THE CDC45 DNA REPLICATION FACTOR: BIOCHEMICAL STUDIES AND BIOTECHNOLOGICAL PERSPECTIVES

Vincenzo Sannino

Dottorato in Scienze Biotecnologiche – XXV ciclo
Indirizzo Biotecnologie industriali e molecolari
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XXV ciclo
Indirizzo Biotecnologie industriali e molecolari
Università di Napoli Federico II

THE CDC45 DNA REPLICATION FACTOR: BIOCHEMICAL STUDIES AND BIOTECHNOLOGICAL PERSPECTIVES

Vincenzo Sannino

Dottorando: Vincenzo Sannino

Relatore: Prof. Giovanni Sannia

Co-relatore: Prof. Mosè Rossi

Coordinatore: Prof. Giovanni Sannia

A Vincenzo Sannino (1925-2013)

Index

RIASSUNTO

La replicazione del DNA è un processo fondamentale alla base della proliferazione cellulare. Di fatto nelle molecole di DNA, che fanno parte dei cromosomi, sono contenute tutte le informazioni necessarie alla cellula per crescere, duplicarsi e portare a termine tutti i compiti che le sono richiesti all'interno di un organismo altamente evoluto. Il meccanismo di replicazione del DNA in eucarioti si è progressivamente sviluppato e specializzato, al fine di evitare una duplicazione non corretta della informazione genica e per evitare che, in seguito ad una perdita o parziale mutazione di tale importante informazione, la cellula o addirittura l'intero organismo possano subire danni irreparabili.

Cdc45 è un fattore fondamentale coinvolto nel processo di replicazione del DNA in tutti gli organismi eucariotici. Esso è implicato sia nella fase di inizio della replicazione sia nella successiva fase di elongazione e rimane costantemente associato alla forcina replicativa per tutta la durata del processo di duplicazione del DNA. E' stato dimostrato che la proteina umana Cdc45 è un antigene associato alla proliferazione cellulare, in accordo con il suo coinvolgimento nel processo di replicazione del DNA. Negli esseri umani tale fattore è quasi totalmente assente nei normali tessuti cellulari adulti ed è presente solo nei tessuti con cellule attive dal punto di vista della proliferazione. La proteina è altamente espressa in tutte le cellule cancerose. In normali cellule proliferanti i livelli di Cdc45 restano costanti durante tutta la durata del ciclo cellulare, anche se un picco di espressione dell'mRNA codificante per la proteina è stato osservato al confine tra la fase G1 e la fase S.

E' stato osservato che il caricamento di Cdc45 sulla cromatina coincide con il *firing*, ovvero con l'attivazione, delle origini di replicazione. Le origini di replicazione sono dei punti specifici sul genoma dove la duplicazione del DNA può avere inizio. Un complesso specializzato di 6 proteine, noto come complesso ORC (*o*rigin *r*ecognition *c*omplex) è in grado di legare la cromatina specificamente alle origini di replicazione già durante la fine della fase M e durante la fase G1 del ciclo cellulare. Al complesso ORC si associano altri due fattori, Cdc6 e Cdt1, che sono indispensabili per il caricamento sulla forcina replicativa del complesso etero-esamerico MCM2-7. Tale complesso, composto da sei proteine paraloghe che hanno la capacità di legare ed idrolizzare molecole di ATP, rappresenta il motore molecolare dell'attività DNA elicasica, necessaria alla forcina replicativa per svolgere i due filamenti del DNA da duplicare. Il complesso MCM2-7 diventa cataliticamente attivo solo quando altri due co-fattori, il complesso etero-tetramerico GINS e Cdc45, vi si associano, formando il cosiddetto complesso CMG (*C*dc45/*M*CM2-7/*G*INS) che è in grado di fornire l'attività DNA elicasica indispensabile per lo svolgimento dell'elica di DNA. In seguito all'attivazione del complesso CMG la replicazione può avere inizio. Il meccanismo molecolare tramite cui Cdc45 è caricato sulla cromatina e si associa alla forcina replicativa non è ancora chiarito in dettaglio, ma è noto che altri fattori, come ad esempio MCM10, sono coinvolti in tale processo e che l'attività enzimatica delle protein chinasi CDK2 e DDK è indispensabile affinchè questo processo avvenga.

Quando abbiamo iniziato il nostro lavoro sulla proteina umana Cdc45 davvero poco era noto, soprattutto riguardo l'aspetto biochimico di questa proteina. Non era disponibile alcun dato strutturale e non c'erano altri fattori che potessero essere messi in relazione con Cdc45 da un punto di vista strutturale o funzionale. Abbiamo focalizzato la nostra attenzione sul complesso CMG, il complesso DNA elicasico di

cui Cdc45 fa parte, e siamo partiti tracciando un parallelo tra il macchinario replicativo degli organismi eucariotici e quello di Archaea. Infatti è noto che il sistema replicativo di questi ultimi è simile a quello eucariotico di cui costituisce una versione semplificata. Ad esempio per le sei proteine eucariotiche paraloghe che costituiscono l'etero-esamero del complesso MCM2-7 è presente in Archaea come controparte, una singola proteina MCM, omologa alle proteine MCM eucariotiche ed in grado di formare strutture omo-esameriche. Per quanto riguarda GINS, quattro differenti proteine compongono questo complesso negli organismi eucariotici, mentre in Archaea esso è formato solo da due proteine, GINS15 e GINS23, ancora una volta correlate alle proteine GINS eucariotiche, ed in grado di formare strutture tetrameriche. Tuttavia nel sistema replicativo degli Archaea non era nota alcuna proteina omologa o che in qualche modo potesse essere relazionabile al fattore eucariotico Cdc45 e questo risultava alquanto strano dato il ruolo fondamentale giocato da Cdc45 nella replicazione del DNA.

Abbiamo dunque cominciato la nostra attività sperimentale facendo una ricerca tramite l'algoritmo BLAST per individuare eventuali proteine omologhe a Cdc45 in Archaea. La sequenza amminoacidica della proteina umana Cdc45 è stata utilizzata come *query* e la ricerca è stata effettuata in un *database* genomico di Archaea. Il risultato ottenuto indicava un certo grado di similarità tra la proteina umana Cdc45 ed una proteina archeale appartenente alla famiglia delle proteine DHH. Si tratta di una famiglia di fosfoesterasi ad ampio raggio, diffuse in tutti i tipi di organismi: Eucarioti, Archaea e Batteri. Queste proteine sono anche note come proteine *RecJ-like* dal nome della DNA esonucleasi RecJ di *Escherichia coli*, una delle proteine più rappresentative di questa famiglia. Abbiamo comparato tramite allineamento multiplo le sequenze amminoacidiche di varie proteine Cdc45 eucariotiche e varie proteine *RecJ-like* dal mondo archeale e batterico. Il risultato di tale allineamento mostra che la proteina umana Cdc45, ed in generale tutte le omologhe proteine eucariotiche, presentano un buon grado di similarità con quelle batteriche ed archeali, solo nella porzione N-terminale. Confrontando però i motivi di struttura secondaria conservati nel *core* strutturale della proteina *RecJ-like* di *Thermus thermophilus*, di cui è nota la struttura cristallografica ai raggi X, con quelli predetti per la proteina umana, si può osservare un buon grado di similarità per tutta la lunghezza delle proteine.

Abbiamo prodotto e purificato la proteina umana Cdc45 in forma ricombinante da cellule batteriche. Abbiamo analizzato mediante SAXS la struttura della proteina ricombinante in soluzione in collaborazione con la Dr Silvia Onesti (Elettra Sincrotrone, Trieste). I dati raccolti ci hanno permesso di costruire un modello della struttura tridimensionale di Cdc45. La proteina sembra essere strutturata a forma di cornetto, con una porzione centrale più grossa e compatta, e due estensioni laterali che si diramano ai due lati opposti. Abbiamo dimostrato che la struttura del *core* della proteina RecJ-like da *Thermus thermophilus* può essere sovrapposta, in maniera altamente coincidente, con la porzione centrale del modello di struttura da noi predetto per la proteina umana Cdc45. Dunque la porzione centrale sembrerebbe essere realmente relazionabile al *core* proteico delle proteine DHH, mentre le due estensioni laterali sono verosimilmente attribuibili a due inserzioni amminoacidiche, una presente solo nella sequenza delle proteine eucariotiche Cdc45, l'altra presente anche nei fattori *RecJ-like* di Archaea ma non in quelli batterici.

Nelle proteine DHH, una serie di motivi strutturali altamente conservati sono coinvolti nella coordinazione di due ioni metallici bivalenti, che sono necessari a queste

proteine per l'attività enzimatica fosfoesterasica. Nonostante nella sequenza della proteina umana Cdc45 siano conservati solo alcuni dei residui amminoacidici coinvolti nel legame dei metalli nelle proteine DHH, abbiamo ritenuto ugualmente opportuno indagare l'eventuale presenza di ioni metallici ad essa legati. I risultati ottenuti tramite tecniche diverse, indicano che Cdc45 non lega alcuno ione metallico. In accordo con questo dato abbiamo verificato che Cdc45 umana non possiede né attività pirofosfatasica né DNA esonucleasica, questa ultima tipica delle proteine *RecJ-like*. E' stata inoltre valutata la capacità della proteina umana Cdc45 di legare il DNA e si è trovato che essa mostra affinità di legame per DNA a singolo filamento, mentre non ha affinità per il DNA a doppio filamento o per molecole di RNA.

