COMPUTATIONAL INVESTIGATIONS OF MOLECULAR COMPLEXES OF BIOTECHNOLOGICAL INTEREST

Flavia Autore

Dottorato in Scienze Biotecnologiche – XX ciclo
Indirizzo Biotecnologie Industriali
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XX ciclo
Indirizzo Biotecnologie Industriali
Università di Napoli Federico II

COMPUTATIONAL INVESTIGATIONS OF MOLECULAR COMPLEXES OF BIOTECHNOLOGICAL INTEREST

Flavia Autore

Dottoranda: Flavia Autore

Relatore: Prof. Franca Fraternali

Coordinatore: Prof. Giovanni Sannia

Alla mia famiglia

The complete sequencing of the human genome (2003), together with the ones of a number of microorganisms, has resulted in a large amount of new DNA sequences in the genetic databanks. However, the rapidly growing number of sequenced genes and genomes is outpacing by far the number of experimentally determined structures. One of the ultimate goals of genomic and proteomic projects is to determine the biological functions and cellular roles of all genes and proteins. To determine the function, the knowledge of molecular structure of a protein becomes essential. Additionally, recent developments in proteomics technologies such as mass spectrometry, genome-scale yeast two-hybrid experiments and display cloning experiments are uncovering numerous novel protein-protein interactions. X-ray crystallography is rapidly evolving towards the determination of complex large macromolecular structures, nevertheless, limitations of resources and techniques have left many protein complex structures with key cellular role, yet unsolved. Computational techniques can help in filling, where possible, those gaps by predicting the tridimensional structure of newly sequenced proteins, and in predicting the structural arrangement of macromolecular complexes.The general purpose of this thesis was to master and apply methods for the optimization of enzyme design and for the characterization of protein-protein interactions. In particular, we focused on the application of these methods to molecular systems of biotechnological interest.
The originality of the approach relies on the combination of computational biology tools with experimental procedures by correlating genotypic and phenotypic mutation data and by designing new experiments to enhance the functional properties of the studied molecules. The following systems of biotechnological interest have been analysed in this study: 1) Laccases from *Pleurotus ostreatus*. To use laccases in new bio-based processes a deep understanding of the structure/function relationships of native enzymes is required with the aim to design molecules with improved performance, more suited for industrial applications. In this context the research has been aimed at characterizing molecular determinants of the activity of these enzymes, and at developing and characterizing new laccases obtained from rational design and direct evolution using *P.ostreatus* isoenzymes. 2): Interaction of Malaria Parasite-Inhibitory Antibodies with the Merozoite Surface Protein $MSP1_{19}$. More than 30 distinct antigens identified in various life cycle stages of the malaria parasite have been proposed as potential vaccine candidates, based on various observations such as the surface expression of the antigen on one or more life cycle stages. Several monoclonal antibodies (mAbs) directed against the C-terminus of MSP1 ($MSP1_{19}$) have been identified. Structural information on *Plasmodium falciparum* proteins is very scarce and therefore molecular modelling and molecular docking can be particularly useful in elucidating the details of protein-protein interactions in the parasite. The purpose of this part of the thesis was an investigation of the interaction between the protein $MSP1_{19}$ and two inhibitory antibodies Fab 12.8 and Fab 12.10 by means of docking calculations. 3): Interactions of the human antiviral proteins APOBEC3F and APOBEC3G (A3F/G) with HIV-1 Vif. The human proteins APOBEC3F and APOBEC3G mediate this so-called innate immunity by selectively mutating and inhibiting the production of the viral DNA. Unfortunately, HIV has evolved a protein called the Virus Infectivity Factor (VIF), which binds to and subsequently destroys the APOBEC3F/G proteins. The purpose of this part of the study consisted in modelling the structures of APOBEC3F/G proteins and in using these models to rationalize mutation experiments targeted to key residues from each predicted dimer-interface. The native proteins and their mutants were evaluated by a subsequent assay of the antiviral and dimerization properties of such mutants.

L’enorme interesse verso discipline quali la genomica e la proteomica e’ strettamente correlato all’ambizioso scopo di determinare la funzione biologica e il ruolo cellulare di tutti i geni e delle proteine da essi espressi.
Negli ultimi anni, in seguito al sequenziamento del genoma umano e di quello di innumerevoli microrganismi, il divario tra il numero di sequenze note e quelle di cui conosciamo la struttura e la funzione e’ aumentato notevolmente. In seguito a iniziative quali la genomica strutturale, il numero complessivo di strutture risolte mediante tecniche quali NMR e cristallografia ai raggi X ha subito un notevole incremento negli ultimi anni, tuttavia il divario tra la conoscenza della struttura primaria e quella terziaria è ancora enorme.
Da un punto di vista evolutivo la funzione è più conservata della struttura e la struttura più conservata della sequenza; risulta chiaro, dunque, che la conoscenza della struttura tridimensionale di sitemi macromolecolari biologici rappresenta un passo fondamentale nella comprensione dell’informazione contenuta nel genoma. Nel caso in cui la struttura tridimensionale sia di difficile raggiungimento, la costruzione di modelli tridimensionali può compensare in parte il divario esistente tra sequenze disponibili e strutture determinate sperimentalmente. La costruzione di modelli per omologia e la ricostruzione, tramite questi, di modelli macromolecolari complessi, associata all’analisi di mutanti opportunamente disegnati, permette l’ottenimento di una serie di informazioni utili circa la funzione della proteina, l’identificazione di importanti siti di legame, il disegno di nuovi farmaci.
La funzione di una proteina e’ strettamente correlata al numero e al tipo di interazioni con altre proteine. Diventa sempre più importante, quindi, non solo conoscere la struttura tridimensionale della singola proteina, ma anche quella di complessi macromolecolari.
Negli ultimi anni sono stati fatti notevoli progressi in questo campo e strutture complesse quali quella del Ribosoma e del Nucleosoma sono state determinate con un’elevata accuratezza. Tuttavia esistono ancora limitazioni tecniche che ostacolano la risoluzione strutturale su larga scala di complessi macromolecolari che abbiano un ruolo chiave nei processi cellulari.
Il mio lavoro di tesi rappresenta un contributo in questa direzione: l’applicazione di metodi computazionali atti a caratterizzare interazioni tra proteine e tra proteine e ligandi in sistemi molecolari di interesse biotecnologico.
L’originalità dell’approccio scelto sta nell’avere combinato metodi di biologia computazionale con procedure sperimentali, al fine di migliorare la conoscenza delle proprietà funzionali delle molecole in esame mediante la progettazione di mutanti sito-specifici e la loro successiva caratterizzazione. A tal fine, sono stati utilizzati programmi per la costruzione di modelli per omologia, programmi di *Docking* per la predizione di complessi enzima-substrato e anticorpo-antigene, e programmi per l’analisi delle superfici accessibili al solvente, capaci di determinare caratterizzare le relative superfici d’interazione.
I sistemi oggetto di studio di questa tesi sono:

Sistema 1: Laccasi da *Pleurotus ostreatus*

Le laccasi sono enzimi appartenenti alla classe delle fenolo-ossidasi. Esse catalizzano l’ossidazione di fenoli, polifenoli e ammine aromatiche variamente sostituite con relativa riduzione di ossigeno molecolare ad acqua. Risulta dunque chiaro il loro impiego in campo biotecnologico. E’ infatti noto il loro utilizzo per i più

disparati scopi: dalla demolizione della lignina alla *bioremedation*, che li impiega nel risanamento di siti inquinati da effluenti industriali.
L'interesse nella realizzazione di enzimi che presentino una migliore o una nuova attività catalitica è un interessante aspetto che può essere utilizzato per ampliare le potenziali applicazioni biotecnologiche delle laccasi.
Lo sviluppo di laccasi in nuovi bio-processi richiede una profonda conoscenza della relazione struttura/funzione degli enzimi nativi che porterebbe alla realizzazione di "nuovi" e migliori enzimi adatti per le applicazioni industriali.
Differenti strategie possono essere applicate per modificare le proprietà enzimatiche: l'evoluzione guidata e la mutagenesi razionale. La prima richiede la conoscenza della sequenza del gene e ma non necessita della conoscenza della funzione e della struttura tridimensionale della proteina codificata; la seconda, invece, che prevede la pianificazione di esperimenti di mutazione sito-specifica, richiede la conoscenza della sequenza, della struttura tridimensionale, e ove possibile, anche del meccanismo d'azione enzimatica.
Nel Laboratorio in cui si è svolto questo progetto sono state caratterizzate da un punto di vista termodinamico e catalitico tre laccasi, POXC, POXA1b e POXA3 dal fungo basidiomicete *white-rot Pleurotus ostreatus,* le cui strutture tridimensionali sono però ancora sconosciute.
Il mio lavoro di tesi ha previsto la predizione dei modelli tridimensionali delle laccasi da *P.ostreatus* e la successiva caratterizzazione dei determinanti molecolari dell'attività catalitica. L'analisi e la comparazione delle sequenze primarie delle laccasi da *P.ostreatus* con le sequenze delle laccasi da basiodiomiceti a struttura nota, accoppiata con lo studio dei modelli 3D ottenuti, ha permesso di individuare delle singolari caratteristiche delle laccasi in esame. In particolare dall'allineamento è emerso che sia POXC che POXA1b presentano un prolungamento nella sequenza del C-terminale di sei e sedici amminoacidi rispettivamente. E' stato dimostrato che la regione C-terminale influenza le proprietà delle laccasi, suggerendo un suo possibile ruolo funzionale nell'attività laccasica. I modelli tridimenionali ottenuti sono stati utilizzati per predire complessi enzima-substrato permettendo di individuare i residui in diretto contatto con il substrato (acido violurico), probabilmente responsabili dell'affinità dell'enzima per quest'ultimo. Date queste ulteriori analisi è stato identificato un residuo di arginina nel sito del legame al substrato in POXA3 al posto di un residuo di acido aspartico altamente conservato nella maggior parte delle laccasi da basidiomicete. Il residuo di acido aspartico nella zona del sito catalitico risulta probabilmente coinvolto nella stabilizzazione del catione ottenuto dall'ossidazione del substrato fenolico. Noi suggeriamo che la presenza di un'arginina porterebbe alla repulsione del radicale cationico e quindi ad una minore efficienza catalitica.
Al fine di investigare il ruolo della regione C-terminale delle laccasi da *P. ostreatus* sono stati progettati e realizzati i mutanti tronchi POXA1bΔ4, POXA1bΔ16 e POXCΔ6. Inoltre, allo scopo di caratterizzare i determinati molecolari coinvolti nell'interazione con il substrato, sono stati realizzati i mutanti POXA1b(D205R) e POXC(D210R). Successivamente tutti i mutanti sono stati espressi nel lievito *S. cerevisiae.* Per quanto riguarda POXC e i suoi mutanti, è stato possibile effettuare solo un'analisi qualitativa dalla quale è emerso che il mutante POXCΔ6 risulta meno attivo dell'enzima POXC mentre il mutante POXC(D210R) sembrerebbe aver totalmente perso attività enzimatica. Per quanto riguarda POXA1b e i suoi tre mutanti (POXA1bΔ16, POXA1Δ4 and POXA1b(D205R)), è stato possibile purificare le tre proteine ed è stato possibile effettuare una caratterizzazione cinetica e catalitica

dalla quale è emerso che: la costante di Michaelis-Menten (K_M) dei mutanti tronchi è paragonabile a quella dell'enzima *wild-type* rispetto a tutti i substrati saggiati, mentre il valore di K_M di POXA1b(D205R) è circa il doppio rispetto a quella della proteina ricombinante POXA1b. Questi dati evidenziano l'importanza del residuo Asp nell'interazione con il substrato, essendo questa l'unica differenza nelle relative sequenze. Inoltre, dall'analisi della stabilità a differenti pH è risultato che i mutanti tronchi sono più stabili del *wild-type* a pH acidi, mentre tutti i mutanti in esame perdono drasticamente la peculiare stabilità a pH alcalini di POXA1b. Inoltre, la termoresistenza effettuata a 60℃ ha mostrato che P OXA1bΔ16 e POXA1b(D205R) sono sensibilmente meno resistenti di POXA1b.
Dai dati ottenuti si può dedurre che il residuo di acido aspartico presente nel sito attivo è direttamente coinvolto nell'interazione con il substrato e che la sostituzione con un residuo di arginina potrebbe generare un riarrangiamento della struttura che avrebbe ripercussioni sulla stabilità della proteina. Inoltre l'estremità C-terminale sembrerebbe essenziale nello stabilizzare la proteina in condizioni estreme di temperatura e di pH. Un'analisi puramente computazionale è stata invece effettuata su alcuni mutanti generati dalla Dott. G. Festa mediante la tecnica dell'evoluzione guidata (Tesi di dottorato 2006, Università di Napoli "Federico II").
Simulazioni di dinamica molecolare in soluzione sul modello 3D di POXA1b e sui mutanti che presentavano le migliori *performaces* catalitiche hanno permesso di integrare i risultati sperimentali ottenuti e di razionalizzare a livello molecolare la diversa stabilità e la diversa attività.

Sistema 2: Interazionie di anticorpi monoclonali antimalarici con la proteina di superficie $MSP1_{19}$ da *Plasmodium falciparum*

P. falciparum, uno dei quattro potozooi responsabili della malaria, rappresenta la specie non solo più diffusa, ma anche la più virulenta. La resistenza sempre crescente di questo ceppo ai farmaci più diffusi e a basso costo, ha creato un'urgente bisogno di sviluppare nuove strategie per debellare la malattia. Dato l'alto costo dei trattamenti fin'ora adoperati è chiaro che l'interesse per lo sviluppo di un vaccino che possa abbattere non solo la mortalità causata dalla malattia ma anche il suo costo sta crescendo notevolmente.
Il parassita responsabile della malaria presenta svariati antigeni, riconosciuti dal sistema immunitario, che variano a seconda del momento del ciclo di vita del parassita stesso. L'identificazione di questi è fondamentale per lo sviluppo di un vaccino multifasico che potrebbe offrire una completa protezione. Le proteine sulla superficie del merozoite sono dei *targets* eccellenti, in quanto possono essere facilmente riconosciuti dagli anticorpi. Una delle proteine di superficie più studiata è MSP1 (Merozoite Surface Protein 1) da P. falciparum. MSP1 è inizialmente sintetizzata come una preproproteina di circa 200 kDa. Essa subisce due processi di maturazione proteolitica che rilasciano sulla superficie del merozoite un frammento di 19 kDa denominato $MSP\text{-}1_{19}$. La struttura cristallografica della proteina $MSP\text{-}1_{19}$ è stata recentemente risolta mediante NMR e, inoltre, esperimenti di *cross-saturation* effettuati su complessi proteina-anticorpo hanno permesso di individuare il residui dell'antigene in diretto contatto con gli anticorpi monoclonali inibitori Fab12.8 e Fab12.10. In particolare è risultato che l'antigene interagisce con entrambi gli anticorpi utilizzando la stessa faccia.
Lo scopo di tale sessione è stato quello di predire l'interazione antigene-anticopo e di caratterizzare i residui degli anticorpi coinvolti nell'interazione.

Le strutture dei due anticorpi non sono state ancora risolte, ed è stato quindi necessario prima di tutto costruire i modelli tridimensionali di entrambi gli anticorpi Fab12.8 e Fab12.10. Successivamente per la predizione dei complessi anticorpo-antigene è stato utilizzato un programma di *docking* rigido, ZDOCK, che è risultato il migliore per lo studio di questo tipo di interazioni. Per ogni complesso sono state predetti 2000 complessi che sono poi stati selezionati in base ai dati sperimentali NMR a nostra disposizione. La strategia utilizzata è stata verificata utilizzando la struttura cristallografica del complesso $MSP1_{19}$-FabG12.17 per appurare che tale metodo fosse valido. La struttura cristallografica $MSP1_{19}$-FabG12.17 e i complessi selezionati sono stati successivamente sottoposti a simulazioni di dinamica molecolare al fine di migliorare le iterazioni antigene-anticorpo. Dalle analisi effettuate abbiamo dedotto che l'antigene si posiziona in maniera differente rispetto gli anticorpi e sembrerebbe che tale posizione è influenzata dalla presenza di residui positivi sulla superficie degli anticorpi. L'identificazione dei residui responsabili del legame sarà utilizzata per la progettazione di nuovi esperimenti che permetteranno di migliorare l'affinità dei complessi antigene-anticorpo.

Sistema 3: Interazioni di protein uname antivirali APOBEC3F e APOBEC3G (A3F/G) con HIV-1 Vif.

Negli ultimi anni è emerso che le cellule umane presentano una difesa immunitaria contro le infezioni di lentivirus ed in particolare contro il noto virus dell'HIV. Tale difesa immunitaria è ad opera di due proteine omodimeriche, APOBEC3F e APOBEC3G, appartenenti alla famiglia delle citosina-deamminasi (CDA). Questa superfamiglia è largamente distribuita ed ha un ruolo fondamentale in diversi meccanismi enzimatici. Questa classe di enzimi converte la citosina a uracile producendo quindi un effetto sulle funzioni fisiologiche della cellula. In particolare le proteine APOBEC presentano uno o più domini che legano lo zinco mediante una sequenza specifica H-X-E-X_{23-28}-P-C-X_{2-4}-C (dove X sta per ogni amminoacido), denominati CDA. Sfortunatamente, il virus dell'AIDS ha sviluppato una contro difesa: produce infatti una proteina chiamata VIF (Fattore Virale Infettivo) la quale innesca la distruzione dei limitatori retrovirali, impedendo così le mutazioni del suo DNA. Individuare i determinanti molecolari responsabili di tale interazione potrebbe condurre al disegno di nuove proteine A3G/F che presentino una minore affinità verso la proteina VIF.

Solo di recente è stata risolta la struttura cristallografica di un membro della famiglia delle APOBEC, la proteina APOBEC2 (A2). A2 è una proteina tetramerica che presenta 4 domini CDA. I quattro monomeri interagiscono tra di loro formando una struttura allungata e piatta. Tale struttura è certamente atipica paragonata con le altre strutture risolte che presentano un *folding* globulare.
La struttura di A2 è stata quindi utilizzata per predire le strutture tridimensionali dei monomeri e successivamente dei dimeri formati da A3G/F. Ancora non è noto come i monomeri di A3G/F interagiscano per formare il dimero, ma il tipo di dimerizzazione sembra limitato alle possibili interazioni testa-testa o coda-coda dei monomeri. I possibili dimeri molecolari sono stati predetti per entrambi i sistemi utilizzando il programma MODELLER che permette di predire e calcolare l'energia di centinaia di modelli generati. I modelli che presentavano minore energia sono stati selezionati, in particolare due famiglie di modelli sono state generate: una estesa ed allungata (che presenta un folding analogo a quello della proteina stampo) e quella piegata che

presenta un *folding* leggermente diverso, con la superficie di contatto tra i due monomeri piu' estesa, formando una struttura a piegata a V inversa.
Le interfacce tra i monomeri sono state analizzate valutando non solo i residui in contatto tra i monomeri ma anche la natura elettrostatica dell'interazione. I dati ottenuti per la dimerizzazione testa-testa sono in accordo con i dati sperimentali presenti in letteratura, in particolare all'interfaccia è stata riscontrata la presenza di residui responsabili dell'interazione con VIF. I dati ottenuti mediante l'analisi computazionale sono in accordo con recenti risultati sperimentali di mutagenesi sito-specifica effettuati dal gruppo del Prof. Malin del King's College di Londra ed hanno permesso di individuare i residui che influenzano la dimerizzazione; in particolare, tali analisi hanno evidenziato che i monomeri interagiscono mediante i domini N-terminale. Infatti, l'analisi approfondita sui modelli ha messo in evidenza che non solo nella dimerizzazione testa-testa sono coinvolti i residui responsabili dell'interazione con VIF, ma anche che questo tipo di interazione produce all'interfaccia del dimero una regione carica positivamente, particolarmente adatta ad alloggiare il ssDNA/RNA, molecola ben nota per la sua natura acida.
Simulazioni di dinamica molecolare sono state effettuate sui dimeri testa-testa come pure sulla struttura cristallografica di A2 al fine di approfondire la caratterizzazione della natura dell'interazione di tali proteine. Dalle analisi effettuate sulle dinamiche si evince che durante la simulazione i monomeri si avvicinano per ottimizzare l'interazione e che questo parziale movimento, seppur lieve, origina un riarrangiamento delle interfacce. L'analisi dei residui in contatto prima e dopo le simulazioni hanno evidenziato che due nuove regioni dei monomeri sono in contatto e sono stati individuati nuovi residui che stabilizzano l'interazione tra i monomeri. I dati forniti dall'analisi computazionale potranno essere utilizzati per progettare nuovi esperimenti di mutagenesi per validare i dati ottenuti e per caratterizzare ulteriormente i residui responsabili della dimerizzazione.

Chapter 1

Purpose of the thesis

The general purpose of this thesis was to master and apply methods for the optimization of enzyme design and for the characterization of protein-protein interactions. In particular, we focused on the application of these methods to molecular systems of biotechnological interest.

The originality of the approach relies on the combination of computational biology tools with experimental procedures by correlating genotypic and phenotypic mutation data and by designing new experiments to enhance the functional properties of the studied molecules.

The following systems of biotechnological interest have been analysed in this study:

- System 1: Laccases from *Pleurotus ostreatus*.

Laccases belong to the multicopper enzyme family and are phenol-oxidases that catalyze the oxidation of a wide variety of substrates and they have been used in a number of biotechnological applications [1].

The development of new bio-based processes using laccases requires a deep understanding of the structure/function relationships of native enzymes and the design of molecules with improved performance that are more suited for industrial applications. In this context the research has been aimed at characterizing molecular determinants of the activity of these enzymes, and at developing and characterising new laccases using *P.ostreatus* isoenzymes.

- System 2: Interaction of Malaria Parasite-Inhibitory Antibodies with the Merozoite Surface Protein $MSP1_{19}$.

The development of vaccine candidates for the malaria parasite would provide a way of alleviating the devastating effects of this disease in the world. No effective vaccine has been developed yet, mainly because there is no scientific consensus on the best approach to the problem. So far, an enormous amount of resources has been wasted on trials of ineffective vaccines, while one should focus on generating better candidates. This is the outcome of a report released on the 17th September 2007 by the George Institute for International Health in Sydney, Australia. The report documents the status and future prospects of 47 vaccines and 21 drugs in various stages of development. The vaccine being in a more advanced stage has been developed by GlaxoSmithKline (GSK). It should enter phase III trials in 2008 and potentially licensed in 2012. That vaccine nearly halved severe malaria in a 2005 trial on 2,000 children younger than 5 years in Mozambique [2]. Nevertheless the report claimed a need of more basic research on malaria.

More than 30 distinct antigens identified in various life cycle stages of the malaria parasite have been proposed as potential vaccine candidates, based on various observations, such as the surface expression of the antigen on one or more life cycle stages. Recent protection studies in the monkey model system have shown promising results with a 42 kDa MSP1 protein produced in baculovirus [3]. Several monoclonal antibodies (mAbs) directed against the C-terminus of MSP1 ($MSP1_{19}$) have been identified; some inhibit invasion of erythrocytes in vitro [4], while some inhibit both, secondary processing of MSP1 and erythrocyte invasion [5].

Structural information of *Plasmodium falciparum* proteins is very scarce and therefore molecular modelling and molecular docking can be particularly useful in elucidating the details of protein-protein interactions in the parasite.

The purpose of this part of the thesis was an investigation of the interaction between the protein $MSP1_{19}$ and two inhibitory antibodies by means of docking calculations.

- System 3: Interactions of the human antiviral proteins APOBEC3F and APOBEC3G (A3F/G) with HIV-1 Vif.

In recent years it has emerged that the cells of the immune system themselves harbour an ancient defense mechanism against infection by viruses that takes place in the interior of the cell. The human proteins APOBEC3F and APOBEC3G mediate this so-called innate immunity by selectively mutating and inhibiting the production of the viral DNA. Unfortunately, HIV has evolved a protein called the Virus Infectivity Factor (VIF), which binds to and subsequently destroys the APOBEC3F/G proteins. Very recently, the crystal structure of the APOBEC2 (A2) family member has been solved[6]. Although A2 has no known function and certainly does not inhibit HIV, its reported structure provides an ideal template for structural modelling of the A3F/G proteins.
The purpose of this part of the study consisted in modelling the structures of APOBEC3F/G proteins and in using these models to rationalize mutation experiments targeted to key residues from each predicted dimer-interface. The native proteins and their mutants were evaluated by a subsequent assay of the antiviral and dimerization properties of such mutants [7].

Why computational methods?
Whole genome sequencing of a number of organisms, including the complete human genome (2003), have produced a large amount of genomic DNA sequences in the genetic databanks. This enormous amount of data would contribute to a complete picture of biological information only if all the links between genome sequence and protein structure and function were established. However, the rapidly growing number of sequenced genes and genomes is heavily outpacing the number of experimentally determined structures. Genome annotation has produced a larger gap between the number of known protein sequences and the number of solved three-dimensional structures. The latest data from SWISS-PROT-TrEMBL (non-redundant database) contains over 45.000 protein sequences.
Although there is a considerable gap between the number of known sequences and the number of available 3D structures, the knowledge of three-dimensional protein structures is crucial for answering many biological questions. Firstly, structure is more conserved than sequence during the course of evolution [8]. Secondly, function is more conserved than structure, but function is often attributed to a few spatially proximate atoms; therefore, the knowledge of the 3D structure is a fundamental step in the assessment of the function of a new protein.
The quest for more structural knowledge has inspired the formation of a structural genomics consortium that is dedicated to the production of representative experimental structures for all classes of protein folds [9]. Because it is difficult, expensive and time-consuming to obtain experimental structures from methods such as X-ray crystallography and protein NMR for every protein of interest, protein structure prediction methods can provide useful structural models for generating hypotheses about protein function and for directing further experimental work. Therefore, this scenario represents a unique opportunity for theoretical methods to help in filling the sequence/structure gap. The development of efficient structure

prediction methods can contribute to the determination of structures of proteins with high sequence identity (>30%) with other proteins of known 3D-structure.

Comparative homology modeling

Protein structure predictions can be used to obtain accurate protein structure models for evolutionarily related sequences that share similar structures [8, 10]. This technique is called “homology modeling” or “comparative modeling”. It is a computational tool to construct an atomistic model of a protein from its amino acid sequence (the "query sequence" or "target"). Almost all homology modelling techniques use one or more known protein structures (known as "templates" or "parent structures") that are likely to resemble the structure of the query sequence. From the sequences of query and templates a sequence alignment is created that maps residues in the query sequence to the residues in the template sequence. The sequence alignment and the template structure are then used to generate a structural model of the target. As a consequence, a three-dimensional model of a protein can be built from a related protein or proteins of known structure that shows a sequence identity above a certain threshold (around 30%).

Two important factors influence the ability to predict accurate models: the correctness of the alignment and the role of structural conservation between the protein of interest (target) and the template [10, 11]. On the other hand, sequence identity between target and template is not the only important parameter to calculate approximately the quality of a comparative model [11, 12]. A good evaluation of the model quality is the distribution of sequence identity in a multiple sequence alignment comprising the target, template and intermediate homologous sequences [13]. Generally, models based on 30% sequence identity seem to be sufficiently accurate for molecular replacement techniques used in X-ray crystallography [14]. The quality of comparative models tends to decrease with increasing evolutionary distance between target and template [10-12]. However, when template and target share the same function and specificity, usually functional regions are structurally better conserved than other regions [12]. When their function or specificity differs, larger structural changes are expected in functional regions [15].

Accurate modeling of the differences between similar structures (insertions and deletions) is one of the most biologically relevant applications of comparative modelling, because these structural changes usually modify functions and/or specificities.

Very recently [16] protein structure prediction methods have been shown to be helpful in predicting the crystallographic phase with surprising accuracy. These results suggest that the combination of high-resolution structure prediction with state-of-the-art phasing tools could become an integral part of the process of X-ray structure resolution in the future.

The insights gained from comparative models may be used to identify critical residues that are involved in catalysis, binding or structural stability. The 3D model can be a departure point to examine protein–protein interactions, to design new drugs and to devise future experiments.

Used in conjunction with molecular dynamics simulations, homology models can help in generating hypotheses about the kinetics and dynamics of protein interactions, as demonstrated by excellent studies on the ion selectivity of a potassium channel [17]. Large-scale automated modelling of all identified protein-coding regions in a genome has been performed for the yeast *Saccharomyces cerevisiae*, resulting in nearly 1000 quality models for proteins whose structures had not yet been determined at

the time of the study, and identifying novel relationships between 236 yeast proteins and other previously solved structures [18].

Structure Prediction of Protein Complexes

One of the ultimate goals of genomic and proteomic projects is to determine the biological functions and cellular roles of all genes and proteins. Additionally, recent developments in proteomics technology such as mass spectrometry, genome-scale yeast 2-hybrid experiments and display cloning experiments are uncovering numerous novel protein-protein interactions. These interactions influence directly and indirectly the biological networks of which they are a part. X-ray crystallography has provided us with many structures of protein–protein complexes, but limitations of resources and experimental design have left many protein complex structures yet unsolved.

While techniques such as homology modeling and mutational analysis yield some information about binding sites, more information is needed to understand the nature of protein–protein interactions. Thus, it is useful to employ protein docking to predict complex structure. By means of docking it is possible to identify the residues that are directly involved in binding and to reveal the nature of the interface itself.

Protein–protein docking can be defined as the computational determination of the complex structure between two proteins, given the coordinates of the individual proteins. This methodology is often used to predict protein-DNA, protein-RNA and protein-small molecule interactions.

Predictive docking procedures start from the individually determined (unbound) structures of two proteins and they aim at predicting the complex structure. Binding site information rarely accompanies novel protein-protein interactions uncovered by proteomics analyses. All docking programs contain 1. a scoring function to discriminate between near-native docked orientations and incorrect orientations, and 2. a search algorithm that is designed to sample rapidly all possible docking orientations. Proteins undergo conformational changes upon complex formation to varying extent. Even though structural flexibility is mostly restricted to surface side chains [19], the innate flexibility renders protein docking extremely difficult. The algorithm by Abagyan and colleagues [20, 21] allows for surface side chain flexibility; however, such algorithms typically require hundreds of hours of computation for each complex [20], which is not feasible for genome-scale applications. The alternative is to adopt the *rigid-body* approach, which samples only the six rotational/translational degrees of freedom, but uses target functions that are tolerant to some overlap of the two proteins being matched. ZDOCK (rigid-body docking) [22] is an accurate accurate program with a simple target function. In particular, it is essential to have some correct predictions among the complexes that are produced. The terms employed during initial stage docking are shape complementarity, desolvation free energy and electrostatic energy [23]. The ZDOCK program uses two PDB files as input (the two molecules to dock) and creates a list of top-scoring ligand orientations as output. For every rotational orientation, ZDOCK keeps the 10 best translational orientations. The basic search algorithm samples exhaustively the entire rotational and translational space of the ligand with respect to the receptor, which remains fixed at the origin. For every rotation, the algorithm rapidly scans the translational space using fast Fourier transformation (FFT).

Three lists are kept at all times, each containing ligand orientations: the first list with the best shape complementarity score S_{SC}, the second list with the best shape

complementarity and desolvation score $\alpha S_{SC} + S_{DS}$ and a third list with the best shape complementarity, desolvation and electrostatics score $\alpha S_{SC} + S_{DS} + \beta S_{ELEC}$.
Shape complementarity (SC) is the most basic ingredient of all scoring functions for docking. As the name implies, it is a geometric descriptor, stemming from the practical observation that protein surfaces are complementary to each other at the binding interface.
The total Desolvation Free Energy (DS) of complex formation (S_{DS}; DS stands for desolvation) is calculated by summing over the Atomic Contact Energy (ACE) scores of all atom pairs between the receptor and the ligand that are within 6 Å distance [23]. ACE is a desolvation free energy measure, defined as the free energy of replacing a protein-atom/water contact with a protein-atom/protein-atom contact. Combined with an electrostatics term, ACE has been shown to accurately predict the free energy changes associated with transferring amino acids from the protein interior to water, to predict site-specific mutations of hydrophobic amino acids and to assess protease-inhibitor binding [24-27].
The electrostatics energy (ELEC) can be expressed as a correlation between the electric potential generated by the receptor and the charges of the ligand atoms, as described by the Coulomb formula [28]. Additionally, grid points in the core of the receptor are assigned a value of 0 for the electric potential to eliminate contributions from non-physical receptor-core/ligand contacts.
However, in an unbound docking study, surface side chains tend to be at non-optimal positions and therefore the resulting electrostatic energy can be inaccurate and even unrealistic. Artifacts can also be introduced by the grid representation. Therefore the electrostatics energy S_{ELEC} is scaled by a factor β. Also the shape complementarity score has been weighted by a scaling factor α. The final target function is:

$$S=\alpha SSC + SDS + \beta SELEC \quad (1.1)$$

The default values for scaling factors are α= 0.01 and β= 0.06.
One useful feature of ZDOCK is that it allows the user to “block” residues, making it unfavorable for selected residues to be in the interface. This can be used, for example, to constrain the antigen of an antibody–antigen complex to be bound to the CDR loops of the antibody. It is also useful when experimental and/or biological information is known regarding the location of the binding site. Blocking is accomplished by assigning a special atom type to all atoms of the residues to be prohibited from the putative binding site.
Protein docking methods have improved substantially over the past few years. This can be seen from results of the Critical Assessment of Predicted Interactions (CAPRI) – the first community-wide experiment devoted to protein docking [29].

Molecular Dynamics Methods
Molecular Dynamics (MD) is a computer simulation technique in which the time evolution of a set of interacting atoms is computed by integrating their equations of motion. Therefore, MD allows us to view how a molecular system evolves through time and to derive average properties of the system, given a simulation of sufficient length. MD is particularly useful when the system cannot be studied by experimental methods.
The equations of motion can only be solved numerically for a multi-body problem. To calculate the dynamics of the system (*i.e.* the position of each atom as a function of

time), Newton's classical equations of motion are solved for each atom given an empirical force field:

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \tag{1.2}$$

The force on each atom is the negative of the derivative of the potential energy with respect to the position of the atom:

$$F_i = -\frac{\partial V}{\partial r_i} \tag{1.3}$$

Once the coordinates of the atoms of a starting structure and their velocities are defined, the force acting on each atom can be calculated for each point in time t + dt and a new set of coordinates can be generated. The repetition of this procedure generates a molecular trajectory corresponding to the time-dependent fluctuations of the atomic positions. The accuracy of the simulations is directly related to the potential energy function that is used to describe the interactions between particles. In MD a classical potential energy function is used that is defined as a function of the coordinates of each of the atoms.
The potential energy function is separated into terms representing covalent interactions and non covalent interactions.
The covalent interactions may be described by the following terms:

$$V_{bond} = \sum_{i=1}^{N_b} \frac{1}{2} k_i^b \left(r_i - r_{0,i}\right)^2 \tag{1.4}$$

$$V_{angle} = \sum_{i=1}^{N_\theta} \frac{1}{2} k_i^\theta \left(\theta_i - \theta_{0,i}\right)^2 \tag{1.5}$$

$$V_{dihedral} = \sum_{i=1}^{N_\Phi} \frac{1}{2} k_i^\Phi \cos\left(n_i\left(\Phi_i - \Phi_{0,i}\right)\right) \tag{1.6}$$

$$V_{improper} = \sum_{i=1}^{N_\xi} \frac{1}{2} k_i^\xi \left(\xi_i - \xi_{0,i}\right)^2 \tag{1.7}$$

The equations correspond to two, three, four and four body interactions, respectively. These interactions are represented by harmonic potentials for the bond lengths r_i, for the bond angles θi, for the improper dihedral (out of the plane) angle ξi and by a more complex potential for the dihedral angles Φi. The non-covalent (non-bonded) interactions, which correspond to interactions between particles separated by more than three covalent bonds, are usually described by Coulomb's law

$$V_{Coulomb} = \sum_{i<j} \frac{1}{4\pi\varepsilon_0\varepsilon_r} \frac{q_i q_j}{r_{ij}} \tag{1.8}$$

for the electrostatic interactions and by a Lennard-Jones potential:

$$V_{LJ} \sum_{i<j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \tag{1.9}$$

for the van der Waals (vdW) interactions, where rij is the atomic distance between particles i and j. The force field parameters describe the strength of the interactions. For bonded interactions parameters are defined for bond stretching, bond bending and torsional rotation. Another set of parameters determines the strength of non-bonded electrostatic and van der Waals interactions.
Electrostatic interactions are generally represented by point charges located at the center of the atom.
It should be kept in mind, however, that MD is affected by several limitations. Firstly, MD is computationally very demanding and the computational load scales with the square of the system size. Simulation times are currently limited to hundreds of nanoseconds or a few microseconds at most. The phenomena that can be explored must occur with sufficient statistical significance within time scales that are encompassed by the computation.
How the system evolves through time is specified by the force field and by an integration time step that determines where the atoms will be positioned at time t + dt. MD requires the use of a very small time-step (1-2 fs) to achieve accurate results, because small time-steps limit the approximations that are introduced by the numerical integrator. This limits the overall scope of the simulated time and the computable properties.
According to statistical thermodynamics, physical quantities are represented by averages over configurations belonging to a certain statistical ensemble. A trajectory obtained by molecular dynamics provides such an ensemble. Therefore, a measurement of a physical quantity by simulation is simply obtained as an arithmetic average of the various instantaneous values adopted by that quantity during the MD run. Statistical thermodynamics is the link between the microscopic ensembles and and the macroscopic properties. In the limit of an exact force field and very long simulation times, one could expect the phase space to be fully sampled and in that limit the averaging process would yield exact thermodynamic properties. In practice, MD runs are always of finite length and one should exert caution when judging the sampling quality. An important constraint in deriving average properties is to extract configurations only from an ensemble at thermal equilibrium. Therefore, MD simulations start generally with an equilibration phase. Once equilibrium is reached, the simulation enters the production phase. The production run should be long enough to sample the property of interest with sufficient statistical significance.

References:

[1] Rodriguez Couto S, Toca Herrera JL. Industrial and biotechnological applications of laccases: a review. *Biotechnol Adv* (2006);24:500-13.

[2] Alonso PL, Sacarlal J, Aponte JJ, Leach A, Macete E, Aide P, et al. Duration of protection with RTS,S/AS02A malaria vaccine in prevention of Plasmodium falciparum disease in Mozambican children: single-blind extended follow-up of a randomised controlled trial. *Lancet* (2005);366:2012-8.

[3] Singh S, Miura K, Zhou H, Muratova O, Keegan B, Miles A, et al. Immunity to recombinant plasmodium falciparum merozoite surface protein 1 (MSP1): protection in Aotus nancymai monkeys strongly correlates with anti-MSP1 antibody titer and in vitro parasite-inhibitory activity. *Infect Immun* (2006);74:4573-80.

[4] Blackman MJ, Heidrich HG, Donachie S, McBride JS, Holder AA. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. *J Exp Med* (1990);172:379-82.

[5] Blackman MJ, Scott-Finnigan TJ, Shai S, Holder AA. Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein. *J Exp Med* (1994);180:389-93.

[6] Prochnow C, Bransteitter R, Klein M, Goodman M, Chen X. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* (2006).

[7] Huthoff H, Malim MH. Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. *J Virol* (2007);81:3807-15.
[8] Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* (1986);5:823-6.
[9] Williamson AR. Creating a structural genomics consortium. *Nat Struct Biol* (2000);7 Suppl.
[10] Martin-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* (2000);29:291-325.
[11] Kryshtafovych A, Venclovas C, Fidelis K, Moult J. Progress over the first decade of CASP experiments. *Proteins* (2005).
[12] Tress M, Ezkurdia I, Grana O, Lopez G, Valencia A. Assessment of predictions submitted for the CASP6 comparative modelling category. *Proteins* (2005).
[13] Cozzetto D, Tramontano A. Relationship between multiple sequence alignments and quality of protein comparative models. *Proteins* (2005);58:151-7.
[14] Giorgetti A, Raimondo D, Miele AE, Tramontano A. Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics* (2005);21.
[15] Moult J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol* (2005).
[16] Qian B, Raman S, Das R, Bradley P, McCoy A, Read R, et al. High-resolution structure prediction and the crystallographic phase problem. *Nature* (2007).
[17] Capener CE, Shrivastava IH, Ranatunga KM, Forrest LR, G.R. S, Sansom MSP. Homology Modeling and Molecular Dynamics Simulation Studies of an Inward Rectifier Potassium *Channel Biophys J* (2000);78:2929-42.
[18] Sanchez R, Sali A. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. *PNAS* (1998);95:13597-602.
[19] Betts MJ, Sternberg MJ. An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Eng* (1999);12:271-83.
[20] Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 A accuracy. *Nat Struct Biol* (1994);1:259-63.
[21] Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, et al. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nat Struct Biol* (1996);3:233-9.
[22] Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* (2003);52:80-7.
[23] Chen R, Weng Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins* (2002);47:281-94.
[24] Zhang C, Cornette JL, Delisi C. Consistency in structural energetics of protein folding and peptide recognition. *Protein Sci* (1997);6:1057-64.
[25] Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. *J Mol Biol* (1997);267:707-26.
[26] Vasmatzis G, Zhang C, Cornette JL, DeLisi C. Computational determination of side chain specificity for pockets in class I MHC molecules. *Mol Immunol* (1996);33:1231-9.
[27] Zhang C, Chen J, DeLisi C. Protein-protein recognition: exploring the energy funnels near the binding sites. *Proteins* (1999);34:255-67.
[28] Gabb H, Jackson R, Sternberg M. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology* (1997);272:106-20.
[29] Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, et al. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* (2003);52:2-9.

Chapter 2

Introduction: Industrial use of enzymes

The systematic and indiscriminate exploitation of the resources of our planet has caused growing problems for the environment; for this reason the development of environmentally friendly and, at the same time, economically sustainable technologies has acquired increasing interest during the last few years.

Industrial biotechnology, also known as white or environmental biotechnology, is the application of nature's toolset for the sustainable production of biochemicals, biomaterials and biofuels from renewable resources, using living cells and/or their enzymes. This results generally in cleaner processes with minimum waste generation and energy use. Current practice in industrial biotechnology demonstrates that the social, environmental and economic benefits of bio-based processes go hand in hand. The potential economic value of industrial biotechnology for the chemical industry alone is estimated to be € 11-22 billion per annum by 2010. McKinsey & Company estimate that biotechnology could be applied in the production of 10 to 20% of all chemicals sold by 2010, starting from the current level of about 5%.

Industrial biotechnological strategies based on biocatalytic processes represent an area of great interest. The recent sequencing of the human genome and the associated sequencing of bacterial and yeast genomes have also played their part in this area. In addition, the fields of metabolic engineering, bioinformatics and computer-based modelling as well as process optimization are opening up opportunities for new products and cost reductions.

The enzyme industry already from the eighties started to exploit the developments taking place in the fields of biotechnology and genetic engineering. The advantages offered by enzymes compared to conventional chemical catalysts rely on their biodegradability and their extreme selectivity. It is well known that natural enzymes are capable of catalysing reactions with up to 10^{17}-fold rate accelerations [1], and with exquisite control of regio- and stereochemistry. Such control makes the use of enzymes very attractive as an alternative to traditional catalysts in the synthesis of complex molecules, especially when chemical routes are difficult to implement [2, 3]. This meant a rapid and radical development in enzyme production technologies to improve the production efficiency and the product quality of enzymes through genetically modified organisms.

Enzymes have replaced classical reagents in numerous applications, *e.g.* proteases [4] and cellulases [5] in place of phosphates in the detergent sector, and amylases [6] and pectinases [7] instead of sodium hydroxide in the textile industry. The list of enzyme applications grows day by day, bringing with it a list of significant social and environmental benefits: enzymes, in fact, are biodegradable and come from renewable energy sources, they work under relatively mild conditions of pH and temperature and tend to offer selectivity in both reactant and product stereochemistry [8].The research of new, efficient and environmentally benign processes for several industrial purposes has increased interest in the 'green' catalysts.

Laccases represent one of these 'ecofriendly' enzymes. Their uses span from the textile to the pulp and paper industries, from food applications to bioremediation processes; therefore, laccases have received much attention in recent years.

Biotechnological applications of laccases

Laccases (EC 1.10.3.2, p-diphenol: dioxygen oxidoreductase) are part of a larger group of enzymes named multicopper oxidases, which includes among others ascorbic acid oxidase and ceruloplasmin.

Laccase is one of the very few enzymes that have been studied since the end of 19th century. In the 1883 Yoshida was the first to describe a laccase as a component of the Resin ducts of the lacquer tree *Rhus vernicifera* and in 1985 Bertrand characterised a laccase as a metal containing oxidase [9]. This makes laccase one of the earliest enzymes ever described. Although known for a long time, laccases attracted considerable attention only after the beginning of studies of enzymatic degradation of wood by white-rot fungi.

Laccases are commonly found in plants, where they participate in the radical-based mechanisms of lignin polymer formation [10] and in fungi, where laccases probably play several roles including morphogenesis, fungal plant-pathogen/host interaction, stress defense and lignin degradation [11].

Laccases are phenol-oxidases that catalyze the oxidation of a wide variety of substrates, comprising substituted phenols, polyamines and aryl diamines. The one-electron oxidation of the reducing substrate occurs concomitantly with the four-electron reduction of molecular oxygen to water. As these enzymes have low substrate specificity, they can oxidize a variety of substrates, either of natural or synthetic origin. In the presence of small molecules, capable to act as electron transfer mediators, laccases are also able to oxidise non-phenolic structures with redox potentials higher than their own [12].

It is well known that laccases have been applied to numerous processes such as: i) pulp delignification, where laccases can be used in the enzymatic adhesion of fibres in the manufacturing of lignocellulose-based composite materials such as fibreboards. Laccases have been proposed to activate the fibrebound lignin during manufacturing of the composites, thus, resulting in boards with good mechanical properties without toxic synthetic adhesives [13]; ii) textile industry, where the use of laccases is growing very fast, since besides to decolourise textile effluents as commented above, laccases are being used to bleach textiles and even to synthesise dyes [14]. Related to textile bleaching, in 1996 Novozyme (Novo Nordisk, Denmark) launched a new industrial application of laccase enzyme in denim finishing: DeniLite®, the first industrial laccase and the first bleaching enzyme acting with the help of a mediator molecule; iii) food industry, in which the potential applications of laccase [15] have been described recently in relation to bioremediation, beverage processing, ascorbic acid determination, sugar beet pectin gelation, baking and as a biosensor [16].

The white-rot fungus (WRF) *P. ostreatus* produces at least eight different laccase isoenzymes, five of which have been isolated and characterized [17-21]. Although the POXC protein (59-kDa with pI 2.7) is the most abundant in cultures, both, extra- and intracellular, the highest mRNA production was detected in POXA1b, which is probably mainly intracellular or cell wall-associated as it is cleaved by an extracellular protease [19, 21]. POXA1b and POXA1w isoenzymes exhibit a similar molecular weight around 61 kDa. Moreover, *P. ostreatus* produces two heterodimeric enzymes POXA3a and POXA3b consisting of large (61-kDa) and small (16- or 18-kDa) subunits. The production of laccase isoenzymes in *P. ostreatus* is regulated by the presence of copper and the two dimeric isoenzymes have only been detected in the presence of Cu ions [18, 22]. It should be noted that POXA1b exhibits atypical stability at alkaline pH. This is of interest for its use in the elimination of industrial effluents, often characterized by strongly alkaline pH values besides high concentrations of phenols. The oxidative enzymatic system of white-rot fungi is also capable of catalysing the degradation of various xenobiotic substances and dyes [23]. Generally, these xenobiotic compounds are removed slowly and tend to

accumulate in the environment. Due to the high degree of toxicity, their accumulation can cause severe environmental problems. A large number of these molecules, *e.g.* those structurally related to natural laccase substrates, are readily degraded or removed by microorganisms found in soil and water. The detoxification potential of white-rot fungi can be harnessed thanks to emerging knowledge of the physiology of these organisms as well as of the catalytic and stability characteristics of their enzymes.

