curr. Protein Pept. Sci.* **2001**, *2*, 389.
4. Gadsden, M. H.; McIntosh, E. M.; Game, J. C.; Wilson, P. J.; Haynes, R. H. *EMBO J.* **1993**, *12*, 4425.
5. El-Hajj, H. H.; Zhang, H.; Weiss, B. *J. Bacteriol.* **1988**, *170*, 1069.
6. Cedergren-Zeppezauer, E. S.; Larsson, G.; Nyman, P. O.; Dauter, Z.; Wilson, K. S. *Nature* **1992**, *355*, 740.
7. Mol, C. D.; Harris, J. M.; McIntosh, E. M.; Trainer, J. A. *Structure* **1996**, *4*, 1077.
8. Dauter, Z.; Persson, R.; Rosengren, A. M.; Nyman, P. O.; Wilson, K. S.; Cedergren-Zeppezauer, E. S. *J. Mol. Biol.* **1999**, *285*, 655.
9. Prasad, G. S.; Stura, E. A.; McRee, D. E.; Laco, G. S.; Hassekus-Light, C.; Elder, J. H.; Stout, C. D. *Protein Sci.* **1996**, *5*, 2429.
10. Huffman, J. L.; Li, H.; White, R. H.; Tainer, J. A. *J. Mol. Biol.* **2003**, *331*, 885.
11. Chan, S.; Segelke, B.; Lekin, T.; Krupka, H.; Cho, U. S.; Kim, M.; So, M. Y.; Kim, C. Y.; Naranjo, C. M.; Rogers, Y. C.; Park, M. S.; Wald, G. S.; Pashkov, I.; Cascio, D.; Perry, J. L.; Sawaya, M. R. *J. Mol. Biol.* **2004**, *341*, 503.
12. Whittingham, J. L.; Leal, I.; Nguyen, C.; Kasinathan, G.; Bell, E.; Jones, A. F.; Berry, C.; Benito, A.; Turkenburg, J. P.; Dodson, E. J.; Perez, L. M. R.; Wilkinson, A. J.; Johansson, N. G.; Brun, R.; Gilbert, I. H.; Pacanowska, D. G.; Wilson, K. S. *Structure* **2005**, *13*, 329.
13. Harkiolaki, M.; Dodson, E. J.; Bernier-Villamor, V.; Turkenburg, J. P.; González-Pacanowska, D.; Wilson, K. S. *Structure* **2004**, *12*, 41.
14. Moroz, O. V.; Harkiolaki, M.; Galperin, M. Y.; Vagin, A. A.; Gonzalez-Pacanowska, D.; Wilson, K. S. *J. Mol. Biol.* **2004**, *342*, 1583.
15. Hidalgo-Zarco, F.; Camacho, A. G.; Bernier-Villamor, V.; Nord, J.; Ruiz-Perez, L. M.; Gonzalez-Pacanowska, D. *Protein Sci.* **2001**, *10*, 1426.
16. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
17. Böhm, H.-J. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243.
18. Wang, S.-S. *J. Org. Chem.* **1975**, *40*, 1235.
19. Merck Bioscience, (2002/2003): Novabiochem Catalog.
20. Knorr, R.; Trzeciak, A.; Bannwarth, W.; Gillessen, D. *Tetrahedron Lett.* **1989**, *30*, 1927.
21. Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147.
22. Jacobsen, J. R.; Cochran, A. G.; Stephans, J. C.; King, D. S.; Schultz, P. G. *J. Am. Chem. Soc.* **1995**, *117*, 5453.
23. Bernier-Villamor, V.; Camacho, A.; Hidalgo-Zarco, F.; Perez, J.; Ruiz-Perez, L. M.; Gonzalez-Pacanowska, D. *FEBS Lett.* **2002**, *526*, 147.
24. Jones, S. M.; Urch, J. E.; Kaiser, M.; Brun, R.; Harwood, J. L.; Berry, C.; Gilbert, I. H. *J. Med. Chem.* **2005**, *48*, 5932.