Tutti questi risultati sono in accordo con l'ipotesi da noi proposta che Cdc45 si sia evoluta da una proteina DHH ancestore. La proteina nel corso della evoluzione avrebbe perso molti dei residui amminoacidici che erano coinvolti nel legame degli ioni metallici, perdendo di conseguenza anche la propria attività catalitica. Un ricordo delle proteine *RecJ-like* rimane nella capacità di Cdc45 di legare DNA a singolo filamento proprio come molte delle DNA esonucleasi *RecJ-like*. Ad ogni modo, la porzione centrale e compatta del nostro modello per la struttura tridimensionale di Cdc45 non lascia dubbi sulla forte similarità che intercorre col *core* proteico delle proteine *RecJ-like*.

Nella sequenza della proteina umana Cdc45 mancano motivi strutturali che sono solitamente presenti nelle proteine che legano il DNA. Abbiamo ipotizzato che la capactià di Cdc45 di legare DNA a singolo filamento potesse dipendere dalla presenza di una serie di amminoacidi carichi positivamente nella porzione C-terminale della proteina. Per valutare e validare questa ipotesi, sono stati prodotti due costrutti esprimenti due diverse forme tronche della proteina umana. La prima forma tronca corrisponde alla porzione N-termiale della proteina (amminoacidi 1-258) mentre la seconda codifica per la porzione C-terminale di Cdc45 (amminoacidi 394-566). Entrambi i costrutti sono stati espressi in cellule batteriche. Il mutante N-terminale è risultato essere stabile e solubile, ed è per tanto stato purificato in maniera efficiente, mentre il mutante C-terminale è risultato essere altamente instabile e non è stato possibile purificarlo. Ad ogni modo, il mutante esprimente la porzione N-terminale di Cdc45 è stato impiegato per dei saggi di legame al DNA e i risultati ottenuti indicano che questa forma tronca non è più in grado di legare il DNA così come la proteina intera. Questo indicherebbe che effettivamente la regione di Cdc45 impiegata nel legame al DNA è la porzione C-terminale della proteina.

Per studiare il ruolo di Cdc45 quale subunità accessoria della DNA elicasi replicativa abbiamo lavorato a lungo per ottimizzare la produzione del complesso umano CMG in forma ricombinante. Il metodo impiegato a tale scopo è stato quello delle cellule di insetto (nello specifico *Sf9*) infettate con *baculoviruses* ricombinanti. Nei primi tentativi di ottimizzazione di questo sistema di produzione 11 *baculoviruses* differenti, codificanti per le singole 11 subunità che costituiscono il complesso CMG, erano impiegati per un ciclo di infezione delle cellule di insetto. *Baculoviruses* codificanti per i diversi fattori proteici costituenti il complesso CMG, fusi a *tags* diversi, sono stati impiegati nei vari tentativi di produzione, così come diversi approcci sono stati impiegati per il primo passaggio di purificazione che consisteva in una cromatografia d'affinità. Nello specifico si è cercato di sfruttare, per la purificazione del complesso CMG, il *FLAG-tag* fuso alla subunità MCM7 o in alternativa uno *STREP-tag* fuso ad MCM2 o MCM3. Seguendo questo tipo di protocollo di produzione, solitamente si

otteneva l'espressione di tutti gli 11 fattori costituenti il complesso CMG nell'estratto solubile derivante dalle cellule infettate, come indicato dalle analisi di *western blot*. Nonostante ciò nelle frazioni eluite specificamente dalla resina di affinità non era possibile individuare tutte le componenti del complesso CMG. Dunque tutte le proteine risultavano essere espresse in cellule di insetto in seguito all'infezione virale, ma tali fattori non sembravano associarsi a formare un complesso stabile. Verosimilmente ciò dipendeva dal fatto che il sistema basato sulla co-infezione con 11 virus non consentiva la produzione coordinata e contemporanea in tutte le cellule infettate (o almeno in buona parte di esse) degli 11 fattori proteici formanti il complesso CMG. E' inoltre noto che le proteine MCM possono associarsi a formare dei sotto-complessi diversi dal complesso MCM2-7, come ad esempio MCM 2/4/6/7, MCM 4/6/7, MCM 3/5. La formazione di questi sotto-complessi può rappresentare un ulteriore impedimento nei termini di una produzione efficiente per il complesso CMG. Per cercare di ovviare almeno in parte a queste problematiche, abbiamo impiegato un nuovo sistema (Multi Bac^{Turbo} System) che ci ha permesso di ottenere dei *baculoviruses* multi-genici in grado di co-esprimere in cellule di insetto diversi prodotti proteici. In particolare è stato prodotto un *baculovirus* codificante contemporaneamente per MCM2, MCM4 ed MCM6, un secondo *baculovirus* codificante per MCM3 (+FLAG tag), MCM5 ed MCM7, ed infine un *baculovirus* codificante per Cdc45 e le quattro subunità del complesso GINS: Psf1, Psf2, Psf3 e Sld5 (+GST tag). L'impiego di questi *baculoviruses* codificanti per più proteine ci ha consentito di diminuire il numero di virus impiegati per ogni singolo ciclo di infezione, favorendo una espressione contemporanea e maggiormente coordinata delle singole proteine prodotte in ciascuna cellula infettata. Inoltre le sequenze codificanti per le varie proteine del complesso MCM2-7 sono state inserite in due virus separati in una disposizione tale da minimizzare la formazione di sotto-complessi da parte delle proteine MCM. Impiegando questi tre *baculoviruses* per l'infezione di circa 400 x 10^6 cellule *Sf9* siamo riusciti ad ottenere la produzione di un complesso CMG stabile che è stato isolato dall'estratto totale mediante cromatografia d'affinità, sfruttando il *FLAG-tag* fuso alla subunità MCM3. Nonostante tutto, la quantità di complesso ottenuta è davvero esigua e il materiale isolato non può essere impiegato per successivi passaggi di purificazione o per analisi strutturali e/o funzionali del complesso.

Un altro sistema che è stato impiegato per analizzare il ruolo di Cdc45 nel processo di replicazione del DNA è quello degli estratti di uova di *Xenopus laevis*. Si tratta di un sistema molto utile impiegato da anni in particolare per lo studio della replicazione negli organismi eucariotici. Gli estratti ottenuti dalle uova di *X. laevis* sono ricchi di fattori coinvolti nella replicazione del DNA. Se il DNA cromatinico, proveniente dai nuclei spermatici privati di membrana della rana maschio, è aggiunto a questi estratti, la cromatina spermatica comincia a decondensare la sua struttura compatta e una sorta di membrana nucleare chimerica si forma attorno a tale materiale genetico. A questo punto comincia una vera e propria replicazione del DNA *in vitro* che può essere monitorata seguendo l'incorporazione di nucleotidi marcati radioattivamente o con molecole fluorescenti. Inoltre con questo sistema il fattore replicativo di interesse può essere eliminato dagli estratti mediante l'impiego di anticorpi specifici. A questo punto la replicazione può essere saggiata per analizzare l'effetto della eliminazione di quello specifico fattore, oppure mutanti della proteina di interesse possono essere sostituiti alla proteina *wild type* per valutarne l'effetto. Nel mio caso, avendo deciso di impiegare questo sistema per analizzare una proteina eterologa, ho dovuto prima di tutto verificare che la proteina umana Cdc45 potesse essere

impiegata nell'estratto dello *Xenopus*. I saggi di replicazione effettuati hanno dimostrato che la proteina ricombinante umana era in grado di ripristinare, così come la proteina ricombinante di rana, l'attività replicativa negli estratti immuno-privati del fattore endogeno mediante l'utilizzo di anticopri specifici. Allo scopo di indagare, con questo sistema sperimentale, il ruolo svolto dalle due estensioni laterali di Cdc45, non presenti nel *core* conservato delle proteine DHH, ho prodotto tre mutanti tronchi della proteina umana. In un primo mutante, A1_LP_1, circa il 70% dell'inserzione amminoacidica L1, quella presente solo nelle proteine Cdc45 eucariotiche, è stato eliminato. In un secondo mutante, ΔL1 la sequenza L1 è stata eliminata per intero. Nell'ultimo mutante ΔL2, l'altra inserzione amminoacidica L2, presente sia nelle proteine eucariotiche che in quelle archeali, è stata totalmente eliminata. I mutanti ΔL1 e ΔL2 non si sono potuti utilizzare per il lavoro con gli estratti di *X. laevis* in quanto formano aggregati in soluzione e dunque non sono più nello stato fisiologico proprio della proteina monomerica *wild type*. Il mutante A1_LP_1 è stato invece impiegato per dei saggi di replicazione del DNA negli estratti immuno-privati del fattore Cdc45 endogeno. I risultati ottenuti indicano che questo mutante può ripristinare, come la proteina umana *wild type*, la replicazione del DNA negli estratti immuno-privati del fattore endogeno. Questo implicherebbe che la sequenza eliminata nella forma tronca di Cdc45 non è indispensabile per il funzionamento di tale fattore nella replicazione del DNA in condizioni non perturbate.