Structural properties of laccases

Numerous laccase isoenzymes have been identified in many fungal species. Generally, more than one isoenzyme is produced in white-rot fungi. Until now, more than hundred laccases have been purified and characterised from fungi. It is possible to draw some general conclusions about laccases.

Typical fungal laccase is a monomeric protein of approximately 60–70 kDa with an acidic isoelectric point around pH 4.0 (Table I). It seems that there is considerable heterogeneity in the properties of laccases isolated from ascomycetes, especially with respect to molecular weight.

Laccases are glycoproteins like many fungal extracellular enzymes; the extent of glycosylation usually ranges between 10% and 25%. Glycosylation of fungal laccases is one of the main problems for their heterologous production. It was proposed that in addition to the structural role, glycosylation can also participate in the protection of laccase from proteolytic degradation [24].

Laccases contain four copper ions distributed into three sites, defined according to their spectroscopic properties. The catalytic ability of this enzyme family is guaranteed by the presence of these different copper centres in the enzyme molecule; based on spectroscopic analysis, which reflects geometric and electronic features, copper centres are differentiated as Type 1 Cu (T1), or blue copper centres, Type 2 (T2) or normal copper centres, and Type 3 (T3) or coupled binuclear copper centres [25-27] (**Fig. 2.1(a)-(b)**).

The T1 copper, which is the primary oxidation site, is characterized by a strong absorption around 600 nm, which gives rise to the typical blue colour of the copper oxidase. The T2 copper exhibits only weak absorption in the visible region and is electron paramagnetic resonance (EPR)-active, whereas the two copper ions of the T3 site are EPR-silent due to an antiferromagnetic coupling mediated by a bridging hydroxyl ligand (see **Fig.2.1**).

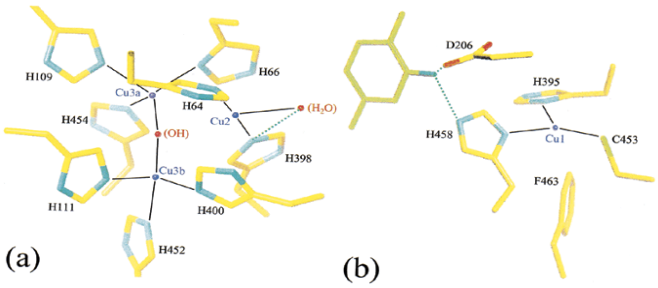


Figure 2.1: Laccase copper sites and their environment. (a) The trinuclear T2/T3 site. Copper coordinations are indicated with dark lines. (b) The T1 site and the neighbouring 2,5-xylidine ligand (green model). Image taken from reference [28].

Table I. Properties of fungal laccases [9].

Property	***n***	**Median**	**Q_{25}**	**Q_{75}**	**Min**	**Max**
Molecular weight (Da)	103	66 000	61 000	71 000	43 000	383 000
pI	67	3.9	3.5	4.2	2.6	6.9
Temperature optimum (℃)	39	55	50	70	25	80
pH optimum						
ABTS	49	3.0	2.5	4.0	2.0	5.0
2,6-Dimethoxyphenol	36	4.0	3.0	5.5	3.0	8.0
Guaiacol	24	4.5	4.0	6.0	3.0	7.0
Syringaldazine	31	6.0	4.7	6.0	3.5	7.0
K_M (μM)						
ABTS	36	39	18	100	4	770
2,6-Dimethoxyphenol	30	405	100	880	26	14 720
Guaiacol	23	420	121	1600	4	30 000
Syringaldazine	21	36	11	131	3	4307
k_{cat} (s^{-1})						
ABTS	12	24 050	5220	41 460	198	350 000
2,6-Dimethoxyphenol	12	3680	815	6000	100	360 000
Guaiacol	10	295	115	3960	90	10 800
Syringaldazine	4	21 500	18 400	25 500	16 800	28 000
n, number of observations; Q_{25}, lower quartile; Q_{75}, upper quartile.						

Substrates are oxidised close to the mononuclear site, and the electrons are transferred to the trinuclear site, where the molecular oxygen is reduced. Neither the electron transfer mechanism nor the oxygen reduction to water is fully understood. The reactio follows a radical type mechanism: substrates (phenols and aromatic or aliphatic amines) are oxidized by the T1 copper to produce radicals, which produce subsequently dimers, oligomers and polymers. The extracted electrons are probably transferred through a strongly conserved His-Cys-His tripeptide motif to the T2/T3 site, where molecular oxygen is reduced to water [27] (**Fig. 2.2**).

Kinetic data suggest a mechanism of reaction “two site ping-pong bi bi” type; in this mechanism the products are released before a new substrate molecule is ligated [29].

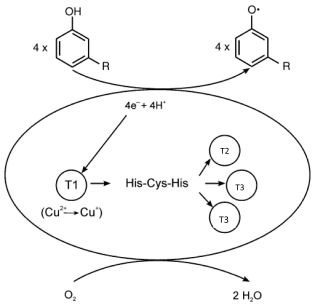


Figure 2.2: Catalytic cycle of laccase**.**

Current information about laccase structures is based on the 3D structures of the five fungal laccases mentioned above: *Coprinus cinereus* (in a copper type-2- depleted form) [30], *Trametes versicolor* [28, 31], *Pycnoporus cinnabarinus* [32]], *Melanocarpus albomyces* [33] and *Rigidoporus lignosus* [34], the last four enzymes with a full complement of copper ions. Generally laccase is a monomer, organized in three sequentially arranged domains (**Fig. 2.3**) and the molecule has dimensions of about 65x55x45 $Å^3$. Each of the three domains presents a similar β-barrel type architecture, related to that of small blue copper proteins such as azurin or plastocyanin. Domain 1 comprises two four-stranded β-sheets and four 3_{10}-helices. Three of the 3_{10}-helices are in connecting peptides between the β-strands, and one is in a segment between domain 1 and 2. The second domain has one six-stranded and one five-stranded β-sheet, and like in domain 1, there are three 3_{10}-helices in peptides connecting individual β-strands and domains 1 and 3, respectively. A 3_{10}-helix between domains 2 and 3 forms part of a 40-residue-long extended loop region. Finally, domain 3 consists of a β-barrel formed by two five-stranded β-sheets and a two-stranded β-sheets that, together with a α-helix and a β-turn, form the cavity in which the type-1 copper is located. The trinuclear copper cluster (T2/T3) is embedded between domains 1 and 3 with both domains providing residues for the coordination of the coppers. The third domain has the highest helical content with one 3_{10}-helix and two α-helices located in the connecting regions between the strands of the different β-sheets. Finally, at the C-terminal of domain 3, three sequentially arranged α-helices complete the fold. An α-helix formed by 13 residues at the C-terminal end is stabilized by a disulfide bridge to domain 1 (Cys-85–Cys-488), and a second disulfide bridge (Cys-117–Cys-205) connects domains 1 and 2. Both N-terminal and C-terminal amino acids benefit from hydrogen bonding networks to the rest of the protein. Despite the amount of information on laccases as well as on other blue multicopper oxidases, neither the precise electron transfer pathway nor the details of dioxygen reduction in blue multicopper oxidases are fully understood [34].

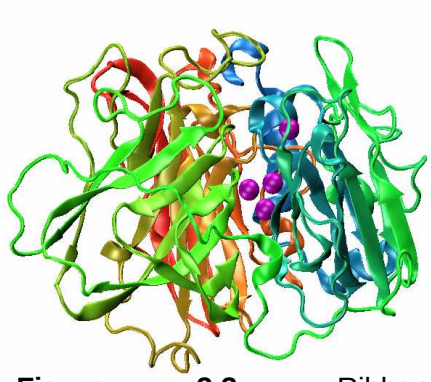


Figure 2.3: Ribbon representation of *T. versicolor* laccase X-ray structure. Domain 1 is showed in red, domain 2 in green and domain 3 in bleu. Copper ions are show in vdW representation [31].

The aim of the present work

Evolutionary design approaches have devoted considerable attention in modifying native proteins. The generation of stable enzymes with improved or novel catalytic activities is a fascinating topic of modern protein biochemistry. This goal is relevant for basic research purposes as well as for applications in biotechnology. In addition, the search of new biocatalysts can have a major impact on the applications of enzymes in industrial processes. Enzyme engineering can, for instance, force enzymatic reactions to proceed in a desired direction, enhance their selectivity and their stability. Two different strategies can be applied to change enzymatic properties at wish. The first one is the directed evolution, which requires the gene (or genes) of interest, but does not require a detailed knowledge of structure and function of the coded protein. The second strategy is rational design, which is the planned redesign of the protein sequence by site-directed mutagenesis. In this case, the design of a new protein requires the knowledge of both structure and sequence as well as the

mechanism of action of the enzyme. Selected residues are targeted for site-directed mutagenesis and after expression and purification the properties of the new enzymes are assessed by comparison with those of the native protein. Further residues may be targeted in further rounds of site-directed mutagenesis. This strategy, which has been applied extensively with variable success, depends on detailed structural and mechanistic information on the parent enzyme. Site-directed mutagenesis has been successful, for example, in redesigning the substrate specificity of a large number of common classes of enzymes, such as oxidoreductases (dehydrogenases and reductases) [35], hydrolases [36], transferases and restriction enzymes [37].

The development of new bio-based processes using laccases requires deeper understanding of the structure/function relationship of native enzymes and the design of novel an improved molecules that are better suited for industrial applications.

The aim of this part of my thesis was to model and to characterize molecular determinants in the mechanism of functioning of laccases form *P. ostreatus.*

Rational mutagenesis of laccases POXA1b and POXC from *Pleurotus ostreatus*

This section is aimed at the characterization of molecular determinants in the function of laccases, and at the development and characterization of new laccases using *P. ostreatus* isoenzymes (POXA1b and POXC) as templates for rational mutagenesis.

Results and Discussion

Sequence alignment analysis

Several laccase genes have been isolated and characterized, and in some cases, the sequences have been deposited in the appropriate gene register. The comparison of primary protein sequences of *P. ostreatus* laccases, whose structures are unknown, with any sequences of basidiomycetes laccases whose structures are known, were performed using PRALINE program [38] **(Fig. 2.4)**. As shown below, all fungal laccases contain several highly conserved ungapped regions, distributed almost throughout the entire length of the proteins. In particular, four ungapped sequence regions (1–4) that are clustered around the catalytic regions can be considered as the overall sequence fingerprint that can be used to identify the laccases, distinguishing them from the broader class of multi-copper oxidases.

```
TvL1KYA          .GIGPVADLT ITNAAVSPDG F.SRQAVVVN GG...........TPGPLIT
TvL1GYC          .AIGPAASLV VANAPVSPDG F.LRDAIVVN GV...........FPSPLIT
POXC             .AIGPAGNMY IVNEDVSPDG F.ARSAVVAR SVPATDPTPATASIPGVLVQ
POXA1b           ASIGPRGTLN IANKVIQPDG F.YRSTVLAG G...........SYPGPLIK
RlG1V10          ATV..ALDLH ILNANLDPDG TGARSAVTAE G...........TTIAPLIT
CcL1A65          QIVNSVDTMT LTNANVSPDG F.TRAGILVN G...........VHGPLIR
POXA3            ..ATKKLDFH IRNDVVSPDG F.ERRAITVN GI...........FPGTPVI

                                                          Region 1
TvL1KYA          GNMGDRFQLN VIDNLTNHTM LKSTSIHWHG FFQKGTNWAD GPAFINQCPI
TvL1GYC          GKKGDRFQLN VVDTLTNHTM LKSTSIHWHG FFQAGTNWAD GPAFVNQCPI
POXC             GNKGDNFQLN VVNQLSDTTM LKTTSIHWHG FFQAGSSWAD GPAFVTQCPV
POXA1b           GKTGDRFQIN VVNKLADTSM PVDTSIHWHG LFVKGHNWAD GPAMVTQCPI
RlG1V10          GNIDDRFQIN VIDQLTDANM RRATSIHWHG FFQAGTTEMD GPAFVNQCPI
CcL1A65          GGKNDNFELN VVNDLDNPTM LRPTSIHWHG LFQRGTNWAD GADGVNQCPI
POXA3            LQKNDKVQIN TINELTDPGM RRSTSIHWHG LFQHKTSGMD GPSFVNQCPI

                                                Region 2
TvL1KYA          SSGHSFLYDF QVPDQAGTFW YHSHLSTQYC DGLRGPFVVY DPNDPAADLY
TvL1GYC          ASGHSFLYDF HVPDQAGTFW YHSHLSTQYC DGLRGPFVVY DPKDPHASRY
POXC             ASGDSFLYNF NVPDQAGTFW YHSHLSTQYC DGLRGPFVVY DPSDPHLSLY
POXA1b           VPGHSFLYDF EVPDQAGTFW YHSHLGTQYC DGLRGPLVVY SKNDPHKRLY
RlG1V10          IPNESFVYDF VVPGQAGTYW YHSHLSTQYC DGLRGAFVVY DPNDPHLSLY
CcL1A65          SPGHAFLYKF TPAGHAGTFW YHSHFGTQYC DGLRGPMVIY DDNDPHAALY
POXA3            PPNSTFLYDF DTAGQTGNYW YHSHLSTQYC DGLRGSFIVY DPNDPLKHLY

TvL1KYA          DVDNDDTVIT LVDWYHVAAK ...L...GP. AFPLGADATL INGKGRSPST
TvL1GYC          DVDNESTVIT LTDWYHTAAR ...L...GP. RFPLGADATL INGLGRSAST
POXC             DIDNADTVIT LEDWYHIVAP ...Q...NA. AIPT.PDSTL INGKGRYAGG
POXA1            DVDDESTVLT VGDWYHAPSL ...S...LT. GVP.HPDSTL FNGLGRSLNG
RlG110           DVDDASTVIT IADWYHSLST ...VLFPNPN KAPPAPDTTL INGLGRNSAN
CcLA65           DEDDENTIIT LADWYHIPAP ...SI..... QGAAQPDATL INGKGRYVGG
POA3             DVDDESTIIT LADWYHDLAP HAQNQFFQTG SVPI.PDTGL INGVGRFKGG

TvL1KYA          TTA.DLSVIS VTPGKRYRFR LVSLSCDPNY TFSIDGHNMT IIETDSINTA
TvL1GYC          PTA.ALAVIN VQHGKRYRFR LVSISCDPNY TFSIDGHNLT VIEVDGINSQ
POXC             PTS.PLAIIN VESNKRYRFR LVSMSCDPNF TFSIDGHSLL VIEADAVNIV
POXA1b           PAS.PLYVMN VVKGKRYRIR LINTSCDSNY QFSIDGHTFT VIEADGENTQ
RlG1V10          PSAGQLAVVS VQSGKRYRFR IVSTSCFPNY AFSIDGHRMT VIEVDGVSHQ
CcL1A65          PAA.ELSIVN VEQGKKYRMR LISLSCDPNW QFSIDGHELT IIEVDGELTE
POXA3            PLV.PYAVIN VEQGKRYRFR LIQISCRPFF TFSIDNHTFD AIEFDGIEHD
```

```
TvL1KYA          PLVVDSIQIF AAQRYSFVLE ANQAVDNYWI RANPNFGNVG ....FTGGIN
TvL1GYC          PLLVDSIQIF AAQRYSFVLN ANQTVGNYWI RANPNFGTVG ....FAGGIN
POXC             PITVDSIQIF AGQRYSFVLT ANQAVDNYWI RANPNLGSTG ....FVGGIN
POXA1b           PLQVDQVQIF AGQRYSLVLN ANQAVGNYWI RANPNSGDPG ....FENQMN
RlG1V10          PLTVDSLTIF AGQRYSVVVE ANQAVGNYWI RANPSNGRNG ....FTGGIN
CcL1A65          PHTVDRLQIF TGQRYSFVLD ANQPVDNYWI RAQPNKGRNG LAGTFANGVN
POXA3            PTPAQNIDIY AAQRASIIVH ANQTIDNYWI RAPLTGGNPA GNPNLDISLI

TvL1KYA          SAILRYDGAA AVEPTTTQTT STAPLNEVNL HPLVATAVPG SPVAGGVDLA
TvL1GYC          SAILRYQGAP VAEPTTTQTT SVIPLIETNL HPLARMPVPG SPTPGGVDKA
POXC             SAILRYAGAT EDDPTTTSST ST.PLLETNL VPLENPGAPG PPVPGGADIN
POXA1b           SAILRYKGAR SIDPTTPEQN ATNPLHEYNL RPLIKKPAPG KPFPGGADHN
RlG1V10          SAIFRYQGAA VAEPTTSQNS GTA.LNEANL IPLINPGAPG NPVPGGADIN
CcL1A65          SAILRYAGAA NADPTTSANP NPAQLNEADL HALIDPAAPG IPTPGAADVN
POXA3            RAILRYKGAP AVEPTTVATT GGHKLNDAEM HPIAQEG.PG NLGTGPPDMA

TvL1KYA          INMAFNFNGT NF..FINGAS FTPPTVPVLL QIISGAQNAQ DLLPSGSVYS
TvL1GYC          LNLAFNFNGT NF..FINNAS FTPPTVPVLL QILSGAQTAQ DLLPAGSVYP
POXC             INLAMAFDFT TFELTINGVP FLPPTAPVLL QILSGASTAA SLLPSGSIYE
POXA1b           INLNFAFDPA TALFTANNHT FVPPTVPVLL QILSGTRDAH DLAPAGSIYD
RlG1V10          LNLRIGRNAT TADFTINGAP FIPPTVPVLL QILSGVTNPN DLLPGGAVIS
CcL1A65          LRFQLGFSGG RF..TINGTA YESPSVPTLL QIMSGAQSAN DLLPAGSVYE
POXA3            ITLNIAQPNP PF.FDINGIS YLSPSVPVLL QMLSGARKPQ DFLPSEQVII

                                        Region 3
TvL1KYA          LPSNADIEIS FPATAAAPGA PHPFHLHGHA FAVVRSAGST VYNYDNPIFR
TvL1GYC          LPAHSTIEIT LPATALAPGA PHPFHLHGHA FAVVRSAGST TYNYNDPIFR
POXC             LEANKVVEIS MP..ALAVGG PHPFHLHGHT FDVIRSAGST TYNFDTPARR
POXA1b           IKLGDVVEIT MP..ALVFAG PHPIHLHGHT FAVVRSAGSS TYNYENPVRR
RlG1V10          LPANQVIEIS IPG.....GG NHPFHLHGHN FDVVRTPGSS VYNYVNPVRR
CcL1A65          LPRNQVVELV VP..AGVLGG PHPFHLHGHA FSVVRSAGSS TYNFVNPVKR
POXA3            LPANKLIEVS IPG.....AG AHPFHLHGHT FDIVRTSNSD VVNLVNPPRR
                                                 Region 4
TvL1KYA          DVVSTGTPAA GDNVTIRFRT DNPGPWFLHC HIDFHLEAGF AVVFAE...D
TvL1GYC          DVVSTGTPAA GDNVTIRFQT DNPGPWFLHC HIDFHLEAGF AIVFAE...D
POXC             DVVNTGT.GA NDNVTIRFVT DNPGPWFLHC HIDWHLEIGL AVVFAE...D
POXA1b           DVVSIGDDPT .DNVTIRFVA DNAGPWFLHC HIDWHLDLGF AVVFAE...G
RlG1V10          DVVSIG..GG GDNVTFRFVT DNPGPWFLHC HIDWHLEAGL AVVFAE...D
CcL1A65          DVVSLGV..T GDEVTIRFVT DNPGPWFFHC HIEFHLMNGL AIVFAE...D
POXA3            DVLPIN.... GGNTTFRFFS GNSGAWFLHC HIDWHLEAGL AVVFAERPAE

TvL1KYA          IPDVASANPV PQAWSDLCPT YDARDPSDQ. .......... .....
TvL1GYC          VADVKAANPV PKAWSDLCPI YDGLSEANQ. .......... .....
POXC             VTSISA..P. PAAWDDLCPI YNALSDNDKG ......GIVP S....
POXA1b           VNQTAAANPV PEAWNNLCPI YNSSNPSKLL MGTNAIGRLP APLKA
RlG1V10          IPNIPIANAI SPAWDDLCPK YNANNPDS.. ......GLA. .....
CcL1A65          MANTVDANNP PVEWAQLCEI YDDLPPEA.. ..TSIQ.TVV .....
POXA3            VNEGEQAQIV TQDWRTLCPA YDGLAPEFQ. .......... .....
```

Figure 2.4: Sequence alignment performed using the PRALINE program [38]. In purple are highlighted C-terminal regions of POXA1b and POXC laccases. The Asp residues highly conserved in all laccases are in blue and in red are shown the conserved histidine and cysteine**.**

Moreover, it is possible to observe that the copper-ligating residues in laccases are present in regions that are conserved across all analyzed laccases. A total of 11 histidine residues and 1 cystein residues are conserved across all multi-copper oxidases and are thought to serve as the copper ligands.

T. versicolor has a 459Leu-Glu-Ala461 sequence, typical for the high E^0 laccases [39], whereas CcL has Leu-Met-Asn in the equivalent position. Glu-460 forms a hydrogen bond (H-bond) between and Ser-113, the latter being situated in the opposite domain 1. In particular, this serine residue is presumably forced by the H-bond into an unfavourable main chain conformation. As a consequence of the attractive H-bond interaction, the whole helix, which contains His-458, is pulled toward domain 1, thereby increasing the Cu-N distance that seems to be responsible for the high E^0 of this laccase (800 mV) [31] (**Fig 2.5**). In POXA1b the positions corresponding to Ser-113 and Glu-460 are taken by a glycine and an asparagine residue, respectively. These two residues cannot form an H-bond; a similar situation is found also in CcL (*Coprinus cinereus* laccase), a laccase with a redox potential of 550 mV [30].

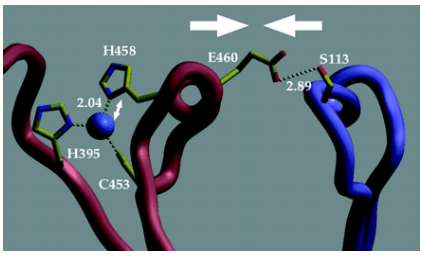


Figure 2.5: The hydrogen bond between Glu-460 and Ser-113 which would subsequently cause an elongation of the Cu1-N (His-458) bond at the T1 site. The bond lengths are given in Angstroms. Image taken form [31].

Moreover, an unusual protruding extension present in the C-terminal region, normally absent in other laccases, has been identified in POXC (six amino acids extra) and in POXA1b (sixteen amino acids extra).

3D models of POXA1b and POXC **(Fig. 2.6)** were generated using *T. versicolor* laccase (PDB code 1GYC) that exhibits about 60% identity with both enzymes. To build the last sixteen residues of POXA1b, the coordinates of the reported C-terminal protruding extension (available only for the ascomycete laccase from *M. albomyces PDB code* 1GWO) were used [33]. Hakulinen *et al.* (2000) proposed that proteolytic cleavage of the C-terminus allows the opening of the tunnel for the entrance of the oxygen molecule and the exit of water [33]. Moreover, by replacing 11 amino acids at the C-terminus with a single cysteine residue, Gelo-Pujic *et al.* (1999) managed to change the redox potential of the type-1 Cu of laccase from *T. versicolor*, produced in *Pichia pastoris*, [40].

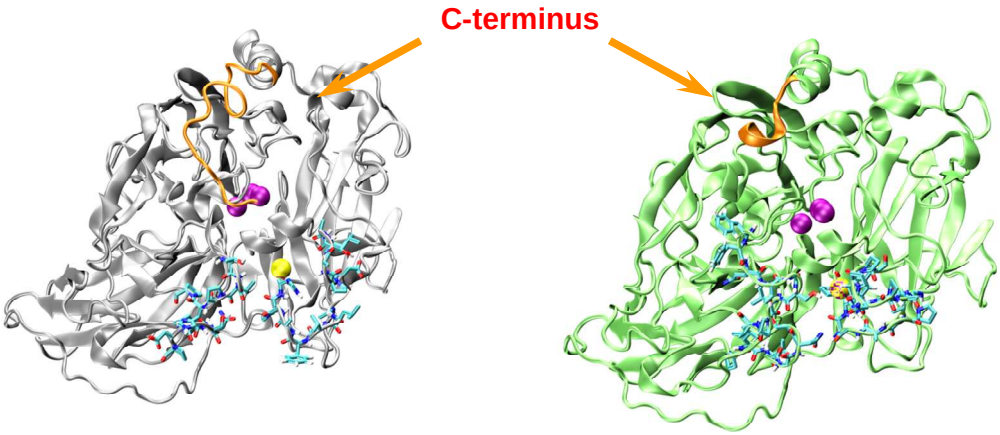


Figure 2.6: POXA1b (right) and POXC (left) models

From all these observations it seems reasonable to suspect that C-terminus may play a key role in the function of the laccase enzyme.
In order to investigate the role of the C-terminus in *P. ostreatus* laccases, site-directed mutagenesis experiments have been performed in this region. Mutants of C-terminal truncated laccases were designed: particularly POXA1bΔ4 and POXA1bΔ16 and POXCΔ6 have been created.
It is widely accepted that in an evolutionarily related family of proteins fewer mutations occur in the functionally important residues. Many of the residues of catalytic regions are conserved in the analyzed laccases. The study of the 3D models of POXA1b and POXC has allowed for the identification of those residues constituting the catalytic site and involved in the interaction with substrates. The crystal structure of *T. versicolor* with 2,5-xylidine revealed two important residues for the interaction between the amino group of the reducing substrate and the enzyme: the first is a histidine residue (458) that also coordinates the T1 copper and acts as the primary electron acceptor from the substrate. This His458 is highly conserved among all laccases. The second, aspartate 206, is hydrogen bonded via the terminal oxygen of its side chain to the amino group of 2,5-xylidine. Asp206 appears relatively buried in the cavity, surrounded by hydrophobic residues which lie in a more exposed region [28]. The comparison of 3D models of POXA1b and POXC with POXA3 models and docking with violuric acid, a phenolic substrate (**Fig. 2.6**), show that the arginine present in POXA3 could generate a repulsion that could lead to a decrease in the affinity for the substrate.

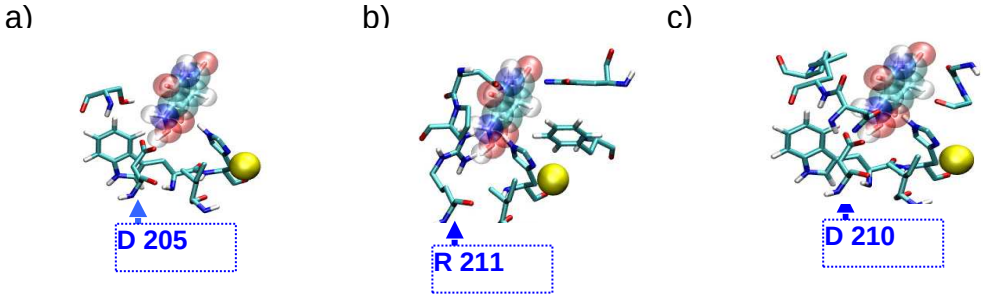


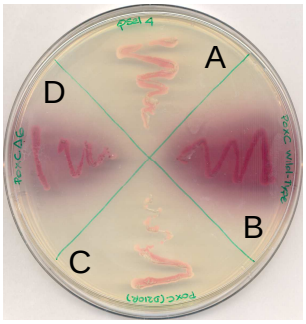
Figure 2.6: Docking with violuric acid, the T1 Copper ion (yellow) and the violuric acid are shown in vdW representation. In licorice representation are highlighted the residues in close contact with the substrate for a) POXA1b, b) POXA3 and c) POXC**.**

The Aspartic Acid, negatively charged, could stabilize the radicalic cation formed during the catalytic reaction; indeed the Arginine cannot stabilize the radicalic intermediate being positively charged. To characterize the molecular determinants involved in the interaction with substrates, mutants of POXA1b (D205R) and POXC(D210R) have been designed.
The cDNAs of these two mutants have been introduced into the expression vectors by using host homologous recombination system in the yeast *S. cerevisiae.*

POXC mutants: a qualitative evaluation

The recombinant expression represents a good strategy in order to obtain sufficient recombinant protein for the characterization analyses. For the heterologous expression of POXC and of its mutants POXCΔ6 and POXC (D210R), the expression vector pSAL4 has been used. This plasmid is characterized by the presence of the inducible promoter of metallothionein (CUP1) from *S. cerevisiae* and of the signal peptide of *P. ostreatus*.

Production of recombinant laccases was first assayed on plate using ABTS (2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid)) as substrate as shown in Figure 2.7. The purple halo obtained from the oxidation of substrate proves the presence of laccase activity. The marked difference of the halo intensity has led to the supposition that the mutants are less active than POXC. In particular, as shown in the Figure 2.7, mutant POXCΔ6 turns out to be less active than the wild-type protein, while mutant POXC(D210R) seems to be inactive.



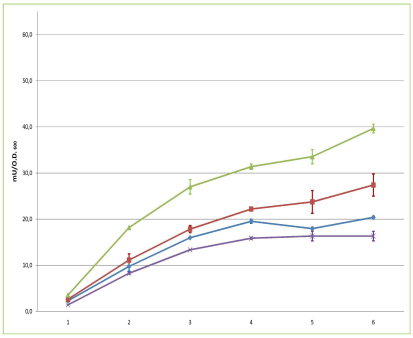
However, when a similar analysis was conducted on liquid culture, activity production was too low to allow for the purification and subsequent characterization of these new laccases.

Figure 2.7 : Plate assay of POXC wild-type and mutant laccases, using ABTS. (A) pSAL4; (B) pSAL4+*poxC*; (C) pSAL4+*poxC(D210R)*; (D) pSAL4+*poxCΔ6*.

POXA1b

In order to obtain heterologous expression of recombinant wild-type POXA1b and of its mutant proteins (POXA1bΔ4, POXA1bΔ16 and POXA1b(D205R)), the respective cDNAs were cloned in psal4 (see Methods) under the control of a strong promoter that is inducible by Copper. The yeast *S. cerevisiae* was transformed with each construct and cellular growth was obtained in selective medium (SD).

Optical density (O.D.) and laccase activity production measurements were performed during the time course of the preparative culture. Mutants and wild-type expressing yeast growth curves did not differ significantly (data not shown).



The analysis of the activity production has highlighted a higher specific activity production for the mutants POXA1bΔ4 and POXA1bΔ16 than for the wild-type protein; this could be attributed to a more effective enzymes production for these mutants. Instead the specific activity production of POXA1b(D205R) is comparable to that of the wild-type protein (**Fig 2.8**).

Figure 2.8: Laccases' specific activity production in preparative cultures with ABTS; POXA1b (blue), POXA1bΔ4 (red), POXA1bΔ16 (green) and POXA1b(D205R) (purple).

Purification of wild-type and mutant laccases

Culture broths collected at the 6th growth day were used for laccase purifications. The recombinant POXA1b protein and the three mutants (POXA1bΔ16, POXA1bΔ4 and POXA1b(D205R)) were purified by ammonium sulphate selective precipitation followed by gel filtration chromatography Superdex 75 prep grade as described in Materials and Methods. Similar amounts of total protein were obtained in all cases. The active fractions were pooled and concentrated on an Amicon PM-30 membrane.

In Table II the yields after protein purification are given, obtained following the described procedures (see Materials and Methods).

Table II: Wild type and mutant proteins purification from *S. cerevisiae* cultures.

	Purification step	Total activity (U)	Total protein (mg)	Specific activity (U/mg)	Recovery (U%)	Purification fold
POXA1b Wild-type	Broth	170	471.6	0.36	100	1
	$(NH_4)_2SO_4$* and SEC	100	5.8	17.2	59	47
POXA1b Δ4	Broth	225	509.6	0.44	100	1
	$(NH_4)_2SO_4$* and SEC	135	6.3	21.6	60	49
POXA1b Δ16	Broth	107	527.4	0.20	100	1
	$(NH_4)_2SO_4$* and SEC	66	5.8	11.3	62	56
POXA1b (D205R)	Broth	128	463.3	0.28	100	1
	$(NH_4)_2SO_4$* and SEC	82	9.33	8.8	64	31

*Laccase activity was recovered from supernatant of 100% ammonium sulphate precipitation

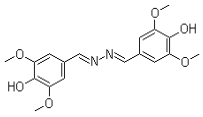
Kinetic and catalytic parameters determination

Some kinetic and catalytic parameters for three different substrates : 1) 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) at pH 3; 2) 2,6-dimethoxyphenol (DMP) at pH 5; 3) syringaldazine (SGZ) at pH 6) were determined; they are reported in Table III.

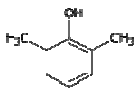
Table III: Kinetic and catalytic parameters of wild

Laccases	**K_M (mM)**	**Specific activity (U/mg)**	**Catalytic efficiency (U/mg*mM)**
	ABTS		
Wild-type	0.16±0.01	37±1	231±2
POXA1bΔ4	0.11±0.01	23±2	209±3
POXA1bΔ16	0.14±0.02	9±1	63±1
POXA1b (D205R)	0.33±0.05	9±1	27±1
	DMP		
Wild-type	0.57±0.01	18±1	32±1
POXA1bΔ4	0.43±0.04	22±2	52±2
POXA1bΔ16	0.43±0.05	15±1	35±2
POXA1b (D205R)	1.26±0.07	0.4±0.2	0.3±0.1
	SGZ		
Wild-type	0.04±0.01	2,0±0.8	49±1
POXA1bΔ4	0.03±0.01	2,3±0.5	77±1
POXA1bΔ16	0.03±0.01	1,6±0.5	53±1
POXA1b (D205R)	N.D.	-------	-------

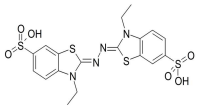
ABTS structure



DMP structure



SGZ structure



In general only slight differences have been observed in the Michaelis-Menten constant (K_M), toward all substrates, of mutants POXA1bΔ4, POXA1bΔ16 compared with that of the wild-type protein. On the other hand, the K_M value of POXA1b (D205R) toward ABTS and 2,6-DMP as substrates is 2-fold the one of the wild-type protein toward the same substrates. This reduced affinity for both substrates could be attributed to the introduced mutation. This mutation (Asp205 → Arg), located in the active site, could cause an alteration of the substrate-enzyme complex stabilization. This supports the hypothesis of a key role played by this aspartate residue in the interaction with the substrate.

Effect of pH and Temperature on laccase activity

The effect of pH on the activity of laccases towards DMP has been studied in the pH range 3,0÷6,4. In general, a bell-shaped profile of the activity as a function of pH has been observed for all the tested laccases. The optimum reaction pH of laccases for DMP oxidation was exceptionally high around pH values of 5.4-5.6.

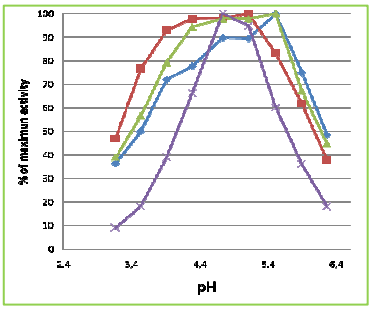


Figure 2.9: Effect of pH on the activity of wild-type and mutated laccases. POXA1b (blue), POXA1bΔ4 (red), POXA1bΔ16 (green) and POXA1b(D205R) (purple).

In particular, both POXA1bΔ16 and POXA1bΔ4 show a broader range of optimum pH than the wild-type protein. The mutant POXA1b(D205R) shows its maximum activity at lower value of pH (~ 5) (**Fig 2.9**).

Optimal reaction temperature was determined by varying the reaction temperature, by 10℃ increments, in the range 20–70℃.

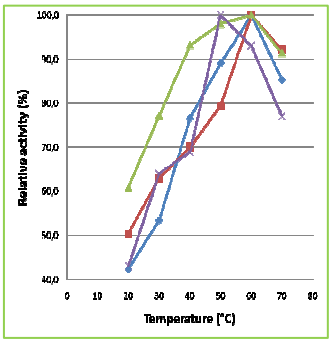
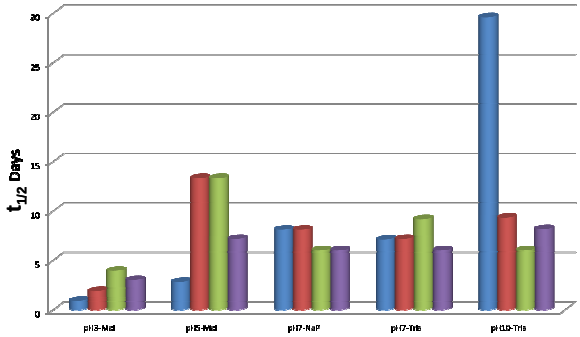


Figure2.10: Effect of temperature on the activity of wild type and mutated laccases. POXA1b (blue), POXA1bΔ4 (red), POXA1bΔ16 (green) and POXA1b(D205R) (purple).

POXA1b and the truncated mutants show their maximum activity at 60℃, whereas POXA1b (D205R) at lower temperature (50℃) (**Fig. 2.10**).

Stability as function of pH and temperature



The laccase stability at different pH values has been examined by pre-incubating the enzymes at various pH values varying from 3.0 to 10.0 and by measuring the residual activity under standard assay conditions.

As shown in Figure 2.11, the mutant POXA1b(D205R) exhibits a similar trend under all examined

Figure 2.11: Stability of POXA1b (blue) and mutated laccases POXA1bΔ4 (red) and POXA1bΔ16 (green) and POXA1b(D205R) (purple) at different pH values.

conditions, except for pH 3, which is an extreme condition for all the examined proteins. Therefore, the pH value seems not to influence strongly the protein stability. The generally low stability of POXA1b(D205R) could be attributed to a conformational change in the highly conserved region due the presence of the mutated residue with a positive charge (Asp-Arg).
An unusual behavior was observed for the truncated mutants. In particular, at pH 5 both show a $t_{1/2}$ values more than 3-fold with respect to that of POXA1b, in addition at alkaline pH the mutants loose drastically the high characteristic stability of POXA1b (**Figure 2.11** and **Table IV**). At neutral pH POXA1b and its mutants are similarly stable.
These data suggest that C-terminal tail could be essential to stabilize the laccase structure at alkaline pH. In fact is well known that POXA1b is one of most stable laccases at this extreme pH and it is the only one that presents a protruding C-terminal.

Table IV: Protein half life values at different pH values

	$t_{1/2}$ (Days)				
Laccases	pH3, McIlvaine buffer	pH5, McIlvaine buffer	pH7, NaP buffer	pH7, Tris-HCl buffer	pH10, Tris-HCl buffer
Wild-type	0.9	2.8	8.1	7.1	29.6
POXA1bΔ4	1.9	13.3	8.1	7.2	9.3
POXA1bΔ16	3.9	13.3	6.0	9.2	6.0
POXA1b(D205R)	3.0	7.2	6.0	6.0	8.2

In order to determine their thermostability, all enzymes were incubated at 60° at neutral pH and cooled down to room temperature prior to measuring the residual activity using ABTS as substrate.
The enzymes appeared to be reasonably stable with respect to temperature; in particular POXA1bΔ4 displays similar stability compared with that of wild type (**Fig. 2.12** and **Table V**) whereas mutants POXA1b(D205R) and POXA1bΔ16 show a considerable loss of stability.

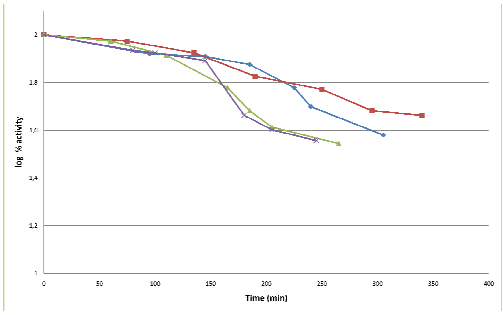


Figure 2.12: Stability of wild type and mutated laccases at 60℃. POXA1b (blue), POXA1bΔ4 (red), POXA1bΔ16 (green) and POXA1b(D205R) (purple).

Table V: Proteins half life values at 60℃

Laccases	**$t_{1/2}$ (hours)**
Wild-type	4.5
POXA1bΔ4	5.1
POXA1bΔ16	3.3
POXA1b (D205R)	3.3

The obtained results have highlighted that POXA1b(D205R) shows not only diminished affinity for the substrate (**Table III**) but also loss of stability under all tested condition. These data suggest that the residue Asp205 is crucial for the affinity to the substrates, and that substitutions occurring in the catalytic region influence the stability of the protein, probably as a result of a rearrangement of the phenolic-substrate cavity.

The truncated mutants exhibit a small dissimilarity in the Michaelis-Menten constant (K_M) but a considerable difference of stability with respect to the wild-type.

The unusual C-terminal extension seems to be directly responsible for the extraordinary stability of POXA1b at alkaline pH.

Materials and Methods

Strains, media and plasmids

The *Escherichia coli* strain Top 10 (F-mcrA D (mrr-hsdRMS-mcrBC) f80lacZDM15 DlacX74 deoR recA1 araD139 D (ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG) was used in all DNA manipulation. The medium used for growing *E. coli* was the rich medium Luria-Bertani (**LB**). Bacterial cultures were conducted at 37℃. Bacte rial cells competent to chemical transformation were obtained from cells grown on SOB medium.

LB

Bacto tryptone (Difco) 1% (*w/v*)
Yeast extract (Difco) 0.5% (*w/v*)
NaCl 1% (*w/v*)
$MgSO_4$

Selective medium was supplemented with 100 µg ml-1 of ampicillin; solid medium contained agar 15% (*w/v*) (Difco).

The *S. cerevisiae* strain used for heterologous expression of mutant laccases was W303-1A (MAT ade2-1, his3-11, 15, leu2-3, 112, trp1-1, ura3-1, can1-100). The medium used for growing wild-type *S. cerevisiae* was the rich medium

YPD

Yeast extract (Difco) 1% (*w/v*)
Bacto tryptone (Difco) 2% (*w/v*)
Glucose 2% (*w/v*)

Cultures were conducted at 28℃. Solid medium conta ined 15% (*w/v*) of agar (Difco). Plasmids used for cloning procedures were pUC13 (Amersham Biosciences) for the gene *poxc* and pUC18 (Amersham Biosciences) for the gene *poxa1b*.

Competent cells preparation and chemical transformation

E. coli TOP 10 cells were picked up from a solid culture and grown in 5 ml of SOB medium at 37° C for 16 hours on a rotary shaker (25 0 rpm). The culture was then diluted 1:100 in the same medium and grown up to an optical density at 600 nm of 0.6-0.7. Growth is stopped by putting the culture on ice for 20 min., and cells are sedimented by centrifugation at 3000 x g at 4℃ for 10 min. Cells were washed twice by incubation with cold $CaCl_2$ 0.1 M for 20 min on ice. Cell pellet was then resuspended in $CaCl_2$ 0.1 M, and Glycerol 15% was added to the suspension. After vigorous mixing, the suspension was aliquoted and promptly frozen in dry-ice/acetone bath. Competent cells was thawed and incubated with transforming DNA plasmid (up to 100 ng) for 30 min. on ice. Mixture was then shocked by incubation at 42℃ for 90 sec. and on ice for the same time. Cell s were spread on selective solid

LB medium after 1 hour incubation at 37 ℃ in LB me dium.

Gel electrophoresis of DNA

The PCR products were separated and analysed on agarose gel 1% (*w/v*). The electrophoresis was conducted in TAE (40 mM tris-acetate, 1mM EDTA pH 8) at 100 V. To visualize DNA 0.5 µg ml-1 ethidium bromide was added to the agarose gel. This substance contains a planar group that intercalates between the stacked bases of DNA. DNA fragments were visualized with ultraviolet light (365 nm).

PCR products and linear plasmid were extracted and purified from agarose gel with the kit Quiaex from Quiagen as specified by the manufacturer.

Mini-preparation of plasmid DNA (alkaline lysis method)

Bacterial cells grown on solid LB medium were inoculated at 37℃ over night in 3 ml of selective LB on a rotary shaker (250 r. p. m.). 1.5 ml of culture was centrifuged for 30 sec. at 13000 x g at room temperature. The cellular pellet was resuspended in 100 µl of solution 1 (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl pH 8). 200 µl of solution 2 (0.2 N NaOH, 1% SDS) were added to the cellular suspension. After a vigorous shaking, the solution was incubated on ice for 5 min to denaturate the cellular proteins. Then 150 µl of solution 3 (potassium acetate 5M pH 5.2, CH_3COOH 11.5ml, H_2O 28.5ml) were added, and the solution was incubated on ice for 5 min to allow the precipitation of genomic DNA. The suspension was centrifuged for 5 min at 13000 x g at room temperature. The supernatant, containing plasmidic DNA, was extracted with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1). The plasmidic DNA was then precipitated by adding 1ml of 100% EtOH at room temperature and rinsed with 200 µl of cold 70% EtOH. The plasmidic DNA was resuspended in 50 µl of TE 1X/RNase pH 8 (10 mM Tris HCl pH 8, 1 mM EDTA), RNAaseA (10 mg/ml) and incubated at 37℃ for 30 min .

Mini-preparation of plasmid DNA from S. cerevisiae [41]

Yeast cells were grown in 3 ml of liquid SD medium at 28℃ over night on a rotary shaker (150 r. p. m.). 1.5 ml of culture was harvested for 5 min. at 5000 x g at room temperature, and the pellet was re-suspended in 100 µl of cold STET (8% sucrose, 50 mM Tris-HCl pH 8, 50 mM EDTA, 5% Triton X-100). Cells were lysated by adding 0.2 g of 0.45 mm glass beads and vigorously vortexing. Following steps were: addition of another 100 µl of STET, briefly vortexing and incubation in a boiling water bath for 3 min. The suspension was then cooled on ice and centrifuged for 10 min. at 5000 x g at 4℃. To 100 µl of the supernatant was added 2.5 M ammonium acetate: the sample was incubated at – 20℃ for 1 h and cent rifuged for 10 min. at 12000 x g at 4℃. At this stage, residual chromosome DNA, lar ge RNA species, and the putative impurities which inhibit *E. coli* transformation are precipitated. To 100 µl of the supernatant were added 200 µl of ice-cold ethanol, and DNA was recovered by centrifugation. The pellet was washed with 70% ethanol and re-suspended in 20 µl of water. 10 µl of this solution were used to transform competent bacteria.

Transformation and growth of S. cerevisiae

S. cerevisiae transformation was done by using a lithium acetate protocol (Gietz et al. 1992). The cell pellet was then incubated in 1 ml of 0.1 M LiAc for 5 min at 30 ℃. Cells were sedimented by centrifuging at 12000 x g for 5 sec, resuspended in a mixture containing: 240 µl of PEG (50% w/v), 36 µl LiAc 1 M, 25 µl SS-DNA (2.0 mg/ml), transforming DNA plasmid (up to 5 µg), 45 µl of H_2O and incubated at 42° C

for 20 min. Cells were then centrifuged, resuspended in H_2O, and spread on selective medium: plates were incubated inverted for 4 days at 28℃ and checked for the development of green or orange colour.

Experiments of homologous recombination were performed following the previous protocol and transforming the yeast cells with different ratios (1: 1 and 1:3) of linearized plasmid (1.5 µg) and mutated cDNA.

Laccase production by transformants was first assayed on SD plates supplemented with 0.6 mM $CuSO_4$ and 0.2 mM ABTS or 0.1 mM 2,6-DMP. Inocula grown on selective **SD** medium were used to seed flasks (250 ml) containing 50 ml of selective medium, starting from 0.5 OD_{600}. Cultures were grown at 28℃ on a rotary shaker at 150 rpm. Preparative culture was done in 1-l flasks containing 200 ml of medium. Culture aliquots (1 ml) were daily collected, centrifuged at 12000 g for 2 min at 4℃ and assayed for optical density and laccase activity determination. Preparative cultures were performed in 1L flasks containing 200 ml of medium.

Transformed yeast was grown in a selective medium:

Selective medium SD

Yeast nitrogen base w/o AA (Difco) 6.7 g/l
Casaminoacids (Difco) 5 g/l
Glucose 2% (*w/v*)
Adenine 30 mg/l
Tryptophane 40 mg/l
Succinate buffer pH 5.3, 50 mM
Copper sulphate 0,6 mM

Homologous recombination

Expression of mutated laccases in *S. cerevisiae* was performed taking advantage of the high level of homologous recombination in yeast, which makes *in vivo* approaches fast, efficient, and non-mutagenic. The optimal length of the homologous overhangs is usually over 100 bp on both 5' and 3' –ends. Overhangs can be generated by restriction of the cDNAs on the basis of the restriction map of the recombinants vectors. This study was performed starting from the most promising construct for both cDNAs, C-pSAL4 and B-pSAL4. Appropriate enzymes must cut the cDNA once or twice (if a deletion is wanted) in the latter case leaving at least 40 bp on each edge, and must not cut into the plasmid. The B-pSAL4 construct was cut with SmaI and BglII restriction enzymes leaving 155 bp at 5'-end and 475 bp at 3'-end of the original *poxa1b* cDNA sequence. The overhangs will allow homologous recombination between the plasmid and the mutated cDNAs within the yeast cells allowing the closed plasmid to be maintained and propagated again. *poxa1b* cDNA was cloned in B-pSAL4 expression vector, digested with SmaI and BglII restriction enzymes, and *poxc* cDNA was cloned in C-pSAL4 expression vector, digested with AatII restriction enzyme, by using homologous recombination expression system of *S. cerevisiae.*

Mutant *poxa1b* inserts were cotransformed with open plasmid that contains sequences homologous to the ends of the poxa1b cDNAs on both ends. Homologous recombination combines these to form complete plasmids. Negative control was performed transforming the yeast cells only with the B-pSAL4 open plasmid carrying *poxa1b* cDNA termini. Yeast cells, where homologous recombination between B-pSAL4 plasmid and mutated cDNAs takes place will be able to grow on selective medium, since they carry out a closed plasmid.

Digestion with restriction enzymes

The PCR products, the cloning vectors and cDNAs were hydrolysed with opportune restriction enzymes (Promega). The reactions were performed using 5U of enzyme each µg of DNA incubating at restriction nuclease optimum temperature for 2 h and 30 min.

Ligation of cDNA in a plasmid

The ligation was performed with ligase from T4 bacteriophage (Boehringer Mannheim) at 16℃ over night. A maximum of 100 ng of plasmid was used for each reaction, and several molecular ratio plasmid: insert were tested (1:2, 1:3, 1:4), in a mix containing 1mM ATP, ligase buffer 1x and 1 U of enzyme.

Polymerase chain reaction (PCR)

To insert the cDNA of truncated *poxc* into pSAL4, was amplified in the presence of an upstream primer Fwpoxc using pUC13 containing *poxc* cDNA, as template, downstream primer RevpoxcΔ6 contain restriction sites for EcoRI and HindIII have been used to remove last eighteen nucleotides. To gain *poxa1bΔ4* and *poxa1bΔ16*, cDNAs of *poxa1b* were amplified in the presence of an upstream primer Fwpoxa1b, using pUC18 containing *poxa1b* cDNA, as template. Different downstream primers contain restriction sites for EcoRI and XhoI were used to remove last twelve and forty-eight nucleotides to obtain the two truncated genes (Table I). To obtain *poxa1b (D205R)* and *poxa1b (D210R)*, a point mutation was chosen so that the new codon fitted the peculiar *S. cerevisiae* codon. The PCR methodology has been used to gain a site-directed mutagenesis

Table VI: List of the primers used in amplification experiments.

Fwpoxc	TTTGAATTCAAGCTTATGTTTCCAGGCGCACGG
Revpoxc Δ6	TTATCTAGAAAGCTTTCATTGCGTTGTCG
Fwpoxa1b	TGTTGCAGATCTTGTCGG
Revpoxa1b Δ4	TAAGCATCCCTCGAGTCACGCAGGCAGACG
Revpoxa1b Δ16	TAAGGATCCCTCGAGTCAGAGTTTCGATGGG
Revmutpoxa1b(dr)	TTGATAGTTGGATCTGCAGGAAGT
FwA(dr)	ACTTCCTGCAGATCCAACTATCAA
Revmutpoxc(dr)	CGTGAAATTGGGTCTACAAGACAT
Revpoxc(dr)	ATGTCTTGTAGACCCAATTTCACG

Laccase purification from S. cerevisiae

Culture media were harvested on the optimal laccase production day (6° day), cells were sedimented by centrifugation at 1600 g at 4℃ for 15 min. Secreted proteins were filtered throw whatmann sheet. The sample was concentrated with Amicon stirred cells with ultrafiltration membranes with cut-off 30000 NMWLC. Then, most secreted proteins were precipitated by the addition of $(NH4)_2SO_4$ up to 100% saturation at 4℃ and centrifuged at 10000 g for 40 min. The supernatant from ammonium sulphate precipitation was loaded onto a 26/60 Superdex 75 prep grade (Amersham Biosciences) column equilibrated with 0,15M NaCl 50 Mm Na-phosphate

buffer pH7. The active fractions were pooled and concentrated on an Amicon PM-30 membrane.

Enzymatic assay

Culture aliquots (1 ml) were collected daily and cells were pelleted by centrifugation (12000 g for 2 min 4℃). Laccase activity in the culture supernatant was assayed at 25℃, monitoring the oxidation of ABTS spectrophotometrically at 420 nm (ε_{420}= 36 mM^{-1} cm^{-1}): the assay mixture contained 2 mM ABTS, 0.1 M Na-citrate buffer pH 3.0. Laccase activity towards 2,6-DMP was assayed in a mixture containing 1 mM 2,6-DMP and the McIlvaine's citrate-phosphate buffer adjusted to pH 5.3. Oxidation product was followed by an absorbance increase at 477 nm (ε_{477}= 14.8 mM^{-1} cm^{-1}). Laccase activity towards syringaldazyne (SGZ) was assayed in a mixture containing 0,1 mM SGZ and 0,2 M Na-phosphate buffer adjusted to pH 6.0. Oxidation of SGZ was followed by an absorbance increase at 526 nm (ε_{526}= 6.5 x 104 M-1 cm-1).

Plate assay was performed on solid selective medium supplemented with enzyme substrate, 0.2 mM ABTS.

Protein determination

Protein concentration was determined allow Bradford method and using commercial reactive of BioRad, with BSA (bovine serum albumin) as standard.

Kinetic parameters determination

Laccase activity in the culture supernatant was assayed utilizing increasing substrate concentrations (ABTS, 2,6-DMP and SGZ), until obtaining the complete saturation of the enzyme. K_M values were estimated using the software GraphPad Prism, on a wide range of substrate concentrations. The values of ΔA/min according to values of concentration of the substrate have been brought back in diagram and have been determine the parameters to you of the curve of Michaelis-Menten, for that substrate.

Effect of pH and temperature

The effect of pH on laccase activity towards 2,6-DMP was measured using a McIlvaine's citrate-phosphate buffer adjusted to different pH level in the range 3.0-6.2. The effect of temperature on laccase activity was measured in the temperature range 20-70℃ in 50 mM Na-phosphate buffer adjusted to pH 7.0.

Stability at pH and temperature

Phenol oxidase stability at 60℃ was measured in 50 mM Na-phosphate buffer adjusted to pH 7.0. Stability at pH values was measured using a McIlvaine's citrate-phosphate buffer adjusted at pH 3 and 5; 50 mM Na-phosphate buffer adjusted at pH 7.0; Tris-HCl buffer adjusted at pH 7 and 10.

Molecular Modelling and construction of enzyme substrate complexes

The structure of POXA1b and POXC were obtained by homology modelling from the crystal structure of *Trametes versicolor* (1GYC pdb entry) [31], with which it shares about 60% sequence identity. It is well known that a sequence identity threshold for obtaining a reasonable model is 30% [42]. The last sixteen residues

were modelled using the coordinates of the C-terminus of the crystal structure of the *M. albomyces* laccase (1GWO pdb entry). This choice is justified in consideration of the fact that the fold of the *M. albomyces* and of the *T. versicolor* laccases is very similar, in particular in the C-terminal region the fold is almost identical. 3D models were generated using the SWISS-MODEL web server by means of the project mode option that allow to select the template and control the gap placement in the alignment [43].

To obtain a reasonable starting model of the complexes between laccases models and violuric acid, the crystal structure of *T. versicolor* laccase complexed with an arylamide [28] (PDB code: 1KYA) was chosen as template based on the high sequence identity (60%) of proteins. The backbone and the common residues of laccase models were superimposed onto that of *T. versicolor* laccase. The violuric acid was also superimposed onto the 2,5-xylidine and then was merged with structure of each models.

Development of new laccases by directed evolution: functional and computational analyses

The aim of this section is to integrate the results of the generated mutants by Festa G. (PhD thesis 2006, University of Naples “Federico II”) through structural and computational analyses to investigate some of the reasons, at a molecular level, for the enhanced activity shown by specific mutations.

Development of new laccases by directed evolution: functional and computational analyses

Giovanna Festa[1], Flavia Autore[1,2], Franca Fraternali[2], Paola Giardina[1*], and Giovanni Sannia[1]

[1]Dipartimento di Chimica Organica e Biochimica, Università di Napoli "Federico II", Complesso Universitario Monte S. Angelo, via Cintia, 80126 Napoli, Italy

[2]King's College, Randall Division of Cell and Molecular Biophysics, New Hunt's House, SE1 1UL, London, U.K.

Materials and Methods

Strains, media and plasmids.

The *Escherichia coli* strain Top 10 (*F-mcrA D* (*mrrhsdRMS- mcrBC*) *f80lacZDM15 DlacX74 deoR recA1araD139 D (ara-leu) 7697 galU galK rpsL* (*StrR*) *endA1 nupG*) was used in all DNA manipulations. *E. coli* was grown in Luria–Bertani (LB) medium (in g l^{-1}: 10 bacto tryptone, 10 NaCl, 5 yeast extract), supplemented, when required, with 100 µg ml^{-1} of ampicillin.

The *S. cerevisiae* strain used for heterologous expression was W303-1A (MATade2-1, his3-11, 15, leu2-3, 112, trp1-1, ura3-1, can1-100). The plasmids used for *S. cerevisiae* expression were C-pSAL4 and B-pSAL4,[15] both carrying URA3 gene for auxotrophic selection. *S. cerevisiae* was grown on a selective medium [6.7 g l^{-1} yeast nitrogen base without amino acids and ammonium sulfate, 5 g l^{-1} casamino acids, 30 mg l^{-1} adenine, 40 mg l^{-1} tryptophan, 50mM succinate buffer (pH 5.3), 20 g l^{-1} glucose].

The *K. lactis* strain used for heterologous expression was CMK5 (a *thr lys pgi1 adh3 adh1URA3 adh2URA3*).[16] The plasmids used for *K. lactis* expression were pYCps and pYA1bps.[15] *K. lactis* was grown in YPPG medium (in g l^{-1}: 10 yeast extract, 40 bacto tryptone, 20 glucose), supplemented with ethanol (0.5%), $CuSO_4$ (0.1mM) and when required, with 100 µg ml^{-1} of geneticin G418.

Random mutagenesis.

DNA shuffling of *poxc* and *poxA1b* cDNAs was performed using the protocol described by Van der Veen[17] based on the shuffling method described by Zhao and Arnold[18]. The cDNAs were fragmented using the following restriction enzymes: NlaIV and TaqI for *poxc* and HinfI and TaqI for *poxa1b.* Fragmented genes were subjected to a PCR without primer (PCR1) using Taq DNA polymerase (Promega Corp., Madison, Wisconsin, USA). Cycling parameters were 40 cycles of 95℃ for 1 min, 55℃ for 1 min, 72℃ for 1min + 5 s/cycle. Aliquots of the product of PCR1 were used as template in a conventional PCR with four different combination of primers: 1)POXA1bfw and POXa1brev; 2) POXA1bfw and POXCrev; 3) POXCfw and POXA1Brev; 4) POXCfw and POXCrev (Table I). Cycling parameters were 10 cycles of 95℃ for 30 s, 55℃ for 30 s, 72℃ for 45 s and 14 cycles of 95℃ for 30 s, 55℃ for 30 s, 72℃ for 45 s + 5 s/cycle. The full-length sh uffled cDNAs were purified using QIAquick Gel Extraction Kit (Qiagen GmbH, Hilden, Germany).

Random mutations were introduced with low, medium and high frequency of mutation, into the POXC and POXA1b encoding cDNAs using GeneMorph™ PCR Mutagenesis Kit (Stratagene, La Jolla, CA, USA). Error-Prone PCR was performed with primers POXA1bfw and POXA1brev for *poxa1b* and POXCfw and POXCrev for *poxc* (Table I).

Construction of a mutant library.

The cDNAs resulting from DNA shuffling were digested EcoRI/HindIII and ligated in B-pSAL4 vector digested with EcoRI/HindIII restriction enzymes. The cDNA resulting from EP-PCR on *poxa1b* cDNA were cloned in B-pSAL4 expression vector, digested

with SmaI and BglII restriction enzymes, by using homologous recombination expression system of *S. cerevisiae*. The cDNA resulting from EP-PCR on *poxc* cDNA were cloned in C-pSAL4 expression vector, digested with AatII restriction enzyme, by using homologous recombination expression system of *S. cerevisiae.* Yeast transformation was done by using the lithium acetate protocol.[19] The cells were spread on selective medium supplemented with 0.6 mM $CuSO_4$ and 0.2 mM ABTS and the plates incubated upside down for 4 days at 28°C.

Library screening.

Single clones grown on plate were picked and transferred into 96-well plates containing 30 µl of selective medium per well. Plates were incubated at 28°C, 250 rpm for 24 h. After 24 h 130 µl of selective medium was added to each well and the plates were incubated at 28°C, 250 rpm for 24 h. Thirty microliters of each culture was transferred to a new 96-well plate to measure the OD_{600} value. The plates were then centrifuged for 10 min at 1,500×g, 4°C and a suitable volume of supernatant was transferred to a new 96-well plate to perform laccase assay. Phenol oxidase activity was assayed at 25°C using 2 mM 2,2'-azino- bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) in 0.1 M sodium citrate buffer, pH 3.0. Oxidation of ABTS was followed by absorbance increase at 420 nm (ε 36,000 M^{-1} cm^{-1}), using Benchmark Plus microplate spectrophotometer (BioRad, Hercules, CA, USA). Enzyme activity was expressed in international units (U).

Cultures in shaken flasks were also performed. Pre-cultures (10 ml) were grown on selective medium at 28° C on a rotary shaker (150 rpm). A volume of suspension sufficient to reach a final OD_{600} value of 0.5 was then used to inoculate 250 ml Erlenmeyer flasks containing 50 ml of selective medium and cells were then grown on a rotary shaker. Optical density and laccase activity determination were daily assayed.

DNA sequencing.

Sequencing by dideoxy chain-termination method was performed by the Primm Sequencing Service (Naples, Italy) using specific oligonucleotide primers.

Protein purification.

Preparative cultures were performed in 1L flasks containing 200 ml of medium. Culture media were harvested on the optimal laccase production day, cells were sedimented by centrifugation at 1600×g at 4°C for 15 min. Secreted proteins were filtered throw Whatmann sheet. The sample was concentrated using Amicon stirred cells with PM30 ultrafiltration membranes (Millipore, Billerica, MA, USA). Most secreted proteins were precipitated by the addition of $(NH_4)_2SO_4$ up to 100% saturation at 4°C and centrifuged at 10000×g for 40 min. The supernatant from ammonium sulphate precipitation (containing laccases) was loaded onto a 26/60 Superdex 75 prep grade (GE healthcare Bio-Sciences, AB, Uppsala, Sweden) column equilibrated with 0.15M NaCl, 50 mM Na-phosphate buffer pH7. The active fractions were pooled and concentrated on an Amicon PM-30 membrane.

Protein concentrations were determined using the BioRad Protein Assay (BioRad), with bovine serum albumine as standard.

Laccase activity assays.

Phenol oxidase activity was assayed at 25°C using ABTS, 2,6-dimethoxyphenol (DMP), and syringaldazine (SGZ) as substrates as follows. (*a*) The assay mixture contained 2 mM ABTS and 0.1 M sodium citrate buffer, pH 3.0. Oxidation of ABTS was followed by absorbance increase at 420 nm (ε_{420} 36,000 M^{-1} cm^{-1}). (*b*) The assay mixture contained 1 mM DMP and the citrate-phosphate buffer was adjusted to pH 5. Oxidation of DMP was followed by an absorbance increase at 477 nm (ε_{477} 14,800

M^{-1} cm^{-1}). *(c)* The assay mixture contained 0.5 mM syringaldazine (dissolved in ethanol) and 50 mM phosphate buffer, pH 6. Oxidation of syringaldazine was followed by an absorbance increase at 525 nm (ε_{525} 65,000 M^{-1} cm^{-1}). Enzyme activity was expressed in IU.
Phenol oxidase activity as a function of pH was measured at room temperature using citrate-phosphate buffer adjusted to different pH levels in the range 3.0–6.2. Stability at pH values was measured using citrate-phosphate buffer adjusted at pH 3, 5 and 7.0, Tris-HCl buffer was adjusted at pH 10.
Molecular Modelling and Molecular Dynamics simulations.
The structure of POXA1b was obtained by homology modelling from the crystal structure of *Trametes versicolor* (1GYC pdb entry), with which it shares 60% sequence identity. It is well known that a sequence identity threshold for obtaining a reasonable model is 30%.[20] The last sixteen residues of POXA1b were modelled using the coordinates of the corresponding residues at the C-terminus of the crystal structure of the *Melanocarpus albomyces* laccase (1GWO pdb entry). This choice is justified in consideration of the fact that the fold of the *M. albomyces* and of the *T. versicolor* laccases is very similar, in particular in the C-terminal region the fold is almost identical. As can be seen in Fig. S2, the additional tail can be easily allocated onto this region of the protein
3D model and *in silico* mutants were generated using the SWISS-MODEL web server by means of the project mode option that allow to select the template and control the gap placement in the alignment.[21]
Refinement of the models has been performed by Molecular Dynamics simulations.[22,23]

Simulations on the wild type POXA1b and on the *in silico* generated mutants 1M9B and 3M7C were performed with the GROMACS package[24] using the GROMOS96 force field.[25] The systems were neutralized and were solvated in a box of SPC water,[26] in a solution of 50mM NaCl (17 Na^+ and 18 Cl^-). All the simulated boxes contained about 16,500 water molecules. Simulations were carried out at a constant temperature of 300K. The Berendsen algorithm was applied for the temperature and pressure coupling.[27] After a first steepest descent energy minimization with positional restraints on the solute, the LINCS algorithm was used to constrain the bonds[28] and to carry out an initial 200ps simulation with the positions of the solute atoms restrained by a force constant of 3000 kJ/(mol_*nm^2) in order to let the water diffuse around the molecule and to equilibrate. The particle mesh Ewald method (PME)[29] was used for the calculation of electrostatic contribution to non bonded interactions (grid spacing of 0.12 nm). The systems were run for 4 ns each and the last 500 ps of simulations were used for the energy decomposition analysis. GROMACS package and self-written programs have been used for the analysis of the data. Images were produced with visual molecular dynamics (VMD 1.8.5.).[30]

Results

Construction and screening of mutant laccase libraries
Random mutations were introduced into *poxa1b* and *poxc* cDNAs by error-prone PCR. Three different PCR conditions were used to yield low, medium and high average of mutation frequency (0÷3 mut/kbases, 3÷7 mut/kbases and more than 7mut/kbases respectively). Moreover a shuffled cDNA library was constructed by *poxc* and *poxa1b* cDNA digestion followed by *in vitro* recombination.
Both *K. lactis* and *S. cerevisiae* were used as expression systems for *P. ostreatus* laccases heterologous production, although *K. lactis* is a much more effective

producer than *S. cerevisiae*[15]. Mutated cDNAs were introduced into the expression vectors by using host homologous recombination system. When *K. lactis* was used, several mutants (70%) able to grow on selective medium but unable to express any laccase activity were obtained. Even if *K. lactis* was transformed with the open vector (control) several clones growing on selective medium were obtained. As a fact, the budding yeast *K. lactis*, in contrast to *S. cerevisiae*, is known to show, variable, but generally low, gene targeting efficiency.[31,32] Therefore random recombination of the resistance gene could explain the achievement of clones able to grow on selective medium even if transformed with the open vector. On this basis *S. cerevisiae* was preferred to *K. lactis* as host for expressing laccase mutated cDNAs, because no problem related to low gene targeting efficiency was faced in this case. 1100 transformants were obtained by *S. cerevisiae* transformation with pSAL4 vector and mutated cDNAs. Thirty-eight mutants were selected, by multiwell plate screening, having activity towards ABTS higher than that of the controls (yeasts expressing wild type POXC and POXA1b). The second screening, performed following growth rate and activity production of the clones in multiwell plates for three days, allowed us to confirm positivity of only 7 mutants out of the 38 previous isolated. These seven clones were then grown in shaken flasks for six days for further analyses. The mutant 1M9B, originated from error-prone PCR with medium frequency using *poxa1b* as template, showed significantly higher activity (about 50% increase) than that of wild type POXA1b (Fig. 1).

1m9b cDNA was used as template for a second round of error-prone PCR at low and medium frequency of mutation. A second generation library of 1200 clones was obtained and screened using the same procedure described before. Three mutants, 1L2B, 1M10B and 3M7C, were selected showing 70÷100% activity increase with respect to 1M9B towards the substrate ABTS. Figure 1 shows the laccase specific production (mU/OD) of these clones compared to that of the wild type.

Characterization of mutated and wild type laccases.

The recombinant POXA1b protein and the four selected mutants (1M9B, 1L2B, 1M10B and 3M7C) were partially purified by ammonium sulphate selective precipitation followed by gel filtration chromatography as described in Materials and Methods. Similar amounts of total proteins were obtained in all the cases.

A non phenolic substrate, ABTS, and two phenolic substrates, SGZ and DMP, were used to analyse the catalytic properties of mutants compared with those of the wild type enzyme (Table II). All the mutants show Michaelis constant (K_M) values similar to that of wild type towards ABTS, while their affinity towards DMP is higher than that of wild type protein. Only 1L2B mutant showed K_M value towards SGZ lower than that of the wild type. The specific activities of the mutants towards ABTS increased (1.7÷2.5-fold) with respect to the wild type protein. 1L2B, 1M10B and 3M7C mutants also exhibit 1.5 to 1.8-fold specific activity increase towards DMP, while no significant changes of their ability in SGZ oxidization were detected. On the other hand 1M9B shows a specific activity value lower than that of wild type protein towards both phenolic substrates.

All proteins, including the wild type, are fully active in the temperature range 40-65℃.

The effect of pH on the activity of recombinant laccases towards DMP was studied in the pH range 3.0÷6.2 (Fig. 2). All the laccases tested show maximum activity at pH value of ~4.2, but all the mutants are slightly more responsive to the pH increase above pH 4.4 value than the wild type protein.

Stability of the enzymes with respect to temperature at neutral pH was studied. All the mutants, except for 1M9B, showed higher half life at 60℃ with respect to the wild

type recombinant protein (Table III). Stability of the laccases in different buffers and pH values was also studied at room temperature (Table III). Mutant 1M9B is the less stable one in all the analyzed conditions, while 1L2B, 1M10B and 3M7C show $t_{1/2}$ values similar or higher than that of the wild type protein.

Sequence analysis of the mutants led to the identification of nucleotide substitutions producing the mutations reported in brackets: 1M9B (L112F); 1L2B (L112F/N248Y/N261K/V350I); 1M10B (L112F/K37Q/K51N); 3M7C (L112F/P494T).

In silico structural analyses

Models of the 3D structure of POXA1b was constructed by homology modelling from the crystal structure of the *Trametes versicolor* laccase [33] as a template, with which shares about 60% of identity (alignment of the two proteins is shown in Figure S1). The last sixteen residues [L498-A513] of POXA1b were modelled using the coordinates of the corresponding residues at the C-terminus of the *Melanocarpus albomyces* laccase,[34] the only laccase whose structure has been resolved showing a protruding C-terminus, like POXA1b. Inspection of the sequence alignments of the laccase family revealed that the position 112, which is mutated in all the isolated mutants, follows the consensus region 98-111 (WYHSH) where the histidine residues coordinating type 3 copper ions are located. In position 112 there is a generally conserved leucine in all laccases from basidiomycetes, while a phenylalanine seems to be conserved in laccase sequences from ascomycetes. POXA1b 3D model shows that residue 112 is located in the channel through which the solvent has access to the oxygen-reducing T3 site (Fig. 3).

As far as 1L2B mutant is concerning, position 248 is quite variable in laccase sequences: N, E, D, T, H or K residues can occur at this position but a Y residue is never found. Differently, the two residues N261 and V350 are highly conserved and no laccase showing a K residue in position 261 or an I residue in position 350 can be found. From the 3D model we can observe that position 261 is located in a loop potentially interacting with the substrate (Fig. 3).[35]

The two additional substitutions observed in 1M10B mutant occur in positions 37 and 51 that are generally conserved in laccase sequences and occupied by amidic residues.

The last mutant analysed, 3M7C, presents only one substitution (P494T) besides the L112F. Position 494 is located in a variable and mobile loop at the C-terminus that, in accordance with modelling data, forms a sort of “plug” blocking the channel through which the solvent has access to type-3 copper sites (Fig. 3B).

Molecular Dynamics simulations

Two selected mutants generated by the first and second screenings, 1M9B and 3M7C respectively, have been further characterized and compared to the wild type POXA1b by means of MD simulations. The simulations performed on the three systems have reached equilibration after 500 ps as can be seen from the RMSD (Root Mean Square Deviation) of the Cα atoms as a function of time (Fig. 4A). The molecules are stable under the simulated conditions and converge to similar energy values as shown in Table IV.

The major differences between the systems are highlighted by the analysis of their flexibility exhibited in solution. POXA1b is mostly rigid and does not show any major concerted motion. In Figure 4 (panels B and C) is shown the plot which represents the difference of the Root Mean Square Fluctuation (Δ-RMSF) of the Cα atoms between POXA1b and the 1M9B and 3M7, respectively. 1M9B experiences a large fluctuation in the region 300-380 while 3M7C shows a highest fluctuation in the regions 265-270 and 330-350.

The Principal Component Analysis (PCA) method is used to reduce the complexity of the data obtained by MD simulations and to highlight the principal components responsible of the observed motion. The total motion is decomposed in the principal components to which an eigenvector and an eigenvalue are associated. We decided to present only the analysis of the first two eigenvectors because they account more than 45% of the total amount of the motion of the systems.
1M9B shows a rearrangement of the residues in proximity of the F112 giving rise to an increased flexibility of the entire subdomain. On the other hand 3M7C presents more rigidity in that subdomain, while showing a major rearrangement in proximity of the substrate binding region (Fig. 5), leading to a more accessible T1 copper site. Nevertheless the secondary structure remains generally preserved for both mutants and the fold is not affected by the enhanced flexibility.
A significant effect generated by the mutations is observed in the permeability to water of the T2/T3 channel. Residue F112 is located at the entrance of the channel and its steric hindrance affects the passage of water molecules towards the trinuclear copper cluster. A detailed analysis of the water molecules in close contact with the channel residues has been performed, in terms of both their distances from these residues and their residence times in the channel (Fig. 6). Moreover, snapshots from the MD simulations (Fig. 6 D, E, F) show the location of the water molecules in the channel T2/T3 for each system. In the case of 1M9B an average of five water molecules are trapped near the cluster, while only three were found in the POXA1b channel (Fig. 6 A, B). The F112 residue of 1M9B acts as a gating residue for the water molecules preventing their exit from the channel as shown in Figure 6E. A different behaviour was observed in 3M7C because the mutation P494T affects the position of F112. As a consequence, F112 is not obstructing the channel anymore, and although an average of three water molecules remain locked in the channel, as for POXA1b (Fig. 6 A, C), more water molecules can enter in the cavity remaining coordinated for a shorter time then in 1M9B and POXA1b.

Discussion

In this paper we describe the *in vitro* evolution of laccases from *Pleurotus ostreatus* leading to enzymes endowed with increased specific activity towards phenolic and non-phenolic substrates as well as increased stability. We performed random mutagenesis experiments on *poxc* and *poxa1b* cDNAs, using Error Prone PCR (EP-PCR), the most frequently used non recombinative method, and DNA shuffling, the most traditional among recombinative methods.[36] Applications of these methods resulted in collections of shuffled and mutated sequences with low, medium and high frequency. After transformation into *Saccharomyces cerevisiae,* a first library of 1100 clones was obtained and screened.
As far as DNA shuffling is concerning, only 40% of the clones were able to express active laccases. Although the most significant changes in enzyme functions can be created using the DNA shuffling technique, cDNAs coding for inactive proteins are mainly produced during “gene reconstruction” confirming the idea that this methodology leads to frequent damages of essential structural protein components.[36] As a matter of fact, no “better performing” variants were obtained from DNA shuffling.
When EP-PCR was used, the number of mutants able to produce laccase activity drastically dropped down, increasing the frequency of mutation (90, 30 and 10% for low, medium and high frequency of mutation, respectively). This is consistent with the generally accepted concept that the probability of discovering beneficial mutations

decreases at high mutation rate, since the majority of random mutations are either neutral or deleterious.[37]

The POXA1b mutant named 1M9B selected from the screening of the first library, is more active towards ABTS but less stable with respect to POXA1b in all the analyzed conditions. To elucidate the role played by the single mutation L112F of 1M9B, MD simulations were performed and compared to those of POXA1b. The analyses show a movement of the subdomain around position 112 as a consequence of a conformational rearrangement due to the presence of the bulkier residue of phenylalanine. Moreover a larger number of water molecules in the T2/T3 channel has been observed for 1M9B (Fig. 6B). These data could suggest an increased affinity of this mutant towards oxygen molecules, thus justifying its improved specific activity.

This mutant has been used as a template for a second round of EP-PCR. From the library of 1200 clones obtained, three mutants have been selected (1L2B, 1M10B and 3M7C), partially purified, and characterised. These mutants, besides improved activity towards phenolic and non-phenolic substrates, exhibit higher stability then POXA1b in all the conditions analysed. Therefore the additional substitutions found in the second generation mutants, compensate for the destabilizing effect of the L112F substitution either in a correlated or uncorrelated fashion.

The most stable and active mutant, 3M7C, presents, besides L112F, only one substitution (P494T) that has to be responsible of its improved performances. The position 494 is located in the C-terminal loop that has already been ascertained to affect the function of fungal laccases. Hakulinen *et al.*[34] proposed that a conformational change in the C-terminus allows opening of the tunnel for the entrance of oxygen molecule and the exit of water in the laccase from the Ascomycete *Melanocarpus albomyces*. Moreover Gelo-Pujic *et al.*[38] have changed the redox potential of the type-1 Cu of laccase from *Trametes versicolor*, produced in *Pichia pastoris*, replacing 11 amino acids at the C-terminus with a single Cys residue. During the MD simulations of all the studied systems, POXA1b, 1M9B and 3M7C, high flexibility of the C termini and involvement of these regions to direct water molecules towards the T2/T3 channel has been observed.

Furthermore MD simulations of the 3M7C mutant protein reveal a lower flexibility of the subdomain around position 112 with respect to that of 1M9B, this could be related to the recovered stability of 3M7C. On the other hand, an increased mobility of loops forming the reducing substrate binding site has been observed leading to higher accessibility of water molecules to the T1 copper site and possibly leading to an increased activity of the enzyme.

Laccases are proteins with highly conserved and functionally essential wide regions; therefore not many positions can be mutated without loosing activity. The identified serendipitous mutations, although not directly involved in catalysis, have been proved to influence the efficiency and the stability of the enzyme. Our mutagenesis experiment has led to more efficient enzymes and to an increase of our knowledge of the structure-function relationships of this class of enzymes.

Table I: List of the primers used in amplification experiments. Nucleotides in bold are complementary to the laccase cDNA sequences. Nucleotides in italics are recognized by restriction enzymes.

Primer	Nucleotide sequences	Ann. T
POXA1b*fw*	ATAA*AAGCTTGAATTC***ATGGCGGTTGCATTCG**	90℃
POXA1b*rev*	TAA*GGATCCAAGCTT***TTATAATCATGCTTC**	82℃
POXC*fw*	TTT*AAGCTTGAATTC***ATGTTTCCAGGCGCACGG**	96℃
POXC*rev*	AAAA*GAATTCAAGCTT***TTAAGAAGTAGGAATAAGAAG**	94℃

Table II. Catalytic parameters of the wild type protein and the selected mutants.

	ABTS			DMP			SGZ		
Laccases	K_M (mM)	Specific activity (U mg^{-1})	Catalytic efficiency (U $mg^{-1}mM^{-1}$)	K_M (mM)	Specific activity (U mg^{-1})	Catalytic efficiency (U $mg^{-1}mM^{-1}$)	K_M (mM)	Specific activity (U mg^{-1})	Catalytic efficiency (U $mg^{-1}mM^{-1}$)
Wild type	0,085±0.007	183±1	2153±15	0,54±0.02	187±1	346±7	0,0 48±0.006	23.2±0.8	483±2
1M9B	0,067±0.005	315±1	4701±23	0,36±0.02	168±1	467±9	0,0 38±0.007	9,3±0.5	245±1
1L2B	0,107±0.009	420±2	3925±71	0,31±0.01	295±2	952±19	0, 025±0.004	24,3±0.5	972±2
1M10B	0,100±0.009	421±2	4210±76	0,35±0.01	281±1	803±8	0,0 40±0.006	24,3±0.6	607±2
3M7C	0,077±0.009	454±2	5896±106	0,24±0.01	337±2	1404±28	0,055±0.011	26,9±0.9	489±5

Table III Half-life of the wild type protein and the selected mutants at different pH values and at 60℃.

Laccases	$t_{½}$ (days)				$t_{1/2}$ (hours)
	pH3	pH5	pH7	pH10	60℃, pH7
Wild type	0.9	2.8	7.6	29.6	2.2
1M9B	0.6	1.9	3.6	24.0	1.7
1L2B	1.1	4.0	6.3	34.4	2.8
1M10B	1.1	3.9	8.6	44.6	2.5
3M7C	1.5	5.7	13.5	31.9	3.1

TABLE IV: Energy decomposition analysis of MD simulations

The first and the last 500 ps of simulations were used for the energy decomposition analysis.

POXA1b			
Energy (kJ mol^{-1})	**E initial**	**Energy (kJ mol^{-1})**	**E final**
Protein-Protein (El)	-38626	Protein-Protein (El)	-38819
Protein-Protein (LJ)	-18286	Protein-Protein (LJ)	-18822
Protein-Protein (total)	-56912	Protein-Protein (total)	-57641
Protein-Solvent (total)	-32272	Protein-Solvent (total)	-32101
1M9B			
Energy (kJ mol^{-1})	**E initial**	**Energy (kJ mol^{-1})**	**E final**
Protein-Protein (El)	-38502	Protein-Protein (El)	-38790
Protein-Protein (LJ)	-18235	Protein-Protein (LJ)	-18735
Protein-Protein (total)	-56737	Protein-Protein (total)	-57525
Protein-Solvent (total)	-32325	Protein-Solvent (total)	-32544
3M7C			
Energy (kJ mol^{-1})	**E initial**	**Energy (kJ mol^{-1})**	**E final**
Protein-Protein (El)	-38805	Protein-Protein (El)	-38873
Protein-Protein (LJ)	-18414	Protein-Protein (LJ)	-18681
Protein-Protein (total)	-57219	Protein-Protein (total)	-57554
Protein-Solvent (total)	-31650	Protein-Solvent (total)	-32730

REFERENCES

1. Xu F. Oxidation of phenols, anilines, and benzenethiols by fungal laccases: correlation between activity and redox potentials as well as halide inhibition. Biochemistry 1996;35(23):7608-7614.
2. Xu F, Shin W, Brown SH, Wahleithner JA, Sundaram UM, Solomon EI. A study of a series of recombinant fungal laccases and bilirubin oxidase that exhibit significant differences in redox potential, substrate specificity, and stability. Biochim Biophys Acta 1996;1292(2):303-311.
3. Thurston CF. The structure and function of laccases. Microbiology 1994;140:19-26.
4. Baldrian P. Fungal laccases - occurrence and properties. FEMS Microbiol Rev 2006;30(2):215-242.
5. O' Malley DM, Whetten R, Bao W, Chen CL, Seedorf RR. The role of laccase in lignification Plant J;4:751–757.
6. Claus H. Laccases and their occurrence in prokaryotes. Arch Microbiol 2003;179(3):145-150.
7. Dittmer NT, Suderman RJ, Jiang H, Zhu YC, Gorman MJ, Kramer KJ, Kanost MR. Characterization of cDNAs encoding putative laccase-like multicopper oxidases and developmental expression in the tobacco hornworm, Manduca sexta, and the malaria mosquito, Anopheles gambiae. Insect Biochem Mol Biol 2004;34(1):29-41.
8. Kramer KJ, Kanost MR, Hopkins TL, Jing H, Zhu YC, Xhu R, Kerwin JL, Turecek F. Oxidative conjugation of catechols with proteins in insect skeletal systems. Tetrahedron 2001;57:385–392.
9. Gianfreda L, Xu F, Bollag JM. Laccases: a useful group of oxidoreductive enzymes Bioremediation J 1999;3:1-25.
10. Solomon E, Sundaram U, Machonkin T. Multicopper Oxidases and Oxygenases. Chem Rev 1996;96(7):2563-2606.
11. Palmieri G, Giardina P, Marzullo L, Desiderio B, Nitti G, Cannio R, Sannia G. Stability and activity of a phenol oxidase from the ligninolytic fungus Pleurotus ostreatus. Appl Microbiol Biotechnol 1993;39(4-5):632-636.
12. Palmieri G, Giardina P, Bianco C, Scaloni A, Capasso A, Sannia G. A novel white laccase from Pleurotus ostreatus. J Biol Chem 1997;272(50):31301-31307.
13. Giardina P, Palmieri G, Scaloni A, Fontanella B, Faraco V, Cennamo G, Sannia G. Protein and gene structure of a blue laccase from Pleurotus ostreatus1. Biochem J 1999;341 (Pt 3):655-663.
14. Palmieri G, Cennamo G, Faraco V, Amoresano A, Sannia G, Giardina P. Atypical laccase isoenzymes from copper supplemented Pleurotus ostreatus cultures. Enzyme Microb Technol 2003;33:220–230.
15. Piscitelli A, Giardina P, Mazzoni C, Sannia G. Recombinant expression of Pleurotus ostreatus laccases in Kluyveromyces lactis and Saccharomyces cerevisiae. Appl Microbiol Biotechnol 2005;69(4):428-439.
16. Saliola M, Mazzoni C, Solimando N, Crisà A, Falcone C, Jung G, Fleer R. Use of the KlADH4 promoter for ethanol-dependent production of recombinant human serum albumin in Kluyveromyces lactis. Appl Environ Microbiol 1999;65(1):53-60.
17. van der Veen BA, Potocki-Véronèse G, Albenne C, Joucla G, Monsan P, Remaud-Simeon M. Combinatorial engineering to enhance amylosucrase performance: construction, selection, and screening of variant libraries for increased activity. FEBS Lett 2004;560(1-3):91-97.
18. Zhao H, Arnold FH. Optimization of DNA shuffling for high fidelity recombination. Nucleic Acids Res 1997;25(6):1307-1308.
19. Gietz D, St Jean A, Woods RA, Schiestl RH. Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res 1992;20(6).
20. Rost B. Twilight zone of protein sequence alignments. Protein Eng 1999;12(2):85-94.
21. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 2003;31(13):3381-3385.
22. Flohil JA, Vriend G, Berendsen HJ. Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. Proteins 2002;48(4):593-604.
23. Fan H, Mark AE. Mimicking the action of folding chaperones in molecular dynamics simulations: Application to the refinement of homology-based protein structures. Protein Sci 2004;13(4):992-999.

24. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Computer Physics Communications 1995;91(1-3):43-56.
25. Daura X, Mark AE, van Gunsteren WF. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. Journal of Computational Chemistry 1998;19:535-547.
26. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J, Pullman B. Interaction models for water in relation to protein hydration. Intermolecular Forces; 1981. p 331-342.
27. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. Journal of Chemical Physics 1984;81(8):3684-3690.
28. Hess B, Bekker H, Berendsen HJC, Fraaije J. LINCS: a linear constraint solver for molecular simulations. Journal of Computational Chemistry 1997;18(12):1463-1472.
29. Darden T, Perera L, Li L, Pedersen L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. Structure Fold Des 1999;7(3).
30. Humphrey W, Dalke, A. and Schulten, K. VMD—visual molecular dynamics. J Mol Graph 1996;14:33–38.
31. Zeeman AM, Steensma HY. The acetyl co-enzyme A synthetase genes of Kluyveromyces lactis. Yeast 2003;20(1):13-23.
32. Kooistra R, Hooykaas PJ, Steensma HY. Efficient gene targeting in Kluyveromyces lactis. Yeast 2004;21(9):781-792.
33. Piontek K, Antorini M, Choinowski T. Crystal structure of a laccase from the fungus Trametes versicolor at 1.90-A resolution containing a full complement of coppers. J Biol Chem 2002;277(40):37663-37669.
34. Hakulinen N, Kiiskinen LL, Kruus K, Saloheimo M, Paananen A, Koivula A, Rouvinen J. Crystal structure of a laccase from Melanocarpus albomyces with an intact trinuclear copper site. Nat Struct Biol 2002;9(8):601-605.
35. Larrondo LF, Salas L, Melo F, Vicuna R, Cullen D. A novel extracellular multicopper oxidase from Phanerochaete chrysosporium with ferroxidase activity. Appl Environ Microbiol 2003;69(10):6257-6263.
36. Williams GJ, Domann S, Nelson A, Berry A. Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution. Proc Natl Acad Sci U S A 2003;100(6):3143-3148.
37. Valetti F, Gilardi G. Directed evolution of enzymes for product chemistry. Nat Prod Rep 2004;21(4):490-511.
38. Gelo-Pujic M, Kim HH, Butlin NG, Palmore GT. Electrochemical studies of a truncated laccase produced in Pichia pastoris. Appl Environ Microbiol 1999;65(12):5515-5521.

LEGEND AND FIGURES

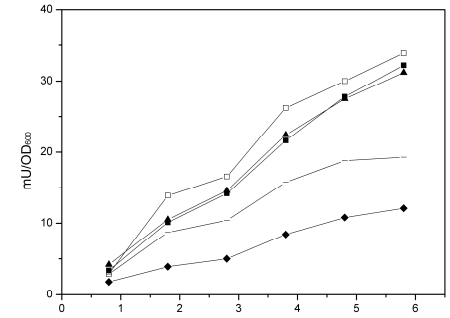


Figure 1. Time course of laccase activity in the extracellular medium of *S. cerevisiae* cultures producing wild-type and mutated enzymes (♦ wild type POXA1b; - 1M9B; ▲3M7C; ■ 1M10B; □ 1L2B)

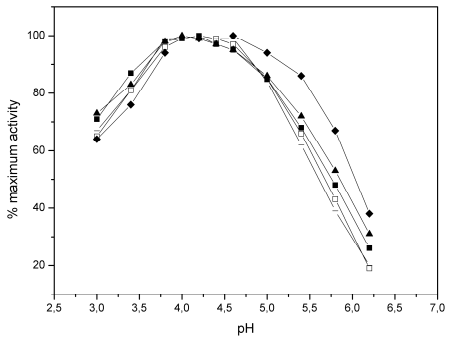


Figure 2. Effect of pH on the activity of wild type and mutated laccases towards DMP as substrate (♦ wild type POXA1b; - 1M9B; ▲3M7C; ■ 1M10B; □ 1L2B)

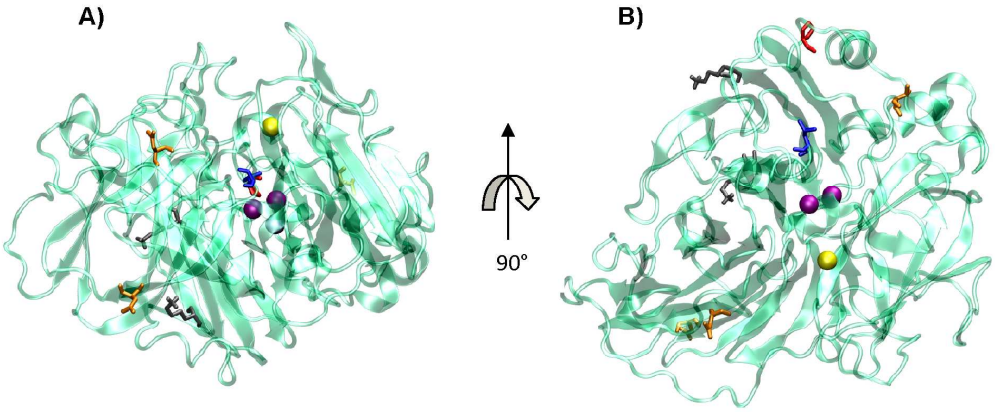


Figure 3 (A) and (B) show ribbon representation of POXA1b model. The mutated residues are in licorice rendering (L112 blue; N248, N261, V350 orange; K37, K51 gray; P494 red). Trinuclear coppers and T1 copper are highlighted in van der Waals representation (purple and yellow respectively).

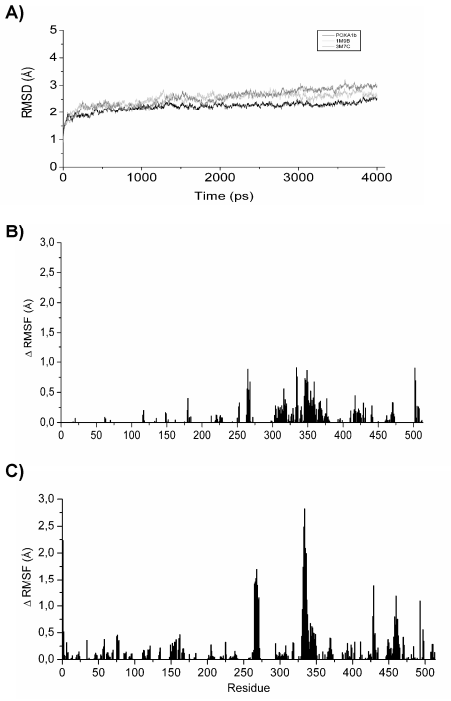


Figure 4. Panel A: RMSD of the Cα atoms from their initial coordinates as a function of time, for the three simulated systems. Black line, POXA1b; light gray, 1M9B; gray, 3M7C.
Panels B and C: ΔRMSF of 1M9B (B) and 3M7C (C) of the Cα atoms with respect to their average position along their entire simulation compared to the Cα atoms from the simulation of POXA1b.

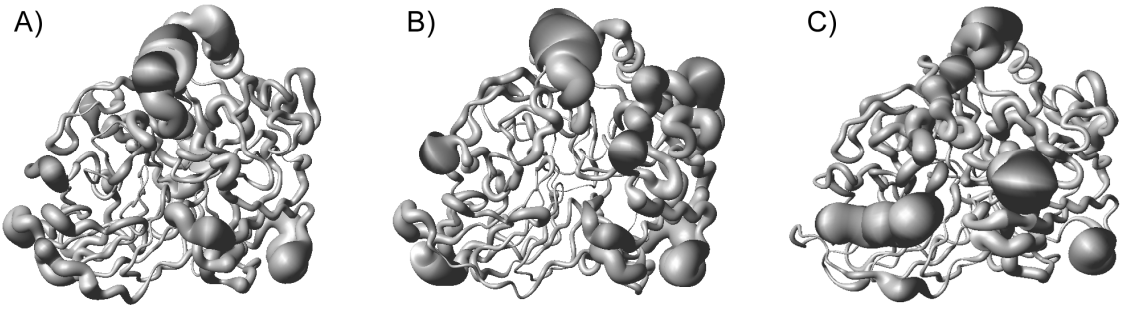


Figure 5. 'Sausage' plots representing the motions observed during simulation trajectories along the first eigenvector. (A) POXA1b; (B) 1M9B; (C) 3M7C.