Data la specificità di espressione di Cdc45 in cellule proliferanti, come le cellule tumorali, negli ultimi anni la comunità scientifica sta focalizzando l'attenzione su questo fattore e lo si comincia a valutare come possibile *target* per eventuali terapie molecolari anti-cancro. Dato il ruolo fondamentale di Cdc45 nella regolazione dell'inizio del processo di replicazione del DNA, una terapia molecolare che inibisca in maniera specifica l'associazione di Cdc45 alla forcina replicativa, causando un blocco altamente specifico delle cellule proliferanti, potrebbe rappresentare una valida alternativa alle metodiche farmacologiche attualmente in uso che prevedono l'impiego di agenti chemioterapici.

SUMMARY

DNA replication is a crucial step for cell survival. All the information useful for the cells to propagate and complete their tasks is indeed kept in the DNA molecules that are needed to be duplicated in a proper way. Cdc45 is a fundamental factor involved both in the initiation and in the elongation phases of the DNA replication process. According to its function in the DNA replication mechanism it has been shown to be a proliferation-associated antigen and in fact it is expressed only in proliferating cells. Even if a peak for the mRNA codifying for Cdc45 can be observed at the G1/S border, levels of Cdc45 are stable throughout the cell cycle in proliferating cells.

The loading of Cdc45 onto the chromatin has been shown to coincide with the *origin firing* (i.e. replication origins activation). During late M and G1 phases of the cell cycle, replication origins, the peculiar sequences on the genome where DNA replication process is supposed to start, are bound by the *O*rigin *R*ecognition *C*omplex (ORC). This latter is itself bound by two other factors, Cdc6 and Cdt1, which are involved in the loading onto the chromatin of the MCM2-7 complex. This complex represents the molecular motor of the replicative DNA helicase, but at this stage it remains in an inactive state. At the G1/S border the GINS complex and Cdc45 are recruited at the replicative forks and associate with the MCM2-7 complex, bringing to the formation of the so called CMG complex, which is finally able to start working as DNA helicase, unwinding the DNA helix which can now be duplicated by the DNA polymerases. Even if the peculiar mechanism by which Cdc45 is loaded onto DNA is not fully understood, it is known that the activity of CDK2 and DDK kinases is required to this purpose. Cdc45 is the last factor to be loaded onto chromatin at the replicative fork and it is actually used by cells to switch on the DNA replication process. Once DNA replication has started, Cdc45 will keep on moving with the replicative fork co-operating together with GINS and MCM2-7 to provide DNA helicase activity.

When we started our analyses on the human Cdc45 factor, not much was known about the biochemistry of this protein. Indeed, no structural data were available for Cdc45 and no enzymatic activity was known to be associated with it. We carried out a Blast search, using the human Cdc45 aminoacidic sequence as query, looking into archaeal genome databases. We started from the observation that all the other factors forming the eukaryotic CMG complex were found to have a counterpart in the archaeal DNA replication machinery, while no homologous was known for Cdc45 in these organisms. The result we got was actually highlighting a match between the human Cdc45 protein and an archaeal protein belonging to the DHH protein super-family, which includes a wide range of phosphoesterases present in all organisms (i.e. Bacteria, Archaea and Eukarya). Comparing the human Cdc45 primary sequence with the ones from many other bacterial and archaeal DHH super-family members, a good degree of similarity can be observed in the N-terminal portion of these sequences. Comparison of the secondary structure elements of the core of DHH proteins with the predicted secondary structure motifs for the human Cdc45 protein, a good degree of similarity can be observed along the whole sequences. We produced and purified human Cdc45 from *E. coli* cells in recombinant form and we carried out SAXS analyses on this protein, in collaboration with Dr. Silvia Onesti (Sincrotrone, Trieste). The reconstruction model based on the results of these analyses shown that human Cdc45 seems to be folded in a kind of cornet shape, with a more compact central core and two long lateral extensions. The crystal

structure of the *Thermus thermophilus* RecJ-like protein (a DHH phopshoesterase) can be properly superimposed with the central compact core of the hCdc45 predicted structure. Instead the two lateral extensions can be attributed to two long aminoacidic insertions which are not present in the bacterial proteins. Since DHH phosphoesterases usually bind two metal ions that are required for their catalytic activity, we checked if also the human Cdc45 protein was able to bind any kind of metal ion. According to the absence of most of the aminoacidic residues responsible for binding metal ions in the DHH proteins, Cdc45 was found not to bind any metal ion and, in agreement with this finding, it was found to show neither phyrophosphatase nor DNA exonuclease activity. We demonstrated that human Cdc45 has the ability to bind single-stranded DNA, while it shows no affinity for double-stranded DNA or RNA molecules. All these results were in agreement with the idea that Cdc45 has evolved from an ancestor DHH protein, that lost the enzymatic activity and just retained the ability to bind single-stranded DNA.

We demonstrated that the N-terminal region of Cdc45 is not involved in the DNA binding activity, which can be likely ascribed to the C-terminal portion of the protein. Moreover, by using the *Xenopus laevis* egg extract experimental system we demonstrated that the “L1 loop”, a long aminoacidic insertion which is unique to the eukaryal Cdc45 proteins and is absent in the related archaeal and bacterial proteins, is not needed for the role played by Cdc45 in DNA replication in not perturbed conditions.

To study the role played by Cdc45 in association with GINS and MCM2-7, we finally set up a method to produce the CMG complex in insect cells co-infected with multi-genic baculoviruses.

INTRODUCTION

Cell proliferation is the paramount mechanism for the maintenance and the propagation of life. Depending on their tasks, cells can differentiate, get older and die or they can undergo cell division to proliferate. *Omnis cellula e cellula*: as the famous quotation summarizes, a cell can only originate from another cell which has divided itself to generate two daughter cells.

Daughter cells need to possess all the material that was present in the mother cell to have a proper development and life cycle. The most important thing is to stably maintain the genetic information contained in the DNA molecules. Indeed the information coming from DNA is the base for life, providing instructions to the cells on how to live, how to grow up, how to interact with the other cells, how to divide into two daughter cells, how to replicate DNA before cell division and so how to keep the informations contained in the same molecular structure which is the base for this information.

The series of events actually leading a cell during all its life are tightly regulated by means of a number of proteins. These proteins can be factors involved in mechanisms dictating the cell life progression timing, factors that work as molecular motors for the progression of cell life's phases or these can be proteins which actually work as molecular controllers on this progression of phases. All these factors are indeed involved in the regulation of what we call *Cell Cycle*, which is tightly regulated and shows different timing on the basis of the cell type and the moment of cell life.

The eukaryotic cell cycle

The series of events by which cells grow and prepare all the material to proliferate is named *Cell Cycle*. In eukaryotic organisms this process is the basis for the propagation of unicellular species. In complex multi-celluar organisms it is indispensable not only for the propagation of these species, but also for the correct growing and physiology of the living organism. In complex organisms some cells, such as skin cells, continuously go trough cell cycle undergoing a number of cell divisions, while other cells, more differentiated, such as neurons, just stop growing and dividing once they are mature and so definitively differentiated.

In active proliferating cells mitosis, the process by which a cell eventually divides into two identical daughter cells, represents only the final step of the cell cycle. All the remaining phases of the cell cycle are included in the so called *interphase* (Fig. 1).

The interphase is characterized by a continuous growing process and is itself divided into three different sequential steps. The first one is called G1 (gap 1) phase and it starts just after the previous mitotic phase. It is characterized by active RNAs and proteins production, which are the basis for cell growing and sustenance. Differentiated cells, such as neurons, are in a prolonged G1 phase also named G0 phase.

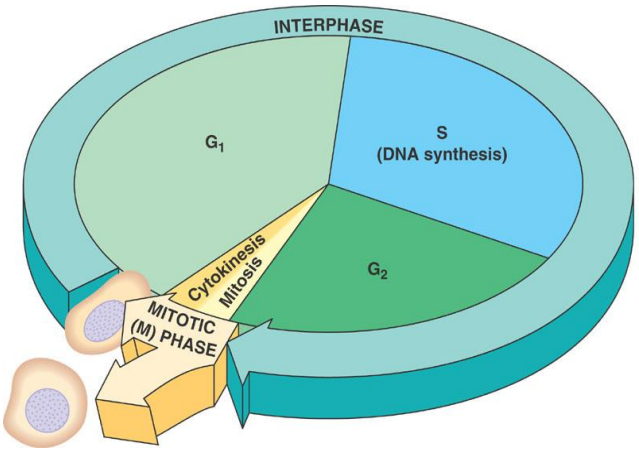

Figure 1 – Schematic representation of the cell cycle. During interphase cells grow and duplicate their DNA by three main steps. At the end they divide into two new daughter cells during the mitotic phase.

In this case cells are not growing anymore and they just keep on producing all the RNA molecules and proteins needful for their life and to complete their tasks. Then there is the S (synthesis) phase during which all the chromatin, so the genetic information present in cells, is completely duplicated. Eventually a G2 (gap 2) phase separates the S phase from the next mitosis step. It is actually a preparation phase for the cell before it undergoes mitosis. In general it does not last for long.

After the G2 step cells eventually enter in the M (mitotic) phase. Chromosomes, which have been previously duplicated, and all the other materials indispensable for cell's life are equally divided into the two nascent daughter cells. The mother cell divide itself through a process named *cytokinesis*.

The DNA replication process

DNA replication process is a critical step in cell's life because all the information contained in the chromosomes has to be correctly transferred to the daughter cells. Indeed cells have evolved a number of strictly controlled mechanisms to ensure a *high-fidelity* and complete replication of the genome and to avoid re-replication of the same DNA sequences. Here below I am going to focus on the eukaryotic DNA replication machinery.

DNA replication only starts at very specific points on the genome named *replication origins*. Although in humans no recognizable sequences are known to be usually present in the replication origins so far, factors involved in the recognition of origins are known. Starting from the late M phase and during the G1 phase a complex of 6 proteins, named ORC (*o*rigin *r*ecognition *c*omplex), is able to bind replication origins on chromatin. This complex shows ATP-ase activity and it is indispensable for the loading on chromatin of all the subsequent replicative factors [1-3]. Indeed the ORC complex directly binds the DNA and it is bound itself by two other factors: Cdc6 (*C*ell *d*ivision *c*ycle protein 6) and Cdt1 (*C*dc10-*d*ependent *t*ranscript 1). Cdc6, as the ORC complex subunits, is an ATP-binding protein belonging to the AAA^+ super-family (*A*TPases *a*ssociated with a variety of cellular *a*ctivities). Cdc6 and Cdt1 are directly involved in the chromatin loading of the hetero-hexameric MCM2-7 complex, formed

by the six paralogous *M*ini-*C*hromosome *M*aintenance proteins (MCM 2,3,4,5,6 and 7) [2,4] (Fig. 2).

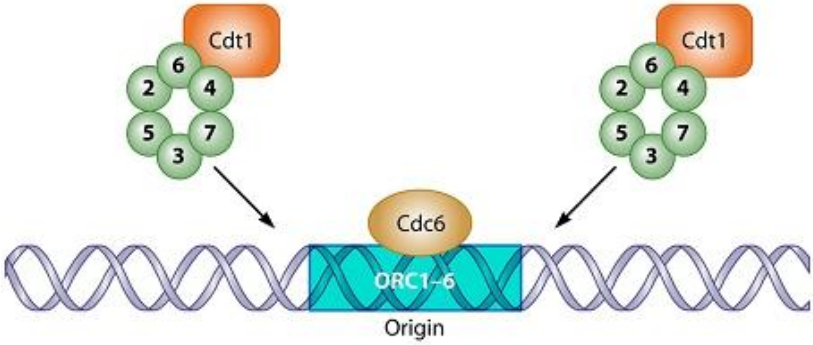

Figure 2 – Pre-RC formation. During the G1 phase Cdc6 and Cdt1 recruit the MCM2-7 complex onto chromatin at a replication origin which has been recognized and bound by the ORC complex.

This complex represents the molecular motor for the DNA unwinding activity but in this phase it remains in an inactive state. The super-complex, which has just formed at this step and comprises ORC, Cdc6, Cdt1 and the MCM2-7 complex, is named pre-*R*eplicative *C*omplex (pre-RC) [2,4]. The peculiar mechanism by which Cdc6 and Cdt1 bring MCM2-7 onto chromatin is called *origin licensing*, because now that specific replication origin is authorized to go on with the next steps of the replication process [5].

The next step consists in the *origin firing* (i.e. the real activation of the replicative fork). At the G1/S border of the cell cycle the pre-RC is converted to a pre-IC (pre-*I*nitiation *C*omplex) by the adding of several factors to the previously formed pre-RC. The MCM2-7 complex will be activated through the association with two other important factors. The first one is the hetero-tetrameric GINS complex composed by 4 different proteins: Psf1, Psf2, Psf3 (*P*artner of *S*ld *f*ive) and Sld5 (*S*ynthetically *l*ethal with *D*pb11). The second one is the Cdc45 (*C*ell *d*ivision *c*ycle 45) protein. Eventually a CMG complex is formed, composed by the MCM2-7 complex, the GINS complex and Cdc45 and it comes to be the real replicative DNA helicase. Other factors are thought to intervene in the formation of the pre-IC and a full map has not yet been drawn [2,6]. What is clear is that the formation of the pre-IC requires the action of two serine/threonine kinases: the *C*yclin-*D*ependent *K*inase 2 (CDK2) and the *D*bf4 (*D*umb*b*ell *f*ormer 4)-*D*ependent *K*inase (DDK) (Fig. 3) [7-9].

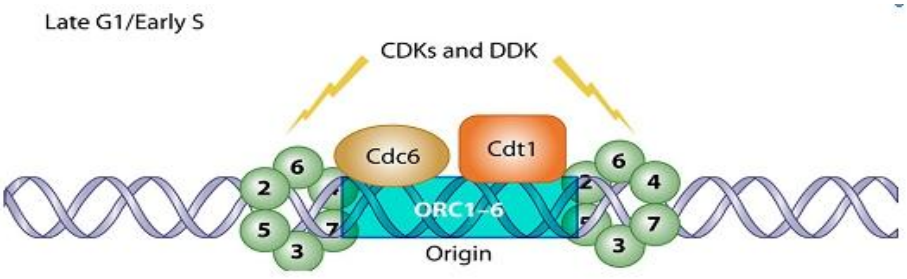

Figure 3 – Pre-IC formation. At the G1/S border the pre-RC complex undergoes phosphorylation by CDKs and DDK. This allows the association of GINS and Cdc45 bringing to the formation of the pre-IC.

In particular the action of the CDKs is fundamental to dictate the timing of these mechanisms. Indeed CDKs are functional only when associated with their activating counterparts known as *cyclins*. Cyclins are proteins whose level in the cell is tightly regulated depending on the phase of the cell cycle. The combination of a specific cyclin with a CDK precisely determines the phase of the cycle [10]. Hence the activity of CDK2 and DDK is indispensable for the loading of Cdc45 and the GINS complex onto chromatin even if the complete map of the factors targeted by these protein kinases is not known. Once Cdc45 and the GINS complex are associated with the MCM2-7 complex at the replicative fork a stable and active CMG complex is formed which can start unwinding the DNA helix [11,12]. After this step DNA polymerases are loaded onto chromatin and they can start duplicate the two strands of the DNA. More precisely, three different DNA polymerases (DNA pols) are involved in the elongation step. The first one is the DNA pol α-primase composed by 4 different subunits. It is able to produce short RNA primers subsequently elongated with short RNA-DNA chains known as *Okazaki fragments* [13]. These fragments are indispensable for the activity of the other two DNA polymerases involved in the elongation phase: DNA pol ε and DNA pol δ. Indeed these two latter DNA polymerases are not able to start DNA polymerization without a primer already present on the template. During elongation of the nascent DNA, DNA pol ε will specifically work duplicating the *leading strand* of the DNA while DNA pol δ is involved in the elongation of the *lagging strand* [14] (Fig. 4).

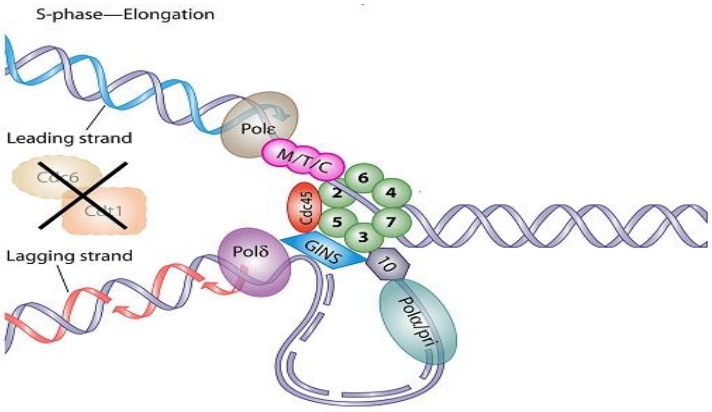

Figure 4 – The elongation phase. When Cdc45 and GINS associate with MCM2-7 to form an active CMG complex, DNA is unwound and DNA replication can start by the action of three different DNA polymerases.