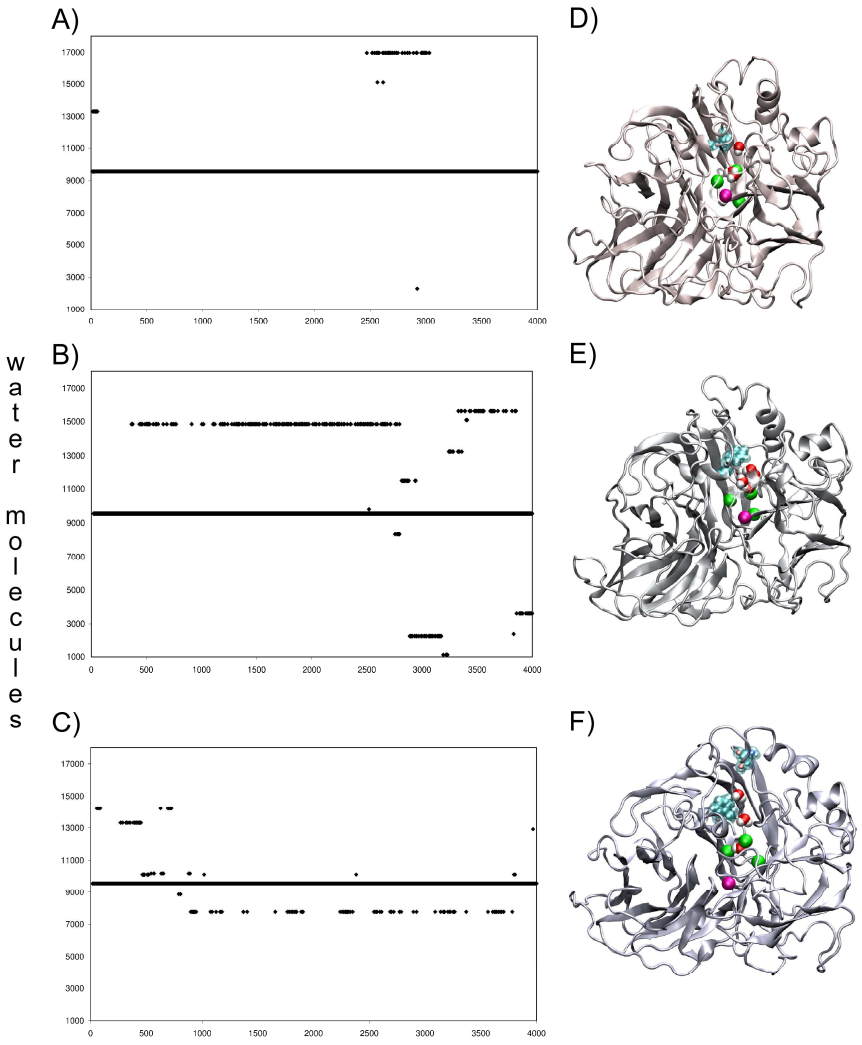


Figure 6. Analysis of the water molecules that are in close contact (0.4 nm) with the channel residues of (A) POXA1b, (B) 1M9B and (C) 3M7C during the entire simulated time. Ribbon representation of models, POXA1b (D), 1M9B (E) and 3M7C (F). Cu ions (T2/T3 cluster in green and T1 in magenta) and the water molecules in close contact with the channel of T2/T3 cluster are shown in van der Waals rendering. In panel D residue L112 is highlighted, in panel E residue F112, and in panel F both residues F112 and T494 are shown.

Structural characterization of heterodimeric laccases from *Pleurotus ostreatus*

The aim of this section was to generate the models of the two isoforms of POXA3 generated by a different sliced and to analyses the sequence diversity of POXA3 with respect to other laccases.

Appl Microbiol Biotechnol (2007) 75:1293–1300
DOI 10.1007/s00253-007-0954-4

BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

Structural characterization of heterodimeric laccases from *Pleurotus ostreatus*

Paola Giardina · Flavia Autore · Vincenza Faraco · Giovanna Festa · Gianna Palmieri · Alessandra Piscitelli · Giovanni Sannia

Received: 21 February 2007 / Revised: 15 March 2007 / Accepted: 16 March 2007 / Published online: 12 April 2007

Abstract The subfamily of POXA3 laccase isoenzymes produced by the fungus *Pleurotus ostreatus* has been characterized as an example of the complexity and heterogeneity of fungal isoenzyme patterns. Two isoenzymes, POXA3a and POXA3b, were previously purified, exhibiting an unusual heterodimeric structure composed of a large (67 kDa) and a small (18 or 16 kDa) subunit. A unique gene encodes the large subunit of both POXA3a and POXA3b, but alternative splicing produces two variants—differing for an insertion of four amino acids—for each isoenzyme. Two genes encoding POXA3a and POXA3b small subunits have been identified, and the corresponding amino acid sequences show only two amino acid substitutions. The 18- and 16-kDa subunits of both POXA3a and POXA3b differ for *N*-glycosylation at Asn150 of the 16-kDa subunit. The POXA3 large subunit 3D model allows us to highlight peculiarities of this molecule with respect to the laccases whose 3D structures are known.

Keywords Phenol oxidase · White rot fungi · Quaternary structure

P. Giardina (✉) · F. Autore · V. Faraco · G. Festa · A. Piscitelli · G. Sannia
Dipartimento di Chimica Organica e Biochimica, Università di Napoli "Federico II," Complesso Universitario Monte S. Angelo, via Cintia,
Naples 80126, Italy
e-mail: giardina@unina.it

G. Palmieri
IBP, Consiglio Nazionale delle Ricerche,
via P. Castellino 111,
Naples 80131, Italy

Introduction

Laccases are multicopper oxidative enzymes, useful for biotransformation of environmental organic pollutants and exploited industrially in various oxidative processes. Multiple isoforms of laccases are usually secreted by each fungus depending on species and environmental conditions. Up to 17 laccase-encoding genes have been found in the genome of *Coprinopsis cinerea* (Kilaru et al. 2006). Laccases are the major extracellular components of the lignin-degrading system of the white rot fungi belonging to the *Pleurotus* genus (Baldrian et al. 2005). The isoenzymes produced by *Pleurotus ostreatus* have been extensively studied. POXC laccase is the most abundantly produced under all the growth conditions examined (Palmieri et al. 1993); other isoenzymes secreted by the mycelium have also been purified and characterized (POXA1w and POXA1b) (Palmieri et al. 1997; Giardina et al. 1999).

Two closely related laccase isoenzymes (POXA3a and POXA3b) produced by *P. ostreatus* in copper supplemented cultures were purified (Palmieri et al. 2003). They exhibit unusual structural features. Unlike most of the known laccases that are monomeric proteins, both native isoenzymes are heterodimers constituted of a large (67 kDa) and a small (18- or 16-kDa) subunit (Palmieri et al. 2003). Some laccases endowed with quaternary structure have already been found in *Phellinus ribis* (Min et al. 2001), *Trametes villosa* (Yaver et al. 1996), *Pleurotus pulmonarius* (De Souza and Peralta 2003), and *Rhizoctonia solani* (Wahleithner et al. 1996), but all of them are homodimeric proteins. A few heterooligomeric laccases were found in the fungi *Monocillium indicum* (Thakker et al. 1992), *Agaricus bisporus* (Perry et al. 1993) and *Armillaria mellea* (Curir et al. 1997), but only limited characterization of these proteins was reported.

Despite their remarkably different chromatographic behavior, POXA3a and POXA3b isoenzymes show similar catalytic properties (Palmieri et al. 2003). Their large subunits share the same MALDI-MS peptide maps, corresponding to the amino acid sequence encoded by the unique *poxa3* gene. This protein sequence is clearly homologous to other known laccase sequences, and contains all the four putative copper-binding residues and the four Cys residues forming disulfide bridges in most of other laccases (Cullen 1997). Each small subunit, either from POXA3a or POXA3b, is produced as two isoforms of 18- or 16-kDa molecular mass. Peptide-mapping analysis of the small subunits shows close relationships among them. The amino acid sequence of some tryptic peptides from the 18-kDa subunit of POXA3a did not show significant homology with other known proteins (Palmieri et al. 2003), leaving the nature and the role of these subunits still obscure.

The singular properties of POXA3a and POXA3b isoenzymes led us to deeply investigate their structural features, also taking into account their biotechnological potential. Among *P. ostreatus* laccases, POXA3 isoenzymes are able to efficiently decolorize the anthraquinonic dye Remazol Brilliant Blue R, showing a catalytic efficiency sixfold higher than that of POXC (Palmieri et al. 2005).

In this paper, the structural characteristics of various POXA3 forms are analyzed. The heterogeneity observed for this laccase can be an example of the complexity of fungal isoenzyme patterns.

Materials and methods

Organism and culture conditions Dikaryotic strain of the white rot fungus, *P. ostreatus* (Jacq.: Fr.) Kummer (type: Florida) (ATCC no. MYA-2306) was maintained through periodic transfer at 4°C on potato dextrose agar plates in the presence of 0.5% yeast extract (Difco, Detroit, MI, USA). Mycelia were grown in shaking flasks containing potato dextrose (24 g/l) broth with 0.5% yeast extract. Fifty milliliters of a 5-day-old culture were transferred in 1-l flasks containing 450 ml of broth supplemented with 150 μM of $CuSO_4$. Enzyme purification was performed from 10 days of culture broth.

Enzyme purification Secreted proteins were precipitated from the filtered medium by addition of $(NH_4)_2SO_4$ up to 80% saturation and, after extensive dialysis, loaded onto a DEAE Sepharose Fast Flow (GE Healthcare, Milan, Italy) column. Two fractions containing laccase activity, recovered with the equilibrating buffer, were separately pooled, concentrated on an Amicon PM-10 membrane (Millipore, Billerica, MA, USA), and equilibrated in Tris–HCl, 50 mM, pH 8.0. Each pool was loaded onto an anion exchange Mono Q HR5/5 (GE Healthcare) column in a fast protein liquid chromatography system (GE Healthcare) equilibrated with the same buffer. The enzyme was eluted with a linear gradient (0–0.3 M NaCl). Ammonium sulfate at 1 M of the final concentration was added to the active fractions pooled and desalted. These fractions were loaded onto a Phenyl Superose PC 1.6/5 (GE Healthcare) at 0.05 ml/min. Elution was performed by saline linear gradient (1–0 M ammonium sulfate).

Protein deglycosylation was performed by incubating laccase samples with 1 mU of endoglycosidase H (Boehringer Ingelheim GmbH, Ingelheim, Germany) in 20 mM sodium acetate, pH 5.2, 0.01% sodium dodecyl sulfate (SDS), and 5 mM dithiothreitol overnight at 37°C.

Isolation and sequencing of POXA3 small subunit genes and cDNAs Amplification experiments of *P. ostreatus* genomic DNA were performed at 44°C annealing temperature, using the following oligonucleotide couples: 5′-CAR ATYCAYGTBAAYATYCC-3′ and 5′-CATCATN GTNGTRTANGT-3′ (Y = T/C, R = G/A;V = G/C/A, B = G/T/C, D = G/A/T, n = G/A/T/C). A 400-bp fragment was obtained, cloned in pGEM T Easy vector, and sequenced. This fragment, ^{32}P-labeled by random priming, was used as probe to screen a *P. ostreatus* genomic library. Colony hybridization experiments were carried out in 5×SSC (0.75 M NaCl, 0.075 M sodium citrate) at 65°C.

Total RNA was extracted from lyophilized mycelia, harvested from 3 days of culture as already described (Lucas et al. 1977). Reverse transcription reaction was performed using Superscript II (Gibco BRL, Carlsbad, CA, USA) and following the manufacturer's instructions. Amplification experiments of specific cDNA were performed at 58°C annealing temperature using the following oligonucleotide couple: 5′-GGAGCTAAATACGCTACAAGC-3′ and 5′-AATTCGCGGCCGCTTTTTTTTTTTTTT-3′. The amplified fragment was cloned in the pGEM T Easy vector (Promega, Madison, WI, USA) and sequenced.

In situ digestion Mass spectrometric analyses were performed on the Coomassie blue-stained proteins excised after preparative SDS electrophoresis on a 12% polyacrylamide gel (Laemmli 1970). Excised bands were washed with acetonitrile and then with 0.1 M of ammonium bicarbonate. Protein samples were reduced by incubation in 10 mM dithiothreitol for 45 min at 56°C and carboxamidomethylated by using 55 mM iodoacetamide in 0.1 M of NH_4HCO_3 for 30 min, in the dark, under nitrogen atmosphere at room temperature. The gel particles were then washed with ammonium bicarbonate and acetonitrile (J.T. Baker, Phillipsburg, NJ, USA). Enzymatic digestions were carried out with 15 mg/ml trypsin (Sigma-Aldrich, St. Louis, MO, USA) in 50 mM of ammonium bicarbonate, pH 8.5 at 4°C for 4 h. The buffer solution was then

removed and a new aliquot of enzyme/buffer solution was added for 18 h at 37°C. A minimum reaction volume, sufficient for complete rehydration of the gel was used. Peptides were then extracted washing the gel particles with 20 mM of ammonium bicarbonate and 0.1% trifluoroacetic acid (Carlo Erba, Milan, Italy) in 50% acetonitrile at room temperature and then lyophilized. Aliquots of the digests were directly analyzed by MALDI-MS or separated on a narrow bore Vydac C_{18} column (25×0.21 cm, 5 μm) (The Separation Group, Hesperia, CA, USA) using 0.1% trifluoroacetic acid (Sigma-Aldrich) as solvent A and 0.07% trifluoroacetic acid in 95% acetonitrile as solvent B. A linear gradient of solvent B from 5 to 65% in 60 min at flow rate of 0.2 ml/min was employed. The UV absorbance of the eluent was monitored at 220 nm.

Mass spectrometry analysis MALDI mass spectra were recorded using a Voyager DE Pro MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA, USA). A mixture of analyte solution and α-cyano-4-hydroxycinnamic acid as matrix was applied to the sample plate and air-dried. Mass calibration was obtained using the quasimolecular ions (MH^+) from an external calibration mixture provided by the manufacturer. Raw data were analyzed by using the computer software provided by the manufacturer and are reported as monoisotopic masses.

Alignment and modeling Amino acid sequence alignment was generated by using the T-coffee program (Notredame et al. 2000) using as option the Lalign algorithm (Huang and Miller 1991) from the FASTA package (Pearson and Lipman 1988). The template was selected on the basis of high alignment score and resolution of the X-ray structure. Following these criteria, the template chosen was *Trametes versicolor* laccase (1GYC: pdb code). 3D models were generated using the SWISS-MODEL web server by means of the project mode option that allows us to submit a manually optimized modeling request. In this way it is possible to control a wide range of parameters, e.g., selection of template and gap placement in the alignment (Schwede et al. 2003).

Nucleotide sequence accession numbers The EMBL accession numbers of the sequences of the *P. ostreatus* genes encoding the small subunits of POXA3a and POXA3b laccases reported in this paper are AM409318 and AM409319, respectively.

Results

Gene cloning The amino acid sequence of some peptides from the tryptic digest of POXA3a small subunit had been previously determined (Palmieri et al. 2003). On the basis of these sequence data, degenerate oligonucleotide primers were designed and used in amplification experiments on *P. ostreatus* genomic DNA. A *P. ostreatus* genomic library (Giardina et al. 1995) was screened using the amplified fragment. Two different genes (1 and 2) encoding the small subunits were identified, whose sequences only differed for six bases. The corresponding cDNAs were amplified, using specific oligonucleotides. One of the isolated gene clones enclosed a 5′-untranslated region (UTR) extending about 200 nucleotides upstream the ATG (EMBL accession no. AM409319). In this region, some stretches closely matching consensus sequences of regulatory elements have been recognized: one putative MRE (115–121), two putative XREs (64–69 and 74–79), and one putative HSE (89–102).

The encoded amino acid sequences are reported in Fig. 1 and resulted to be different for two residues at positions 68 and 86. The deduced protein sequences did not share significant homology with any protein sequence in data banks.

As far as POXA3 large subunit is concerned, some cDNA fragments with sequences differing from that already published (Palmieri et al. 2003) were found by analyzing several clones. An insertion of 12 bp (5′-CTTGTTCAC TAG-3′) at the 3′ terminus of intron XII was detected, thus indicating that alternative splicing could occur at the junction between exons XII and XIII, with canonical splicing sequences conserved in both cases. The encoded amino acid sequence linear function test replacement (LFTR) is inserted between P267 and L268 of the mature large subunit sequence. MALDI-MS tryptic peptide maps of the large subunits showed the presence of both forms, with and without the four amino acid insertion, in both POXA3a and POXA3b purified isoenzymes. As a matter of fact, besides the peak at 1990.7 *m/z* corresponding to peptides 266–285 of the POXA3 sequence, a peak at 1821.1 *m/z* was recorded in both MALDI spectra. The

```
MFFRPSLLVSVAALLVSVVASPLHPRQSSN   30

TNPAIYQAISVLSRQIHVNIPELNTLQASG   60

GATDLTVRNELNELTDAFTLAAATITNTAV   90
     G                   A
SSGDTTNFPTNDDISITYAVALQLVASTAA  120

GLKQVNSLTTYTTMMSDLDPAIAALHVALN  150

RTLPNSINLVRVMMLDAQQFLTQAGLTQSR  180

ASLGFA
```

Fig. 1 Amino acid sequence of POXA3a and POXA3b small subunits. Amino acid substitutions in POXA3b small subunit are indicated. The putative signal peptide is *underlined*, and the six amino acid residues cleaved at the N termini are *doubly underlined*. The *N*-glycosylation site is indicated by an *arrow*

additional peak was assigned to peptides 272–289 of the POXA3_LFTR sequence (according to Fig. 4 sequence numbering), due to the occurrence of the additional tryptic site of this variant.

Small subunit characterization On the basis of the protein sequences deduced from the corresponding cDNAs (Fig. 1), MALDI-MS tryptic peptide maps of POXA3a and POXA3b small subunits (18- and 16-kDa) were analyzed, allowing the amino acid sequences differing for two residues to be assigned (Table 1). In fact, mass signals at *m/z* 2546.6 and 2546.8 corresponding to the sequenced peptides 45–68 are only present in POXA3a 18- and 16-kDa subunits, respectively, indicating that gene 1 encodes the protein containing R at position 68, corresponding to the POXA3a small subunits.

It has been previously reported that none of the POXA3 small subunits could be directly sequenced because they have all blocked N termini. Furthermore, when the peptide at *m/z* 1932.5 was analyzed by Edman degradation, no sequence was obtained, thus suggesting that this fragment has a blocked N terminus (Palmieri et al. 2003).

The mass signals at about *m/z* 1932 present in all the tryptic maps (Table 1) occurred 17 Da lower than the mass value expected for peptides 27–44, suggesting that Q27 had been converted into a pyrrolidone carboxyl residue. A putative signal peptide, corresponding to fragments 1–20, was recognized submitting the cDNA deduced amino acid sequences to SignalIP software (http://www.cbs.dtu.dk/services/). Therefore, further six amino acid residues at the N terminus of each small subunit (from S21 to R26) should have been cleaved and Q27 cyclization should have occurred, giving rise to the blocked N termini.

A single putative *N*-glycosylation site was inferred at N150 in both small subunit sequences. The signal at *m/z* 4261.0 found in the POXA3a 18-kDa small subunit (Table 1) was assigned to peptides 124–151 on the basis of sequence data. Therefore, N150 has to be modified in this peptide by high-mannose-type glycans with two mannose residues linked to the pentasaccharide core. On the other hand, the signals at 3046 *m/z* observed in the mass spectra of POXA3a and POXA3b 16-kDa small subunits (Table 1) were assigned to the unmodified peptides 124–151. Therefore, the mass differences between the 16- and 18-kDa subunits of both POXA3a and POXA3b are due to the presence of a glycosidic moiety attached to N150 in the 18-kDa subunits (Fig. 1). These results were confirmed by incubation of POXA3a and POXA3b isoenzymes with endoglycosidase H followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Fig. 2). After deglycosylation treatment, the 18-kDa small subunit signals disappeared and only bands with mobility corresponding to the 16-kDa subunits were observed.

Taking into account the identified posttranslational modifications, the expected molecular mass values of the small subunits are in agreement with those experimentally determined (Palmieri et al. 2003).

Isoenzyme heterogeneity When POXA3a or POXA3b were analyzed on Phenyl Superose chromatography, two active peaks from each isoenzyme have been eluted with the salt gradient, corresponding to the complexes formed by the large subunit and either the 18- or 16-kDa small subunit, respectively (Fig. 3). Therefore, four heterodimeric POXA3 isoenzymes have been separated (POXA3a16, POXA3a18, POXA3b16, and POXA3b18), as demonstrated by SDS-PAGE analysis. Peptides originated from both splicing variants described above were found in the MALDI MS maps of all the four isoforms. When kinetic parameters or stabilities at different temperatures or pH values of the four complexes were analyzed, no significant differences could be observed. Moreover, two further peaks eluted at the end

Table 1 MALDI-MS analyses of tryptic peptides from POXA3a and POXA3b small subunits

Peptides (sequence 1)	MH^+ theoretical	POXA3a 16 kDa sub. MH^+ exper.	POXA3a 18 kDa sub. MH^+ exper.	Peptides (sequence 2)	MH^+ theoretical	POXA3b 16 kDa sub. MH^+ exper.	POXA3b 18 kDa sub. MH^+ exper.
21–26	706.4			21–26	706.4		
27–44	1949.0	1932.2	1931.0	27–44	1949.0	1932.5	1932.0
45–68	2547.3	2546.8	2546.6	45–123	8015.1		
69–123	5615.8						
124–151	3046.5	3045.4		124–151	3046.5	3047.0	
N-glyc. 124–151			4261				
152–161	1126.7	1126.6	1126.5	152–161	1126.7	1126.8	1126.0
162–180	2138.1	2138.0	2136.9	162–180	2138.1	2138.3	2138.3
181–186	565.3			181–186	565.3		

sub. Subunit, *exper.* experimental

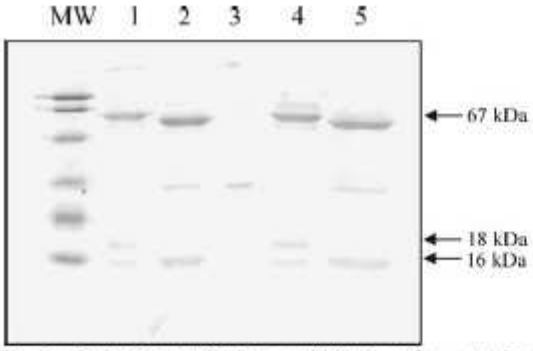


Fig. 2 SDS-PAGE of POXA3a and POXA3b before and after treatment with endoglycosidase H (*EndoH*). *Lane 1* POXA3a, *lane 2* POXA3a after EndoH treatment, *lane 3* EndoH, *lane 4* POXA3b, *lane 5* POXA3b after EndoH treatment

of the gradient of the POXA3b Phenyl Superose chromatography, corresponded to the free 18- and 16-kDa small subunits as demonstrated by mass mapping analysis, accounting for about 10% (mol/mol) of the complexes. On the other hand, large subunits have never been found isolated, and are only present in an associated form.

POXA3 large subunit model The structure of POXA3 large subunit was modeled onto the *T. versicolor* laccase structure 1GYC. POXA3 large subunit exhibits 51% identity with the template, while POXA3 small subunits do not show adequate identity with any protein in the Protein Data Bank database to enable us to build a 3D model.

The multiple alignment of POXA3 large subunit sequence with the other laccases whose structures had been resolved (Fig. 4) highlights some differences among them. According with the 3D model, the majority of the amino acid substitutions with respect to these other laccases are randomly distributed on the protein surface. Some interesting insertions or deletions are located in the loops delineating the phenolic substrate cavity (Fig. 5). One of these insertions (H157-Q161) is located in L154-I170 loop involved in substrate interaction. A similar insertion was only found in *Rigidoporus* laccase but its structure was not resolved, probably because of the high flexibility of this long loop. Therefore, positions of these amino acids and their effect on the overall structure cannot be accurately deduced on the basis of the data nowadays available.

Another loop (P267–L279) on the same side of the cavity shows an insertion corresponding to the stretch G275–N278. A similar insertion was only found in *Coprinopsis* laccase; however, in this case it is positioned far from the substrate cavity. On the other hand, the LFTR insertion found in the POXA3 variant is located in this loop next to P267. Because the residues of this loop delineating the cavity are T269–P273, the LFTR insertion should affect the position of the following residues belonging to the cavity.

On the opposite side of the cavity the G395–A398 loop is shorter than that found in most of the other laccases because of a deletion between G395 and A396. A similar structure was only found in the *Rigidoporus* laccase where this deletion seems to make T1 copper more accessible to the solvent.

Three residues belonging to the substrate cavity are remarkably different with respect to the other laccases. The S167 residue, located on the boundary of the cavity, substitutes a hydrophobic residue, which should be directly involved in protein–ligand interaction. In position 211 of the POXA3 sequence the residue D, conserved in most of the other laccases, is replaced with R. In the crystallographic structure of *T. versicolor* laccase (pdb code 1KYA) complexed with the arylamine (Bertrand et al. 2002), this aspartate is hydrogen-bonded via the terminal oxygen of its

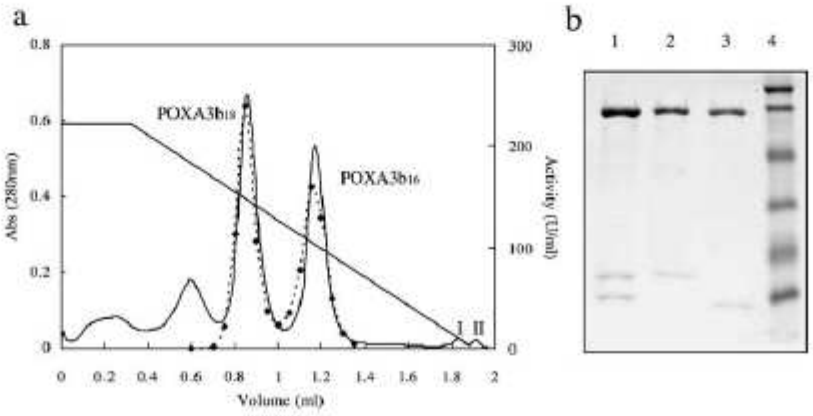


Fig. 3 a Chromatography profile of POXA3b loaded on Phenyl Superose. Peaks marked with *I* and *II* are 18- and 16-kDa small subunits, respectively. *Continuous line*, Abs at 280 nm; *dotted line*, laccase activity. b SDS-PAGE of POXA3b (*lane 1*), POXA3b18 (*lane 2*), POXA3b16 (*lane 3*), and molecular weight markers (*lane 4*)

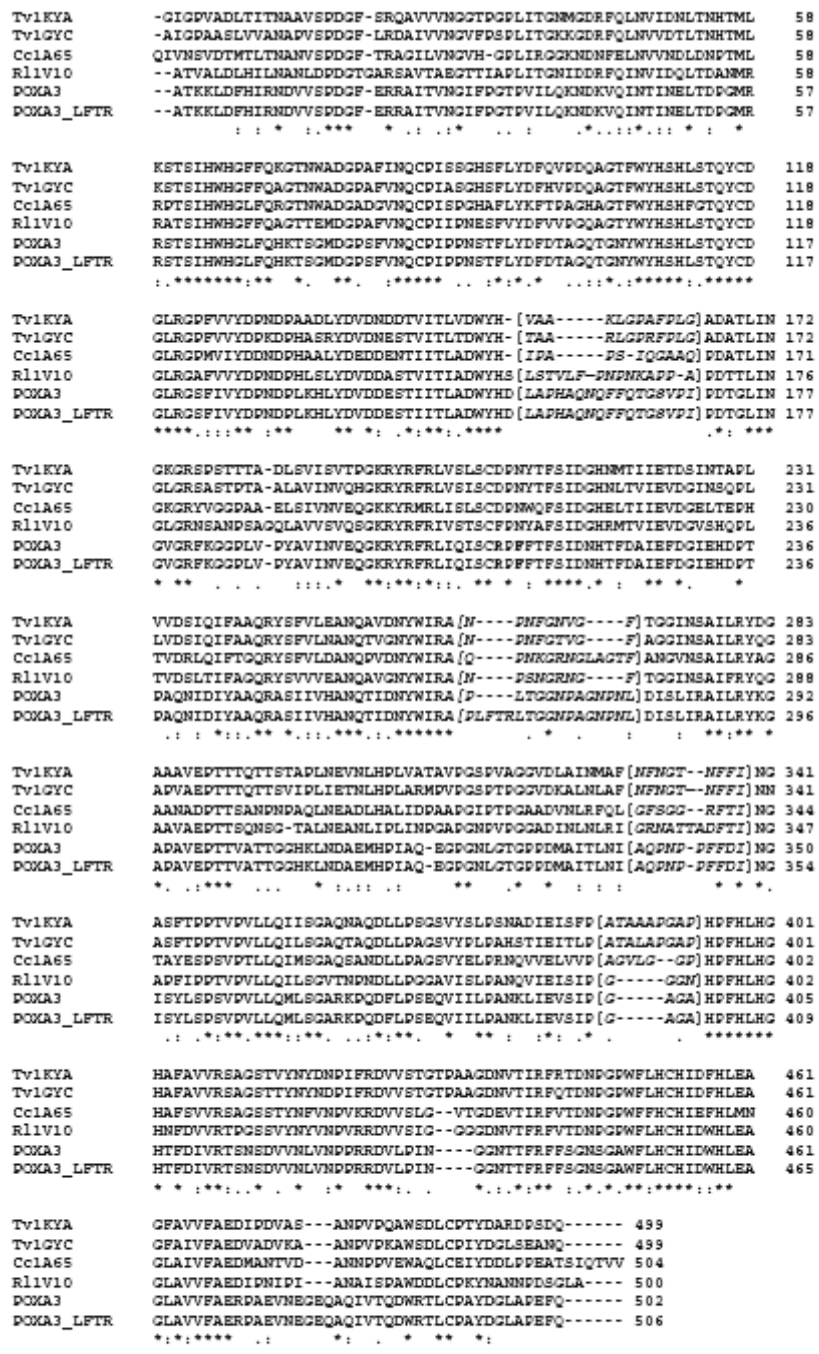


Fig. 4 Multiple sequence alignment of laccases whose 3D structures had been resolved and the two POXA3 large subunit forms (POXA3 and POXA3_LFTR). *Tv1KYA T. versicolor* laccase (pdb code 1KYA), *Tv1GYC T. versicolor* laccase (pdb code 1GYC), *Cc1A65 C. cinerea* laccase (pdb code 1A65), *Rl1V10 Rigidoporus lignosus* laccase (pdb code 1V10). The loops involved in substrate binding (Larrondo et al. 2003) are in *italic* and in *brackets*. The three residues of the substrate cavity different from the other laccases are in *bold*

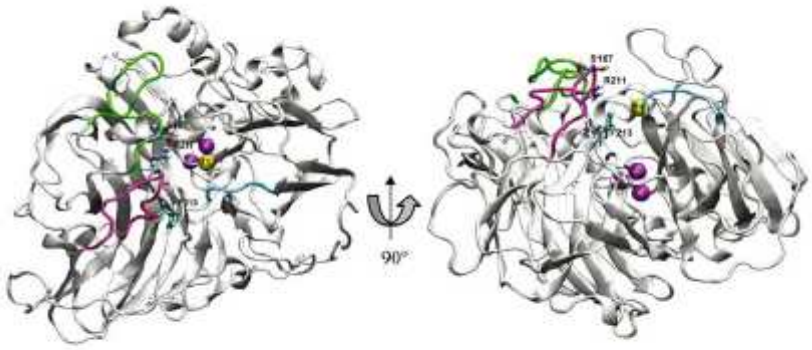


Fig. 5 Ribbon representation of POXA3 large subunit (without LFTR sequence insertion) model. Loops forming T1 copper cavity and showing insertions or deletions with respect to the other analyzed laccases are highlighted (L154-I170 *green*, P267-L279 *purple*, and G395-A398 *cyan*). S167, R211, and F213 are shown in *stick representation*. Copper atoms are depicted as *spheres* (T1 *yellow*, T2 and T3 *purple*)

side chain to the amino group of 2,5-xylidine. This residue is also substituted in the *Rigidoporus* laccase, but in this case by the hydrophobic residue F. Moreover, the apolar residue F in position 213 of the POXA3 replaces the conserved polar residue N found in most of the other laccase sequences.

Discussion

P. ostreatus produces a wide variety of laccase isoenzymes (Giardina et al. 1999; Palmieri et al. 1993, 1997, 2003). This heterogeneity is certainly due to the multiplicity of the corresponding encoding genes, but the isoenzymatic pattern is made more complex by the different posttranslational modifications (proteolytic processing, glycosylation, etc.) that each isoenzyme can undergo. Moreover, alternative splicing can occur, further increasing the isoform number. The subfamily of POXA3 isoenzyme is emblematic of this complexity, consisting of heterodimers showing further heterogeneity on both subunits. In this paper, a deep analysis of all POXA3 forms is reported as an example of complexity of fungal isoenzyme pattern.

A unique gene codes for both POXA3a and POXA3b large subunits. Among the large subunit transcripts, the presence of two splice variants has been observed. In both cases, the splicing occurs at consensus splice junction sequences. A variant, 12 nucleotides longer than the previously published nucleotide sequence, encodes a protein with an insertion of four amino acids between P287 and L288. Nowadays, several authors have reported alternative splicing in fungi (Larrondo et al. 2004; Lodato et al. 2003; Stuardo et al. 2005), but, at the best of our knowledge, no information is available on the encoded protein variants. We have demonstrated that both splicing variants are translated and the two corresponding proteins were found in the purified POXA3a and POXA3b complexes. Further investigations are needed to evaluate the effect of this insertion on the enzymatic activity.

Only two genes differing for six nucleotides encoding POXA3 small subunits have been identified. Because of the dikaryotic state of the analyzed fungal strain, these genes could be allelic forms of the same gene. Analysis of the 5'-UTR of *poxa3b* small subunit gene showed the presence of putative MRE and XRE, as found in the corresponding region of the large subunit gene, thus suggesting the hypothesis of transcriptional coregulation of the two genes. The structural characteristics of POXA3 small subunits and the differences among them were analyzed by comparison of the sequences deduced from the encoding genes and the tryptic maps of the proteins. These analyses allowed us to ascertain that (1) POXA3a and POXA3b small subunits differ for two amino acid substitutions, (2) differences between 18- and 16-kDa subunits of both POXA3a and POXA3b are only due to posttranslational modification of the 16 kDa subunit by *N*-glycosylation at Asn150, (3) all the small subunits undergo a proteolytic cleavage at their N termini followed by Gln27 cyclization. Therefore, the unique difference between POXA3a and POXA3b enzymes seems to be related to the occurrence of two amino acid substitutions in their small subunits. The different chromatographic behavior of the two isoenzymes on anionic exchange chromatography (Palmieri et al. 2003) can be caused by the substitution of the positively charged amino acid R in POXA3a small subunit with a neutral one in POXA3b. Moreover, each isoenzyme can be further split into two forms differing for the presence of the glycosydic moiety in the small subunits. All of these isoenzymes contain both variants of the large subunit originating from the alternative splicing.

POXA3 large subunit is clearly homologous to fungal laccases and shows all known *consensus* sequences involved in copper binding (Palmieri et al. 2003). The reason why only this protein, among several other laccase isoenzymes produced in the same culture, is found associated with another polypeptide chain is to elucidate. The sequence of the small subunit does not show significant homology with any sequence in data banks, therefore, no indication on the

function of this subunit can be inferred from its primary structure. A 3D model of the large subunit has been built to investigate on the peculiar characteristics of this enzyme. On the other hand, no model of the POXA3 small subunit can be built because of the absence of known homologous proteins. Analysis of the POXA3 large subunit protein surface does not indicate atypical areas that could be pointed out as putative small subunit binding regions. Major differences shown by POXA3 large subunit 3D model with respect to the other laccase structures are located on one side of the substrate cavity. One of these substitutions concerns the residue R211, replacing the D residue conserved in most of the other laccases and shown to be involved in interaction with substrate (Bertrand et al. 2002; Madzak et al. 2006). The differences observed in the POXA3 large subunit structure, compared with other laccases, could be responsible of the complex formation and therefore of the singleness of these proteins.

Acknowledgment This work was supported by the European Commission, Sixth Framework Program (SOPHIED contract NMP2-CT2004-505899), by grants from the Ministero dell'Università e della Ricerca Scientifica (Progetti di Rilevante Interesse Nazionale, PRIN), INTAS (International Association for the promotion of cooperation with scientists from the New Independent States of the former Soviet Union, ref. no. 03-51-5889), and from Centro Regionale di Competenza BioTekNet. The authors thank Prof. Piero Pucci and Dr. Franca Fraternali for helpful discussions.

References

Baldrian P, Valaskova V, Merhautova V, Gabriel J (2005) Degradation of lignocellulose by *Pleurotus ostreatus* in the presence of copper, manganese, lead and zinc. Res Microbiol 156:670–676

Bertrand T, Jolivalt C, Briozzo P, Caminade E, Joly N, Madzak C, Mougin C (2002) Crystal structure of a four-copper laccase complexed with an arylamine: insights into substrate recognition and correlation with kinetics. Biochemistry 41:7325–7333

Cullen D (1997) Recent advances on the molecular genetics of ligninolytic fungi. J Biotechnol 53:273–289

Curir P, Thurston CF, Daquila F, Pasini C, Marchesini A (1997) Characterization of a laccase secreted by *Armillaria mellea* pathogenic for Genista. Plant Physiol Biochem 35:147–153

De Souza CGM, Peralta RM (2003) Purification and characterization of the main laccase produced by the white-rot fungus *Pleurotus pulmonarius* on wheat bran solid state medium. J Basic Microbiol 43:278–286

Giardina P, Cannio R, Martirani L, Marzullo L, Palmieri G, Sannia G (1995) Cloning and sequencing of a laccase gene from the lignin-degrading basidiomycete *Pleurotus ostreatus*. Appl Environ Microbiol 61:2408–2413

Giardina P, Palmieri G, Scaloni A, Fontanella B, Faraco V, Cennamo G, Sannia G (1999) Protein and gene structure of a blue laccase from *Pleurotus ostreatus*. Biochem J 341:655–663

Huang X, Miller W (1991) A time-efficient, linear-space local similarity algorithm. Adv Appl Math 12:337–357

Kilaru S, Hoegger PJ, Kües U (2006) The laccase multi-gene family in *Coprinopsis cinerea* has seventeen different members that divide into two distinct subfamilies. Curr Genet 50:45–60

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage *T4*. Nature 227:680–685

Larrondo LF, Salas L, Melo F, Vicuna R, Cullen D (2003) A novel extracellular multicopper oxidase from *Phanerochaete chrysosporium* with ferroxidase activity. Appl Environ Microbiol 69:6257–6263

Larrondo LF, González B, Cullen D, Vicuña R (2004) Characterization of a multicopper oxidase gene cluster in *Phanerochaete chrysosporium* and evidence of altered splicing of the *mco* transcripts. Microbiology 150:2775–2783

Lodato P, Alcaino J, Barahona S, Retamales P, Cifuentes V (2003) Alternative Splicing of Transcripts from *crtI* and *crtYB* genes of *Xanthophyllomyces dendrorhous*. Appl Environ Microbiol 69:4676–4682

Lucas MC, Jacobson JW, Giles NH (1977) Characterization and in vitro translation of polyadenylated messenger ribonucleic acid from *Neurospora crassa*. J Bacteriol 130:1192–1198

Madzak C, Mimmi MC, Caminade E, Brault A, Baumberger S, Briozzo P, Mougin C, Jolivalt C (2006) Shifting the optimal pH of activity for a laccase from the fungus *Trametes versicolor* by structure-based mutagenesis. Protein Eng Des Sel 19:77–84

Min KL, Kim YH, Kim YW, Jung HS, Hah YC (2001) Characterization of a novel laccase produced by the wood-rotting fungus *Phellinus ribis*. Arch Biochem Biophys 392:279–286

Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217

Palmieri G, Giardina P, Marzullo L, Desiderio B, Nitti G, Cannio R, Sannia G (1993) Stability and activity of a phenol oxidase from the ligninolytic fungus *Pleurotus ostreatus*. Appl Microbiol Biotechnol 39:632–636

Palmieri G, Giardina P, Bianco C, Scaloni A, Capasso A, Sannia G (1997) A novel white laccase from *Pleurotus ostreatus*. J Biol Chem 272:31301–31307

Palmieri G, Cennamo G, Faraco V, Amoresano A, Sannia G, Giardina P (2003) Atypical laccase isoenzymes from copper supplemented *Pleurotus ostreatus* cultures. Enzyme Microb Technol 33:220–230

Palmieri G, Cennamo G, Sannia G (2005) Remazol Brilliant Blue R decolourisation by the fungus *Pleurotus ostreatus* and its oxidative enzymatic system. Enzyme Microb Technol 36:17–24

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Perry CR, Matcham SE, Wood DA, Thurston CF (1993) The structure of laccase protein and its synthesis by the commercial mushroom *Agaricus bisporus*. J Gen Microbiol 139:171–178

Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31:3381–3385

Stuardo M, Larrondo LF, Vásquez M, Vicuña R, González B (2005) Incomplete processing of peroxidase transcripts in the lignin degrading fungus *Phanerochaete chrysosporium*. FEMS Microbiol Lett 242:37–44

Thakker GD, Evans CS, Rao KK (1992) Purification and characterization of laccase from *Monocillium indicum* Saxena. Appl Microbiol Biotechnol 37:321–323

Wahleithner JA, Xu F, Brown KM, Brown SH, Golightly EJ, Halkier T, Kauppinen S, Pedersen A, Schneider P (1996) The identification and characterization of four laccases from the plant pathogenic fungus *Rhizoctonia solani*. Curr Genet 29:395–403

Yaver DS, Xu F, Golightly EJ, Brown KM, Brown SH, Rey MW, Schneider P, Halkier T, Mondorf K, Dalboge H (1996) Purification characterization, molecular cloning, and expression of two laccase genes from the white rot basidiomycete *Trametes villosa*. Appl Environ Microbiol 62:834–841

References

[1] Radzicka A, Wolfenden R. A proficient enzyme. *Science* (1995);267:90-3.

[2] Schoemaker HE, Mink D, Wubbolts MG. Dispelling the myths--biocatalysis in industrial synthesis. *Science* (2003);299:1694-7.

[3] Straathof AJ, Panke S, Schmid A. The production of fine chemicals by biotransformations. *Curr Opin Biotechnol* (2002);13:548-56.

[4] Gupta R, Beg QK, Khan S, Chauhan B. An overview on fermentation, downstream processing and properties of microbial alkaline proteases. *Appl Microbiol Biotechnol* (2002);60:381-95.

[5] Ito S. Alkaline cellulases from alkaliphilic Bacillus: enzymatic properties, genetics, and application to detergents. *Extremophiles* (1997);1:61-6.

[6] Pandey A, Nigam P, Soccol CR, Soccol VT, Singh D, Mohan R. Advances in microbial amylases. *Biotechnol Appl Biochem* (2000);31 (Pt 2):135-52.

[7] Kashyap DR, Vohra PK, Chopra S, Tewari R. Applications of pectinases in the commercial sector: a review. *Bioresour Technol* (2001);77:215-27.

[8] Zaks A. Industrial biocatalysis. *Curr Opin Chem Biol* (2001);5:130-6.

[9] Baldrian P. Fungal laccases - occurrence and properties. *FEMS Microbiol Rev* (2006);30:215-42.

[10] Hoopes JT, Dean JF. Ferroxidase activity in a laccase-like multicopper oxidase from Liriodendron tulipifera. *Plant Physiol Biochem* (2004);42:27-33.

[11] Thurston C. The structure and function of fungal laccases. *Microbiology* (1994);140:19-26.

[12] Bourbonnais R, Paice MG. Oxidation of non-phenolic substrates. An expanded role for laccase in lignin biodegradation. *FEBS Lett* (1990);267:99-102.

[13] Huttermann A, Mai C, Kharazipour A. Modification of lignin for the production of new compounded materials. *Appl Microbiol Biotechnol* (2001);55:387-94.

[14] Setti L, Giuliani S, Spinozzi G, Pifferi P. Laccase catalyzed-oxidative coupling of 3-methyl 2-benzothiazolinone hydrazone and methoxyphenols. *Enzyme Microb Technol* (1999);25:285-9.

[15] Minussi RC, Pastore GM, Duràn N. Potential applications of laccase in the food industry. *Trends Food Sci Technol* (2002);13:205–16.

[16] Kuznetsov, Shumakovich GP, Koroleva OV, Yaropolov AI. On applicability of laccase as label in the mediated and mediatorless electroimmunoassay: effect of distance on the direct electron transfer between laccase and electrode. *Biosens Bioelectron* (2001);16:73-84.

[17] Sannia G, Giardina P, Luna M, Rossi M, Buonocore V. Laccase from *Pleurotus ostreatus* *Biotechnol Lett* (1986);8:797-800.

[18] Palmieri G, Giardina P, Bianco C, Fontanella B, Sannia G. Copper induction of laccase isoenzymes in the ligninolytic fungus Pleurotus ostreatus. *Appl Environ Microbiol* (2000);66:920-4.

[19] Palmieri G, Giardina P, Bianco C, Scaloni A, Capasso A, Sannia G. A novel white laccase from Pleurotus ostreatus. *J Biol Chem* (1997);272:31301-7.

[20] Palmieri G, Giardina P, Marzullo L, Desiderio B, Nitti G, Cannio R, et al. Stability and activity of a phenol oxidase from the ligninolytic fungus Pleurotus ostreatus. *Appl Microbiol Biotechnol* (1993);39:632-6.

[21] Giardina P, Palmieri G, Scaloni A, Fontanella B, Faraco V, Cennamo G, et al. Protein and gene structure of a blue laccase from Pleurotus ostreatus1. *Biochem J* (1999);341 (Pt 3):655-63.

[22] Palmieri G, Cennamo G, Faraco V, A. A, Sannia G, Giardina P. Atypical laccase isoenzymes from copper supplemented Pleurotus ostreatus cultures. *Enzyme Microb Technol* (2003);33:220-30.

[23] Wesenberg D, Kyriakides I, Agathos SN. White-rot fungi and their enzymes for the treatment of industrial dye effluents. *Biotechnol Adv* (2003);22:161-87.

[24] Yoshitake A, Katayama Y, Nakamura M, Iimura Y, Kawai S, Morohoshi N. N-Linked carbohydrate chains protect laccase-III from proteolysis in Coriolus versicolor. *J Gen Microbio* (1993);139:179-85.

[25] Solomon E, Sundaram U, Machonkin T. Multicopper Oxidases and Oxygenases. *Chem Rev* (1996);96:2563-606.

[26] Messerschmidt A, Steigemann W, Huber R, Lang G, Kroneck PM. X-ray crystallographic characterisation of type-2-depleted ascorbate oxidase from zucchini. *Eur J Biochem* (1992);209:597-602.

[27] Messerschmidt A. In: Messerschmidt A, editor. Multi-Copper Oxidases. Singapore, 1997.

[28] Bertrand T, Jolivalt C, Briozzo P, Caminade E, Joly N, Madzak C, et al. Crystal structure of a four-copper laccase complexed with an arylamine: insights into substrate recognition and correlation with kinetics. *Biochemistry* (2002);41:7325-33.
[29] Petersen LC, Degn H. Steady-state kinetics of laccasse from Rhus vernicifera. *Biochim Biophys Acta* (1978);526:85-92.
[30] Ducros V, Brzozowski AM, Wilson KS, Brown SH, Ostergaard P, Schneider P, et al. Crystal structure of the type-2 Cu depleted laccase from Coprinus cinereus at 2.2 A resolution. *Nat Struct Biol* (1998);5:310-6.
[31] Piontek K, Antorini M, Choinowski T. Crystal structure of a laccase from the fungus Trametes versicolor at 1.90-A resolution containing a full complement of coppers. *J Biol Chem* (2002);277:37663-9.
[32] Antorini M, Herpoel-Gimbert I, Choinowski T, Sigoillot JC, Asther M, Winterhalter K, et al. Purification, crystallisation and X-ray diffraction study of fully functional laccases from two ligninolytic fungi. *Biochim Biophys Acta* (2002);1594:109-14.
[33] Hakulinen N, Kiiskinen LL, Kruus K, Saloheimo M, Paananen A, Koivula A, et al. Crystal structure of a laccase from Melanocarpus albomyces with an intact trinuclear copper site. *Nat Struct Biol* (2002);9:601-5.
[34] Garavaglia S, Cambria MT, Miglio M, Ragusa S, Iacobazzi V, Palmieri F, et al. The structure of Rigidoporus lignosus Laccase containing a full complement of copper ions, reveals an asymmetrical arrangement for the T3 copper pair. *J Mol Biol* (2004);342:1519-31.
[35] Dementin S, Arnoux P, Frangioni B, Grosse S, Leger C, Burlat B, et al. Access to the active site of periplasmic nitrate reductase: insights from site-directed mutagenesis and zinc inhibition studies. *Biochemistry* (2007);46:9713-21.
[36] Seitz C, Ameres S, Forkmann G. Identification of the molecular basis for the functional difference between flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase. *FEBS Lett* (2007);581:3429-34.
[37] Penning TM, Jez JM. Enzyme redesign. *Chem Rev* (2001);101:3027-46.
[38] Simossis VA, Heringa J. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* (2005);33:W289-94.
[39] Xu F, Berka RM, Wahleithner JA, Nelson BA, Shuster JR, Brown SH, et al. Site-directed mutations in fungal laccase: effect on redox potential, activity and pH profile. *Biochem J* (1998);334 (Pt 1):63-70.
[40] Gelo-Pujic M, Kim HH, Butlin NG, Palmore GT. Electrochemical studies of a truncated laccase produced in Pichia pastoris. *Appl Environ Microbiol* (1999);65:5515-21.
[41] Robzyk K, Kassir Y. A simple and highly efficient procedure for rescuing autonomous plasmids from yeast. *Nucleic Acids Res* (1992);20:3790.
[42] Rost B. Twilight zone of protein sequence alignments. *Protein Eng* (1999);12:85-94.
[43] Schwede T, Koop J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* (2003);31:3381-33815.

Chapter 3

Introduction: Malaria as a public health problem

Still in the 21st century malaria represent one of the major threats for health in many parts of the world. It is estimated that between 400 and 900 million febrile episodes occur every year only for African children, with a minimum of 750,000 deaths (probably up to 3 million). Sub-Saharan Africa is the most affected region in the world, but malaria is also a serious problem in several other places, such as South-East Asia, Oceania, Middle East and Latin America. Although transmission of the parasites via the *Anopheles* mosquitoes was discovered in 1880 by Charles Alphonse Louis Laveran, malaria remains a serious public-health problem for roughly 40% of the world's population.Human malaria is caused by one of four parasites: *Plasmodium falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*. Of these the two species *P. falciparum* and *P. vivax* are the most widely distributed, both causing a vast amount of human pain. All trials to develop a new vaccine have paid considerably more attention to *P. falciparum* than to *P. vivax*. This is because *P. falciparum* is more virulent and is responsible for more than 95% of the malaria deaths worldwide; it is also the only human malaria parasite resistant to chloroquine and other antimalarial drugs. Humans are the intermediate hosts in the plasmodia life cycle and are usually infected when bitten by a female mosquito carrying the parasite. Transmission can be prevented by the use of repellents and insecticides, but nevertheless according to World Health Organization (WHO) the prevalence of malaria is on the rise.

The cost of malaria is particularly heavy for developing countries, because the economic effects of preventing and treating malaria are very high and present therefore a large burden on the health budget. The economics of developing new pharmaceuticals for different diseases, including malaria, are such that there is a great disparity between the public health importance of the disease and the amount of resources invested in developing new cures.

Due to the lack of cost-effective treatments and the emergence of resistance, a malaria vaccine is likely to be crucial in reducing both the morbidity and the mortality of this disease. Drug resistance has been implicated in the increase of malaria to new areas and re-emergence of malaria in areas where the disease had been eliminated. Population movement has introduced resistant parasites to areas previously free of drug resistance. In general, resistance appears to happen through spontaneous mutations that confer reduced sensitivity to a class of drugs. For some drugs, only a single point mutation is necessary to give resistance, while for other drugs, multiple mutations appear to be required. Vaccination is a successful method of disease control. Many are the factors that make the development of a malaria vaccine an incredibly difficult challenge. In fact it is the complex biology of the life cycle of the parasite that impedes vaccine development. The malaria parasite presents thousands of antigens to the human immune system that vary throughout its life cycle. Identifying those that may prove to be vaccine targets is complicated. Most vaccines are targeted at individual stages of the malaria life cycle, although it is likely that only the development of a multistage vaccine will offer complete protection to both visitors to, and residents of, a malaria-endemic area. Even if the development of a vaccine would result successful, other issues such as cost, distribution, education, and compliance will have to be addressed.

Life cycle of the malaria parasite

The life cycle of the malaria parasite is very complex. It can be divided into two distinct stages, one occurring in the mosquito host and another in the human host (**Fig 3.1**). The sporozoites are transmitted into the subcutaneous tissue via the bite of infected mosquito. The sporozoites enter hepatocytes shortly after inoculation into the blood circulation; it's demonstrated that this process involves surface proteins of the sporozoites and host cell surface molecule. Once inside hepatocytes, the sporozoites develop into schizonts containing thousands of merozoites. The merozoites are released into the blood flowing through the sinusoids of the liver after 6–15 days, depending on the *Plasmodium* spp. The circulating merozoites infect erythrocytes within a few seconds and begin the asexual cycle. Within the erythrocyte, the parasite passes through the ring and trophozoite stages before the production of the erythrocytic schizont; each mature schizont produced about 20 merozoites. The merozoites are released after 48 h, with the destruction of the erythrocyte, and are then free to infect further cells. Finally some merozoites within the erythrocytes differentiate into immature gametocytes. On biting an infected intermediate host the mosquito ingests blood containing the parasite and the normally asexually dividing bloodstream forms (merozoites) die. The gametocytes are stimulated and the sexual stage of the cycle then continues in the mosquito.

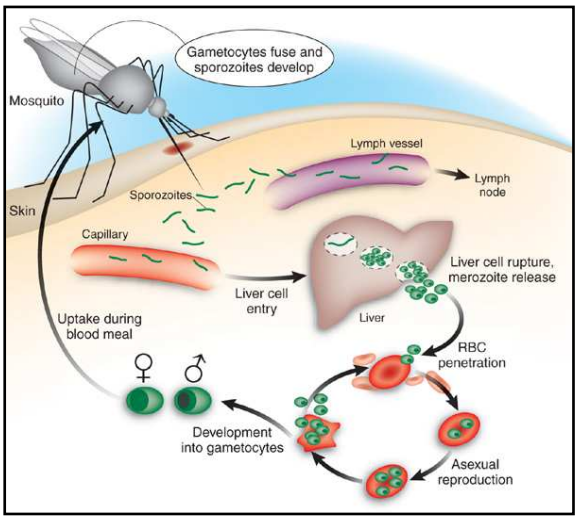


Figure 3.1: Schematic illustration of the life cycle of malaria parasites. Image from [1]

The aim of the present work

Vaccination is a useful tool to control the malaria disease. Studies of the immune response to the malaria parasite in humans and in animal models have provided a wealth of information on possible protective mechanisms. Antigen identification have produced many candidates for subunit vaccines [2]. Single proteins or peptides have been shown to be at least in part protective when are used in vaccination studies [3]. While different antigens epitopes have been identified, such as those recognized by neutralizing antibodies [4], multiple target vaccine have not yet been successfully designed.Proteins on the surface of the merozoite are excellent targets for development of vaccines against malaria, because they are exposed to antibody. One of these proteins is the merozoite surface protein 1 (MSP-1). Its C-terminal sequence is the target of protective antibody for this reason it can be a candidate for vaccine development. This protein is involved in the initial recognition of the erythrocyte; MSP-1 interacts with spectrin on the cytoplasmic face of the erythrocyte membrane [5].

Initially MSP-1 is synthesized as a large (~200 kDa) precursor during intracellular merozoite development; it is present on the surface of the released merozoite and it's anchored by a glycosyl phosphatidyl inositol tail [6]. MSP-1 is cleaved by proteases in two processing steps; the second step releases the bulk of the protein from the surface and goes to completion during successful red blood cell invasion. The precursor is first cleaved into four pieces in a primary processing step that occurs on merozoite release; at some point between merozoite release and completion of erythrocyte invasion, the membrane-bound component ($MSP\text{-}1_{42}$) of this surface complex is further cleaved at a single site to form two fragments, $MSP\text{-}1_{33}$ and $MSP\text{-}1_{19}$ (secondary processing). After that the majority of the complex is being shed from the parasite surface, leaving only a 96 amino acid residue C-terminal fragment ($MSP\text{-}1_{19}$) onto the surface of the parasite. The structure of *P. falciparum* $MSP1_{19}$ has been solved by NMR [7] (**Fig 3.2**), and its homologues from *Plasmodium cynomolgi* [8] and *Plasmodium knowlesi* [9] have been solved by X-ray crystallography.

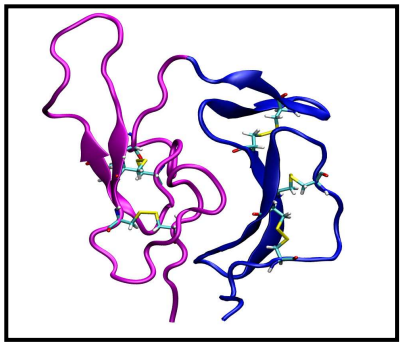


Figure 3.2: Ribbon representation of $MSP1_{19}$. Domain 1 is shown in blue and in magenta the domain 2. In licorice representation in highlighted the cysteines forming disulphide bridge [7].

This fragment is composed of two epidermal growth factor (EGF)-like motifs. Each domain presents a major stretch of antiparallel β-sheet containing the third and fourth Cys residues of each domain, as expected for an EGF-like fold, as well as an additional minor antiparallel β-sheet at the C-terminal end of domain-1, similar to almost all EGF family members. The C-terminal fragment of MSP1, is a leading candidate antigen for development of a vaccine against the blood stages of the malaria parasite. In fact, antibodies binding to the C-terminus of *P. falciparum* MSP-1 can inhibit both the processing and erythrocyte invasion (inhibitory antibodies). On the other hand, other antibodies that bind to either the C-terminal sequence or elsewhere in the molecule are termed blocking antibodies, because on binding prevent the binding of the inhibitory antibodies [10, 11].

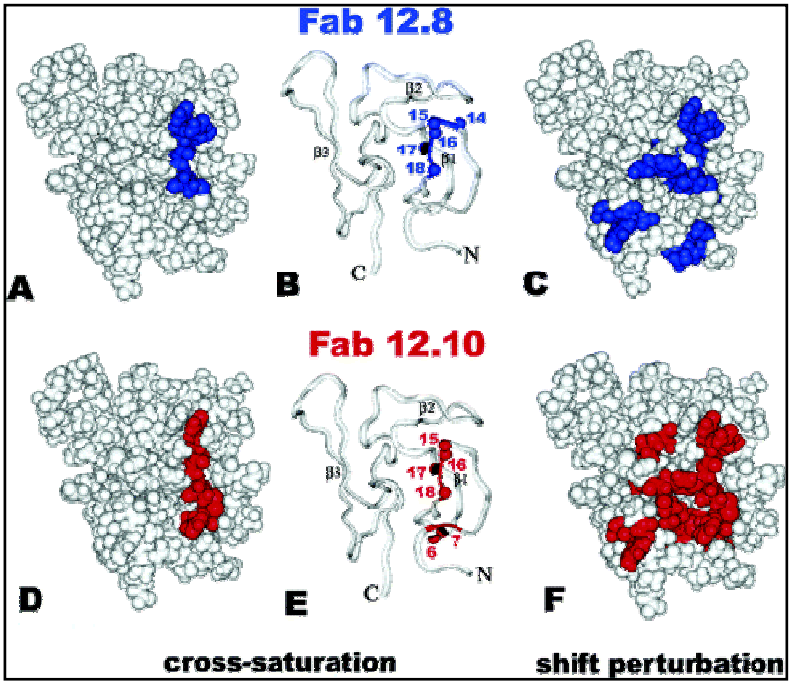


Figure 3.3: Interface residues for the *P. falciparum* $MSP1_{19}$-Fab complexes mapped by TROSY NMR. The residues most sensitive to cross-saturation are shown by colour (panel A and D) and residues that experienced a large chemical shift perturbation are displayed in colour (panels C and F) on the antigen surface. The amide groups of the same residues are also highlighted on the ribbon diagram (B and E) [11].

Morgan *et al* [12, 13] have studied by TROSY NMR experiments, the complexes of *P. falciparum* $MSP1_{19}$ with Fab fragments from three monoclonal antibodies. Two of them, Fab12.8 and Fab12.10, have shown a parasite-inhibitory activity *in vitro* [6]. NMR epitope mapping of Fab12.8 and Fab12.10 have shown a close relationship between binding sites for the two inhibitory antibodies, distinct from the location of the non-inhibitory antibody. Cross-saturation mapping of the epitopes recognized by these two parasite inhibitory anti-$MSP1_{19}$ monoclonal antibodies exhibited a remarkable degree of similarity. The most affected residues identified for each of the inhibitory monoclonal antibodies cluster in a region located on one of the two broad surfaces of the disk-shaped $MSP1_{19}$ molecule (**Fig 3.3**). From these experiments, six residues have been individuated for antibody 12.10 (Gln 6, Cys 7, Asn 15-Cys 18) and five residues for antibody 12.8 (Gln 14-Cys 18). These results are in good agreement with previous binding studies showing that antibody 12.8 bounded to $MSP1_{19}$ domain 1 by itself, while binding of mAb 12.10 required the presence of both domains [14]. This was consistent with the greater degree of perturbation of domain 2 residues by Fab12.10. It is important to remember that, one limitation of this method is that it provides no information on contacts to the side chains of antigen residues. In order to investigate this further, we decided to model the antibody structures, and to investigate protein-protein interactions with the antigen (including potential side chain contacts) by means of docking calculations and therefore to refine the resulting complexes from the docking procedure by performing Molecular Dynamic simulations. Moreover ,in order to validate the docking procedure the X-ray structure of a Fab complex formed with *Pf*$MSP1_{19}$ [15] was used as test case.

PROTEINS: Structure, Function, and Bioinformatics 66:513–527 (2007)

Interaction of Malaria Parasite-Inhibitory Antibodies With the Merozoite Surface Protein $MSP1_{19}$ by Computational Docking

Flavia Autore,[1] Sara Melchiorre,[1] Jens Kleinjung,[3] William D. Morgan,[2] and Franca Fraternali[1,4]*

[1]*Dipartimento di Chimica Organica e Biochimica, Università di Napoli Federico II, Complesso Universitario Monte Sant'Angelo, via Cinthia, 80126, Naples, Italy*

[2]*Division of Parasitology, National Institute for Medical Research, Mill Hill, London, NW7 1AA, United Kingdom*

[3]*Division of Mathematical Biology, National Institute for Medical Research, Mill Hill, London, NW7 1AA, United Kingdom*

[4]*Bioinformatics Unit, Randall Division of Cell and Molecular Biophysics, New Hunt's House, London, SE1 1UL, United Kingdom*

ABSTRACT **Merozoite surface protein 1 (MSP1) of the malaria parasite *Plasmodium falciparum* is an important vaccine candidate antigen. Antibodies specific for the C-terminal maturation product, $MSP1_{19}$, have been shown to inhibit erythrocyte invasion and parasite growth. Specific monoclonal antibodies react with conformational epitopes contained within the two EGF-like domains that constitute the antigen $MSP1_{19}$. To gain greater insight into the inhibitory process, the authors selected two strongly inhibitory antibodies (designated 12.8 and 12.10) and modeled their structures by homology. Computational docking was used to generate antigen–antibody complexes and a selection filter based on NMR data was applied to obtain plausible models. Molecular Dynamics simulations of the selected complexes were performed to evaluate the role of specific side chains in the binding. Favorable complexes were obtained that complement the NMR data in defining specific binding sites. These models can provide valuable guidelines for future experimental work that is devoted to the understanding of the action mechanism of invasion-inhibitory antibodies. Proteins 2007;66:513–527.** © 2006 Wiley-Liss, Inc.

Key words: antibody–antigen complex; docking; molecular dynamics; NMR; malaria; *Plasmodium falciparum*; MSP1 (merozoite surface protein 1)

INTRODUCTION

Malaria is one of the most widely spread diseases, with more than 40% of the world population at risk of being infected. The development of effective malaria vaccines is therefore a mandatory public health challenge.[1] Acute clinical malaria from infection with *Plasmodium falciparum* is associated with replication of the asexual blood-stage parasite (merozoite) in circulating erythrocytes. After invasion, the merozoite multiplies within the host cell causing its rupture and to liberate several newly formed merozoites. These, in turn, invade other erythrocytes thus dramatically increasing the level of parasitemia that causes the typical symptoms of malaria. The merozoite expresses a number of surface proteins, one or more of which are thought to mediate the initial interaction between the parasite and host erythrocyte.[2,3] Merozoite surface protein 1 (MSP1) is one of the most extensively studied molecules of *P. falciparum*[4] and is present on the surface of the released merozoite in the form of a multicomponent protein complex derived via proteolytic processing.[5,6] In between the merozoite release and the completion of erythrocyte invasion, the membrane-bound component ($MSP1_{42}$) of this surface complex is further cleaved at a single site to form two fragments, $MSP1_{33}$ and $MSP1_{19}$. As a consequence, the majority of the complex is shed from the parasite surface, leaving only $MSP1_{19}$ exposed. Therefore, this protein is considered a very promising malaria vaccine candidate.[7] Several monoclonal antibodies (mAbs) directed against $MSP1_{19}$ have been identified, and some inhibit invasion of erythrocytes in vitro,[8] while some inhibit both secondary processing of MSP1 and erythrocyte invasion.[9] Those antibodies that are specific for $MSP1_{19}$ but are not inhibitory can be classified as blocking antibodies (that interfere with the binding of inhibitory antibodies) and neutral antibodies (which bind to $MSP1_{19}$ but neither inhibit invasion nor block the binding to inhibitory antibodies).[10,11]

Abbreviations: CDR, complementarity determining region; El, electrostatic potential; LJ, Lennard–Jones potential; mAb, monoclonal antibody; MD, molecular dynamics; VH, variable heavy chain; VL, variable light chain; SR, short range.

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsors: Medical Research Council, European Commission through the EUROMALVAC consortium and CEINGE Advanced Biotechnology consortium

*Correspondence to: Franca Fraternali, Dipartimento di Chimica Organica e Biochimica, Università di Napoli Federico II, Complesso Universitario Monte Sant'Angelo, via Cinthia, 80126, Naples, Italy. E-mail: franca.fraternali@kcl.ac.uk

Received 10 April 2006; Revised 4 August 2006; Accepted 9 August 2006

Published online 15 December 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21212

The structure of *P. falciparum* $MSP1_{19}$ has been solved by NMR,[12] and its homologues from *Plasmodium cynomolgi*[13] and *Plasmodium knowlesi*[14] have been solved by X-ray crystallography. The molecule is composed of two EGF-like domains forming a disk-shaped structure. The binding mode to specific antibodies was subject of different structural studies. Recently, the crystal structure of $MSP1_{19}$ in complex with the Fab-fragment of mAb G17.12 has been solved.[15] This antibody is certainly not inhibitory, but it has not yet been determined whether it is a neutral or a blocking antibody. A recent study has been devoted to the discrimination between blocking, inhibitory, and neutral antibodies for $MSP1_{19}$ mapping *via* mutagenesis the epitopes for inhibitory mAbs 12.8 and 12.10[16] and for blocking mAbs such as 1E1 and 7.5.[17]

NMR epitope mapping has been performed with several $MSP1_{19}$-specific antibodies using chemical shift perturbation as well as the more precise cross-saturation method.[18,19] From this experiment, six residues have been individuated for antibody 12.10 (Gln 6, Cys 7, Asn 15-Cys 18) and five residues for antibody 12.8 (Gln 14-Cys 18). This technique is convenient for mapping protein–protein interactions, since there is one observable probe for each residue in the antigen. In practice, a fraction of the backbone amide signals may not be observed because of exchange line broadening or signal overlap. One limitation of this method is that it provides no information on contacts to the side chains of antigen residues. To investigate this further, we decided to model the antibody structures, and to investigate protein–protein interactions with the antigen (including potential side chain contacts) by means of docking calculations. Prior knowledge of the epitope on the antigen, as obtained from NMR cross-saturation, provides reliable restraints to guide these calculations.[20]

The rigid body docking algorithm ZDOCK proved to be very effective in predicting protein–protein complexes in previous studies.[21] The effectiveness of the method relies on a scoring function that properly takes into account the major determinants of protein–protein association, such as shape, electrostatic complementarities and desolvation energies.[22,23] The method profits from a very efficient search engine, based on a Fast Fourier transformation algorithm,[21] and was found to be one of the best protein–protein interaction predictors for antibody–antigen prediction.[24] Precise mapping of the epitopes that are recognized by the mAbs 12.8 and 12.10 has important implications for rational vaccine design and development of antibody-based therapeutics. Recently, the importance of local protein flexibility on protein–protein association has been shown,[25] and therefore we decided to refine and optimize the resulting complexes from the docking procedure by performing Molecular Dynamic simulations in explicit solvent. The use of MD simulations for the unbound state of protein–protein complexes has shown that core interface residues are less mobile than the rest of the surface and therefore rigid docking procedures do not account for different mobilities.[26] This implies that while rigid docking is acceptable for core residues, it is not ideal for the residues on the surface of the protein. We observed that the docked complexes are not optimally positioned. The two proteins tend to be in general more distant in the model than in the reference complex, probably because of ZDOCK's weak electrostatic term (see also Discussion). Therefore, we opted for an MD refinement that should correct for the relative positioning through a more accurate force field and added residue mobility at the interface.

The methodology used here may prove to be a valuable general approach to the problem of antibody–antigen interactions, as in cases where experimental structure resolution is difficult.

METHODS

Sequence Alignments and Modeling of the Antibodies

Sequence alignments were generated by using the T-Coffee method[27] that uses the Lalign algorithm[28] from the FASTA package.[29] The best alignments were selected according to not only the alignment score but also the length of the hypervariable loop gaps in the complementarity determining region (CDR). When multiple choices were possible, we opted for the template structure with the highest resolution. Three-dimensional models were generated using the SWISS-MODEL web server by means of the project mode option that allows for submitting a manually optimized modeling request. This way it is possible to control a wide rage of parameters, for example, selection of template and gap placement in the alignment.[30] Since only the variable regions of the antibody were sequenced, the antibodies were modeled in the form of single chain variable fragments (scFv), with the light and heavy chains connected by a flexible linker segment.

NMR Filter

The cross-saturation technique involves excitation of protons in one macromolecule (in this case the antibody), followed by transfer of magnetization to protons within the second macromolecular component of the complex (the antigen in this case). Cross-saturation is detected by monitoring the backbone amide cross peaks in a HSQC or TROSY correlation spectrum of the antigen, which is perdeuterated and uniformly labeled with ^{15}N. The maximum range of saturation transfer has been estimated at 7 Å, indicating that for each antigen residue where a strong cross-saturation effect is observed, at least one antibody proton is located within this distance from the backbone amide proton of that residue. The nomenclature used for the models is the following: models 12.10_2FBJ and 12.10_1JHL were built using templates 2FBJ[31] and 1JHL,[32] respectively; model 12.8_1IND was built using template 1IND[33] and model 12.8_1GGI using template 1GGI.[34]

Complex Selection and Clustering

The mAbs were submitted to a docking procedure with antigen $MSP1_{19}$ using the program ZDOCK 2.3.[21,22,35] Only the framework of the antibody was excluded from the docking searches.

Complexes were selected on the basis of experimental NOE couplings (see NMR filter) and classified by cluster analysis as explained in the following. From NMR experiments lists of cross-saturation antigen HN atoms in possible contact with ANY of the antibody's H atoms are given.

For antibody 12.10 the HN of residues 6, 7, 15, 16, 17, 18 (List 1), while for antibody 12.8 the HN of residues 14, 15, 16, 17, 18 (List 2) have been identified by these NMR experiments. These few restraints are clearly not sufficient for an unambiguous complex selection; therefore, a number of analyses (distances from the hypervariable regions, solvent accessibilities, manual inspections) are necessary to be able to propose models.

An HN-H (antigen–antibody) distance below the cut off of 7 Å (see NMR filter) was counted as a restraint satisfaction. The total number of restraint satisfactions for each of the 2000 model structures was taken as the rank of the structure and a list sorted by structure rank was compiled. The top 17 models (selected structures) for 12.8 and the top 9 models for 12.10 of the sorted list were analyzed in detail by clustering. Clustering was performed in two dimensions: structural similarity as first dimension and constraint satisfaction similarity as second dimension. Structural similarity values were obtained from structural alignments of all pairs of selected structures, using a modified contact-based alignment program.[36] The root mean square deviation (RMSD) values of all $n(n-1)/2$ pairs form a triangular matrix (excluding diagonal elements). Constraint satisfaction similarity was calculated the same way by calculating the difference of restraint satisfactions between all pairs, yielding another triangular matrix. The two triangular matrices were combined in a nonsymmetrical square matrix, and two-dimensional clustering was performed using the "heatmap" function of the R environment.[37] This function performs hierarchical clustering in two dimensions based on the distance matrix values and generates a false color image of the cluster matrix.

The qualitatively same result was obtained when the triangular difference matrices were both scaled to the value range [0:1], but the grouping of favorable solutions was more scattered.

The docking of mAb G17.12 with antigen $MSP1_{19}$ was performed as a test case to check the reliability of the filtering through external restraints and the applied clustering. To test the validity of using a model for the antibody in the docking procedure, we have additionally constructed a model for the mAb G17.12 sequence from the template 1IAI with 83% sequence identity. We have also performed the docking on the modeled structure that we will call MG17.12 and have compared the results with the ones from the original complex (see Results).

For the analysis of the data relative to G17.12, an artificial dataset of restraints had to be created (in this case there are no NOE experiments but one can assume to use the restraints from the crystal structure in an analogous way). Specifically, the distance between 20 HN atoms in the antigen (residues 8–15, 21–28, and 37–41) and 58 hydrogen atoms of the antibody's heavy chain loops (11 on H1, 11 on H2, and 36 on H3) was calculated. In this case the top 50 models were considered for further analysis.

Clustering allowed us to divide the solution space into subgroups and to investigate whether the grouping is similar in terms of structural similarity and restraint satisfaction. We expect favorable subsolutions to be structurally similar and to have similar rank. Moreover, we expect favorable solutions to have a high number of restraint satisfactions, meaning a low rank. Therefore, in Figure 1 and Figure S4(A,B), Supplementary Materials, we were selecting for clusters of intense red cells (low distance) with low-ranking labels.

Molecular Dynamics Simulations

All simulations were performed with the GROMACS package[38] using the GROMOS96 force field.[38] The structures were solvated in a box of SCP water,[39] in a solution of 50 m*M* NaCl. All simulated boxes contained about 17,000 water molecules. Simulations were carried out at a constant temperature of 300 K. The Berendsen algorithm was applied for temperature and pressure coupling.[40] After a first steepest descent energy minimization with positional restraints on the solute, the LINCS algorithm was used to constrain the bonds[41] and to carry out an initial 200 ps simulation with the positions of the solute atoms restrained by a force constant of 3000 kJ/(mol nm^2) to let the water diffuse around the molecule and to equilibrate. The particle mesh Ewald method (PME)[42] was used for the calculation of electrostatic contribution to non bonded interactions (grid spacing of 0.12 nm). A reaction field method[43] was used for the calculation of electrostatic contributions in the simulations of the X-ray structure, and ions were added to neutralize the system.

During a number of simulation trials we observed the flexible loop of the antigen encompassing residues 65–75 to be rapidly attracted toward the antibody and to directly interact with the side chains of loops L1 and L2 of antibody 12.10 and with loops H1 and H2 of antibody 12.8. This attraction force would govern the antigen positioning at the binding surface. In some cases, as for antibody 12.10, it would even obstruct a proper docking to the CDRs. Considering that the NMR data rule out significant interactions with this loop, we decided to keep the residues 65–75 restrained to the NMR conformation during the simulation with a force constant of 3000 kJ/(mol nm^2).

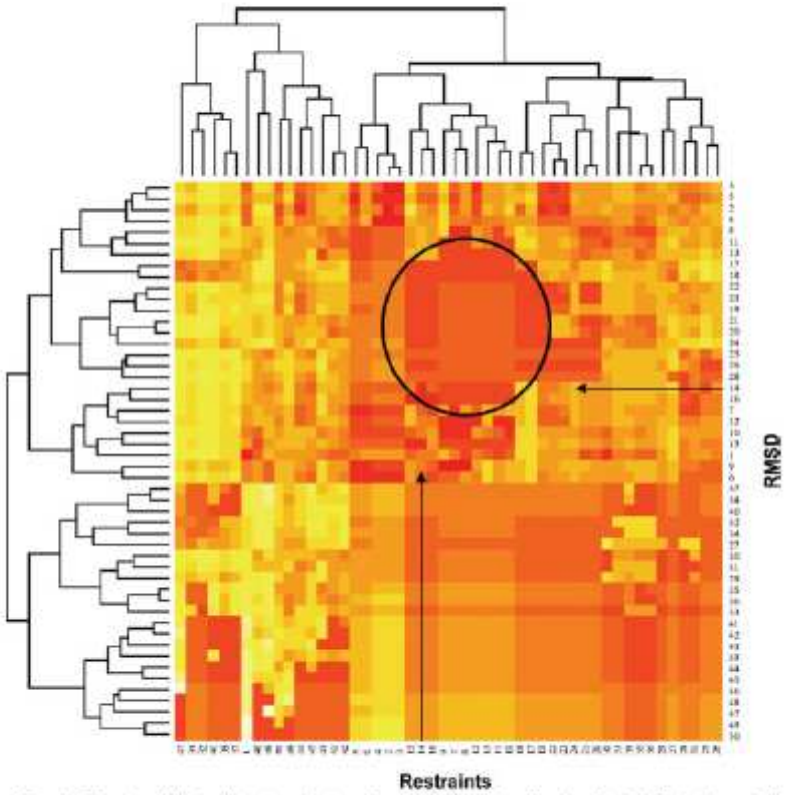


Fig. 1. Selection of docked structures by two-dimensional clustering. Number labels indicate the rank of the complex (see text). The root means square deviation dimension was obtained from a distance matrix of structural alignments of all pairs of the 50 top ranking structures, while the restraints dimension was obtained by calculating the difference matrix of restraint satisfactions between all pairs. Clustering was performed by using the "heatmap" function of the R environment,[37] which generates a false color image of the clustering matrix. The circle indicates the region of the subclusters with low pairwise distances (dark red).

RESULTS

Test of the Docking and Complex Selection Procedure

To test the adopted docking procedure, in particular the filtering of structures based on external restraints, we chose as a test case a known complex of the antigen $MSP1_{19}$ with the Fab-fragment of mAb G17.12, recently solved by X-ray crystallography.[35]

We separated antigen and antibody and generated 2000 docked structures, starting from randomly chosen relative orientations. Favorable complexes were preselected by a combination of restraint filtering and two-dimensional-clustering as follows: structures were ranked by the number of HN-H (antigen–antibody) distances below 7 Å for a given list of 20 HN atoms and 58 H atoms (see Methods for details). These restraints are chosen in analogy with the NMR data that are available for the unknown complexes of antigen $MSP1_{19}$ with antibodies 12.10 and 12.8. The 50 top-ranked complexes were clustered according to their pairwise structural differences (first cluster dimension) and their pairwise restraint–satisfaction differences (second cluster dimension). The cluster map in Figure 1 highlights a favorable subgroup within those top ranking complexes that show low pairwise differences in both dimensions.

One favorable representative of this selection was submitted to 5 ns of MD simulation. As a reference, we also performed 2 ns of MD on the native structure, to confirm that the structure would be stable under the adopted conditions (data not shown).

In Figure 2(A), a ribbon representation of the X-ray structure of complex mAbG17.12-$MSP1_{19}$ is shown. The binding region as calculated with POPS[45,46] is shown in dark grey. In Figure 2(B), a superimposition of the selected docked structure (black) and the X-ray structure (grey) is shown, while in Figure 2(C) the superim-

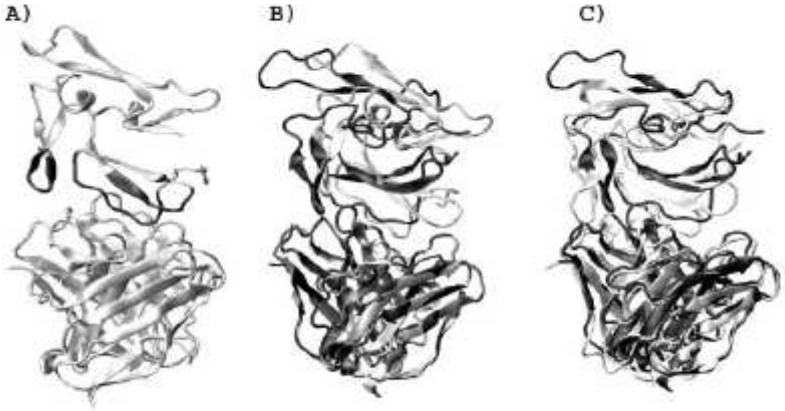


Fig. 2. (A) Ribbon representation of the X-ray structure. The residues of the binding-site buried upon complex formation as calculated by POPScomp[44] are displayed in black. (B) Superimposition of the X-ray structure (grey) and the docked complex (black). (C) Superimposition of the docked complex before (black) and after (grey) 5 ns of MD.

position of the docked structure (black) and the final conformer after 5 ns of MD (grey) is displayed.

The docked complex, although very similar to the X-ray structure, is initially not in its optimal configuration, as can be seen from the comparison of the distances between the centers of mass of antibody and antigen [Fig. 3(A)]. These are about 30 Å because of the size of the antibody.

The X-ray structure shows a value of 31 Å for this distance, while the docked complex shows an initial value of 33 Å. During the MD simulation, antigen and antibody move closer to each other and the distance between their centers of mass shortens by 2.0 Å, after about 3800 ps becomes very close to the X-ray value, and toward the end stabilizes around 30 Å.

In Table I, the solvent accessible surface areas buried upon formation of the two complexes (X-ray and final conformer from MD) are reported. By comparison with the X-ray structure we found that 50% of the original contacts formed by the antibody and 60% of those in the antigen were preserved. As shown already in Figure 2(A), the antibody contains contacts in all surface regions that are buried in the X-ray complex, and it manages to draw also the loop of the second EGF domain [residues 65–73, in black in Fig. 2(A)] closer to the antibody with respect to the structure from ZDOCK. The time evolution of the number of hydrogen bonds that are formed between the antigen and the CDRs of the light (grey line) and heavy (black line) chains of the antibody is shown in Figure 3(B). During MD, we observe an increase of the number of hydrogen bonds at the interface, reaching a total number of eight at the end of the simulation. For the X-ray structure the total number of antibody–antigen hydrogen bonds was reported to be 11.[15]

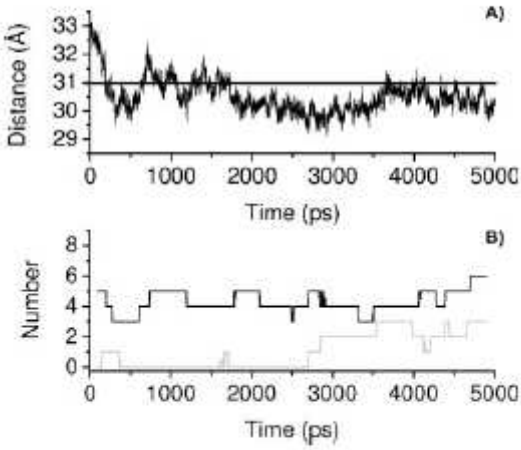


Fig. 3. (A) Time evolution of the distance between the centers of mass of antibody and antigen of the docked complex. The distance of centers of mass in the X-ray structure is represented by a continuous line at 31 Å. (B) Number of hydrogen bonds formed between the antigen and the CDRs of the antibody in the docked complex. The hydrogen bonds formed with the heavy chain and the light chain are displayed in black and in grey, respectively.

For the modeled structure of G17.12 onto the template FAB 730.1.4 (MG17.12), the docking performed with the modeled structure (MG17.12) gave rise to a best complex very similar to the one obtained with the original crystal structure. The superimposition between the complex from the modeled structure and the X-ray structure is reported in Figure S1, Supplementary Materials. The complexes from G17.12 and MG17.12 have a similar

TABLE I. Solvent Accessible Surface Areas Buried Upon Formation of the Complex Between Antigen $MSP1_{19}$ and Antibody G17.12

Complex ZDock	ΔSASA Å^2	X-ray	ΔSASA Å^2	CDR
(a) Comparison of the selected conformer from ZDOCK with the X-ray structure				
Antibody				
THR H30	25.6			H1
THR H31	**43.9**	**THR H31**	**14.3**	**H1**
TRP H50	**13.4**	**TRP H50**	**38.3**	**—**
		ASN H52	19.7	H2
HISH 54	**56.1**	**HIS H54**	**38.4**	**H2**
		GLN H59	31.8	—
ASN H99	13.3			H3
TYR H100	**34.9**	**TYR H100**	**21.7**	**H3**
TYR H100a	**66.4**	**TYR H100a**	**70.5**	**H3**
ARG H100b	**51.1**	**ARG H100b**	**42.1**	**H3**
		SER L30	34.2	L1
		SER L31	22.3	L1
TYR L33	**40.9**	**TYR L33**	**52.5**	**—**
		HIS L93	23.9	L3
		HIS L94	19.1	L3
		TYR L97	17.3	L3
Antigen				
		VAL 8	47.1	
LYS 9	**108.5**	**LYS 9**	**69.0**	
LYS 10	**16.4**	**LYS 10**	**56.6**	
GLN 11	**144.0**	**GLN 11**	**168.0**	
CYS 12	12.0			
PRO 13	**52.7**	**PRO 13**	**21.8**	
GLN 14	**132.4**	**GLN 14**	**37.7**	
		ASP 23	32.9	
		GLU 24	45.5	
		ARG 25	27.3	
GLN 36	35.2			
ASP 39	**100.7**	**ASP 39**	**110.5**	
LYS 40	36.9			
(b) Comparison of the last conformer after 5 ns of MD simulations of the X-ray structure				
Antibody				
THR H28	23.8			H1
THR H30	49.7			**H1**
THR H31	**23.9**	**THR H31**	**14.3**	H1
		TRP H50	38.3	H2
		ASN H52	19.7	H2
HIS H54	**54.9**	**HIS H54**	**38.4**	**H2**
		GLN H59	31.8	—
TYR H100	**20.8**	**TYR H100**	**21.7**	**H3**
TYR H100a	**73.5**	**TYR H100a**	**70.5**	**H3**
ARG H100b	**23.1**	**ARG H100b**	**42.1**	**H3**
ASP H101	18.0			H3
		SER L30	34.2	L1
		SER L31	22.3	L1
SER L32	16.3			L1
TYR L33	**79.6**	**TYR L33**	**52.5**	**—**
PHE L92	13.4			L3
HIS L93	**11.5**	**HIS L93**	**23.9**	**L3**
		HIS L94	19.1	**L3**
TYR L97	**20.0**	**TYR L97**	**17.3**	**L3**
Antigen				
		VAL 8	47.1	
LYS 9	**77.3**	**LYS 9**	**69.0**	
LYS 10	**76.8**	**LYS 10**	**56.6**	
GLN 11	**12.3**	**GLN 11**	**168.0**	
CYS 12	47.8			
PRO 13	**67.1**	**PRO 13**	**21.8**	
GLN 14	**101.5**	**GLN 14**	**37.7**	
		ASP 23	32.9	
GLU 24	**101.5**	**GLU 24**	**45.5**	
		ARG 25	27.3	
CYS 28	12.7			
ASP 39	**106.8**	**ASP 39**	**110.5**	
LYS 40	29.6			

In bold: matching proton restraints of the ZDOCK complex and the X-ray structure.

distance in rms from the X-ray complex: 2.8 and 3.2 Å, respectively. We decided therefore to use the original complex for further analysis.

On the basis of the above results, we considered the adopted complex modeling procedure to be reliable in detecting plausible complexes, and therefore we applied it to the modeling of the unknown complex structures of MSP_{19} and mAbs 12.8 and 12.10.

Modeling of the Antibodies

In order to model the structures of single chain variable fragments (scFv) of the two mAbs 12.8 and 12.10 (see Methods section), we selected the templates according to the guidelines of Morea et al.[47] The key role of loop H3 in the antigen recognition specificity restricts the choice of templates to the ones with the same length of this loop. In Table II, the PDB codes of these potential template structures are reported together with the H3 loop sequence and sequence identities. The structures that were finally selected as modeling templates are shown in bold. For mAb 12.10 we selected templates 1JHL and 2FBJ, because they share the most similar sequence in the central part of the H3 loop, which should be critical to the interaction with the antigen. The nomenclature used for the models is specified in the Methods section. Complete sequence alignments are reported in Figure 4. In Figure 5 the electrostatic surfaces of the modeled antibodies and of the antigen are displayed. The chosen view is from the top of the binding site of the CDR loops. The surfaces are quite different in the binding region: 12.8_1IND presents a more pronounced positive surface (blue) in the central region, while 12.8_1GGI presents a more negative surface (red) on the right side of this surface. For mAb 12.10, both modeled surfaces are on average more hydrophobic than 12.8, with 12.10_1JHL presenting a cavity with negative electrostatic potential in the central binding site region.

TABLE II. Parameters of Potential Template Structures for the Modeling of Antibodies 12.8 and 12.10

PDB code	Resolution Å	Identity %	Length (H3)	Sequence (H3)
12.8	**—**	**—**	**10**	CVQSVIDNWG
1IND[33]	**2.2**	**41% (30%)**	**10**	CASHRFVHWG
1GGI[34]	**2.8**	**48% (50%)**	**10**	CVQEGYIYWG
12.10	**—**	**—**	**14**	CARNWAYWFEDVWG
1BBD[51]	2.8	66% (36%)	14	CDGYYSYYDMDYWG
1JHL[32]	**2.4**	**66% (43%)**	**14**	CTRDDNYGAMDYWG
2FBJ[31]	**1.9**	**52% (43%)**	**14**	CARLHYYGYNAYWG
7FAB[52]	2.0	45% (57%)	14	CARNLIAGGIDVWG
1MFB[53]	2.1	51% (36%)	14	CTRGGHGYYGDYWG
1JEL[54]	2.8	65% (57%)	14	CARVMGEQYFDVWG

The length and the sequences of H3 loops are reported for each potential template structure. Percentages of identity with 12.8 and 12.10 given relative to the entire sequence; identities in parentheses are given relative to the H3 loop only. In bold: structures that were selected as templates.

The surface of 12.10_2FBJ is on the other hand more hydrophobic. The negatively charged surface of the $MSP1_{19}$ region, as identified by epitope mapping, should prove to be more complementary for interactions with either more hydrophobic surfaces or surfaces with fewer negative spots in the electrostatic potential.

We have also modeled the sequence of G17.12 mAb with the template FAB 730.1.4, as a control test (see methods section) and the sequence alignments are reported in Figure S2, Supplementary Materials. The rms deviation on the C^{α} atoms between the model and the X-ray mAbs is 1.39 Å, and the CDRs loops are very similar in sequence; therefore, we expect a similar behavior in terms of binding/selectivity.

Docked Structures

Docking of 12.10 and $MSP1_{19}$

We performed ZDOCK calculations with the antigen $MSP1_{19}$ and the two modeled antibodies 12.10_2FBJ and 12.10_1JHL. For the latter, none of the 2000 docked complexes that were passing the NMR filter (see Methods section) presented the antigen in a suitable position for binding, that is, located more or less at the centre of the binding CDR loops. Therefore, no further analysis was performed on this complex. For 12.10_2FBJ a total of nine complexes were satisfying the NMR data (Fig. S3 and Table SI, Supplementary Materials).

The clustering procedure described in the Methods section of these nine complexes is shown Figure S4(A) and we selected complex 1 (low distance, intense red cells with low ranking labels).

For this structure, 5 ns of MD simulations were performed, and the final conformer was used for further analysis. In Figure 6(A,B) the selected ZDOCK complex and the final MD conformer are shown superimposed. As for the case of the X-ray complex, the MD simulation has the effect of pulling the antigen closer to the antibody. In Table IIIa the areas that were buried by formation of the complex as calculated by POPScomp[44] are reported. All of the residues indicated by the NMR cross-saturation data (Gln 6, Cys 7, Asn 15-Cys 18) are in contact with the heavy chain of the antibody, and as can be seen from Table IIIb, four of the six residues of the NMR data set are close to the H3 loop. Specifically, for the antigen residues with strong cross-saturation, the model shows interaction of the side chain of Arg H57 (flanking loop H2) with residues Gln 6 and Cys 7, of side chains Tyr H100b and Trp H100c (loop H3) with residues Asn 15 to Cys 18, and residues Asn 15 and Ser 16 also near hydrophilic side chains of loops L1 and L3. These results are summarized in Figure 6(C): the NMR-predicted interaction surface is colored in blue, while the residues identified by POPScomp are plotted in orange. In addition, the residues that did not show sufficient burial despite respecting the NMR filter are colored in yellow. In Table IVa, Lennard–Jones (LJ) and electrostatic contributions (El) are reported for the intramolecular energies of the antibody and the antigen (mAb-mAb and $MSP1_{19}$-$MSP1_{19}$ contributions respectively), together with the intermolecular antigen–antibody contributions. For the latter term, which reflects the relative strength of the interaction, the El and LJ contributions become increasingly more favorable during the MD simulation, suggesting that the system relaxes toward a more favorable structure. The energetic contributions of the loops contain most of the entire binding interaction of the antibody: therefore, while these become more favorable, the antigen manages to position itself better at the CDRs. The final binding mode shows one side of the discoidal $MSP1_{19}$ facing the CDR loops, with the antigen loop comprising residues 65–75 pointing toward the left of the central binding site, facing antibody loops L2 and L1 [Fig. 6(A)].

Docking of 12.8 and MSP119

As for antibody 12.10 we performed ZDOCK calculations with the antigen $MSP1_{19}$ and the two modeled antibodies 12.8_1IND and 12.8_1GGI. Both gave rise to a series of complexes respecting the NMR filter: 16 complexes for 12.8_1IND and 15 complexes for 12.8_1GGI (Fig. S5 and Table SII, Supplementary Materials).