The MCM2-7 complex

The genes codifying for the 6 MCM proteins (MCM2 to MCM7) were first discovered in budding yeast during extensive genetic screenings by a group of researchers looking for genes involved in plasmids stability. They are conserved in all eukaryotes [15,16]. All six proteins are homologous and belong to the large family of AAA^+ ATPases, as several ORC subunits and Cdc6 [17,18]. These proteins contain a region of homology which is called the MCM box [17,19] (Fig. 5).

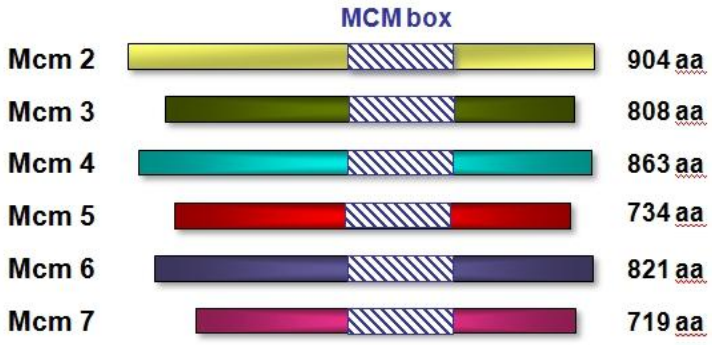

Figure 5 – MCM proteins. Schematic alignment of the six paralogous human MCM proteins from 2 to 7.

Walker A and *Walker B* ATPase motifs together with an *Arginine finger* motif are located in this conserved region.

The *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* MCM2 to MCM7 proteins are nuclear during G1 and S phases and are actively transported from the cell nucleus to the cytosol during the G2 and M phases. In contrast, in higher eukaryotes MCM2 to MCM7 proteins are constitutively in the nucleus and their association with chromatin is regulated during the cell cycle diminishing with progression of these cells through S phase [17].

The eukaryotic MCM proteins form a hetero-hexameric complex (MCM2,3,4,5,6,7), which is also named MCM complex [17,18]. In addition they can associate with each other in different ways forming various sub-complexes such as those containing the proteins MCM2,4,6,7, MCM3-5 and MCM4,6,7 [17,18,20]. The activities of the MCM proteins are essential for the initiation and elongation steps of eukaryiotic DNA replication, but the molecular mechanism of their functions in DNA replication is still unsolved [21-23]. The recombinant MCM2-7 complexes from *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Homo sapiens*, reconstituted *in vitro*, show intrinsic ATPase activity [24,12,25]. ATPase hydrolysis is a requirement for helicase activity. Indeed all helicases show DNA-dependent ATPase activity. Although the MCM2-7 complex exhibits several features expected for a replicative DNA helicase only the sub-complex containing MCM4, MCM6 and MCM7 has been shown to contain a weak DNA helicase activity *in vitro* [17,26]. The helicase activity of these three proteins, which also form a hetero-hexamer (two trimers), is stimulated by DNA molecules with a fork-like structure. These latter molecules are also the preferred substrates for strand displacement activity [27]. However, genetic, cell biological and biochemical data in various model systems indicate that all six subunits of the MCM2-7 complex are necessary for the initiation and elongation phases of DNA replication [28-31]. Despite the progress of investigations on the MCM proteins and their enzymatic activities, these studies have been complicated by the abundance of the MCM proteins both in free and chromatin-bound state. In the cell-free *Xenopus* replication system 20 to 40 MCM2-7 complexes *per* each ORC complex are loaded onto chromatin although only two MCM complexes are sufficient to give full replication activity [32]. A similar situation can be found in yeast where more than 10 MCM complexes seem to be associated with each replication origin at the beginning of the S phase [17]. Despite all this, the stoichiometry of Cdc45, which is actually involved in the activation of the replicative helicase, seems to be of about two

molecules per each ORC complex in both these systems [17] suggesting that Cdc45 works as a molecular trigger for the activation of the MCM2-7 complex. These evidences are consistent with the model according to which MCM2-7 together with Cdc45 and GINS forms the real replicative DNA helicase [31-33]. This model is supported by the finding that in *Drosophila melanogaster* a complex called CMG consisting of Cdc45, MCM2-7 and GINS possesses DNA unwinding activity. This suggests that GINS and Cdc45 would regulate the helicase activity of the MCM proteins *in vivo* [34].

More recently, the *D. melanogaster* CMG complex, and the human one as well, have been reconstituted in recombinant form using the *baculovirus*-insect cells system. These complexes were found to possess DNA helicase activity and increased ATPase and DNA binding affinity if compared with the Mcm2-7 complex alone [12,25].

The Cdc45 DNA replication factor

The Cdc45 gene was first identified by Botstein's group with a genetic screening for cold-sensitive cell cycle mutants in budding yeast together with Cdc46 and Cdc47 (synonyms for MCM5 and MCM7) [35]. The replication factor Cdc45 is essential both for the initiation and the elongation phase of eukaryotic DNA replication. It associates with chromatin in the G1 phase of the cell cycle when the pre-IC complex is established [2,20,23,29]. The association of Cdc45 with replication origins coincides with origin activation (i.e. origin firing) [36]. It was recently shown in yeast and *Drosophila* that Cdc45 forms a multi-subunit complex with MCM2-7 and GINS (i.e. CMG complex), as also found in various other eukaryotic organisms including fission yeast, *Xenopus laevis*, *Mus muris* and *Homo sapiens* [31,33,34,37,38].

Cdc45 levels are stable throughout the cell cycle in human cells even if the mRNA codifying for the protein peaks at the G1/S border. On the basis of western blot analyses after cellular fractionation and immuno-fluorescence-based localization studies it has been observed that Cdc45 is present in both the cytosolic and nuclear fractions [42]. A significant fraction of the protein starts to be nuclear-localized at the G1/S border with a sort of diffuse staining into the nucleus. During the S-phase a different pattern can be observed where Cdc45 seems to associate with replication *foci* (i.e. active DNA replication sites) co-localizing with other factors known to be present at these sites [37,42,58].

The physical and functional interaction of Cdc45 and the MCM2-7 complex is highly conserved in eukaryotes but interactions of Cdc45 with other important replication factors have not yet been clearly established. The interaction between Cdc45 and RPA, the most abundant single strand binding protein involved in the stabilization of the single-stranded DNA during replication, has been detected in a variety of organisms from lower to higher eukaryotes. The same is true for the interaction with DNA Pol ε, whereas the physical interaction between Cdc45 and DNA polα-primase is still under discussion. Cdc45 from fission yeast and *X. laevis* has been shown to physically interact with DNA pol α*,* while these factors do not associate in budding yeast, *D. melanogaster* and *H. sapiens* [37,39,40]. Moreover, in several organisms an interaction has been shown between Cdc45 and DNA pol δ, MCM10, TopBP1,

Claspin (Mrc1 in budding yeast) and other factors involved in the S-phase checkpoint / DNA damage repair pathways [6,37,41,42,43].

The interaction between Cdc45 and DNA pol δ can be detected later in S-phase respect to the one with DNA pol ε and the extended association of Cdc45 and DNA pol δ during the cell cycle could suggest an involvement of these proteins in the maturation of Okazaki fragments or in post-replicative DNA repair reactions [37].

MCM10 was first discovered in the same genetic screening by which all the other MCM proteins were identified [71]. Despite this, there is no structural and functional similarity between MCM10 and all the other MCM proteins. Comparing the function of this protein with Cdc45, studies in yeast suggest that Cdc45 and MCM10 are involved in the same cellular pathway and both proteins interact with the same regions of MCM5 and MCM7 [6]. MCM10 seems to be involved in the loading of Cdc45 onto DNA replication origins and in the recruitment of DNA pol α at the same sites [44,45].

TopBP1 (DNA *Top*oisomerase II *b*inding *p*rotein 1) is a fundamental factor for several aspects of cellular life, such as DNA replication and DNA damage checkpoint activation. Thus it interacts with Cdc45 intervening somehow in the first steps of DNA replication process.

Claspin is a protein known to work as a mediator of the S-phase checkpoint. Eukaryotic cells have evolved very complex control systems to stably maintain the integrity of their genomes. These control mechanisms stop the cell cycle, if a DNA damage has occurred, until the lesion will be repaired. These DNA damage-dependent signal transduction pathways are called *checkpoints* [46,47]. Importantly they seem to be conserved from yeast to humans, where they may have a role in the control of cancer development [48,49]. Depending on the type of DNA damage, different checkpoint pathways become activated in the cell: DNA double-strand breaks formation mainly triggers the ATM-Chk2 pathway, whereas ultraviolet irradiation and stalled replication forks primarily promote the ATR-Chk1 pathway [52,53]. ATM (*A*taxia *t*elangiectasia *m*utated) and ATR (*A*taxia-*t*elangiectasia and *R*ad3-related protein) are members of the phosphoinositides kinase-related family of protein-kinases [54]. ATR and ATM are responsible for phosphorylation-mediated activation of the downstream kinases Chk1 and Chk2, respectively [55-57]. Recently it has been shown that Cdc45 is involved in the regulation of DNA replication after the activation of the S-phase checkpoint due to DNA damage [6,41,50,51]. It has been shown by means of experiments carried out in the *X. leavis* egg extracts experimental system that after DNA damage, replication is prevented by the inhibition of Cdc45 loading onto chromatin [50]. This control mechanism seems to be conserved also in humans. Indeed it was shown that in *H. sapiens* there is a DNA damage-dependent regulation of Cdc45 loading onto chromatin and this occurs in a Chk1 (*ch*eckpoint *k*inase 1)-dependent manner [51]. Chk1 is one of the factors involved in the final steps of the S-phase checkpoint activation (i.e. checkpoint effectors) [52].

Not much is known about the biochemistry and the physiology of Cdc45 so far, despite its fundamental importance in various DNA transactions.

The GINS complex

The GINS complex is an hetero-tetrameric complex composed by four proteins: Sld5, Psf1, Psf2 and Psf3 [61]. They are conserved in all eukaryotic organisms and the complex they form has been named after the acronym of the Japanese words "*G*o-*I*chi-*N*i-*S*an" which mean "five-one-two-three" representing the final numbers of the names of the four proteins.

The genes codifying for these factors were first identified in *S. cerevisiae* by means of several genetic screenings. Sld5 (*S*ynthetically *l*ethal with *D*pb11) was discovered as a novel interactor of Dpb11, an initiation factor responsible for the recruitment of DNA pol ε at the replication fork through the interaction with Dpb2, one of the subunits of this polymerase. By means of two-hybrid screenings Psf1, Psf2 and Psf3 were then identified as *P*artners of *S*ld *F*ive [59]. In the same period Kubota and co-workers identified the homologous complex in *X. laevis*. These authors proposed that the GINS complex may have a role in DNA replication, because a reduction of the dNTPs incorporation could be observed in DNA replication after the immuno-depletion of the GINS complex from the extract used for the *in vitro* assays [60]. Either in *S. cerevisiae* or in *X. laevis* the GINS complex was shown to be associated with the chromatin and to interact with Cdc45 and the MCM2-7 complex [59,60].

The crystallographic structure of the complex has been solved by three different groups, showing a trapezoidal shape with Sld5 and Psf1 associated in the top layer and Psf2 and Psf3 associated at the bottom. This confirmed the subunit arrangement proposed on the basis of two-hybrid screenings and mapping with monoclonal antibodies [59,62]. The structure retains a central hole which is actually funnel-shaped because its diameter is about 70Å on a side and about 25Å on the other side [62-65]. The human GINS complex has been shown to preferentially bind single-stranded DNA (ssDNA) [62].

Several studies revealed that the GINS complex is indispensable not only for the initiation but also for the elongation steps of the DNA replication process. Indeed GINS has been shown to be part of the so called *R*eplisome *P*rogression *C*omplex (RPC) together with MCM2-7, Cdc45 and other factors associated to the replication fork during S phase [30,31]. The interaction between GINS and MCM2-7 has also been demonstrated in *S. cerevisiae* and these two complexes have been shown to be part of the RPC together with Cdc45 by mass spectrometry analysis [33].

Moreover, a complex containing Cdc45, GINS and MCM2-7 was isolated from *D. melanogaster* embryos and more recentely, as I have stated above, the *D. Melanogaster* CMG complex, and the human one as well, have been reconstituted in recombinant form using the *baculovirus*-insect cells system [12,25,34].

The CMG complex

DNA unwinding is a fundamental activity of the DNA replication machinery. DNA helix has to be opened to let DNA polymerases work. In eukaryotic cells the CMG complex has been definitively identified as the real replicative DNA helicase. Within this multi-subunit complex, the six MCM proteins form a hexameric ring-structure that represents the molecular motor of the DNA helicase. Indeed all the MCM proteins

retain ATPase activity that is needed for the DNA unwinding function [17,18]. Thus Cdc45 and the GINS complex are believed to be auxiliary factors needed to activate the helicase motor [12,25].

The CMG complex was first purified from *Drosophila melanogaster* embryo extracts. Moyer and co-workers carried out an immuno-affinity pull down of Cdc45 from the extracts and they were able to detect in the pulled-down fraction 10 Cdc45-interacting proteins, which were analysed by mass spectrometry and were identified to be the 6 MCM proteins (i.e. the MCM2-7 complex), plus the 4 proteins composing the GINS compex (Sld5, Psf1, Psf2, Psf3). Moreover they demonstrated that the complex purified from these extracts possesses DNA helicase activity *in vitro* [11]. Some years later Ilves and co-workers were able to produce a stable recombinant *D. melanogaster* CMG complex by co-infecting insect cells with 11 baculoviruses codifying for the 11 single subunits of the complex. They purified the recombinant CMG complex and compared its enzymatic properties with those of the MCM2-7 complex produced in insect cells [12]. The ATP hydrolysis rate by the MCM2-7 complex was shown to be enhanced by 300 folds when the latter was associated with Cdc45 and GINS. Since neither Cdc45 nor the GINS proteins are ATPases, these Authors proposed that the increased ATPase activity could be due to an allosteric remodelling of the MCM2-7 complex after the association with these auxiliary factors. Moreover the CMG complex was shown to have an improved DNA binding activity in comparison with the MCM2-7 complex and interestingly its DNA binding activity was found to be ATP-dependent, while the one of the MCM2-7 complex was not. Most important, a robust *in vitro* DNA helicase activity was shown to be associated with the CMG complex, while the MCM2-7 complex alone did not show at all this activity [12].

In a subsequent work the *D. melanogaster* CMG complex obtained with the same procedure was used by Costa and co-workers to get structural data by single-particle electron microscopy analysis [66]. According to their model, the MCM2-7 complex would assume two different basic conformations. The first one is a spiral conformation with a gap occurring between the MCM2 and MCM5 subunits. In the other one, MCM2-7 has a planar ring-like structure and this conformation seems to be favourite when Cdc45 and GINS are associated whit it. It is important to notice that Cdc45 and GINS bind the MCM2-7 complex at the interface with MCM2 and MCM5 closing the gap between these two subunits [66] (Fig. 6).

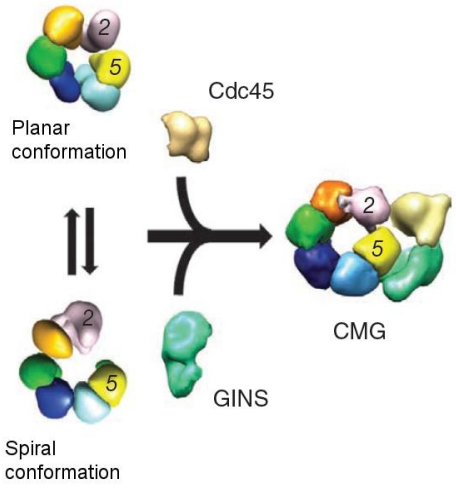

Figure 6 – Model for CMG complex. Schematic representation of the CMG complex according to the model proposed by Costa *et al.* [66].

In 2012 the human CMG complex has been reconstituted in recombinant form by Kang *et al.* [25]. As the *D. melanogaster* complex, the human CMG complex has been produced in recombinant form in insect cells by co-infection with 11 baculoviruses codifying for each subunit of the complex. As the *D. melanogaster* counterpart the human CMG complex was found to possess increased ATPase and DNA helicase activity. Moreover, with a very elegant experiment it was demonstrated that the helicase activity of the human CMG complex is coordinated with the synthetic function of DNA pol ε on a small circular single-stranded DNA template, which was replicated several times *in vitro* with a rolling circle mechanism [25].

Summarizing, it is well established so far that the CMG complex is the real eukaryotic replicative DNA helicase: the MCM2-7 complex represents the molecular motor of this enzymatic activity, but it needs to be activated by the association of Cdc45 and GINS; nevertheless only a speculation can be made about the mechanism by which these two co-factors activate the MCM2-7 activity. Additional studies are needed to fully understand the mechanism by which the CMG complex exerts the DNA unwinding activity and the specific role played by Cdc45 and GINS.

The DHH family of phosphoesterases

The DHH protein family is a wide spread group of phosphoesterases, including inorganic pyrophosphatases and RecJ ssDNA exonucleases, which are conserved in all kind of organisms (i.e. eukarya, archaea and bacteria) that was first described by Aravind and Koonin [74]. This protein family is defined by a number of conserved sequence motifs and it is usually split into two distinct clusters.

The first one, identified as subfamily 1, is represented so far only in Archaea and Bacteria. The most representative member of this subfamily is the *Escherichia coli* RecJ protein, a 5’ to 3’ single-stranded DNA exonuclease. Therefore, these proteins are usually referred to as RecJ-like proteins.

The subfamily 2 is also found in Eukarya and it includes the *Drosophila melanogaster* Prune protein and the *Saccharomyces cerevisiae* exopolyphosphatase PPX1 [74]. Crystallographic analysis carried out on several DHH proteins [75–77] have highlighted the presence of a catalytic protein core composed by two main domains: domain I and domain II (also named N-terminal and C-terminal domain). In the N-terminal domain invariant residues (aspartic acid and histidine residues) are conserved and involved in the coordination of two metals ions (typically manganese ions). This would suggest a two-metal ions mediated hydrolysis activity for these proteins.