Additional NMR information was obtained by performing 3D-TROSY-NOESY experiments (W. Morgan, unpublished results) for the complexes of the 12.8 and 12.10 mAb

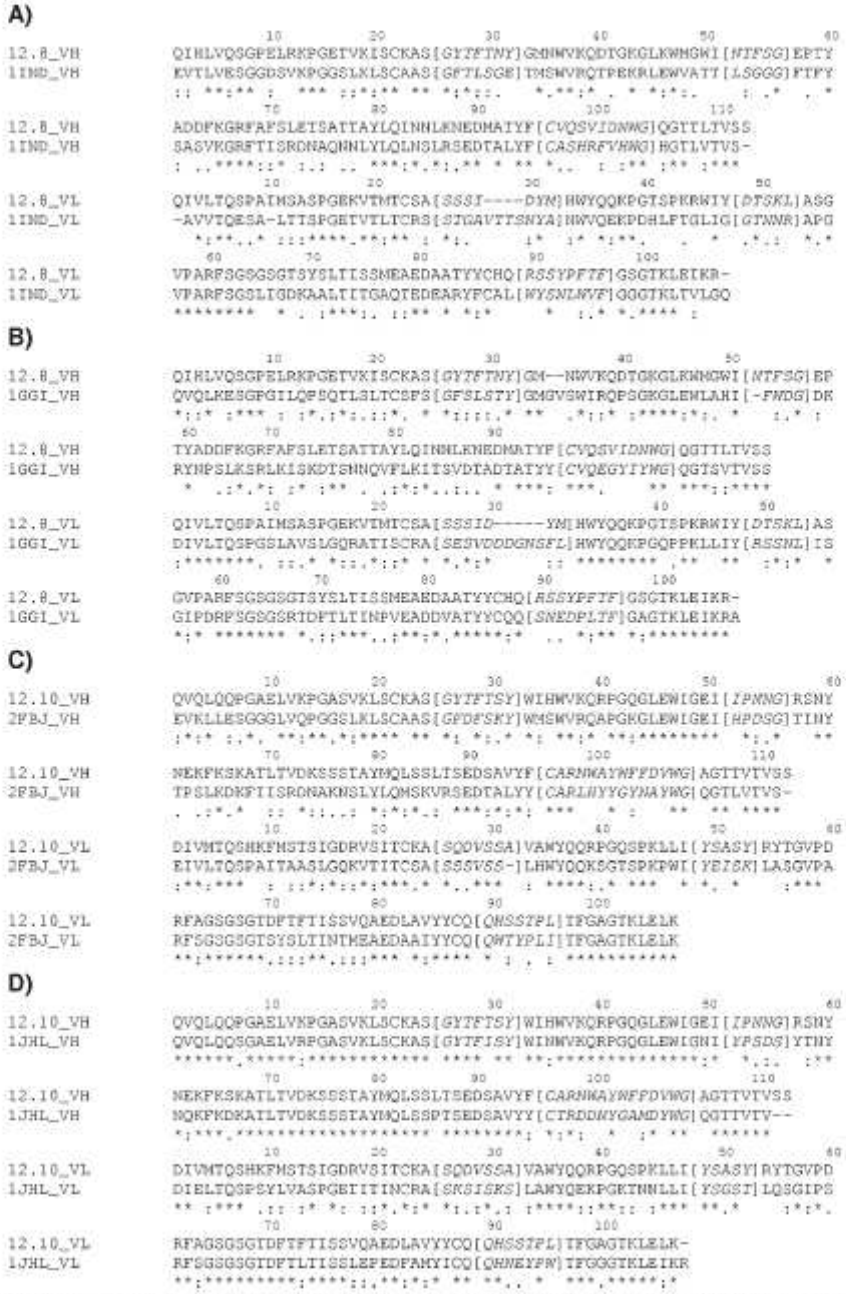


Fig. 4. Sequence alignments of loop H3 of the antibodies 12.8 and 12.10 with loop H3 of the selected templates. The alignments were obtained using T-Coffee.[27] Residues are numbered according to Chothia et al.[52] The CDR regions are highlighted in parentheses and in italic format.

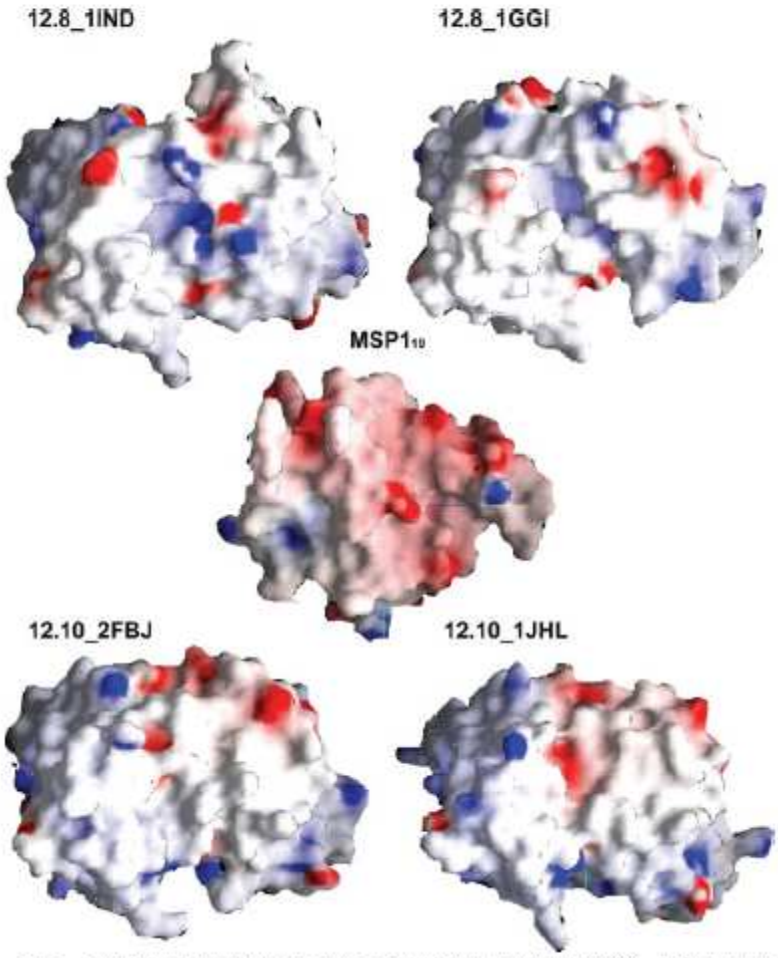


Fig. 5. Comparison of the electrostatic potential of the studied antibodies and MSP1$_{19}$. Each molecule is viewed from the top of the binding-site. The potential is ranged from −10 kT (red) to the maximal positive value +9 kT (blue).

fragments with perdeuterated, $^{13}C/^{15}N$-labeled MSP1$_{19}$, under conditions similar to those used previously.[18] Intermolecular NOEs from signals showing methyl group-region chemical shifts to the antigen could be identified in the spectrum of the mAb 12.8 complex. These must originate in the mAb fragment, since the aliphatic side chains in the antigen are perdeuterated, although as the mAb fragment cannot be isotope-labeled, there is no assignment information for this component (no NOEs with methyl region chemical shifts were observed with the Fab 12.10 complex). Two antigen residues that showed strong cross-saturation in the 12.8 complex, Asn 15 and Ser 16, showed backbone amide NOE cross peaks with a pair of methyl signals. One of these methyl signals showed an additional NOE to the side chain amide protons of a residue in the antigen, tentatively identified as Gln 14. From these data, we conclude that there is a close contact between antigen residues Asn 15 and Ser 16 and an aliphatic residue of the antibody, probably Val, Ile, or Leu. The 12.8 docking results were manually inspected according to this hypothesis. Among the selected complexes, none was found to have valine, isoleucine, or leucine in short reach from the amides of residues 15 and 16 and we therefore generated 1000 additional complexes for each antibody. On the surface of the CDRs, we can identify Ile 29 in loop L1, Leu 53 in loop L2, and Val 100 in loop H3 as possible candidates for the interaction in question. It should be noted that Ile L29 and Leu L53 are situated at

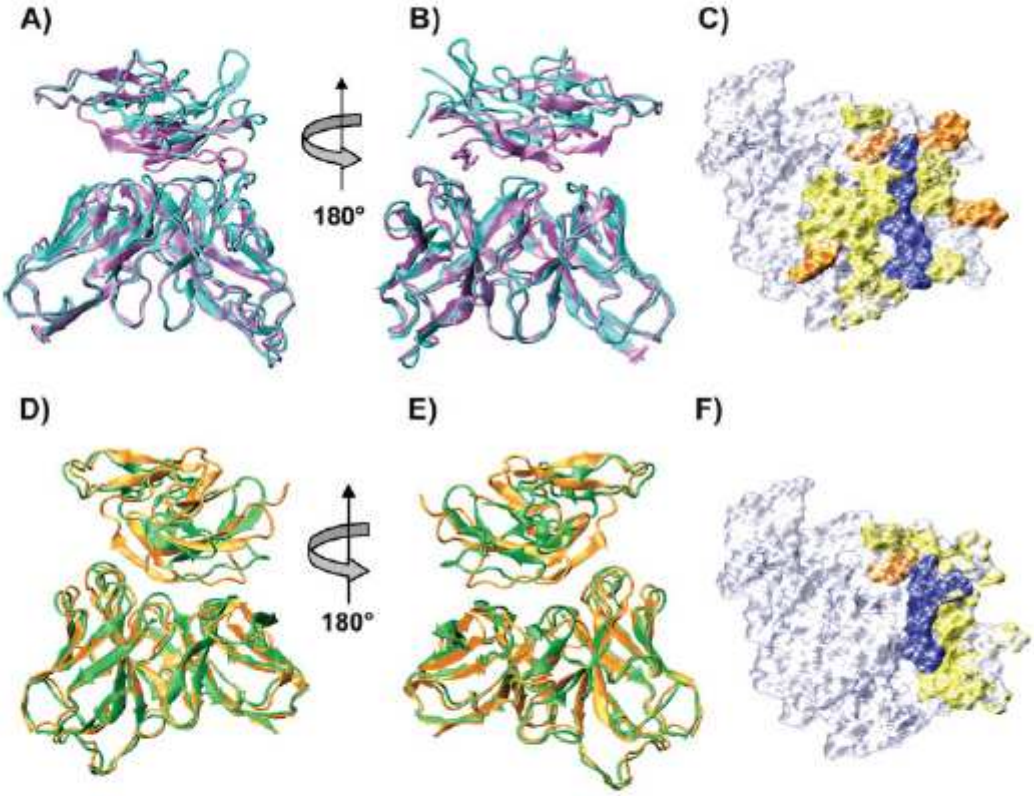


Fig. 6. Panels (A), (B), and (C) refer to complex mAb 12.10-MSP1$_{19}$. (A) and (B) show the superimposition of the complex before (light blue) and after 5 ns of MD simulation (purple). Panel (C) displays the surface of the antigen buried upon complex formation, as calculated by POPS-comp.[44] The residues predicted by NMR experiments are displayed in blue, in yellow are shown the residues that, after MD simulation, still pass the NMR filter. In orange are displayed the remaining residues that bury a $\Delta SASA \geq 10$ Å^2 due to the formation of the complex (values shown in Table II). Panels (D), (E), and (F) refer to complex mAb 12.8-MSP1$_{19}$. (D) and (E) show the superimposition of the complexes before (orange) and after 5 ns of MD simulation (green). Panel (F) is colored like panel (C) (values shown in Table V).

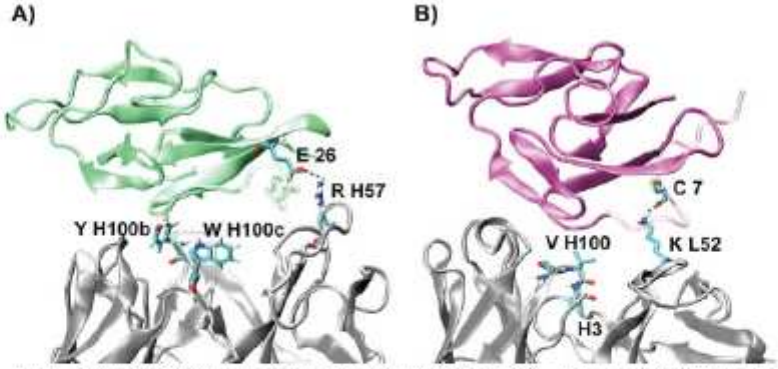


Fig. 7. Panel (A): Ribbon representation of complex mAb 12.10-MSP1$_{19}$. In licorice rendering Arg H57 (H2 loop) forming a salt bridge with Glu 26 of MSP1$_{19}$, and Trp H100c that shields the cavity in the central binding site region with negative electrostatic potential. Residues Tyr H100b and Trp H100c are found to be in contact with the antigen. Panel (B): Ribbon representation of complex mAb 12.8-MSP1$_{19}$. In licorice rendering, Lys L52 forming a hydrogen bond with the carbonyl oxygen of Cys 7 of MSP1$_{19}$. The central residues in loop H3 (Ser H99-Val H100-Ile H100a) are highlighted.

TABLE III. Residues of Antigen $MSP1_{19}$ in Contact With mAb 12.10_2FBJ After 5 ns of MD Simulation

(a) Accessible surface areas buried upon formation of the antigen-antibody complex

Antigen		Antibody		
Complex 12.10-$MSP1_{19}$	ΔSASA Å^2	Complex 12.10-$MSP1_{19}$	ΔSASA Å^2	CDR
ASN 1	29.1	SER H31	24.8	H1
ILE 2	52.6	TYR H33	25.8	H1
GLN 4	83.2	ILE H52	25.6	H2
IHS 5	10.2	ASN H54	56.9	H2
GLN 6	96.0	ASN H55	60.2	H2
GLN 11	20.8	ARG H57	172.9	—
CYS 12	22.1	ASN H59	63.7	—
GLN 14	50.8	TYR H100b	34.4	H3
ASN 15	39.4	ASP L28	27.8	L1
TYR 34	19.8	SER L30	23.2	L1
LYS 40	14.0	SER L67	22.0	—
GLU 43	34.4	SER L92	32.9	L3
TYR 84	16.0	SER L93	45.8	L3
PHE 87	18.9	THR L94	54.7	L3

(b) Residues complying to the NMR filter

Antigen	Antibody	
Residue	Residue	CDR
ASN 1	SER H31	H1
ILE 2	SER H31	H1
	TYR H33	—
SER 3	TYR H33	—
GLN 4	ARG H57	—
HIS 5	ARG H57	—
GLN 6	ARG H57	—
	ASN H59	—
CYS 7	ARG H57	—
VAL 8	ARG H57	—
	LYS H65	—
CYS 12	SER L93	L3
	THR L94	L3
PRO 13	THR L94	L3
GLN 14	SER L93	L3
	THR L94	L3
ASN 15	TYR H100b	H3
	SER L30	L1
	SER L92	L3
	SER L93	L3
	THR L94	L3
SER 16	TYR H100b	H3
	SER L30	L1
	SER L92	L3
	SER L93	L3
GLY 17	TYR H100b	H3
	TRP H100c	H3
	SER L93	L3
CYS 18	TRP H100c	H3
	SER L93	L3
LEU 31	TYR H100b	H3
GLU 43	SER L67	—
PRO 85	TYR H33	H1
LEU 86	TYR H33	H1
PHE 87	TYR H100b	H3
ASP 88	TYR H100b	H3

TABLE IV. Energy Decomposition Analysis of the Docked Complexes

(a) Energy decomposition for mAb12.10-$MSP1_{19}$

Complex 12.10-$MSP1_{19}$	Loops ZDOCK	Antibody ZDOCK	Loops 1 ns MD	Antibody 1 ns MD	Loops 5 ns MD	Antibody 5 ns MD
E1-SR: mAb-mAb	−4152.5	−9797.2	−4840.3	−10267.9	−5053.8	−10860.5
LJ: mAb-mAb	−3533.8	−7542.7	−3519.9	−7856.1	−3491.7	−7945.2
E1-14: mAb-mAb	12464.4	26577.5	13028.8	27635.1	12936.9	27557.7
LJ-14: mAb-mAb	−142.1	−317.6	206.3	401.8	146.4	356.4
E1-SR: mAb-$MSP1_{19}$	**−8.3**	**−8.6**	**−83.2**	**−84.4**	**−104.6**	**−104.0**
LJ: mAb-$MSP1_{19}$	**−85.0**	**−87.0**	**−248.7**	**−256.5**	**−329.7**	**−332.1**
E1-SR: $MSP1_{19}$-$MSP1_{19}$		−3528.9		−4342.9		−4180.7
LJ: $MSP1_{19}$-$MSP1_{19}$		−1637.8		−2811.6		−2929.7
E1-14: $MSP1_{19}$-$MSP1_{19}$		10646.7		10085.0		10067.6
LJ-14: $MSP1_{19}$-$MSP1_{19}$		1575.7		109.5		110.6

(b) Energy decomposition for mAb12.8-$MSP1_{19}$; same as for (a)

E1-SR: mAb-mAb	−1915.8	−10626.7	−2041.4	−10626.8	−1775.0	−10335.7
LJ: mAb-mAb	−1218.7	−6968.4	−1308.8	−7701.1	−1375.4	−7895.5
E1-14: mAb-mAb	5473.8	26280.2	5806.1	27300.3	5651.5	27255.6
LJ-14: mAb-mAb	−69.7	−356.4	157.5	498.3	93.4	344.1
E1-SR: mAb-$MSP1_{19}$	**1.7**	**28.6**	**−33.4**	**−35.0**	**−96.6**	**−116.6**
LJ: mAb-$MSP1_{19}$	**40.2**	**14250.8**	**−125.6**	**−131.2**	**−212.8**	**−220.8**
E1-SR: $MSP1_{19}$-$MSP1_{19}$		−3512.0		−4346.9		−4448.5
LJ: $MSP1_{19}$-$MSP1_{19}$		−1641.2		−2792.7		−2923.5
E1-14: $MSP1_{19}$-$MSP1_{19}$		10646.4		10173.9		10183.1
LJ-14: $MSP1_{19}$-$MSP1_{19}$		1576.8		97.7		68.8

The CDR loops have been decomposed from the entire antibody, in order to estimate their contribution to the binding. Values have been reported for the starting configuration (ZDOCK), after 1 ns of MD and at the end of the 5 ns of MD. The intermolecular energy contributions (antigen-antibody) are given in bold. Energies are given in units of kJ/mol.

the edge of the binding region. Therefore, to observe a contact between these residues and residues 15 and 16 of the antigen, the antigen itself would not be anymore centrally located with respect to the binding site, and in particular it would not be in contact with the H3 loop. With these concepts in mind we were able to select only one complex (from 12.8_1IND) with Asn 15 and Ser 16 of $MSP1_{19}$ at a short distance from residue Val H100 of the antibody. From the clustering procedure (as described in the Methods section) of the total of 17 acceptable complexes, our selected complex is number 10 (Fig. S4, Supplementary Materials). All complexes between 1 and 11 are equally good with respect to the clustering, but complex 10 is the only one among them that has the antigen centered relative to the antibody's binding loops.

For this complex, 5 ns of Molecular Dynamics simulations were performed and further analyses were done on the final complex. In Figure 6(D,E) the selected ZDOCK complex and the final MD conformer are shown superimposed. As in the previous cases, the MD simulation has the effect of pulling the antigen closer to the antibody. In Table Va the areas that are buried upon formation of the complex as calculated by POPScomp[44] are reported. Also in this case most of the antigen residues indicated by the NMR data are in contact with the antibody's heavy chain CDRs. Because loop H3 of antibody 12.8 is particularly short and lacks long side-chains, most of the interactions are found to form with loops H1 and H2. Table Vb shows that residue Gln 14 is in contact with Val H100, and this should represent one of the key interactions for this complex. The Val H100 side chain is still too far from the backbone amides of antigen residues Asn 15 and Ser 16 to generate a strong NOE, while residues Gly 17 and Cys 18, where cross-saturation was observed, are not sufficiently close to the antibody protons. These residues were close to the antibody in the ZDOCK complex, but moved away during the MD refinement. This effect could be ascribed to some repulsion between residues Lys 9 and Lys 10 of the antigen and residue Lys L52 of the antibody. The electrostatic interaction as described in the MD force field could amplify such an effect. The overall results are summarized in Figure 6(F). In Table IVb LJ and electrostatic contributions are reported analogously to Table IVa. As for the 12.10 complex, the electrostatic and intermolecular LJ contributions become increasingly more favorable during the simulated time (Table IVb), with larger contributions to the binding from CDR loops. We observe for this antibody a slightly different binding mode when compared with 12.10, with the antigen binding primarily to the first EGF-like domain to loops L2, H1 and H3. The antigen loop comprising residues 65–75 points toward loop H2 and is tilted with a larger distance from H2 than from L2.

CONCLUSIONS

We present here the docking refinement and analysis of two strongly inhibitory antibodies (designated 12.8

TABLE V. Residues of Antigen $MSP1_{19}$ in Contact With mAb 12.8 After 5 ns of MD Simulation

(a) Accessible surface areas buried upon formation of the antigen-antibody complex

Antigen		Antibody		
Complex 12.8-$MSP1_{19}$	ΔSASA Å^2	Complex 12.8-$MSP1_{19}$	ΔSASA Å^2	CDR
VAL 8	24.8	GLY H26	17.7	H1
LYS 9	34.0	TYR H27	10.4	H1
GLN 11	65.5	THR H28	57.3	H1
PRO 13	14.2	ASN H31	29.8	H1
GLN 14	113.9	TYR H32	47.5	H1
ASN 15	19.2	THR H53	10.7	H2
TYR 34	18.7	PHE H54	50.4	H2
ASP 39	20.3	VAL H100	35.1	H3
LYS 40	46.5	SER L51	18.9	L2
CYS 41	10.2	LYS L52	96.2	L2

(b) Residues complying to the NMR filter

Antigen	Antibody	
Residue	Residue	CDR
CYS 7	LYS L52	L2
VAL 8	LYS L52	L2
LYS 9	SER L51	L2
	LYS L52	L2
	LEU L53	L2
LYS 10	LEU L53	L2
GLN 11	TYR L48	—
	ASP L49	—
	LYS L52	L2
	LEU L53	L2
CYS 12	TYR L48	—
	LYS L52	L2
PRO 13	TYR H32	H1
	TYR L48	—
GLN 14	THR H28	H1
	TYR H32	H2
	VAL H100	H3
ASN 15	THR H28	H1
	TYR H32	H1
SER 16	THR H28	H1
	TYR H32	H1
GLY 38	GLN H1	—
ASP 39	GLN H1	—
	ILE H2	—
LYS 40	GLN H1	—
	ILE H2	—
	GLY H26	H1
	TYR H27	H1
	THR H28	H1
	TYR H32	H1
CYS 41	TYR H27	H1
	THR H28	H1
	TYR H32	L1
	SER H93	H3
VAL 42	TYR H27	H1
	THR H28	H1
GLU 43	THR H28	H1

and 12.10) to the antigen structure $MSP1_{19}$ as solved by NMR.[12] The antibodies were modeled by comparative homology modeling and the docking results were screened by selecting the structures with the help of data from NMR cross-saturation and chemical shift perturbation experiments,[18,19] complemented by additional

NOE experiments. Molecular Dynamics simulations in explicit solvent were performed on the selected complexes to refine their structures to evaluate the role of specific side chains in the binding. The trial exercise using the Fab G17.12-MSP1$_{19}$ complex X-ray structure indicated the feasibility of obtaining successfully docked structures using this NMR-based docking/MD approach. With this exercise we mainly wanted to test our filter of structures based on external restraints. Moreover, the additional docking performed with the modeled Fab G17.12 gave rise to a complex very similar to the one obtained with the original X-ray structure. This finding supports the use of models in the case of absence of X-ray structures for the antibody, but this procedure is of course limited to a sufficiently high sequence identity between the sequences of model and template, in particular in the binding region.

This issue is very important, especially in relation to new structural genomics/system biology initiatives that attempt the structural reconstruction of protein–protein interaction networks.[53]

Our computed complexes show two slightly different binding modes for the two antibodies 12.10 and 12.8: complex 12.10-MSP1$_{19}$ shows antigen binding with one face of its discoidal shape, lying flat on the surface of the CDRs; complex 12.8-MSP1$_{19}$ presents the antigen in a tilted position, anchored to the antibody mainly through its first EGF-like domain. The second EGF domain is more distant from the H2 loop. It is noteworthy that previous experimental evidences had pointed out that antibody 12.10 can only bind in the presence of both MSP1$_{19}$ EGF domains, while for 12.8 only the presence of domain 1 is sufficient to observe binding.[54] For this antibody we stress that loop H3 is particularly short and without long side-chain residues that could be specific for the binding. This could be the cause of the observed binding to mainly one EGF-like domain of MSP1$_{19}$.

From a detailed analysis of our docked complexes we underline the importance of a positively charged residue on the surface of the antibody. In particular antibody 12.10 has residue Arg H57 on the H2 loop that forms a salt bridge with Glu 26 of the antigen [Fig. 7(A)]. Antibody 12.8 contains residue Lys L52 that forms a hydrogen bond with the carbonyl oxygen of Cys 7 of MSP1$_{19}$ [Fig. 7(B)]. Our analysis pointed out the necessity of a positively charged electrostatic spot on the CDRs surface that seems to be a general requirement for the recognition and binding of this antigen. Accordingly also mAb G17.12 has been observed to form a salt bridge with MSP1$_{19}$ through its residue Arg H98.[15]

Each docking experiment produced a limited number of results consistent with the NMR data and the general expectation of typical antibody binding modes (recognition through the CDR loops), also in reason of the few restraints obtained from cross-saturation experiments. In principle a few restraints would be enough to define the binding mode precisely if these were distributed about the vertices of a triangle covering one face of the antigen. Once a binding mode is adopted having one of the discoidal faces lying flat on the surface of the CDRs, we can observe that the NMR data are concentrated on one half of this surface, therefore the docked solutions are only mildly restrained [Fig. 6(C,F)]. For the antibody 12.10 we had two more restraints (on a distant spot from the main ones (HN 6, 7 *vs* HN 15, 16, 17, 18)) and therefore the docking procedure proved to be more straightforward.

Although it would be desirable to crystallize one or both of the 12.8 and 12.10 mAb complexes to verify the modeled structures, useful information can be derived from the current models. They provide a more detailed picture of the binding interface, and predict intermolecular interactions that can be used, for example, as a guide for mutagenesis work. The models will be valuable for understanding the mechanism of action of invasion-inhibitory antibodies and for studying therapeutic effects of antibody reagents derived from these antibody sequences.

ACKNOWLEDGMENTS

This work was supported by Medical Research Council and by European Commission through the EUROMALVAC consortium. F. A. was supported by CEINGE Advanced Biotechnology consortium. The authors thank Annalisa Pastore and Andres Ramos for the use of computer facilities.

REFERENCES

1. Engers HD, Godal T. Malaria vaccine development: current status. Parasitol Today 1998;14:56–64.
2. Holder AA. Proteins on the surface of the malaria parasite and cell invasion. Parasitology 1994;108 Suppl:S5–S18.
3. Barnwell JW, Galinski MR. The adhesion of malaria merozoite proteins to erythrocytes: a reflection of function? Res Immunol 1991;142:666–672.
4. Holder AA, Guevara Patino JA, Uthaipibull C, Syed SE, Ling IT, Scott-Finnigan T, Blackman MJ. Merozoite surface protein 1, immune evasion, and vaccines against asexual blood stage malaria. Parassitologia 1999;41:409–414.
5. McBride JS, Heidrich HG. Fragments of the polymorphic Mr 185,000 glycoprotein from the surface of isolated *Plasmodium falciparum* merozoites form an antigenic complex. Mol Biochem Parasitol 1987;23:71–84.
6. Holder AA, Sandhu JS, Hillman Y, Davey LS, Nicholls SC, Cooper H, Lockyer MJ. Processing of the precursor to the major merozoite surface antigens of *Plasmodium falciparum*. Parasitology 1987;94:199–208.
7. Holder AA, Riley EM. Human immune response to MSP-1. Parasitol Today 1996;12:173,174.
8. Blackman MJ, Heidrich HG, Donachie S, McBride JS, Holder AA. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. J Exp Med 1990;172:379–382.
9. Blackman MJ, Scott-Finnigan TJ, Shai S, Holder AA. Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein. J Exp Med 1994;180:389–393.
10. Guevara Patino JA, Holder AA, McBride JS, Blackman MJ. Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. J Exp Med 1997;186:1689–1699.
11. Nwuba RI, Sodeinde O, Anumudu CI, Omotun YO, Odaibo AB, Holder AA, Nwagwu M. The human immune response to *Plasmodium falciparum* includes both antibodies that inhibit merozoite surface protein-1 secondary processing and blocking antibodies. Infect Immun 2002;70:5328–5331.
12. Morgan WD, Birdsall B, Frenkiel TA, Gradwell MG, Burghaus PA, Syed SE, Uthaipibull C, Holder AA, Feeney J. Solution

structure of an EGF module pair from the *Plasmodium falciparum* merozoite surface protein 1. J Mol Biol 1999;289:113–122.
13. Chitarra V, Holm I, Bentley GA, Pétres S, Longacre S. The crystal structure of C-terminal merozoite surface protein-1 at 1.8 A resolution, a highly protective malaria vaccine candidate. Mol Cell 1999;3:457–464.
14. Garman SC, Simcoke WN, Stowers AW, Garboczi DN. Structure of the C-terminal domains of merozoite surface protein-1 from *Plasmodium knowlesi* reveals a novel histidine binding site. J Biol Chem 2003;278:7264–7269.
15. Pizarro JC, Chitarra V, Verger D, Holm I, Pétres S, Dartevelle S, Nato F, Longacre S, Bentley GA. Crystal structure of a Fab complex formed with PfMSP1-19, the C-terminal fragment of merozoite surface protein 1 from *Plasmodium falciparum*: a malaria vaccine candidate. J Mol Biol 2003;328:1091–1103.
16. Jennings RM.Murine and human antibody responses to *Plasmodium falciparum* merozoite surface protein-1. PhD Thesis, London: University College London; 2004.
17. Uthaipibull C, Aufiero B, Syed SE, Hansen B, Guevara Patino JA, Angov E, Ling IT, Fegeding K, Morgan WD, Ockenhouse C, Birdsall B, Feeney J, Lyon JA, Holder AA. Inhibitory and blocking monoclonal antibody epitopes on merozoite surface protein 1 of the malaria parasite *Plasmodium falciparum*. J Mol Biol 2001;307:1381–1394.
18. Morgan WD, Frenkiel TA, Lock MJ, Grainger M, Holder AA. Precise epitope mapping of malaria parasite inhibitory antibodies by TROSY NMR cross-saturation. Biochemistry 2005;44:518–523.
19. Morgan WD, Lock MJ, Frenkiel TA, Grainger M, Holder AA. Malaria parasite-inhibitory antibody epitopes on *Plasmodium falciparum* merozoite surface protein-1(19) mapped by TROSY NMR. Mol Biochem Parasitol 2004;138:29–36.
20. Wodak SJ, Méndez R. Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. Curr Opin Struct Biol 2004;14:242–249.
21. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
22. Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. Proteins 2003;51:397–408.
23. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun 1995;91:43–56.
24. Chen R, Tong W, Mintseris J, Li L, Weng Z. ZDOCK predictions for the CAPRI challenge. Proteins 2003;52:68–73.
25. Ehrlich LP, Nilges M, Wade RC. The impact of protein flexibility on protein-protein docking. Proteins 2005;58:126–133.
26. Smith GS, Sternberg MJE, Bates PA. The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. J Mol Biol 2005;347:1077–1101.
27. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000;302:205–217.
28. Huang X, Miller W. A time-efficient, linear-space local similarity algorithm. Adv Appl Math 1991;12:337–357.
29. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 1988;85:2444–2448.
30. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003;31:3381–3385.
31. Suh SW, Bhat TN, Navia MA, Cohen GH, Rao DN, Rudikoff S, Davies DR. The galactan-binding immunoglobulin Fab J539: an X-ray diffraction study at 2.6-A resolution. Proteins 1986;1:74–80.
32. Chitarra V, Alzari PM, Bentley GA, Bhat TN, Eisele JL, Houdusse A, Lescar J, Souchon H, Poljak RJ. Three-dimensional structure of a heteroclitic antigen-antibody cross-reaction complex. Proc Natl Acad Sci USA 1993;90:7711–7715.
33. Love RA, Villafranca JE, Aust RM, Nakamura KK, Jue RA, Major JG, Radhakrishnan R, Butler WF. How the anti-(metal chelate) antibody CHA255 is specific for the metal ion of its antigen: X-ray structures for two Fab?hapten complexes with different metals in the chelate. Biochemistry 1993;32:10950–10959.
34. Alzari PM, Spinelli S, Mariuzza RA, Boulot G, Poljak RJ, Jarvis JM, Milstein C. Three-dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. EMBO J 1990;9:3807–3814.
35. Chen R, Weng Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. Proteins 2002;47:281–294.
36. Kleinjung J, Romein J, Lin K, Heringa J. Contact-based sequence alignment. Nucleic Acids Res 2004;32:2464–2473.
37. Team RDC. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2005.
38. Daura X, Mark AE, van Gunsteren WF. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. J Comput Chem 1998;19:535–547.
39. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J, Pullman B.Interaction models for water in relation to protein hydration. In:Pullman B, editor. Intermolecular forces. Dordrecht: Reidel; 1981. pp 331–342.
40. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
41. Hess B, Bekker H, Berendsen HJC, Fraaije J. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–1472.
42. Darden T, Perera L, Li L, Pedersen L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. Struct Fold Des 1999;7:3).
43. Tironi IG, Sperb R, Smith PE, Vangunsteren WF. A generalized reaction field method for molecular-dynamics simulations. J Chem Phys 1995;102:5451–5459.
44. Kleinjung J, Fraternali F. POPSCOMP: an automated interaction analysis of biomolecular complexes. Nucleic Acids Res 2005;33:W342–W346. (Web server issue).
45. Fraternali F, Cavallo L. Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome. Nucleic Acids Res 2002;30:2950–2960.
46. Cavallo L, Kleinjung J, Fraternali F. POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level. Nucleic Acids Res 2003;31:3364–3366.
47. Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Conformations of the third hypervariable region in the VH domain of immunoglobulins. J Mol Biol 1998;275:269–294.
48. Tormo J, Stadler E, Skern T, Auer H, Kanzler O, Betzel C, Blaas D, Fita I. Three-dimensional structure of the Fab fragment of a neutralizing antibody to human rhinovirus serotype 2. Protein Sci 1992;1:1154–1161.
49. Saul FA, Poljak RJ. Crystal structure of human immunoglobulin fragment Fab New refined at 2.0 A resolution. Proteins 1992;14:363–371.
50. Zdanov A, Li Y, Bundle DR, Deng SJ, MacKenzie CR, Narang SA, Young NM, Cygler M. Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-A resolution. Proc Natl Acad Sci USA 1994;91:6423–6427.
51. Prasad L, Sharma S, Vandonselaar M, Quail JW, Lee JS, Waygood EB, Wilson KS, Dauter Z, Delbaere LT. Evaluation of mutagenesis for epitope mapping. Structure of an antibody-protein antigen complex. J Biol Chem 1993;268:10705–10708.
52. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sherriff S, Padlan EA, Davies D, Tulip WR, Colman PM, Spinelli S, Alzari PM, Poljak J. Conformations of immunoglobulin hypervariable regions. Nature 1989;342:877–883.
53. Aloy P, Russel RB. Structural systems biology: modelling protein interactions. Nat Rev Mol Cell Biol 2006;7:188–197.
54. Chappel JA, Holder AA. Monoclonal antibodies that inhibit *Plasmodium falciparum* invasion in vitro recognise the first growth factor-like domain of merozoite surface protein-1. Mol Biochem Parasitol 1993;60:303–311.

References:

[1] Jones MK, Good MF. Malaria parasites up close. *Nat Med* (2006);12:170-1.

[2] Hoffman SL, Franke ED, Hollingdale MR, Druilhe P. Malaria Vaccine Development: A Multi-Immune Response Approach. Washington, D.C: American Society of Microbiology, 1996.

[3] Stoute JA, Kester KE, Krzych U, Wellde BT, Hall T, White K, et al. Long-term efficacy and immune response following immunization with the RTS,S malaria vaccine. *J Infect Dis* (1998);178:1139-44.

[4] Yoshida N, Nussenzweig RS, Potocnjak P, Nussenzweig V, Aikawa M. Hybridoma produces protective antibodies directed against the sporozoite stage of malaria parasite. *Science* (1980);207:71-3.

[5] Holder AA. Proteins on the surface of the malaria parasite and cell invasion. *Parasitology* (1994);108 Suppl.

[6] Blackman MJ, Heidrich HG, Donachie S, McBride JS, Holder AA. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. *J Exp Med* (1990);172:379-82.

[7] Morgan WD, Birdsall B, Frenkiel TA, Gradwell MG, Burghaus PA, Syed SE, et al. Solution structure of an EGF module pair from the Plasmodium falciparum merozoite surface protein 1. *J Mol Biol* (1999);289:113-22.

[8] Chitarra V, Holm I, Bentley GA, Pétres S, Longacre S. The crystal structure of C-terminal merozoite surface protein 1 at 1.8 A resolution, a highly protective malaria vaccine candidate. *Mol Cell* (1999);3:457-64.

[9] Garman SC, Simcoke WN, Stowers AW, Garboczi DN. Structure of the C-terminal domains of merozoite surface protein-1 from Plasmodium knowlesi reveals a novel histidine binding site. *J Biol Chem* (2003);278:7264-9.

[10] Guevara Patino JA, Holder AA, McBride JS, Blackman MJ. Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. *J Exp Med* (1997);186:1689-99.

[11] Nwuba RI, Sodeinde O, Anumudu CI, Omosun YO, Odaibo AB, Holder AA, et al. The human immune response to Plasmodium falciparum includes both antibodies that inhibit merozoite surface protein 1 secondary processing and blocking antibodies. *Infect Immun* (2002);70:5328-31.

[12] Morgan WD, Lock MJ, Frenkiel TA, Grainger M, Holder AA. Malaria parasite-inhibitory antibody epitopes on Plasmodium falciparum merozoite surface protein-1(19) mapped by TROSY NMR. *Mol Biochem Parasitol* (2004);138:29-36.

[13] Morgan WD, Frenkiel TA, Lock MJ, Grainger M, Holder AA. Precise epitope mapping of malaria parasite inhibitory antibodies by TROSY NMR cross-saturation. *Biochemistry* (2005);44:518-23.

[14] Chappel JA, Holder AA. Monoclonal antibodies that inhibit Plasmodium falciparum invasion in vitro recognise the first growth factor-like domain of merozoite surface protein-1. *Mol Biochem Parasitol* (1993);60:303-11.

[15] Pizarro JC, Chitarra V, Verger D, Holm I, Pétres S, Dartevelle S, et al. Crystal structure of a Fab complex formed with PfMSP1-19, the C-terminal fragment of merozoite surface protein 1 from Plasmodium falciparum: a malaria vaccine candidate. *J Mol Biol* (2003);328:1091-103.

Chapter 4

Introduction: Restriction factors of retroviruses replication

The latest data published from UNAIDS/WHO in the 2006 estimated that 39.5 million people are living with Human immunodeficiency virus (HIV). There were 4.3 million new infections in 2006 with 2.8 million (65%) of these occurring in sub-Saharan Africa and important increases in Eastern Europe and Central Asia, where there are some indications that infection rates have increased by more than 50% since 2004. More than 2.9 million people died of AIDS only in 2006. With numbers such as these, AIDS has attracted vast resources from governments, universities, and industry the world over for the single purpose of controlling, preventing, and ultimately curing the disease.

At the present time, no effective cure or vaccine exists for AIDS. However, due to the relative success of Highly Active Antiretroviral therapy (HAART) survival of infected individuals has dramatically improved in the last 20 years. Untreated, AIDS compromises the body's immune system, and a wide variety of opportunistic infections do occur. These AIDS-related complications span the gamut of diseases and include bacterial, viral, protozoal and fungal infections. Current therapy regimens target the viral reverse transcriptase, protease, envelope and most recently integrase proteins, most commonly as a combination treatment of three different drugs. Nonetheless, due to the high level of genetic drift of HIV, (multi)-drug resistance is increasingly interfering with current therapies. Hence there is a continuous demand for novel drugs that target HIV.

In recent years it has emerged that the cells of the immune system themselves exhibit a defence mechanism against infection by viruses that occur in the cell. The proteins that mediate this innate immunity (restriction factors) are the human proteins APOBEC3F and APOBEC3G (A3F/G). These proteins act as a barrier to the replication of viruses that replicate via a single-stranded minus strand DNA intermediate. Unfortunately, the antiviral function of A3G is neutralized in by the HIV-encoded protein VIF (Virus Infectivity Factor) [2, 3]. In the absence of VIF, the host anti-retroviral factor A3G is packaged into the HIV-1 virus and down-regulates HIV-1 infection by causing extensive lethal deoxycytosine-to-deoxyuracil mutation in the newly synthesized HIV-1 negative-strand DNA [4].

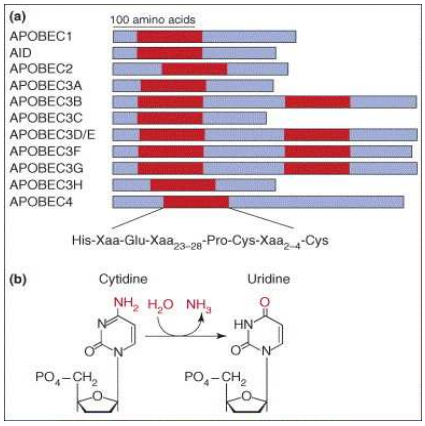
VIF binds specifically A3F/G and then targeting it for ubiquitination and degradation [4], prevents efficiently A3F/G being packed into the newly synthesized virus so that the virus remains infectious. The interaction of Vif with the A3F/G proteins clearly represents an attractive target for the development of new anti_HIV drugs, as therapeutic disruption of this interaction is thought to leave the virus defenceless against the antiviral action of A3F/G.

The APOBEC family of cellular deaminases

Cytidine deaminases are a large superfamily of enzymes involved in pathways such as lipoprotein metabolism, pyrimidine salvage pathways, antibody diversification and the maintenance of genome integrity. This class of enzymes converts cytosine to uracil by deamination. Cytidine deaminases (CDA), are characterized by a Zinc-binding catalytic domain with a specific consensus sequence H-X-E-X_{23-28}-P-C-X_{2-4}-C (X indicates any amino acid) [6]. In Humans the APOBEC sub-family is constituted by: Activation Induced Deaminase (AID), Apolipoprotein B mRNA editing catalytic subunit 1 (APOBEC1 or A1), APOBEC2 (A2), APOBEC3A–APOBEC3H (A3A–A3H) and APOBEC4 (**Fig 4.1**). APOBEC1 is the founding member of this APOBEC family of proteins, having first beeen demonstrated to edit specifically a single codon in the mRNA of Apolipoprotein B, leading to the synthesis of a truncated protein. Other

members have been named after APOBEC1, although it subsequently emerged that these enzymes edit DNA rather than RNA and they do not affect Apoliprotein metabolism. For example, the family member AID hypermutates the variable region of immunoglobulin genes leading to antibody diversification.

More recently, A3F/G were shown to hypermutate single stranded DNA replication intermediates of viruses and mobile genetic elements such as retrotransposons. Human APOBEC3F and 3G are expressed in CD4+ B cells that are the natural target of HIV [4, 7, 8]. The substrate specificities of A3F and A3G are finely different, the former preferring to edit the cytidine in a TC dinucleotide context to TU, while A3G shows a preference for CC dinucleotides [4, 9].



To manage this, the A3G/F protein must be packaged into virions and transferred to target cells but in the presence of VIF this process is inhibited; VIF connects A3G/F to an E3 ubiquitin ligase, inducing polyubiquitylation and the subsequent proteasomal degradation [1].

Figure 4.1: (a) Domain organization of human APOBEC proteins. In red the cytidine deaminase motifs in each protein are highlighted; the consensus amino acid sequence is shown below. (b) The cytidine deamination reaction catalysed by APOBEC enzymes. Image from [1].

Structures of APOBEC proteins

The crystal structures of several nucleotide CDAs have been reported showing a canonical mixed β-sheet motif with the catalytic residues exposed in α-helices at one side of the sheet (**Fig. 4.2 A**). However, these structures are of limited use as templates for the modelling of APOBEC proteins, as they are considerably smaller due to their function of binding a nucleotide rather than polynucleotide substrate.

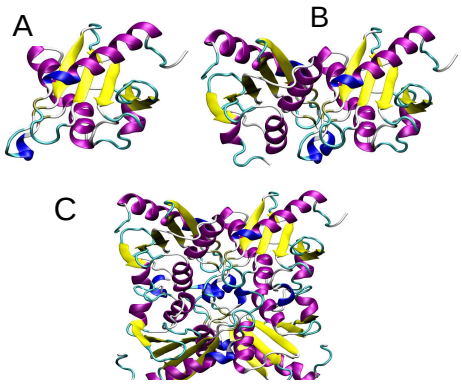


Figure 4.2: Ribbon representation of crystal structure of CDD1: (A) monomer; (B) dimer; (C) tetramer.

Tetramerisation of nucleotide CDAs shows one monomer interacting with other two forming a square arrangement of subunits (**Fig. 4.2**). The interaction between each monomer is through residues of the α-helices (**Fig. 4.2 B-C**) [13, 14]. Recently, the crystal structure of A2 has been reported [12], representing the first structure of an APOBEC protein and showing considerable differences with the nucleotide CDAs. Currently, the function of A2 is unknown, although it has been shown to be specifically expressed in skeletal and cardiac muscle. A2 is a tetrameric protein of 224 amino acids, each monomer contains one CDA domain that consists of five beta-sheet flanked by six helices (**Fig 4.3 A**). The major difference between the structure of A2 and other CDAs is in the mode of multimerization, which in A2 is linear rather than globular. The A2 dimer is formed by joining two β-sheets sideways in order to form one wide β-sheet.

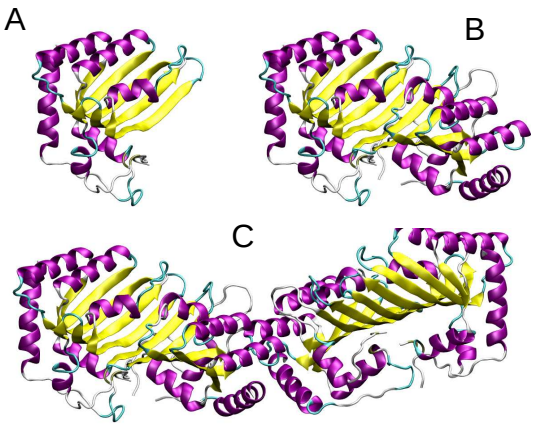


Figure 4.3: Ribbon representation of crystal structure of A2: (A) monomer; (B) dimer; (C) tetramer.

Many hydrogen bonds are formed between the β-sheet of monomers through main-chain atoms, providing the main bonding force between the two monomers (**Fig 4.3 B**). The way in which the two monomers interact is extremely different from the ones previously described. A2 tetramer is formed by two dimers that interact through head-to-head interactions (**Fig 4.3 C**).
There are large number of contacts between the two dimers formed indiscriminately by hydrophobic, aliphatic and charged residues [12].

Aim of the work

The HIV-1 VIF protein by interacting with A3G/F at the time of virus production induces the elimination of these two important restriction factors proteins, and lifts the barrier to viral replication that they cause. The therapeutic blockade of the Vif-A3F/G interaction represents an ideal goal for rational drug design. The disruption of the Vif-A3F/G interaction would leave the virus defenceless, leading A3F/G to inhibit reverse transcription and to hypermutate the viral genome, stopping this way viral replication.
Recently, Huthoff and Malim from the Division of Virology, King's College London have published a prediction of a possible structural arrangement for APOBEC protein domains based on secondary structure predictions [15] and have performed alanine scanning mutagenesis experiments to identify the domain within A3G that mediate sensitivity to HIV-1 Vif [16]. In particular, these data have allowed identifying a three-amino acid motif Asp_{128}-Pro_{129}-Asp_{130} (D-P-D) in A3G that is essential for the interaction with VIF to occur. In addition, a motif of four amino acids YYFW (Tyr_{124}-Tyr_{125}-Phe_{126}-Trp_{127}), immediately precedent to the D-P-D motif was identified to be essential for virion packaging [16].
It is worth mentioning that A3F does not contain the D-P-D motif, suggesting the presence of different sites of interaction between VIF and A3F.
In order to define the molecular determinants playing a role in the interactions between A3F/G and VIF, an interdisciplinary approach has been used to analyze the structure-function characteristics of these complexes comprising the following techniques: molecular genetics, biochemistry, viral evolution, virus infectivity and computational analysis. My contribution in this interdisciplinary effort has focused on the generation of molecular models by means of computational techniques. The resulting models assist rationalization of the results obtained from the other methodologies.

Results and Discussion

Dimerization models:

- ***Human APOBEC3G***

In order to model the 3D structure of APOBEC3G, the structure of A2 was chosen as template. Recently a model of monomeric A3G has been also based on this structure [17]. The difference with our approach is that we extended the modelling to include the dimeric structure of A3G. While A2 is a tetramer formed by four monomers of 224 amino acids with one CDA domain, A3G contains two CDA domains within a 384mer polypeptide (Figure 4.1). Hence we reason that an A3G monomer is equivalent to the dimeric structure of A2 and that dimerization of A3G mimics tetramerization of A2. The amino-terminal CDA domain (N-CDA) of A3G is required for viral encapsidation but not cytosine deamination, whereas the cytosine deamination is performed by the C-CDA domain. [18-20].
To build up the dimer model, each CDA domain of A3G was modelled with respect to the corresponding monomer in the A2 structure.
The dimerization of A3G can be modelled in two possible ways: i) dimerization mediated by N-terminal CDA domains, termed head-to-head (HHd) and ii) dimerization mediated by C-terminal CDA domains, termed tail-to-tail (TTd). Each of this possible dimerization models has been considered as equally probable in a first instance. A2 shows 33% of identity on average with each CDA domain of A3G. In Figure 4.3 the alignment performed with the program praline [21] of the sequence of A2 dimer with the entire sequence of A3G is shown.

```
A2   SGGGMIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHA
A3G  MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYS

A2   A---AHAEEAFFNTILPA--FDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLRLLIL
A3G  EL-KYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIF

A2   VGRLFMWEEPEIQAALKKLKE----AGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWE
A3G  VARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYS---QRELFEPWN

A2   DIQENFLYYEEKLADILK*GSGGGMIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEA
A3G  NLPKYYILLHIMLGEILR*----------HSMDPPTFTFNFNNEPWVRGRHETYLCYEVER

A2   QGKGGQVQASRGYLEDE--HA-------AAHAEEAFFNTILPA-FDPALRYNVTWYVSSS
A3G   MHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWS

A2   PCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEY
A3G   PCFSCAQEMAKFISKNKHVSLCIFTARIYDDQ-GRCQEGLRTLAEAGAKISIMTYSEFKH

A2   VWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK
A3G   CWDTFVDH---QGCPFQPWDGLDEHSQDLSGRLRAILQ
```

Figure 4.4: Alignment between two monomers of A2 and a whole sequence of A3G. The stars indicate the beginning of monomeric unit of A2. The residues of N-CDA and C-CDA are displayed in red and in green, respectively.

In the C-CDA domain there is an insertion of twelve amino acids that precedes the histidine of the catalytic site, while this insertion is shorter in the N-CDA domain.

To build the models we used the program MODELLER [25]. This software generates hundred models for each query (see methods) and performs an optimization of the generated models with respect to a defined objective function.
The selected models were chosen on the basis of the objective function's score and on the basis of the rotamers adopted by the catalytic residues (H-X-E-X_{23-28}-P-C-X_{2-4}-C) coordinating the Zn ions. Two 'families' of dimerization modes were generated by MODELLER: a bent and an elongated model (**FIG 4.5**).

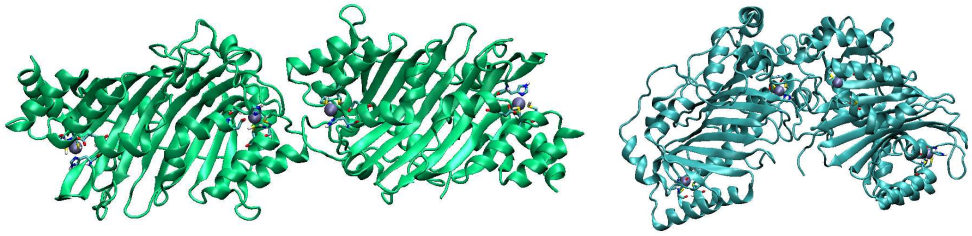


Figure 4.5: The elongate (left)and bent (right) modelobtained from HHd dimerization

The major difference between the 3D models and the template is in the extended loop in the C-CDA that connects strand β2 and helix2 (see **Fig.4.5**), corresponding to an insertion in the sequence. The TTd bent model brings these two loops too close to each other, presenting a distance between the two monomers of 13 Å, whereas the same distance is of 30 Å for the HHd model. The two loops are exposed to the solvent in the HHd bent model.
The fold adopted from TTd bent model disposes the CDA domains at the hinge region of the dimer, making them not easily accessible to ssDNA/RNA, and not in an ideal disposition for the enzyme activity. To characterize the interaction surface, the program POPScomp was used [22, 23] to detect the residues that were hidden from the dimerization (Table I and Table II). POPScomp calculates the solvent accessible surface area (SASA) of each residue before and after the formation of a complex. In this case a difference of SASA of 15 $Å^2$ was used as cut-off to select the residues buried upon complex formation.
The elongated dimers present only few contacts between the two monomers. In particular, the residues involved in the HHd comprise the two motifs essential for the interaction with VIF ($_{128}$DPD$_{139}$) and for virion packaging ($_{124}$YYFW$_{127}$); on the contrary the TTd exposes these residues and places the C-CDA at the interface. In the bent dimerization models there are a larger number of contacts in the hinge region between the two monomers (Table II). Also in this case in the HHd the DPD and YYTW motifs are at the dimer interface. To better characterise the dimer interface, site-directed mutations were introduced at residues 123, 124, 314 and 315 as the HHd predicts subunit interactions at residues 122-130 and the TTd predicts involvement of residues 313-320 (**Fig. 4.6**).

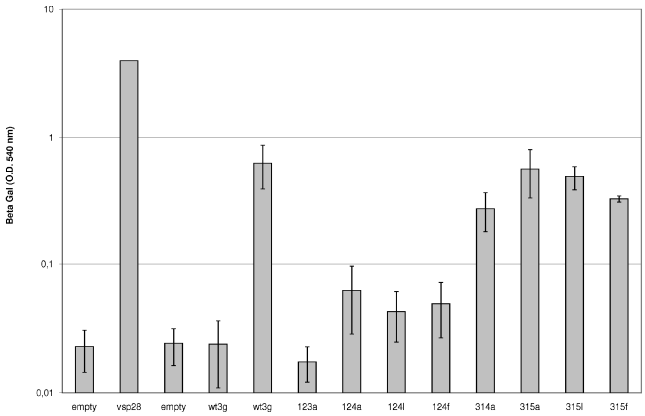


Figure 4.6: yeast two-hybrid assay

Table I: Accessible surface areas buried upon formation of elongated dimer.

HHd			TTd		
Elongated Models					
Residue		ΔSASA (Å^2)	Residue		ΔSASA (Å^2)
Chain A					
ARG	24	41.4	TRP	211	78.5
ILE	26	15.4	VAL	212	45.9
LEU	27	66.5	ARG	213	59.2
SER	28	32.5	TRP	285	31.7
TRP	94	14.5	PHE	289	15.3
LYS	99	28.6	ARG	313	25.9
ARG	122	40.2	ASP	316	34
TYR	125	95.5	GLN	318	28.1
TRP	127	107.5	ARG	320	115
ASP	128	25.3	ARG	374	58.2
PRO	129	16.7	ALA	377	16.8
ASP	130	39.1			
Chain B					
TYR	22	26.2	TRP	211	54.8
LEU	27	83.7	VAL	212	58.9
SER	28	40.7	ARG	213	69.8
ARG	29	40.9	TRP	285	25.5
ARG	122	28.4	ARG	313	28.2
TYR	124	23.6	ASP	316	38.4
TRP	127	54.7	GLN	318	47.3
TYR	181	30.3	ARG	320	24.9
			GLU	366	26.5
			HIS	366	22.5
			ASP	370	71.5
			ARG	374	50.1

Subsequently, yeast two-hybrid assays have been performed with these mutants, measuring the interaction of a mutant subunit (prey) with wild-type A3G (bait) (Huthoff and Malim, unpublished results). In the yeast-two hybrid system, successful interaction is indicated by b-galactidase activity in the cell lysate. Indeed, the wild-type/wild-type (peak 5) control shows a strong signal that is comparable to a positive control (peak 2: the TSG101 and VSP28 proteins from the human endosomal sorting pathway). Negative controls show low signals (peaks 1, 3 and 4). A signal comparable to the wild-type is observed with mutations at residues 314 and 315 (peaks 7-9), whereas mutations at 123 and 124 (peaks 10-13) show a signal that resembles the negative control. Expression of the wild-type and mutant A3G proteins in the yeast cell was confirmed by western blotting (not shown) indicating that loss of signal represents a loss of interaction. Thus, this experiment suggests that dimerization of A3G is mediated by the N-terminal CDA domain, which corresponds to the HHd model.

Table II: Accessible surface areas buried upon formation of bent dimer

HHd				TTd			
Bent model							
Residue	ΔSASA (Å^2)	Residue	ΔSASA (Å^2)				
Chain A		**Chain B**		**Chain A**		**Chain B**	
ARG 24	36.9	ARG 14	43.9	PHE 204	24.7	HIS 195	60.9
ILE 26	47.9	ASP 15	17.7	TRP 211	27.5	SER 196	119.7
LEU 27	26.4	SER 18	36	VAL 212	70.7	MET 197	50.9
SER 28	31.6	TYR 19	66.1	ARG 213	84	PRO 199	43.4
ARG 29	50.8	TYR 22	21	GLY 214	34	PRO 200	19.4
ARG 30	65.4	ILE 26	32	ARG 215	33.1	THR 203	55.6
TYR 59	55.6	LEU 27	17.1	HIS 248	26.3	PHE 204	103.1
LEU 62	43.5	SER 28	33.1	HIS 250	33.9	TRP 211	34.2
TYR 64	29.9	ARG 29	13.9	GLY 251	30.1	VAL 212	13.4
THR 98	22.4	ARG 30	26.3	LEU 253	75.5	ARG 213	82.1
LYS 99	23.2	ASN 31	17	GLU 254	23.3	HIS 216	43.6
ARG 102	65.7	THR 32	22.1	GLY 255	48.3	GLU 217	25.3
THR 106	23.6	VAL 58	33.9	PHE 289	40.1	ARG 313	50.5
ARG 122	40.9	TYR 59	26.3	SER 290	35.7	ASP 317	36.9
TYR 124	15.5	TRP 94	46.4	GLN 293	26.7	GLN 318	97.9
TYR 125	57.4	ARG 122	25.5	GLU 294	35.2	ASP 362	41.8
PHE 126	19.1	PHE 126	37.9	LYS 297	13.6	GLY 363	11.5
TRP 127	114.9	TRP 127	90.6	TYR 315	22.3	GLU 366	38.9
PRO 129	65.7	ASN 177	26	ASP 316	17.5	HIS 367	19.7
GLU 133	66.7	TYR 181	64.5	ASP 317	15.5	ASP 370	67.2
		LEU 184	74.4	GLN 318	72.2	ARG 374	49.3
				GLY 319	67.3		
				ARG 320	63.5		
				GLU 323	51.1		

To better understand the electrostatic contribution to the dimerization, the potential electrostatic surfaces were generated for all systems (**Fig.4.7**).

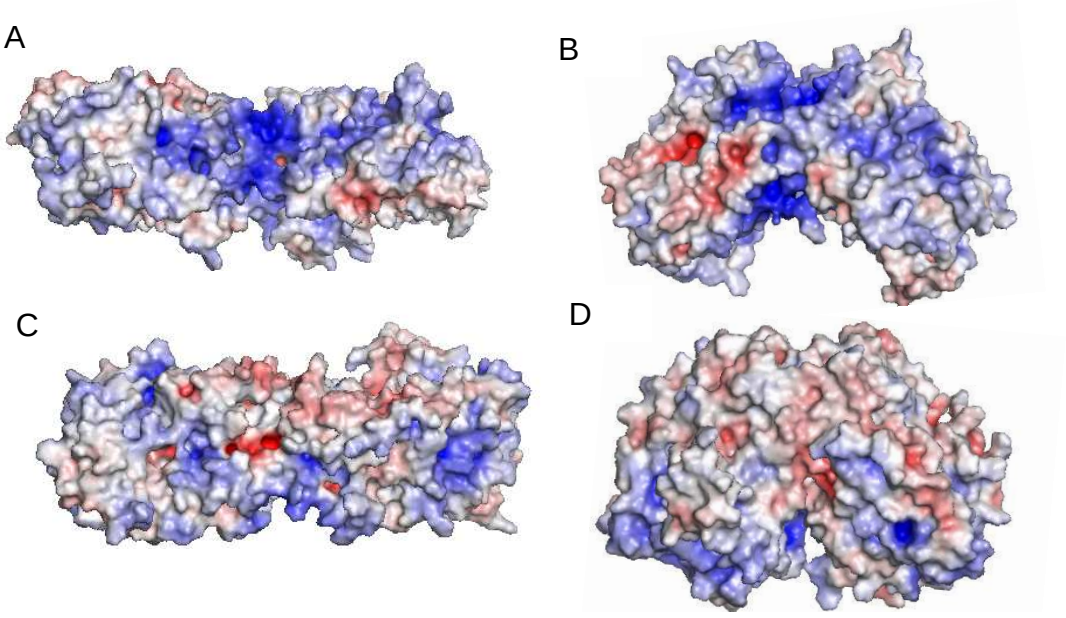


Figure 4.7: Comparison of the electrostatic potential of the predicted modes. In the panels A and B the elongated and the bent models of HHd, respectively, are displayed. In panels C and D the elongated and the bent models of TTd, respectively, are shown. The potential ranges from -10*k*T (red) to the maximal positive value +10*k*T (blue).

The electrostatic surface analysis shows that the contacts in the HHd models are prevalently formed by positive surfaces (blue coloured surfaces), whereas in the TTd models less charged residues are present at the interface. The positive interface in the HHd could represent an ideal surface for the interaction of a strong acid molecule as ssDNA/RNA. Therefore the absence of such a surface in the TTd and the close vicinity of two monomers in the bent model could be used as discrimination criteria against it. Lastly, the C-CDA domains would be too hidden in these models to exert their activity.

Human APOBEC3F

The models of A3F were constructed with the same procedure adopted for A3G. Human APOBEC3F is a protein of 374 amino acids; this provides a shorter protein sequence than A3G.

```
A2    SGGGMIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGKG-GQVQASRGYLED--
A3F    MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYS

A2    --EHAAAHAEEAFFNTILPAFDPALR-YNVTWYVSSSPCAACADRIIKTLSKTKNLRLLI
A3F    QPEH---HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTI

A2    LVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQ
A3F    SAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFV-YSEGQ--PFMPWYKFD

A2    ENFLYYEEKLADILK*GSGGGMIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGK
A3F    DNYAFLHRTLKEILR*NP---M----EAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKH

A2    GGQVQASRGYLE---DEHAAAHAEEAFFNTIL-PAFDPALRYNVTWYVSSSPCAACADRI
A3F    HSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEV

A2    IKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQE
A3F    AEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYND

A2    EGESKAFQPWEDIQENFLYYEEKLADILK
A3F    D---EPFKPWKGLKYNFLFLDSKLQEILE
```

Figure 4.8: Alignment between two monomers of A2 and a whole sequence of A3F.The stars indicate the start of monomeric unit of A2. The residue of N-CDA and C-CDA are displayed in red and in green, respectively.

The 3D models were generated to predict the dimerization of A3F. MODELLER produced only for HHd the two ‘families’ of dimers (**Fig. 4.9**).

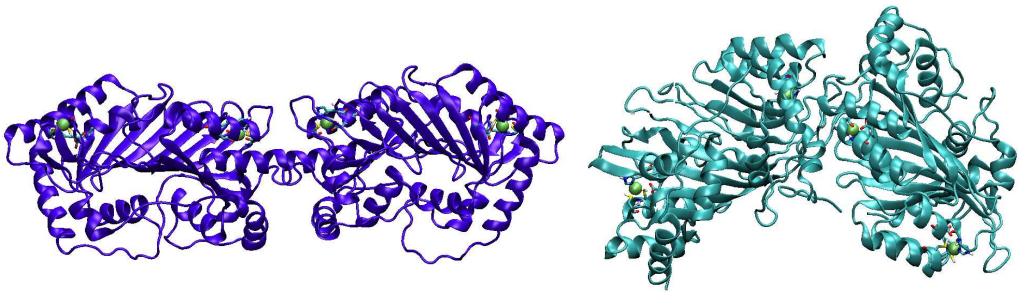


Figure 4.9: The elongate (left)and bent (right) modelobtained from HHd dimerization

As one can see from the analysis of the electrostatic surfaces, the bent TTd would lead to a dimer with two negative surfaces in too close contact; may be the repulsion of this two surfaces have not allowed the construction of this dimer (**Fig.4.10**).
The interface of the dimer HHd appears very similar to that of A3G even though the C-CDA domains of A3F show a larger negative surface respect to those of previously described.

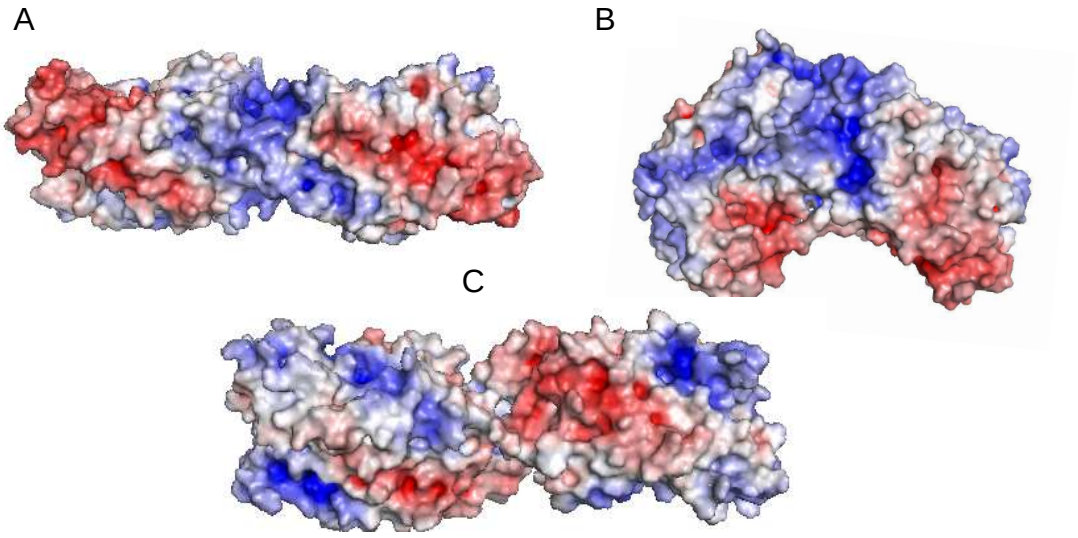


Figure 4.10: Comparison of the electrostatic potential of the predicted modes. (A) Elongated HHd, (B) bent HHd and (C) elongated TTd. The potential is ranged from -10*k*T (red) to the maximal positive value +10*k*T (blue).

In order to characterize the interaction surface, POPScomp was used [22, 23] to detect the residues that were buried upon dimerization (Table III).
Also in this case a difference of SASA of 15 $Å^2$ was used as cut-off to select the residues buried upon formation of the dimer.
For HHd dimerizations modes the residues correspondent to DPD and YYTW motifs are at the interface, in particular the Tyr residue is directly involved in the formation of the dimer. In the HHd bent model the residues of monomer A contribute in large part to the dimerization surface, whereas only few residues of monomer B are in contact with monomer A.
Even though there are very few data indicating details on the possible dimeric form for A3F, the results obtained from homology modelling and the study of interaction surface induce to think that the most plausible dimerization mode is a Head-to-Head one, similarly to A3G.

Table III: Accessible surface areas buried upon formation of dimer

HHd		TTd		HHd	
Elongated Models				**Bent Model**	
Residue	ΔSASA (Å^2)	Residue	ΔSASA (Å^2)	Residue	ΔSASA (Å^2)
		Chain A			
TYR 19	18.4	LYS 205	32.2	TYR 22	17
ARG 24	26.1	LYS 209	51.7	ARG 24	26,6
LEU 27	59.3	ALA 210	27.3	ILE 26	59,4
SER 28	43.5	TYR 211	90.8	LEU 27	71,7
ARG 29	25.8	TRP 277	18.4	SER 28	27,3
ARG 121	31.8	ARG 305	34.7	ARG 29	185,1
TYR 123	19.3	TYR 308	92.2	ARG 30	22,5
TYR 124	39.5	TRP 310	84.2	THR 32	35,7
TYR 125	12.7	ASP 311	46.5	TYR 59	33
TRP 126	81.5	ASP 313	16.6	SER 60	21,2
GLU 127	15.6	PHE 363	15.8	GLN 61	72,8
LYS 172	35.7			PRO 62	27,1
PHE 179	42.7			HIS 65	2,1
	Chain B			ASP 98	73,9
ILE 26	21.3	LYS 209	40.3	ARG 121	18,7
LEU 27	84.4	ALA 210	27.7	TYR 124	37,9
ARG 29	29.8	TYR 211	94.2	TRP 126	70,3
TRP 93	30.7	TRP 277	28.1	GLU 127	29,3
ARG 121	32.5	ARG 305	24.1	ARG 128	99,8
TYR 124	91	TYR 307	21.3	ASP 129	39
TYR 125	23.4	PHE 309	22.1	HIS 227	59
TRP 126	98	TRP 310	82.6	HIS 228	15,8
GLU 127	47.8	ASP 311	15.1	ASN 266	20
ARG 128	35.1	ASP 313	35	**Chain B**	
ASP 129	38.8	PHE 363	29	PHE 5	19,5
PHE 179	30.5			SER 18	35,8
				TYR 19	97,7
				ASN 20	16,5
				TYR 22	13,4
				ILE 26	59,8
				LEU 27	52,3
				SER 28	68,1
				ARG 29	126,1

Molecular dynamic simulations

Molecular Dynamics (MD) simulations is a useful method to refine the 3D models, but also to characterize at the molecular level the effect of specific mutations and their perturbation on the overall structure. This technique may be used to investigate structural consequences of any given mutation of interest (e.g. Vif-resistant and packaging mutants). MD simulations were performed onto the elongated head-to-head dimerization models of A3F/G as well as onto the crystal structure of APOBEC2, used as a reference in our analysis.

In order to evaluate the stability of the elongated structure of A2 and A3G/F models, 5 ns and 10 ns of MD simulations were performed on each of these models. As one can observe from Fig 4.11, all monomers have reached equilibration within the firsts 200 ps and show the same average value of RMSD (Root Mean Square Deviation) around 2 Å for A2 and 3 Å for A3G/F. The molecules are stable under the simulated conditions and converge to similar energy values as shown in Table IV.

During the simulations each monomer moves towards each other, getting in closer contact and improving this way the interaction. As we can see from Table V, the

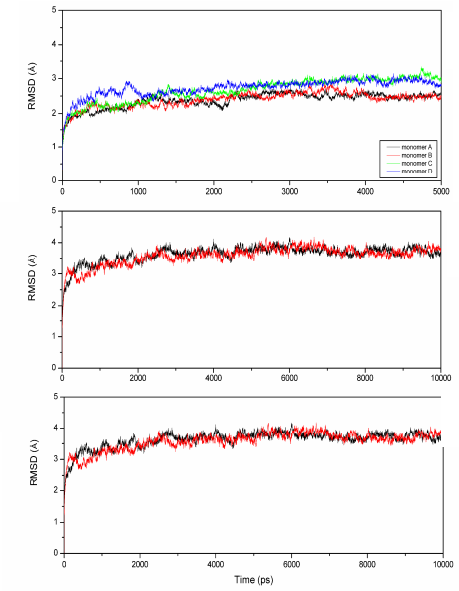


Figure 4.11: RMSD of the Cα atoms from their initial coordinates as a function of time, for the three simulated systems; (A) A2, (B) A3G, (C)

contacts between the monomers are somewhat different, nevertheless in both dimers the residue Tyr124 remains in the contact region for all the simulated time, in good agreement with the experimental data previously described.

Some interactions are lost during the dynamics, in particular for the A3G dimer, where only two (Trp127 and Tyr 181) of the starting interactions, involving the residues of monomer B, remain in contact till the end of the simulation. Visual inspection of A3G simulation revealed that a new secondary structural segment (helix Asn177-Met188) of each monomer causes, after 2 ns, the two subunits to move closer to each other. This movement influences the position of each monomer and causes the loss interactions in another region (Tyr22-Ser28). A similar situation is observed in A3F, where a corresponding helix (Asp174-Leu184) for each monomer comes in close contact after few ns. This motion observed in both dimers could demonstrate that the residues of these helices could play an essential role in the dimerization. This observation would suggest that these residues could represent good candidates for new experiments of site-direct mutagenesis.

Table IV: Energy decomposition analysis of MD simulations

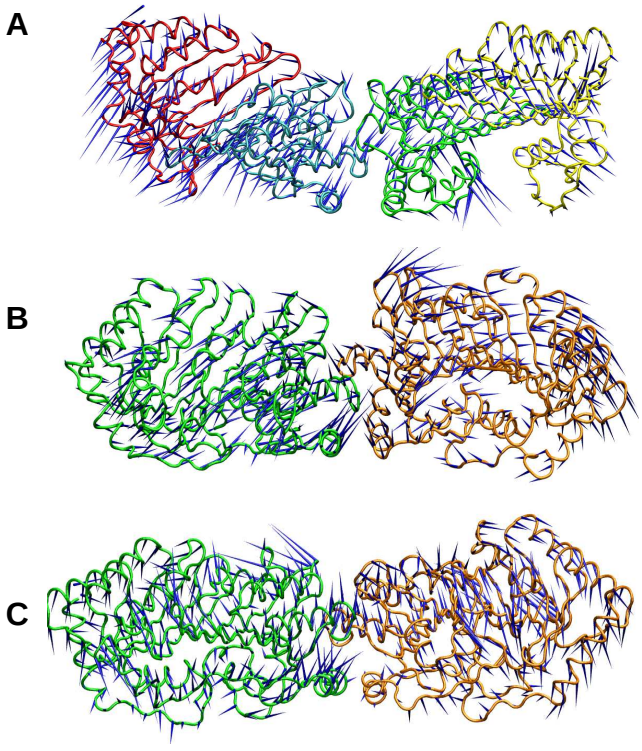
The first and the last 500 ps of simulations were used for the energy decomposition analysis.

Energy (kJ/mol)	APOBEC2 E initial	APOBEC2 E final	A3G E initial	A3G E final	A3F E initial	A3F E final
Pr-Pr (El)	-21319.8	-21958.3	-20446.5	-21682.9	-20457.2	-21785.4
Pr-Pr (LJ)	-26493.6	-27202.9	-27570.3	-29014.9	-27679.3	-28945.4
Pr-Sol (El)	-45859.4	-44159.4	-43159.1	-40086.1	-42958.5	-41345.2
Pr-Sol (LJ)	-3153.1	-2744.9	-4792.0	-3953.2	-3956.6	-3246.2

Principal component analysis (PCA) is a powerful tool for finding global, correlated motions in atomic simulations of macromolecules [24]. PCA is often used to reduce the complexity of the data obtained by MD simulations. The motion is decomposed in the principal components, associated to an eigenvector and an eigenvalue. PCA analysis performed on the modelled dimmers and on the crystal structure showed that in all systems α-carbon fluctuations are almost fully described by the first 10 eigenvectors. In the A2 simulation the first eigenvector accounts for the 50% of the total amount of the motion. In the A3G and A3F the first eigenvector account for the 56% and 36%respectively. Given that, we decided to analyse the first eigenvector of all systems, reducing reasonably the description of the overall motion. In Figure 4.13 the 'porcupine' plots of the first eigenvector of A2, A3G and A3F are shown. This analysis revealed that the large part of the observed motion between the two monomers of A3G/F is due the residues at the interface

Table V: Accessible surface areas buried upon formation of dimers.The residues that remain in contact during all the simulated time are highlighted in bold.

3G				A3F			
Initial		**Final**		**Initial**		**Final**	
Residue	ΔSASA (Å^2)	Residue	ΔSASA (Å^2)	Residue	ΔSASA (Å^2)	Residue	ΔSASA (Å^2)
Monomer A				**Monomer A**			
ARG 24	41.4	**ARG 24**	38.7	TYR 19	18.4	TYR 22	26
ILE 26	15.4	**SER 28**	15.5	ARG 24	26.1	**ARG 24**	32.3
LEU 27	66.5	**TRP 94**	18.7	LEU 27	59.3	TRP 93	15.2
SER 28	32.5	**ARG 122**	17.2	SER 28	43.5	**ARG 121**	25.1
TRP 94	14.5	TYR 124	21.4	ARG 29	25.8	**TYR 123**	17.8
LYS 99	28.6	PHE 126	81.6	ARG 121	31.8	**TYR 124**	70.6
ARG 122	40.2	**TRP 127**	31.8	TYR 123	19.3	**GLU 127**	48.5
TYR 125	95.5	ASN 177	18.2	TYR 124	39.5	ARG 128	15
TRP 127	107.5	LYS 180	93.2	TYR 125	12.7	ASP 129	25.7
ASP 128	25.3	TYR 181	31.2	TRP 126	81.5	TYR 171	88.7
PRO 129	16.7	LEU 184	19.4	GLU 127	15.6	**LYS 172**	69.4
ASP 130	39.1			LYS 172	35.7	ASP 174	11.7
Monomer B				PHE 179	42.7	ASP 175	60.6
TYR 22	26.2	PHE 126	105.8	TYR 19	18.4	**PHE 179**	50.9
LEU 27	83.7	**TRP 127**	92	**Monomer B**			
SER 28	40.7	PRO 129	26.1	ILE 26	21.3	**LEU 27**	148.2
ARG 29	40.9	ASN 179	21.2	LEU 27	84.4	SER 28	52.7
ARG 122	28.4	LYS 180	73.2	ARG 29	29.8	**TYR 124**	48.4
TYR 124	23.6	**TYR 181**	48.7	TRP 93	30.7	**TYR 125**	28.8
TRP 127	54.7	ILE 183	40.7	ARG 121	32.5	**TRP 126**	63.5
TYR 181	30.3	LEU 184	69.3	TYR 124	91	**GLU 127**	26
		MET 188	20.8	TYR 125	23.4	**ARG 128**	126.9
				TRP 126	98	**ASP 129**	45.3
				GLU 127	47.8	ARG 131	71.3
				ARG 128	35.1	ASN 176	48.8
				ASP 129	38.8	**PHE 179**	57.2
				PHE 179	30.5	THR 183	44.5



.

On the other hand A2 appear more rigid and the helices at the interface do not interact with each other.

These very preliminary data could lead at the design of new mutants and are essential to better clarify at molecular level the interactions essential for the dimerization.

Figure 4.12: Images of the motions corresponding to the first eigenvector for simulations of A2 (A), A3G (B) and A3F (C). Each Cα atom has a cone attached pointing in the direction of motion described in the eigenvector for that atom

Methods

Molecular Modelling

The structures of A3G/F were obtained by homology modelling from the crystal structure of human A2 (2NYT pdb entry), with which it shares 30% sequence identity with each monomer. 3D dimer models were generated using MODELLER. MODELLER is used for homology or comparative modelling of protein three-dimensional structures [25]. MODELLER implements comparative protein structure modelling by satisfaction of spatial restraints [26, 27], and can perform many additional tasks, including de novo modelling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function.

Molecular Dynamics simulations.

Refinement of the models has been performed by Molecular Dynamics simulations. [28, 29]Simulations on the crystal structure of A2 and on the models A3G/F were performed with the GROMACS package [30] using the GROMOS96 force field[31]. The systems were neutralized and were solvated in a box of SPC water [32], in a solution of 50mM NaCl (20 Na^+ and 38 Cl^-). All the simulated boxes contained about 20,000 water molecules. Simulations were carried out at a constant temperature of 300K. The Berendsen algorithm was applied for the temperature and pressure coupling.[33] After a first steepest descent energy minimization with positional restraints on the solute, the LINCS algorithm was used to constrain the bonds [34] and to carry out an initial 200 ps simulation with the positions of the solute atoms restrained by a force constant of 3000 kJ/(mol_*nm^2) in order to let the water diffuse around the molecule and to equilibrate. The particle mesh Ewald method (PME) [35] was used for the calculation of electrostatic contribution to non bonded interactions (grid spacing of 0.12 nm). The crystal structure of A2 was run for 5 ns whereas the models A3G/F for 10 ns each. The Dynamite server (www.biop.ox.ac.uk/) was used to produce further PCA analysis of the MD trajectories.

References:

[1] Holmes RK, Malim MH, Bishop KN. APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci* (2007);32:118-28.

[2] Fisher AG, Ensoli B, Ivanoff L, Chamberlain M, Petteway S, Ratner L, et al. The sor gene of HIV-1 is required for efficient virus transmission in vitro. *Science* (1987);237:888-93.

[3] Strebel K, Daugherty D, Clouse K, Cohen D, Folks T, Martin MA. The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* (1987);328:728-30.

[4] Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell* (2003);113:803-9.

[5] Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, Bollman B, et al. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* (2003);114:21-31.

[6] Conticello SG, Harris RS, Neuberger MS. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* (2003);13:2009-13.

[7] Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* (2003);424:99-103.

[8] Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* (2003);424:94-8.

[9] Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol* (2004);14:1392-6.

[10] Xie K, Sowden MP, Dance GS, Torelli AT, Smith HC, Wedekind JE. The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. *Proc Natl Acad Sci U S A* (2004);101:8114-9.

[11] Chung SJ, Friomme JC, Verdine GL. Structure of human cytidine deaminase bound to a potent inhibitor. *J Med Chem* (2005);48:658-60.

[12] Prochnow C, Bransteitter R, Klein M, Goodman M, Chen X. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* (2006).
[13] Betts L, Xiang S, Short SA, Wolfenden R, Carter CW. he 2.3 A crystal structure of an enzyme: transition-state analog complex. *Mol Biol* (1994);235:635-56.
[14] Johanssonn E, Mejlhede N, Neuhard J, Larsen S. Crystal structure of the tetrameric cytidine deaminase from Bacillus subtilis at 2.0 A resolution. *Biochemistry* (2002);41:2563-70.
[15] Huthoff H, Malim MH. Cytidine deamination and resistance to retroviral infection: towards a structural understanding of the APOBEC proteins. *Virology* (2005);334:147-53.
[16] Huthoff H, Malim MH. Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. *J Virol* (2007);81:3807-15.
[17] Zhang KL, Mangeat B, Ortiz M, Zoete V, Trono D, Telenti A, et al. Model structure of human APOBEC3G. *PLoS ONE* (2007);2.
[18] Luo K, Liu B, Xiao Z, Yu Y, Yu X, Gorelick R, et al. Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. *J Virol* (2004);78:11841-52.
[19] Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, et al. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol* (2005);15:166-70.
[20] Navarro F, Bollman B, Chen H, Konig R, Yu Q, Chiles K, et al. Complementary function of the two catalytic domains of APOBEC3G. *Virology* (2005);333:374-786.
[21] Simossis VA, Heringa J. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* (2005);33:W289-94.
[22] Kleinjung J, Fraternali F. POPSCOMP: an automated interaction analysis of biomolecular complexes. *Nucleic Acids Res* (2005);33.
[23] Fraternali F, Cavallo L. Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome. *Nucleic Acids Res* (2002);30:2950-60.
[24] Fujiwara S, Amisaki T. Molecular dynamics study of conformational changes in human serum albumin by binding of fatty acids. *Proteins* (2006);64:730-9.
[25] Martin-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* (2000);29:291-325.
[26] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* (1993);234:779-815.
[27] Fiser A, Do RK, Sali A. Modeling of loops in protein structures. *Protein Sci* (2000);9:1753-73.
[28] Flohil JA, Vriend G, Berendsen HJ. Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. *Proteins* (2002);48:593-604.
[29] Fan H, Mark AE. Mimicking the action of folding chaperones in molecular dynamics simulations: Application to the refinement of homology-based protein structures. *Protein Sci* (2004);13:992-9.
[30] Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* (1995);91:43-56.
[31] Daura X, Mark AE, van Gunsteren WF. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry* (1998);19:535-47.
[32] Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J, Pullman B. Interaction models for water in relation to protein hydration. Intermolecular Forces, 1981. p. 331-42.
[33] Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* (1984);81:3684-90.
[34] Hess B, Bekker H, Berendsen HJC, Fraaije J. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry* (1997);18:1463-72.
[35] Darden T, Perera L, Li L, Pedersen L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure Fold Des* (1999);7.
[36] Humphrey W, Dalke, A. and Schulten, K. VMD—visual molecular dynamics. *J Mol Graph* (1996);14:33–8.

PUBLICATIONS:

1. Madonna S., Papa R., Birolo L., **Autore F.**, Doti N., Marino G., Quemeneur E., Sannia G., Tutino M.L. and Duilio A. The thiol-disulfide oxidoreductase system in the cold-adapted bacterium *Pseudoalteromonas haloplanktis* TAC 125: discovery of a novel disulfide oxidoreductase enzyme. (2006),*Extremophiles*, 10:41-51
2. **Autore F.,** Melchiorre S., Kleinjung J., Morgan W.D., Fraternali F. Interaction of Malaria Parasite-Inhibitory Antibodies with the Merozoite Surface Protein $MSP1_{19}$ by Computational Docking (2007) *Proteins*, 66 (3):513-27.
3. Giardina P., **Autore F.,** Faraco V., Festa G., Palmieri G., Piscitelli A., Sannia G. Structural characterization of heterodimeric laccases from *Pleurotus ostreatus*. (2007), Appl. Microbiol. Biotechnol. 75 (6):1293-300
4. Festa, G., **Autore, F.,** Fraternali, F., Giardina, P., Sannia, G., Development of new laccases by directed evolution: functional and computational analyses, *Protein* (in press)
5. de Pasquale, D., Cusano, A.M., **Autore, F.,** Parrilli, E., di Prisco, G., Marino, G., and Tutino, M.L. The cold-active Lip1 lipase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 is a member of a new bacterial lipolytic enzyme family. *Extremophiles*, (in press)

POSTER PRESENTATIONS

1. **Autore F,**, Melchiorre S., Morgan W.D., Fraternali F. Interaction of Malaria Parasite-Inhibitory Antibodies with the Merozoite Surface Protein $MSP1_{19}$. First European Conference on Chemistry for Life Science Rimini (Italy), October 4-8, 2005
2. **Autore F.**, Fraternali F., Giardina P., Sannia G. Molecular modelling and electrostatic properties of *Pleurotus ostreatus* laccases. First European Conference on Chemistry for Life Science Rimini (Italy), October 4-8, 2005
3. Festa G., Giardina P., Piscitelli A., **Autore F.,** Cestone R.,Sannia G. Directed evolution of *Pleurotus ostreatus* laccases. 3rd European Meeting in Oxizymes, Oeiras, Portugal, September 7-9 2006
4. **Autore F.**, Pagano B., Rittinger K., Fraternali F. *In silico* Phosphorylation of the superSH3 domain of $p47^{phox}$. The 2007 annual meeting of the Italian Bioinformatics Society (BITS) 26-28 April 2007 Naples, Italy
5. Pagano B**., Autore, F.,** Rittinger, K., Fraternali, F. "*In silico* Phosphorylation of the superSH3 domain of $p47^{phox}$". MGMS Silver Jubilee Meeting, London, UK, 13 March, 2007.

ORAL PRESENTATION

6. **Autore, F.**. "Interazionie di anticorpi inibitori del parassita della malaria con la protein di superficie $MSP1_{19}$ mediante *docking* computazionale". *Bioinformatica e Biologia Computazionale in Campania, Istituto di Scienze dell'Alimentazione, CNR, Avellino, Italy, 18 dicembre 2006.*

VISITING APPOINTMENTS

12 Maggio-13 Giugno 2005 National Istitute for Medical Research, Mill Hill London, UK
8 Febbraio-31 Marzo 2006 Randall Division of Cell and Molecular Biophysics, King's College, London UK
10 Luglio-31 Luglio 2006 Randall Division of Cell and Molecular Biophysics, King's College, London UK
1 Febbraio-31 Marzo 2007 Randall Division of Cell and Molecular Biophysics, King's College, London UK
11 Luglio-31 Agosto 2007 Randall Division of Cell and Molecular Biophysics, King's College, London UK
11-31 Ottobre 2007 Randall Division of Cell and Molecular Biophysics, King's College, London UK

SCHOOLS

1. ESF 2nd Training course on Molecular Interactions: New Frontiers for Computational Methods Verona (Italy) July 13th - 16th 2004
2. Second European School on Bioinformatics, CMBI Radboud University Nijmegen, Netherlands. January 22-26 2005

Extremophiles (2006) 10:41–51
DOI 10.1007/s00792-005-0470-3

ORIGINAL PAPER

Stefania Madonna · Rosanna Papa · Leila Birolo
Flavia Autore · Nunzianna Doti · Gennaro Marino
Eric Quemeneur · Giovanni Sannia · Maria L. Tutino
Angela Duilio

The thiol-disulfide oxidoreductase system in the cold-adapted bacterium *Pseudoalteromonas haloplanktis* TAC 125: discovery of a novel disulfide oxidoreductase enzyme

Received: 11 April 2005 / Accepted: 28 June 2005 / Published online: 23 September 2005

Abstract In prokaryotes, protein disulfide bond oxidation, reduction and isomerization are catalyzed by members of the thioredoxin superfamily, characterized by the conserved C–X–X–C motif in their active site. Thioredoxins and glutaredoxins contribute to the reducing power in the cytoplasm, while the Dsb system catalyzes disulfide bonds formation in the periplasmic space.This paper addresses the question of disulfide bonds formation in a cold-adapted micro-organism, *Pseudoalteromonas haloplanktis* TAC 125 (*Ph*TAC125) by characterizing the DsbA system. We found distinctive features respect mesophilic counterparts that highlighted for the first time the occurrence of two adjacent chromosomal *DsbA* genes organised in a functional operon. The sophisticated transcriptional regulation mechanism that controls the expression of these two genes was also defined. The two DsbA proteins, named *Ph*DsbA and *Ph*DsbA2, respectively, were expressed in *Escherichia coli* and characterized.Results reported in this paper provide some insights into disulfide bonds formation in a micro organism isolated in the Antarctic sea water.

Communicated by G. Antranikian

S. Madonna · R. Papa · L. Birolo · F. Autore · N. Doti
G. Marino · G. Sannia · M. L. Tutino · A. Duilio (✉)
Dipartimento di Chimica Organica e Biochimica,
Università di Napoli "Federico II"—Complesso Universitario
Monte Sant'Angelo, Via Cynthia, 80126 Napoli, Italy
E-mail: anduilio@unina.it
Tel.: +39-081-674314
Fax: +39-081-674313

L. Birolo · G. Marino · M. L. Tutino
Facoltà di Scienze Biotecnologiche,
Università di Napoli "Federico II", Napoli, Italy

E. Quemeneur
CEA, Direction des Sciences du Vivant, DIEP/SBTN,
Marcoule, Bagnols-sur-Ceze, France

Keywords *Pseudoalteromonas haloplanktis* · Gene regulation · Cold adaptation · Thiol disulfide oxidoreductase pathways · Protein folding · DsbA

Introduction

A key step in oxidative protein folding is the formation of S–S bonds between correct cysteine pairs. In prokaryotes, oxidation of cysteine residues is a catalyzed process depending on the protein subcellular localization. Generally, cytoplasmic proteins do not contain structural disulfide bonds, although some enzymes like ribonucleotide reductase, thioredoxin peroxidase and methionine sulfoxide reductase form S–S bridges as part of their catalytic cycles (Bardwell et al. 1991; Schallreuter et al. 1991).

In prokaryotes, disulfide bond formation is achieved in the periplasmic space by the Dsb system, comprising a family of disulfide oxidoreductases belonging to the thioredoxin superfamily. This system can be divided into two pathways: an oxidation pathway, consisting of the DsbA and DsbB proteins, and an isomerization pathway that includes DsbC, DsbD and DsbG proteins (Hiniker et al. 2004). The initial oxidative event is catalyzed by DsbA, which interacts with reduced protein substrates and catalyses oxidation of their cysteine residues to disulfide bonds. Homologues of DsbA, characterized by the conserved –C–P–H–C– motif of the catalytic site, have been found in a wide range of bacteria (Hiniker et al. 2004), suggesting the conservation of the mechanisms involved in disulfide bond formation within the Gram negative bacteria.

Recently, besides the chromosomal DsbA, a second disulfide oxidoreductase homologue was identified in *Salmonella enterica* serovar *Typhimurium*, located in a virulence plasmid (Bouwman et al. 2003). This DsbA-like protein, named SrgA, is characterized by a –C–P–P–

C– motif and shows 37% identity with the canonical DsbA. Other SrgA homologues were also identified in *S. enterica* serovar Typhi and *S. enterica* serovar Enteritidis strains (Rodriguez-Penap et al. 1997). It is extremely unusual for organisms to contain more than one chromosomal DsbA; to the best our knowledge the only case reported so far is that of *Neisseria meningitidis* with three DsbAs suggested to have different activities in folding specific target proteins (Sinha et al. 2004; Tinsley et al. 2004).

This paper addresses the question of disulfide bonds formation in a cold-adapted micro-organism, *Pseudoalteromonas haloplanktis* TAC 125 (*Ph*TAC125) (Birolo et al. 2000) by characterizing the DsbA system. Results reported in this paper describe key enzymes of thiol-disulfide oxidoreductase system in a cold adapted micro-organism and demonstrate for the first time the existence of an uncommon DsbA gene organization.

Materials and methods

Bacterial strains, DNA constructs and media

Plasmids are all reported in Table 1. *Pseudoalteromonas haloplanktis* TAC125 (*Ph*TAC125) was collected in 1992 from seawater near the French Antarctic Station Dumont d'Urville (60°40′; 40°01′E) and grown in aerobic conditions at 15°C in TYP broth, pH 7.5 (Birolo et al. 2000).

Escherichia coli strains TOPF'10 and HB1O1 were used as hosts for gene cloning. BL21(DE3) *E. coli* strain (Sambrook et al. 2001) was used as host for heterologous expression.

E. coli cells were routinely grown in LB (Sambrook et al. 2001) containing 100 μg/ml of ampicillin, when transformed. *E. coli* JCB570 and JCB571 were used for motility assays (Bardwell et al. 1991). *E. coli* JCB816 and JCB817 were used for Lac$^-$ phenotype assays (Grauschopf et al. 1995). All these strains were a kind gift from Prof. J.C. Bardwell. pTRC99A expression plasmid (Amersham Biosciences) was used for complementation assay. pUC18 plasmid (Roche, Penzberg, Germany) was used for the construction of the *Ph*TAC125 DNA genomic library. pPLB vector was used for transcriptional analysis (Duilio et al. 2004). pET22b(+) (Novagen) was used for the expression of recombinant proteins (*Ph*DsbA and *Ph*DsbA2).

Table 1 Plasmids constructed in this study

Plasmid	Description[a]
pT(*PhdsbA*)	pTRC99A Δ (*Nco*I- *Eco*RI) Ω (*PhdsbA* gene)
pT(*PhdsbA2*)	pTRC99A Δ (*Nco*I- *Eco*RI) Ω (*PhdsbA2* gene)
pT(*EcdsbA*)	pTRC99A Δ (*Nco*I- *Eco*RI) Ω (*EcdsbA* gene)
pE(*dsbA*)	pET22b Δ (*Nde*I- *Eco*RI) Ω (*PhdsbA* gene)
pE(*dsbA2*)	pET22b Δ (*Nde*I- *Eco*RI) Ω (*PhdsbA2* gene)
P(*yihE*)	pPLB Δ (*BamH*I- *Eco*RV) Ω (*Phdsb* 71–355)
P(*dsbA2*)	pPLB Δ (*BamH*I- *Eco*RV) Ω (*Phdsb* 1640–1990)
P(*dsbA*)	pPLB Δ (*BamH*I- *Eco*RV) Ω (*Phdsb* 1190–1341)
P(*dsbA*^{a})	pPLB Δ (*BamH*I- *Eco*RV) Ω (*Phdsb* 1190–1279)
P(*dsbA*^{b})	pPLB Δ (*BamH*I- *Eco*RV) Ω (*Phdsb* 1279–1341)

[a] Nucleotide coordinates of *Phdsb* locus as deposited in EMBL Databank (AJ634705). Ω, insertion; Δ, deletion.