AIM OF THE THESIS WORK

Dissecting the molecular pathways implied in the DNA replication is the basis for understanding the mechanisms by which cells proliferate. In humans the control of cellular proliferation is a dramatic and fundamental issue for the viability of the organism. If a cell escapes this kind of control it can start growing in an uncontrolled fashion. It means that now this cell is no anymore part of the organism from a physiological point of view and it starts to be lead by altered molecular mechanisms which can lead to tumorigenesis if the cell is still able to divide into two daughter cells.

The Cdc45 replication factor has been shown to be a proliferation-associated antigen [67]. This protein is totally absent in quiescent or terminally differentiated cells while it is actively produced in proliferating cells, such as cancerous cells, during all the phases of the cell cycle [67,68]. Due to its role in the replication forks activation, Cdc45 comes to be one of the fundamental circuit-breakers used by the cell to switch on and off the DNA replication process. Indeed the association of Cdc45 with the pre-RC coincides with *origin firing* [69]. Moreover, after the activation of the DNA-damage dependent checkpoint, Cdc45 is targeted for the dissociation from the replisome and this coincides with the stop of the replication process [50,51]. Thus studying Cdc45 and the molecular mechanisms in which it is involved will provide useful information to understand how DNA replication is switched on or off in normal or perturbed conditions.

Despite the central role of Cdc45 in the DNA replication mechanism not much is known about the biochemistry of this protein. The aim of my thesis work has been to study the Cdc45 protein from a biochemical point of view, both as a single molecule and in association with other proteins in the CMG complex. In fact, for what concerns the single protein, nothing was known about the structure and the putative enzymatic properties of this protein. No characteristic sequence motif was identified in the Cdc45 primary sequence and it was clear that Cdc45 could activate the DNA unwinding activity of the MCM2-7 complex together with GINS, but nothing was known about the specific mechanism. Furthermore the CMG complex had not been yet reconstituted to assay its activity *in vitro*.

By a biotechnological point of view, the biochemical characterization of the Cdc45 replication factor and the study of its role in the cell can provide a basic knowledge that can be exploited to set up novel anti-cancer therapeutics. This would form the basis for a completely new approach for cancer treatment according to a strategy alternative to the traditional ones. Currently, cancer therapy is mainly based on the use of chemotherapeutics, which actually damage the genome of the proliferating cells. A molecular therapy, based on small molecules that could interfere with the activity of Cdc45, would be more powerful and specific. Indeed Cdc45 inactivation would actually stop all cell proliferation and this would happen in a very specific manner due to the fact that the protein is almost absent in differentiated cells and tissues.

EXPERIMENTAL PROCEDURES

Human Cdc45 protein expression and purification

A previously produced plasmid construct was used for the production of the *wild type* recombinant hCdc45 protein with a C-terminal 6xHis-FLAG tag cleavable with tobacco etch virus (TEV) protease (pNIC-CTHF-hCdc45) [70].

Expression of hCdc45 was carried out in *Escherichia coli* BL21 (DE3)-R3-Rosetta cells (SGC, Oxford) grown overnight at 25°C in LB broth, following induction with 0.1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG). A frozen cell pellet (corresponding to 1 litre of culture) was resuspended in 100 ml of buffer A (50mM sodium phosphate buffer, pH 7.4, 1 M NaCl, 10% (v/v) glycerol, 2mM β-mercaptoethanol) containing EDTA-free protease inhibitor mixture (Roche Applied Science), 50 units/ml benzonase nuclease (Novagen), and 1 mg/ml lysozyme (Sigma-Aldrich), and incubated for 2 h at 4 °C. Cells were further disrupted by sonication (Bandelin Sonopuls HD3200 ultrasonic homogenizer).

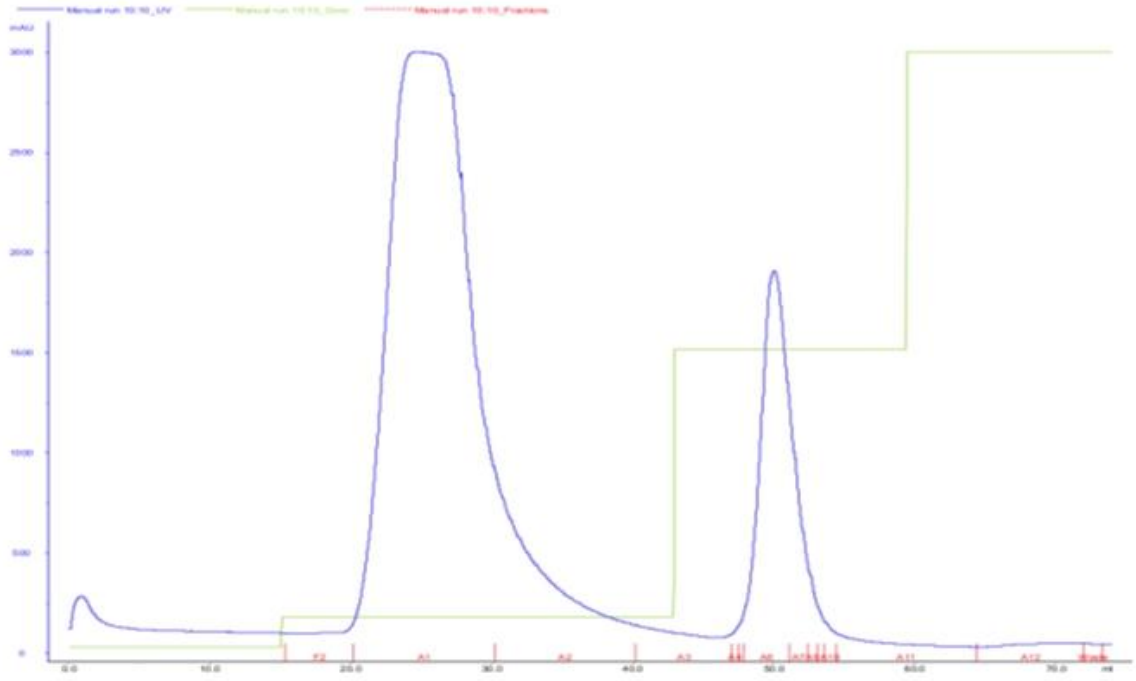

Figure 7 – Nickel affinity chromatography. Chromatogram of the nickel affinity chromatography carried out as first step for the purification of recombinant hCdc45. In *green* the imidazole step-gradient is shown. In the second step of the gradient a sharp peak can be observed that corresponds to hCdc45 protein eluted from the resin.

The insoluble material was removed by centrifugation (17,000 rpm for 1 h at 4 °C, Beckman Allegra 64R), and the supernatant was loaded onto a 5-ml HiTrap chelating column (GE Healthcare) previously charged with Ni^{2+} ($NiSO_4$) and equilibrated with

buffer A. An imidazole step-gradient (25 mM, 250 mM, and 500 mM) was applied, and protein elution was achieved using buffer A plus 250 mM imidazole (Fig. 7).

Fractions containing the hCdc45 protein were pooled and subsequently loaded onto a Superdex 200 16/60 size-exclusion column (GE Healthcare) equilibrated with buffer B (50 mM sodium phosphate, pH 7.0, 150 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol) (Fig. 8).

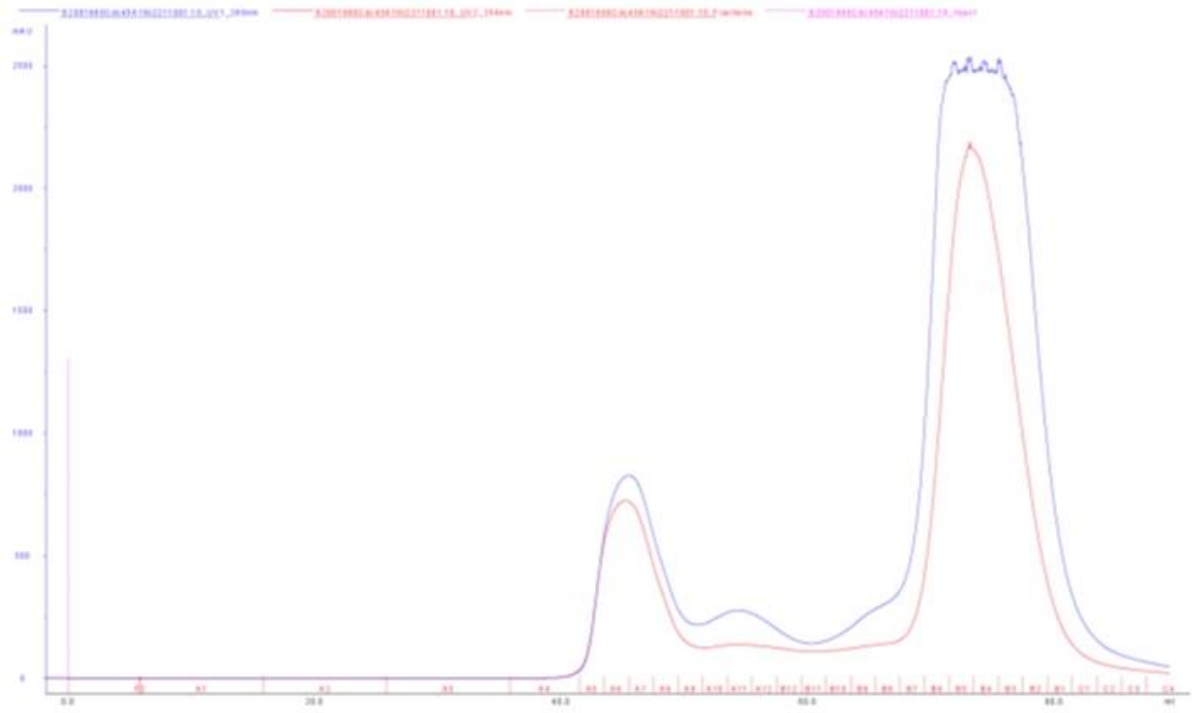

Figure 8 – Size-exclusion chromatography profile for hCdc45. The chromatogram of this purification step is shown. On the right a huge peak can be observed corresponding to hCdc45 exit from the Superdex 200 16/60 column.