Cloning and sequencing of *PhdsbA* locus

Genomic DNA preparation from *Ph*TAC125 was carried out as previously described (Tosco et al. 2003). The *PhdsbA* locus was isolated from a *Ph*TAC125 *Hind*III genomic library by dot blotting screening, by using degenerated oligonucleotides as probes, designed on the *E. coli* amino acidic sequences and from F_{29} to P_{41} *Ec*DsbA for *Ph*DsbA.

The screening was carried out as reported by Georlette et al. (2003), and allowed to identify positive clones for *PhdsbA*. The positive clones were sequenced by the TaqFS dye terminator kit (Perkin Elmer), Norwalk, CT, USA, using the Applied Biosystems Automatic Sequencer model 373A (Perkin Elmer).

In order to complete the sequence of the *PhdsbA* locus, sequencing reactions were performed directly on the genomic DNA by using the Thermo Sequenase radiolabeled terminator cycle sequencing kit (Amersham Pharmacia Biotech., Freiburg, Germany), according to the procedure reported by Krin et al. (2001). In the case of DNA sequences with dyad symmetries containing dG and dC residues, a master mix containing dITP was used for the sequencing. Use of dITP required longer extension times (20 min) at 60°C.

The EMBL Databank accession number for *PhdsbA* locus is AJ634705.

Constructions of the expression plasmids

For the complementation assays, the *PhdsbA*, *PhdsbA2* and *EcdsbA* genes were amplified (Tosco et al. 2003) and cloned into the commercial expression vector pTRC99A, generating the pT(*PhdsbA*), pT(*PhdsbA2*) and pT(*EcdsbA*) plasmids, respectively (Table 1).

For the purification of the recombinant proteins, *PhdsbA* and *PhDsbA2* genes were PCR amplified and cloned into the commercial expression vector pET22b(+). The resulting constructs were indicated as pE(*dsbA*) and pE(*dsbA2*), respectively (Table 1).

For the transcriptional fusion experiments, the promoter regions upstream of *PhyihE* and *PhdsbA2* genes were amplified and cloned into the reporter vector pPLB (Duilio et al. 2001), generating P(*yihE*) and P(*dsbA2*) plasmids (Table 1). The two putative promoters upstream of *PhdsbA* gene were amplified and cloned together and separately, generating P(*dsbA*), P(*dsbA*^{a}), P(*dsbA*^{b}) plasmids, respectively (Table 1).

All the resulting plasmids were used to transform *Ph*TAC125 cells by interspecific conjugation (Duilio et al. 2001) and to perform the transcriptional assays.

Complementation of *E. coli* dsbA mutant strains

Two different phenotypes—dependent upon the presence of active DsbA homologous protein—were used for the complementation assays: motility and Lac^- phenotype of the MalF-β-galactosidase 102 fusion protein (Bardwell et al. 1991).

pT(*PhdsbA*), pT(*PhdsbA2*) and pT(*EcdsbA*) constructs were used for both types of assays. pTRC99A was used for negative control.

For motility assays, the constructs described were used to transform the non-motile *E. coli* strain JCB571 ($dsbA^-$) following standard procedure (Sambrook et al. 2001). *Ec*JCB570 ($dsbA^+$) and *Ec*JCB571(pT *EcdsbA*) strains were used as positive controls, while JCB571(pTRC99A) strain was used as negative control.

The complementation assays with MalF-β-galactosidase 102 fusion protein were performed in according to the procedure reported by Grauschopf et al. (1995), by using *Ec*JCB817 ($dsbA^-$) strain, *Ec*JCB816 and *Ec*JCB817 transformed with pT(*EcdsbA*) construct were used as positive controls.

Transcriptional assays

RNA isolation from *Ph*TAC125 cells (grown at 15°C up to 3 OD_{600}) and Northern blotting analysis were performed as described by Tosco et al. (2003). The P1, P2 and P3 probes used for these experiments were fragments internal to the *PhyihE*, *PhdsbA* and *PhDsbA2* genes, respectively, obtained by PCR amplification.

Primer extension experiments were performed as described by Tosco et al. (2003), by using 18-21 bp specific oligonucleotides as primers.

Reporter assays of transcriptional fusions measured activity of a cold-adapted β-galactosidase as described by Duilio et al. (2004).

Overexpression and purification of *Ph*DsbA and *PhDsbA2*

The recombinant *PhdsbA* and *PhdsbA2* genes were separately expressed in *Ec*BL21(DE3) cells, as follows: fresh cultures (2 ml) were inoculated into 200 ml LB medium containing 100 μg/ml of ampicillin. The recombinant cells were grown at 18°C for 20 h without any induction until the OD_{600} reached 4.5. The bacterial pellets were resuspended in 10 ml of buffer A (30 mM Tris–HCl pH 8, 20% sucrose, 1 mM EDTA pH 8) and incubated at room temperature for 20 min. The shocked cells were collected by centrifugation at 13,000 rpm at 4°C and resuspended in 10 ml ice-cold 5 mM $MgSO_4$. After incubation at 4°C for 20 min and centrifugation at 13,000 rpm, the supernatants (periplasmic fractions) containing the recombinant *Ph*DsbA and *Ph*DsbA2 proteins were collected and extensively dialyzed against 10 mM MOPS pH 7.2 and 10 mM MOPS pH 7.6, respectively.

For *Ph*DsbA protein purification, the sample was loaded on a DEAE-Sepharose *Fast Flow* column (Pharmacia Biotech. Inc., NJ, USA) equilibrated in 10 mM MOPS pH 7.2. Proteins were eluted with a linear NaCl gradient (0–0.3 M in equilibration buffer) and fractions were analysed for reducing activity with insulin as substrate (Holmgren 1979). The fractions containing the active protein were pooled, concentrated, dialyzed against 20 mM Tris–HCl pH 8, 1 M NH_4SO_4 and loaded on a *Phenyl Superose* column (Pharmacia Biotech. Inc.) equilibrated in the same buffer. The proteins were eluted with a linear gradient (20 mM Tris–HCl pH 8, 1 M NH_4SO_4—20 mM Tris–HCl pH 8).

For *Ph*DsbA2 protein purification, the sample was loaded on a SP-Sepharose *Fast Flow* column (Pharmacia Biotech Inc.) equilibrated in 10 mM MOPS pH 7.6 and proteins were eluted with a linear NaCl gradient (0–0.5 M in equilibration buffer). The active fractions were pooled, and loaded on a *Superdex 75* PC 3.2730 column (Pharmacia Biotech. Inc.) equilibrated in 50 mM Na-phosphate pH 7.5, 0.15 M NaCl, buffer.

Protein concentration was determined with the Bio-Rad protein assay (Bradford 1976), using bovine serum albumine as standard. The recombinant *Ph*DsbA and *Ph*DsbA2 proteins were stored at –20°C.

*Ph*DsbA and *Ph*DsbA2 antibodies production and Western blotting analyses

The anti- *E. coli* DsbA rabbit serum already described in Charbonnier et al. (1999) was shown to cross-react with *Ph*DsbA but not with *Ph*DsbA2. Thus, it could be used in Western blotting experiments. An antiserum against *Ph*DsbA2 was raised in rat, using inclusion bodies of the recombinant protein as antigen. The immunisation protocol involved three injections of 330 μg protein each; the first one in complete Freund adjuvant and the two further ones in incomplete Freund adjuvant at days 30 and 60. The optimal titer of antibodies were reached at day 75 where animals were bleeded (Charbonnier et al. 1999).

For Western blotting, protein samples were resolved by using 15% SDS-PAGE gel. Electrophoresis was done under reducing conditions using standard procedure (Sambrook et al. 2001). The proteins were transferred to a PVDF membrane using an electroblotting transfer apparatus (Trans-Blot Semi-Dry Transfer Cell, Bio-Rad, USA). *Ph*DsbA and *Ph*Dsba2 were detected by using the anti-rat polyclonal antibodies with ratio 1:1,000 and 1:3,000, respectively, and peroxidase-conjugated anti-rabbit secondary antisera (1:20,000) (A9169, Sigma, MO, USA). The membranes were developed by using SuperSignal West Femto Maximum Sensitivity Substrate detection kit (Pierce).

Mass spectrometric analysis

The molecular mass of the *Ph*DsbA and *Ph*DsbA2 proteins was determined by electrospray mass spectrometry (ESMS) on a ZQ single quadrupole mass spectrometer (Waters), by injecting protein solutions (10 pmol/μl) into the ion source at a flow of 5 μl/min. Data were elaborated using the Mass Lynx program (Waters).

CD spectroscopy

Spectroscopic characterization

Far-UV CD spectra were recorded on a Jasco J715 spectropolarimeter equipped with a Peltier thermostatic cell holder (Jasco model PTC-348), in a quartz cell of 0.1 cm light path at a protein concentration of 1.0 μM. Temperature was measured directly in the quartz cell, the solutions were filtered just before use on 0.22 μm pore size PVDF membrane (Millipore, Bedford, MA, USA), and data corrected by subtracting a control from which the protein was omitted. Spectra were recorded at 25°C from 280 to 184 nm at 0.2 nm resolution, 16 s response, at a scan rate of 20 nm/min. All data are the averages of three measures, and the results are expressed as mean residue ellipticity (θ), which is defined as $(\theta) = 100\ \theta_{obs}/lc$, where θ_{obs} is the observed ellipticity in degrees, c is the concentration in residue moles per liter, and l is the length of the light path in centimeters.

Thermal denaturation of the protein was followed by recording temperature-induced changes in secondary structure. Ellipticity at 220 nm was measured as the temperature was varied from 20 to 100°C at a rate of 1°C min^{-1}. Enzyme concentration was 1 μM in 10 mM HEPES, 1 mM DTT and 0.15 M NaCl, pH 7.5.

Linear baselines were fitted above and below the transition zone and the apparent fraction of molecules in the unfolded state (F_u) has been derived from the experimental mean residue ellipticity according to Eq. 1:

$$F_u = \frac{(y_n - y_{obs})}{(y_n - y_u)} \quad (1)$$

where y_n and y_u are the pre-transitional and post-transitional baselines, that are assumed to depend linearly on temperature.

Thermal unfolding transitions were analyzed with the two-state N ⇌ D model whose equilibrium constant is given by (2):

$$K_d(T) = \exp\{(-\Delta_d H(T_d)/R)(1/T - 1/T_d)\} \quad (2)$$

where T_d is the denaturation temperature at which $K_d = 1$ and $\Delta_d H(T_d)$ is the denaturation enthalpy change. The denaturation heat capacity change, $\Delta_d C_p$, is considered to equal zero because it cannot reliably be determined from CD measurements.

Correspondingly, the observed molar ellipticity is (3):

$$(\theta) = ((\theta)_N + (\theta)_D K_d)/(1 + K_d) \quad (3)$$

where $(\theta)_N$ and $(\theta)_D$ are the molar ellipticities of the native and denatured states, respectively, which are assumed to depend linearly on temperature. A non-linear least-squares regression was carried out to estimate the unknown parameters associated with the unfolding transition, using Micromath Scientist for Windows.

Results

Cloning and genetic organisation analysis of *PhdsbA* locus

Six thousand clones from a *Ph*TAC125 genomic library were screened using a degenerate primer as probe, designed on the basis of a multiple alignment of several amino-acid sequences from bacterial DsbA proteins available in protein databases. Two identical positive clones were identified, both containing a 1.9 kb insert, whose complete sequencing revealed the presence of three ORFs, two of which (ORF1 and ORF3) were incomplete. The complete ORF2, consisting of 621 bp, encodes a predicted protein of 207 amino acids, with theoretical *M*w of 22991 Da, exhibiting 33% identity with the *E. coli* DsbA protein (*Ec*DsbA) (Bardwell et al. 1991) and 40% identity with a DsbA homologue from *Vibrio parahaemolyticus* (entry Q87GP4).

With the aim of obtaining the complete sequence of the other two genes, the flanking regions of the ORF2 were sequenced by using the genomic DNA direct sequencing method (Krin et al. 2001). As a control, the genomic regions sequenced were amplified by PCR reactions using the *Ph*TAC125 genomic DNA as template, cloned in the pUC18 vector, and sequenced on both strands. Sequence data revealed that ORF1 is located 34 bp upstream ORF2 and consists of 972 bp, encoding a putative protein sharing 53% identity with the *E. coli* YihE protein (Belin et al. 1994), whose biological function is still unknown. ORF3, located 20 bp downstream ORF2 and consisting of 636 bp, encodes a protein of 212 amino acid residues with a theoretical *M*w of 23709 Da, displaying 39% identity with the DsbA homologue from *Vibrio vulnificus* (entry Q8DDF4).

Due to the canonical CPHC sequence of its active site, the protein encoded by the ORF2 was named *Ph*DsbA. The protein product of ORF3 showed a novel sequence at the active site (–C–P–A–C–), and, on the basis of homology considerations, it was tentatively identified as a DsbA-like protein. This protein shares 37% identity and 55% similarity with *Ph*DsbA and thereafter was named *Ph*DsbA2. In Fig. 1, the alignment of *Ph*DsbA and *Ph*DsbA2 to the *Ec*DsbA sequence is shown. Both *Ph*DsbA and *Ph*DsbA2, according to Von Heijne's rules (Von Heijne 1985), contain at the N-terminus a putative translocational signal, typical of the periplasmic proteins.

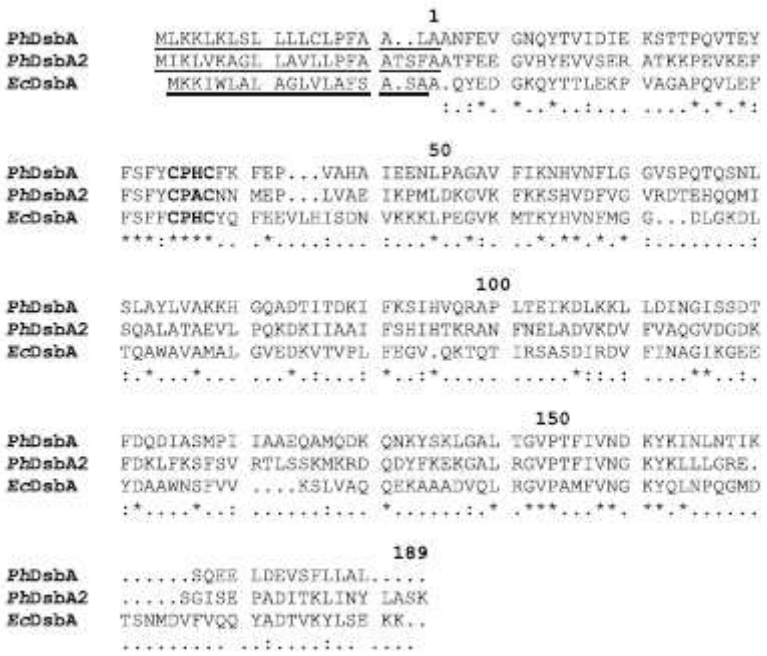


Fig. 1 Sequence alignment of DsbA homologues (*Ph*DsbA and *Ph*DsbA2) proteins from *Pseudoalteromonas haloplanktis* TAC125 with the corresponding counterparts from *E. coli* (*Ec*). Identical residues are labelled with (*), conservative substitutions with (:), and semi-conservative substitutions with (.), conserved translocation signal (_). The numeration of the amino acid residues refers to *Ec*DsbA

The nature of the flanking regions of the *PhdsbA* locus was further investigated by direct sequencing reactions, carried out on *Ph*TAC125 genomic DNA. Divergent oligonucleotides designed on the *PhdsbA* locus sequence were used as primers. Partial sequencing of the left side adjacent region revealed the presence of an ORF encoding a putative protein showing a high degree of similarity with ferredoxin from *Vibrio cholerae* (40%). Partial sequencing on the right side of *PhdsbA* locus revealed the presence of an ORF (ORFb), coding for a putative protein, which showed a significant similarity with the fimbriae-associated adhesion protein Fap1 from *Streptococcus parasanguis* (47%).

Transcriptional analysis of *PhdsbA* locus

Northern analyses were performed to investigate the transcriptional organization of the *PhdsbA* locus. [32] P-labeled DNA fragments internal to *PhyihE, PhdsbA* and *PhdsbA2* genes were used as probes. As shown in Fig. 2A, P1 probe recognized a 1.5 kb transcript corresponding to the co-transcription of the *PhyihE* and *PhdsbA* genes. P2 probe hybridises with two different transcripts: (1) a 1.5 kb transcript corresponding in size to *PhyihE* and *PhdsbA* genes together; (2) a 1.2 kb transcript that was attributed to the co-transcription of the *PhdsbA* and *PhdsbA2* genes. Finally, P3 probe detects two transcripts: (1) a 1.2 kb transcript corresponding in size to the co-transcription of *PhdsbA* and *PhdsbA2* genes; (2) a 0.65 kb transcript corresponding to the *PhdsbA2* monocistronic message. All probes recognize a 2.1 kb transcript whose size is consistent with the length of the mRNA corresponding to the three ORFs together.

Taken together (Fig. 2B), these results suggest the existence of: (1) an active promoter element located upstream of the *PhyihE* gene that is responsible for the full-length locus transcription and *PhyihE–PhdsbA* co-expression; (2) a promoter element upstream of the *PhdsbA* gene, responsible for the *PhdsbA* and *PhdsbA2* co-transcription; (3) a promoter sequence located upstream of the *PhdsbA2* gene, responsible for the synthesis of the *PhdsbA2* monocistronic message. Moreover, these data demonstrate the functionality of the putative transcriptional Rho-independent terminator, located 6 bp downstream of the ORF3 stop codon and indicate the occurrence of a transcriptional terminator, likely Rho-dependent, downstream of the *PhdsbA* gene.

Figure 3 shows primer extension experiments that explained the occurrence of the *PhdsbA– PhdsbA2* messenger. These analyses revealed two different transcriptional start sites for *PhdsbA* gene (Fig. 3A). A multiple distal start site (T/T) was identified 108 bp upstream of the translational start site, while a proximal start site, corresponding to a single thymine base, was identified 29 bp upstream of the *PhdsbA* start codon. The putative −10 and putative −35 boxes identified upstream of the two transcriptional start sites are shown in Fig. 3B. No transcriptional start site was experimentally detected for *PhyihE* and *PhdsbA2* genes although Northern analyses suggested their presence and putative promoter elements had been predicted by computational analysis.

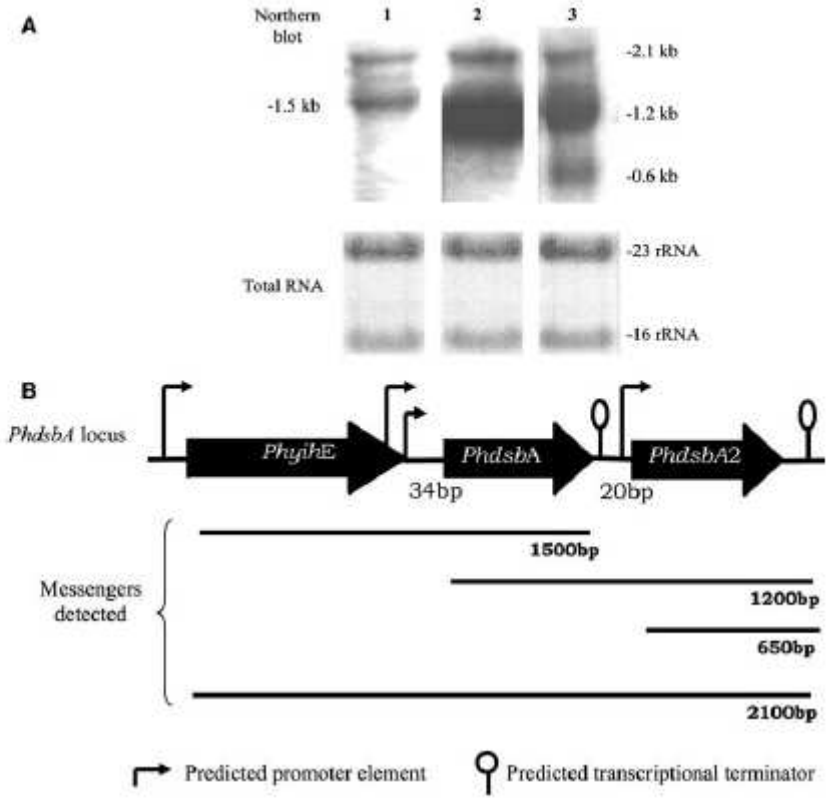


Fig. 2 A Northern blot analysis (*top panel*) and visualization of the same samples in an agarose-formaldehyde gel (*bottom panel*) of PhTAC125 total RNA. Total RNA samples (30 μg), isolated from mid-logarithmic-grown cells, were separated onto the 1.2% agarose gel, blotted to nitrocellulose membrane and hybridized with probes P1 (internal to *Phyi*hE, *lane 1*), P2 (internal to *Phd*sbA, *lane 2*), and P3 (internal to *Phd*sbA2, *lane 3*). B A schematic representation of the *Phd*sbA locus and its transcriptional organization

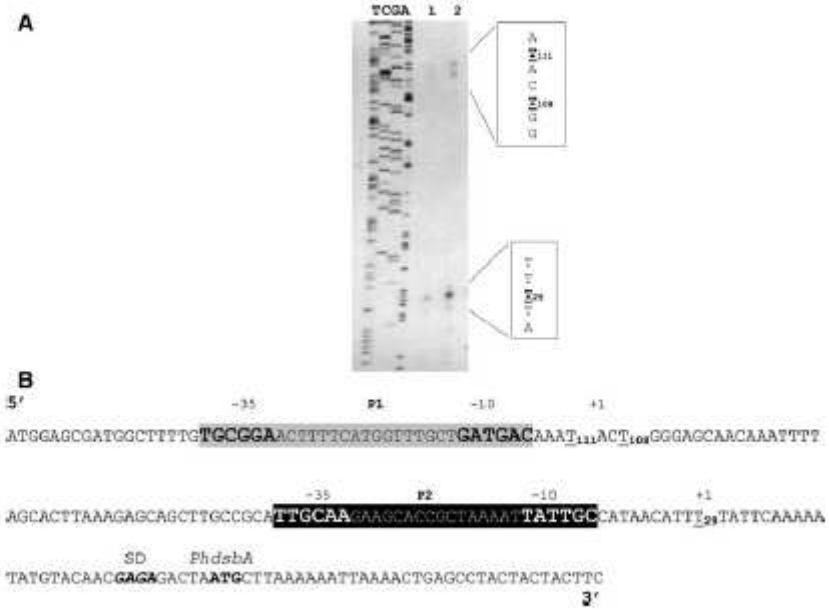


Fig. 3 A Primer extension analysis of the *Phd*sbA transcript. A 18 bp oligonucleotide was annealed to 10 μg (*lane 1*) and 30 μg (*lane 2*) of *PH*TAC125 total RNA and extended using AMV reverse transcriptase. The nucleotide sequence of the upstream region was determined using the same oligonucleotide as a primer (*lanes T, C, G* and *A*). B The putative P1 and P2 promoters are evidenced in *grey* and *black*, respectively. The corresponding −10 and −35 regions are in *bold*. The transcriptional start sites are *underlined*. The ribosome binding site (*SD*) is indicated as *italics*, while the *Phd*sbA start codon is indicated as *bold*

The transcriptional mechanism of the *PhdsbA* locus was further investigated by transcriptional fusion experiments. DNA fragments immediately upstream of *PhyihE* and *PhdsbA2* coding regions (150 and 300 bp, respectively) were individually fused to a promoter-less *lacZ* gene contained in a pPLB plasmid (Duilio et al. 2004), generating the P(*yihE*) and P(*dsbA2*) vectors (Table 1). Both fusion vectors displayed significant β-galactosidase activity (Fig. 4A), compared to the control vector (pPLB), thus confirming the presence of promoter elements within the regions upstream of *PhyihE* and *PhdsbA2* genes.

The transcriptional activity of the *PhyihE* and *PhdsbA2* promoters was further investigated by monitoring β-galactosidase activity during the growth of recombinant *Ph*TAC125 cells. Figure 4A shows that P(*yihE*) transformed cells exhibited a poor promoter activity in the early growth phase, while, during the exponential phase (20–40 h), a rapid increase in β-galactosidase accumulation was observed, with maximum levels at the late exponential phase. As for P(*dsbA2*) transformed cells, β-galactosidase levels reached the maximum value during the stationary phase (Fig. 4A); in both cases, the enzyme levels were constant for at least 12 h. These experiments clearly showed that the regions upstream *PhyihE* and *PhdsbA2* are differently regulated, the former being much more active in the late exponential phase and the latter in the stationary phase.

Similar analyses were performed to investigate the activity of the two promoters (P1 and P2) located upstream of the *PhdsbA* gene. P1 and P2 promoter elements were cloned both together and individually into the pPLB plasmid, generating the P(*dsbA*), P($dsbA^a$) and P($dsbA^b$) plasmids, respectively (Table 1). The corresponding β-galactosidase activity was then measured in *Ph*TAC125 transformed cells during the growth phases, as shown in Fig. 4B. The promoter activity of the P1 region steadily increases up to a maximum value in the stationary phase, while the P2 region provided an essentially constant β-galactosidase activity during the cellular growth. The P(*dsbA*) plasmid, containing the whole promoter region, showed a transcriptional activity strongly growth-phase dependent: in the exponential phase, *Ph*TAC125 cells transformed with P(*dsbA*) exhibited a large increase in β-galactosidase activity reaching the maximum level in the late exponential phase.

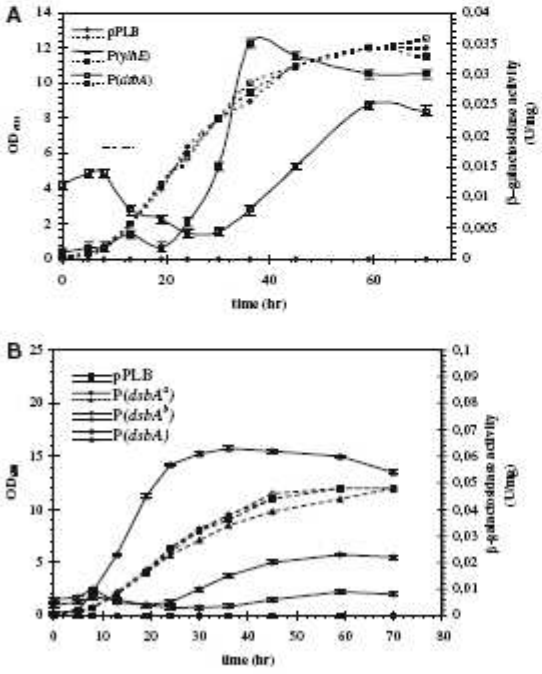


Fig. 4 Profiles of growth in aerated Typ cultures at 15°C and expression of *Ph*dsb::lacZ transcriptional fusion constructs in *Ph*TAC125. *Broken lines* indicate the growth curves, *solid lines* indicate the promoter activity. Each data point represents an average value from two samples. A pPLB (negative control), P(yihE) and P(dsbA2) cells. B pPLB (negative control), P(*dsbA*), P($dsbA^a$), P($dsbA^b$) cells.

*Ph*DsbA and *Ph*DsbA2 thermal stability

The *PhdsbA* and *PhdsbA2* genes were expressed in *E. coli* cells and the corresponding recombinant proteins purified from the periplasmic fractions by using the insulin reductase assay (Holmgren 1979), as described in the Materials and methods section. In both cases, ESMS analysis of the purified proteins showed the presence of a single component with a molecular mass of 20,808.8 ± 0.6 Da, and 21,450.4 ± 0.5 Da, for *Ph*DsbA and *Ph*DsbA2, respectively. These values were per se confirming the correct sequence of the recombinant proteins in agreement with the expected molecular mass of the mature proteins.

The thermal unfolding of *Ph*DsbA and *Ph*DsbA2 was investigated by means of CD measurements in comparison to *Ec*DsbA. The far-UV spectra are qualitatively similar (data not shown), thus suggesting a conserved secondary structure composition. Thermal unfolding of the "active" forms of DsbA was monitored by recording the molar ellipticity at 220 nm as a function of temperature (Fig. 5). Results are presented in Table 2, showing that *Ph*DsbA is less stable than *Ph*DsbA2 which is, in turn, less stable than *Ec*DsbA.

Preliminary analyses by differential scanning calorimetry on *Ph*DsbA and *Ph*DsbA2 revealed that the oxidized forms have denaturation points about 12–16°C lower than their reduced forms (data not shown). This is coherent with the thermodynamical mechanism of oxidative transfer of their disulfide bridge to the substrate protein in both cases, in agreement with the mechanism of *Ec*DsbA (Zapun et al. 1993; Moutiez et al. 1999).

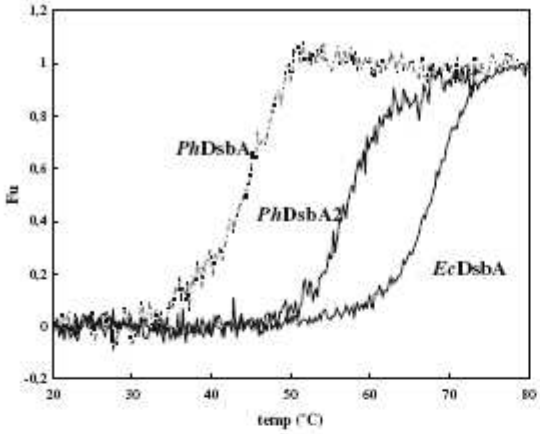


Fig. 5 The melting of secondary structure of DsbAs. Enzyme concentration was 10 μM in 50 mM Tris–HCl pH 7.0. Thermal transitions were followed at 220 nm as temperature was varied at a rate of 1°C min^{-1}. The fraction of unfolded protein was calculated from experimental data as described in the Materials and methods section

Table 2 Thermodynamic parameters of the thermal unfolding of oxidized DsbAs (*Ph*DsbA, *Ph*DsbA2, *Ec*DsbA), obtained by recording the molar ellipticity at 220 nm as a function of temperature

	T_d (°C)	$\Delta_d H(T_d)$ (kJ/mol)$^{-1}$
*Ph*DsbA (ox)	43.5 ± 0.4	4.4 ± 0.4
*Ph*DsbA2 (ox)	57.5 ± 0.5	3.3 ± 0.3
*Ec*DsbA (ox)	67.5 ± 0.3	4.5 ± 0.5

Cellular localization and functional characterization of *Ph*DsbA and *Ph*DsbA2 in *Ph*TAC125

The purified *Ph*DsbA and *Ph*DsbA2 proteins were used to produce-specific polyclonal antibodies in rat as described in the Materials and methods section. Western blotting analyses with anti-*Ph*DsbA and anti-*Ph*DsbA2 antibodies (Fig. 6) clearly showed that both proteins are actually produced by *Ph*TAC125, and totally translocated into the periplasm. Indeed, immunoreaction was only observed in the periplasmic fraction and not in the cytoplasmic portion.

With the aim of investigating the in vivo role of *Ph*DsbA and *Ph*DsbA2 proteins, we set up a complementation test taking advantage of the observation that *E. coli* cells harbouring null mutations in the *dsbA* gene have a pleiotropic phenotype as the correct folding of many proteins is affected. In particular, these mutants lack motility because of the improper assembly of the flagellar motor, due to incorrect disulfide bond formation in the P-ring protein (Dailey et al. 1993). We tested *Ph*DsbA and *Ph*DsbA2 proteins for their ability to restore the cellular motility of *E. coli* mutants.

The cold-adapted genes were PCR amplified, and separately cloned into pTRC99A expression plasmid. The resulting constructs, named pT(*PhdsbA*) and pT(*PhdsbA2*), were used for complementation assay of *E. coli dsbA*$^-$ strain JCB571. A construct containing the *EcdsbA* gene, named pT(*EcdsbA*), was used as positive control, and a non recombinant pTRC99A vector as negative control. As shown in Fig. 7, pT(*PhdsbA*) and pT(*PhdsbA2*) JCB571 recombinant cells (panels 3 and 4) exhibited a restored cellular motility on soft agar plates; panels 1 and 2 show the negative and positive controls, respectively.

A further complementation assay was performed in the *E. coli* strain JCB817, harbouring the *dsbA*:: *Km* null mutation, and encoding the MalF-LacZ102 fusion protein that confers a blue colour to bacterial colonies on plates containing 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-GAL) (Bardwell et al. 1991). In this strain, due to the *dsbA* mutation the β-galactosidase fused protein is able to assemble into an active enzyme, yielding a Lac+ phenotype (blue colonies). In the presence of DsbA-like activity, β-galactosidase is enzymatically inactive (white colonies), since disulfide bond formation causes the fused protein to be entrapped in the cytoplasmic membrane. pT(*PhdsbA*) and pT(*PhdsbA2*) JCB817 transformed cells did not develop

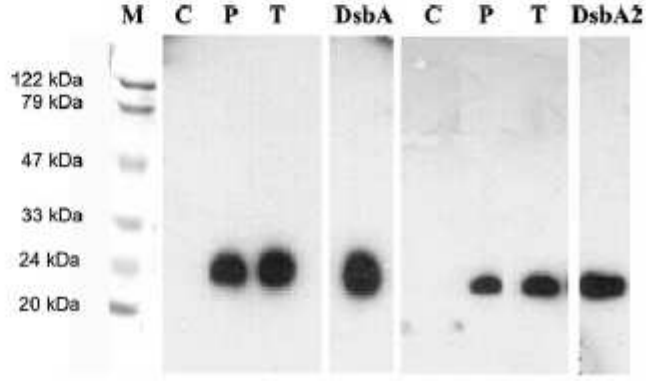


Fig. 6 Western immunoblot showing the cellular localisation of *Ph*DsbA and *Ph*DsbA2 proteins. *C* soluble cytoplasm fraction, *P* periplasmic fraction, *T* whole-cell lysate. DsbA purified protein, DsbA2 purified protein

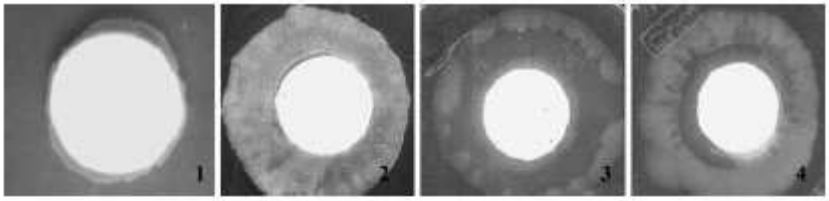


Fig. 7 Complementation assay of *E. coli* dsbA⁻ strain JCB571. 1 non recombinant pTRC99A vector (negative control). 2 pT(EcdsbA) JCB571 recombinant cells (positive control). 3 pT(*Ph*dsbA) and pT(*Ph*dsbA2) JCB571 recombinant cells. 4 pT(*Ph*dsbA2) JCB571 recombinant cells

any blue colour, suggesting that both proteins were able to substitute for *Ec*DsbA. JCB817 transformed with the empty vector, used as negative control, readily developed a blue colour (data not shown).

Discussion

The present study on disulfide isomerases from the cold-adapted Gram-negative bacterium *Pseudoalteromonas haloplanktis* TAC125 provides some insights into disulfide bonds formation in a micro-organism isolated in the Antarctic sea water.

*Ph*TAC125 carries two adjacent chromosomal genes (*PhdsbA – PhdsbA2*), encoding two disulfide oxidoreductases belonging to the DsbA family, named *Ph*DsbA and *Ph*DsbA2. The expression of these two genes is controlled by a sophisticated transcriptional regulation mechanism. To the best of our knowledge, this is the first report of two genes coding for two DsbA-like proteins organised in a functional operon.

A comparison of the amino acid sequences of *Ph*DsbA and *Ph*DsbA2, showing an 33 and 36% identity with *Ec*DsbA, respectively, immediately gives evidence that they are characterized by different motifs of the catalytic site. The *Ph*DsbA contains the canonical –C–P–H–C– motif, highly conserved in other DsbA homologues proteins, while in *Ph*DsbA2 the His in the active site is substituted by an Ala (–C–P–A–C–). We infer *Ph*DsbA2 as a proper DsbA isoenzyme for several reasons: (i) a BLAST run gives DsbA with the highest score, (ii) since the protein is located within the periplasm, it has to be involved at some stage in the oxidative folding pathway, and (iii) finally, *Ph*DsbA2 is not a membrane protein, thus a DsbB or DsbD-like recycling role could be ruled out. Moreover, the protein is monomeric with a molecular mass of 21 kDa: therefore, it is not a DsbC or a DsbG-like protein, since both of them are dimeric (Zapun et al. 1993; Andersen et al. 1997). It is also worth mentioning that a proper homologue of DsbC is present in *Ph*TAC 125 (unpublished results).

Bardwell research group elegantly demonstrated, by in vitro experiments carried out on *Ec*DsbA (Guddat et al. 1997), the importance of the electrostatic contribution of the histidine residue, since mutants in the third position of the catalytic site are less efficient as oxidant than wild type enzyme. Therefore, we would suggest a lower oxidizing power for *Ph*DsbA2. It is worth mentioning that an easy spectroscopic determination of the redox potential of *Ph*DsbA2 is impaired by the absence of Trp125 residue (Sillen et al. 1999). Experiments aimed at determining this parameter by chromatographic analysis (Siedler et al. 1993) are currently in progress.

The occurrence of additional *dsbA* genes in bacteria is rare but well documented, such as in the case of *Shewanella oneidensis* (Entry names: Q8EB18 and Q8EAM7) and *Salmonella enterica* (Bouwman et al. 2003). However, the genes encoding these DsbA-like proteins are either scattered on the chromosome or located on extrachromosomal elements. Here we report the first direct evidence in a Gram-negative bacterium of a functional operon comprising the genes encoding two DsbA proteins.

We observed that the flanking regions of the *dsbA* genes are well conserved and have a high similarity with those of *Salmonella* strains and *E. coli*. Indeed, the presence of *yihE* immediately upstream of the *dsbA* gene is quite widespread among the Gram-negative bacteria (Suntharalingam et al. 2003), as revealed by a computational comparison of bacterial genomes.

Partial sequence data concerning the region downstream of the *PhdsbA2* gene revealed the presence of an ORF (ORFb), encoding a protein homologous to the fimbriae-associated adhesin proteins, generally involved in biofilm formation and in fimbriae assembly. A similar genetic organisation was observed in several *Salmonella* strains, where these DsbA-like proteins are involved in oxidation of specific components of the fimbrial system (Bouwman et al. 2003; Rodriguez-Penap et al. 1997).

Although any involvement of the ORFb in the adhesion of *Ph*TAC125 is still under investigation, this observation suggests a possible function for *Ph*DsbA2. This hypothesis is supported by the ability of *Ph*TAC125 to form biofilm at 4°C (data not shown).

A complex mechanism of transcriptional regulation in *Ph*TAC125 for this operon was highlighted. Northern blotting analysis demonstrated that these genes are transcribed as a tricistronic messenger including the *PhyihE* gene, probably under the control of a promoter region located upstream of it. Additionally, the *PhdsbA* gene can also be transcribed as two different bicistronic

messengers: the first including the *PhyihE* gene, and the second one including *PhdsbA2*. Moreover, the *PhdsbA2* gene is also transcribed as a monocistronic transcript. Two adjacent transcription start points upstream of *PhdsbA*, identifying two distinct promoters (P1 and P2), were located within the 3′-terminal region of *PhyihE*, as also observed in *Salmonella Typhimurium* (Goecke et al. 2002).

The P2 promoter can be classified as a constitutive σ^{70} -dependent promoter, resembling the σ^{70} consensus sequence of *Ph*TAC125 (Duilio et al. 2004). The P1 promoter shows significant differences with constitutive promoters, suggesting the possibility that it is controlled by alternative σ factors, possibly responsible for the transcription under specific conditions. These hypotheses are supported by transcriptional fusion analyses that revealed a fairly constant activity of the P2 promoter region during cellular growth, typical of the σ^{70} -dependent promoters. On the contrary, the P1 promoter region is growth-phase regulated, exhibiting maximum activity during the *Ph*TAC125 stationary phase. The whole promoter region exhibits a marked dependence on the cellular growth, with a maximum activity during the late exponential phase, suggesting a synergic effect of the two promoters.

Although no transcription start site was detected upstream of *PhyihE* and *PhdsbA2* genes, transcriptional fusion analyses clearly demonstrate the presence of two active promoters responsive to growth phase. Their activity increases during the late exponential and the stationary phases, suggesting their dependence on alternative σ factors.

This analysis revealed that *PhdsbA* and *PhdsbA2* expression clearly increases during the late exponential phase, in which oxidoreductase proteins are presumably required for folding of components involved in the physiological changes that occur during this cellular phase. It is possible to suppose that *Ph*TAC125 cells devise a fine-tuning of transcriptional control for each gene, according to different growth conditions and/or different extra-cytoplasmic stimuli.

Both *Ph*DsbA and *Ph*DsbA2 showed the ability to substitute for *Ec*DsbA in *E. coli* mutants lacking this enzymatic activity. These data indicated that both proteins display similar oxidoreductase abilities in vivo. When multiple genes are discovered for proteins with apparently identical functions, the genes are often described as redundant. Taking advantage of the *Pseudoalteromonas haloplanktis* TAC125 genome annotation (to be published), the identification of *Ph*DsbAs protein substrates may, in the near future, be approached by a combination of proteomic tools and site-directed random mutagenesis experiments (Kadokura et al. 2004).

Acknowledgments This work was supported by grants of Ministero dell'Università e della Ricerca Scientifica (Progetti di Rilevante Interesse Nazionale 2002 and 2003; FIRB 2001), of Programma Nazionale di Ricerche in Antartide 2004, and of Regione Campania L.R. 05/03. Support from the National Center of Excellence in Molecular Medicine (MIUR—Rome) and from the Regional Center of Competence (CRdC ATIBB, Regione Campania—Naples) is gratefully acknowledged.We gratefully acknowledge Prof. J.C. Bardwell for providing the *E. coli* strains for complementation assays and to Prof. R.B. Freedman for critical reading of the manuscript.We thank Anne-Helene Davin (CEA Marcoule) for skillful technical assistance.

References

Andersen CL, Matthey-Dupraz A, Missiakas D, Raina S (1997) A new *Escherichia coli* gene, dsbG, encodes a periplasmic protein involved in disulphide bond formation, required for recycling DsbA/DsbB and DsbC redox proteins. Mol Microbiol 26:121–132

Bardwell JC, McGovern K, Beckwith J (1991) Identification of a protein required for disulfide bond formation in vivo. Cell 67:581–589

Belin P, Boquet PL (1994) The *Escherichia coli dsbA* gene is partly transcribed from the promoter of a weakly expressed upstream gene. Microbiology 140:3337–3348

Birolo L, Tutino ML, Fontanella B, Gerday C, Mainolfi K, Pascarella S, Sannia G, Vinci F, Marino G (2000) Aspartate aminotransferase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC 125. Cloning, expression, properties, and molecular modelling. Eur J Biochem 267:2790–2802

Bouwman CW, Kohli M, Killoran A, Touchie GA, Kadner RJ, Martin NL (2003) Characterization of SrgA, a *S. enterica* serovar *Typhimurium* virulence plasmid-encoded paralogue of the disulfide oxidoreductase DsbA, essential for biogenesis of plasmid-encoded fimbriae. J Bacteriol 185:991–1000

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Charbonnier JB, Belin P, Moutiez M, Stura E, Quemeneur E (1999) On the role of the cis-proline residue in the active site of DsbA. Protein Sci 8:96–105

Dailey FE, Berg HC (1993) Mutants in disulfide bond formation that disrupt flagellar assembly in *Escherichia coli*. Proc Natl Acad Sci USA 90:1043–1047

Duilio A, Tutino ML, Metafora V, Sannia G, Marino G (2001) Molecular characterization of a recombinant replication protein (Rep) from the Antarctic bacterium *Psychrobacter* sp. TA144. FEMS Microbiol Lett 198:49–55

Duilio A, Madonna S, Tutino ML, Pirozzi M, Sannia G, Marino G (2004) Promoters from a cold-adapted bacterium: definition of a consensus motif and molecular characterization of UP regulative elements. Extremophiles 8:125–132

Georlette D, Damien B, Blaise V, Depiereux E, Uversky VN, Gerday C, Feller G (2003) Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. J Biol Chem 278:37015–37023

Goecke M, Gallant C, Suntharalingam P, Martin NL (2002) *Salmonella typhimurium* DsbA is growth-phase regulated. FEMS Microbiol Lett 206:229–234

Grauschopf U, Winther JR, Korber P, Zander T, Dallinger P, Bardwell JC (1995) Why is DsbA such an oxidizing disulfide catalyst? Cell 83:947–955

Guddat LW, Bardwell JC, Glockshuber R, Huber-Wunderlich M, Zander T, Martin JL (1997) Structural analysis of three His32 mutants of DsbA: support for an electrostatic role of His32 in DsbA stability. Protein Sci 6:1893–1900

Hiniker A, Bardwell JC (2004) In vivo substrate specificity of periplasmic disulfide oxidoreductases. J Biol Chem 279:12967–12973

Holmgren A (1979) Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. J Biol Chem 254:9113–9119

Kadokura H, Tian H, Zander T, Bardwell JC, Beckwith J (2004) Snapshots of DsbA in action: detection of proteins in the process of oxidative folding. Science 303:534–537

Krin E, Hommais F, Soutourina O, Ngo S, Danchin A, Bertin P (2001) Description and application of a rapid method for genomic DNA direct sequencing. FEMS Microbiol Lett 199:229–233

Moutiez M, Burova T, Haertle T, Quemeneur E (1999) On the non-respect of the thermodynamic cycle by DsbA variants. Protein Sci 8:106–112

Rodriguez-Penap JM, Alvarez I, Ibanez M, Rotger R (1997) Homologous regions of the Salmonella enteritidis virulence plasmid and the chromosome of Salmonella typhi encode thiol: disulphide oxidoreductases belonging to the DsbA thioredoxin family. Microbiology 143:1405–1413

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Schallreuter KU, Wood JM (1991) New aspects in the pathophysiology of cutaneous melanoma: a review of the role of thioproteins and the effect of nitrosoureas. Melanoma Res 1:159–167

Siedler F, Rudolph-Bohner S, Doi M, Musiol HJ, Moroder L (1993) Redox potentials of active-site bis(cysteinyl) fragments of thiol-protein oxidoreductases. Biochemistry 32:7488–7495

Sillen A, Hennecke J, Roethlisberger D, Glockshuber R, Engelborghs Y (1999) Fluorescence quenching in the DsbA protein from *Escherichia coli*: complete picture of the excited-state energy pathway and evidence for the reshuffling dynamics of the microstates of tryptophan. Proteins 37:253–263

Sinha S, Langford PR, Kroll JS (2004) Functional diversity of three different DsbA proteins from *Neisseria meningitidis*. Microbiology 150:2993–3000

Suntharalingam P, Spencer H, Gallant CV, Martin NL (2003) *Salmonella enterica* serovar *typhimurium* rdoA is growth phase regulated and involved in relaying Cpx-induced signals. J Bacteriol 185:432–443

Tinsley CR, Voulhoux R, Beretti JL, Tommassen J, Nassif X (2004) Three homologues, including two membrane-bound proteins, of the disulfide oxidoreductase DsbA in *Neisseria meningitidis*: effects on bacterial growth and biogenesis of functional type IV pili. J Biol Chem 279:27078–27087

Tosco A, Birolo L, Madonna S, Lolli G, Sannia G, Marino G (2003) GroEL from the psychrophilic bacterium *Pseudoalteromonas haloplanktis* TAC 125: molecular characterization and gene cloning. Extremophiles 7:17–28

Von Heijne G (1985) Signal sequences. The limits of variation. J Mol Biol 184:99–105

Zapun A, Bardwell JC, Creighton TE (1993) The reactive and destabilizing disulfide bond of DsbA, a protein required for protein disulfide bond formation in vivo. Biochemistry 32:5083–5092