The eluted protein was incubated at 4 °C overnight with the *t*obacco *e*tch *v*irus (TEV) protease for tag cleavage. SDS-PAGE (12% polyacrylamide:bis 29:1) was carried out to check that the cleavage was complete. The hCdc45-TEV protease mix was loaded onto a 5-ml Heparin column (GE Healthcare), previously equilibrated with buffer B. The cleaved hCdc45 was eluted from the column applying a linear (150 mM to 1 M) NaCl gradient (Fig. 9). Finally the protein was subjected to size-exclusion chromatography in buffer D (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol) (Fig. 10).

The same procedure was used to express and purify all the other mutants of the hCdc45 protein.

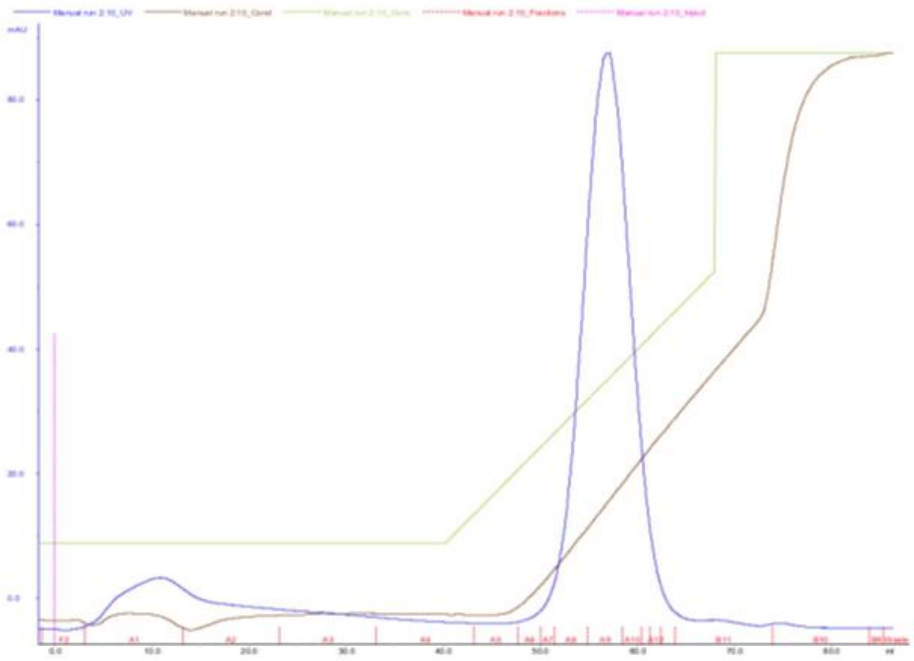

Figure 9 – Heparin chromatography. Chromatogram of the Heparin column purification step: in *blue* the peak corresponding to hCdc45 is shown, while the *green* line depicts the salt gradient applied to the column.

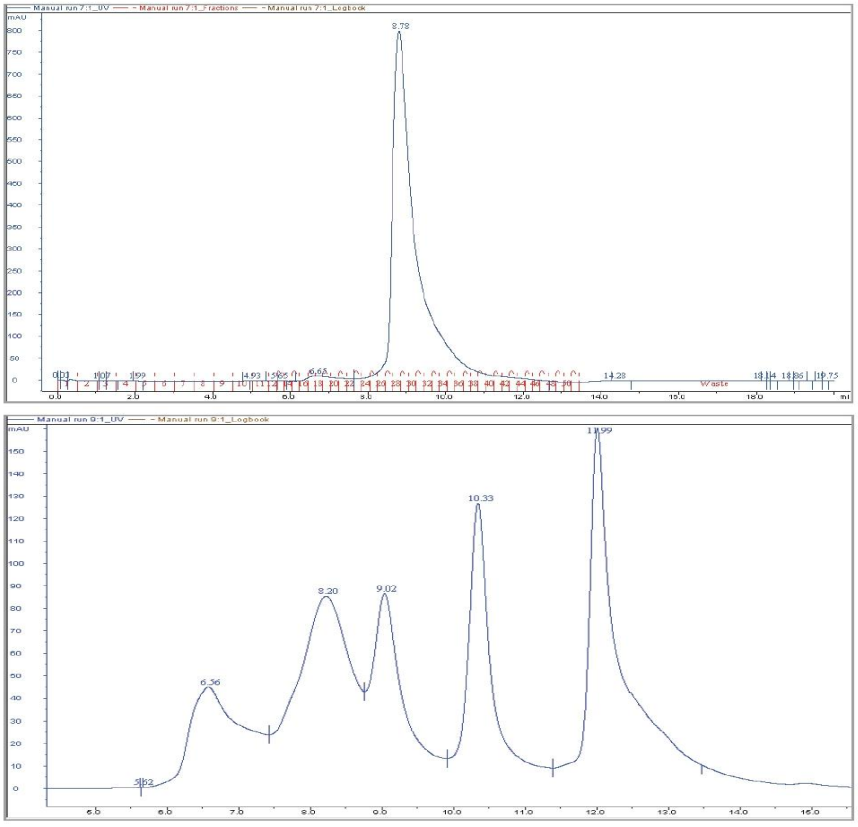

Figure 10 – Size-exclusion chromatography analysis of recombinant hCdc45. The chromatographic profile of hCdc45 on a Bio-Sil SEC-250 (Bio Rad) column is shown (*upper part*). Column calibration (with the following markers: tyroglobulin, bovine γ-globulin, chicken ovalbumin, equin myoglobin, vitamin B12) is shown in the lower part. Cdc45 peak eluted at a volume (8.78 ml) that corresponds to the molecular mass of the protein (68 kDa).

Production of human Cdc45 mutant derivatives

Plasmid constructs for the expression of deletion mutants of the protein were produced by using three sequential PCRs (Fig. 11).

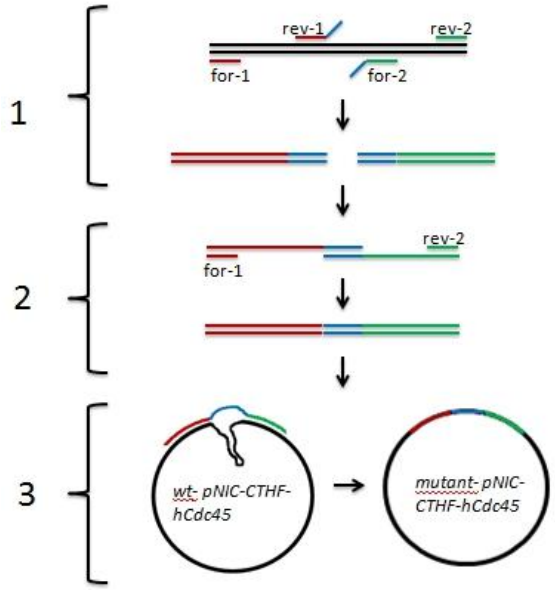

Figure 11 – Schematic representation of the three sequential PCRs used for constructing the Cdc45 deletion mutants.

In the first PCR two separate portions of the Cdc45 CDS (coding sequence) were amplified excluding the region codifying for the portion to be deleted. These sequences were amplified by using four different DNA primers. The reverse primer used for the amplification of the first region of the Cdc45 CDS was partially complementary to the forward primer used for the amplification of the second region. The complementary region shared by these two primers encodes a linker aminoacidic sequence (STGSSGTSGG) that substitutes for the deleted portion of the protein.

In the second PCR the two previously amplified sequences were fused by means of the complementary regions. The forward primer used in the previous PCR for the amplification of the first sequence and the reverse primer used for the amplification of the second sequence were used in this reaction.

The third PCR was used to clone the sequence amplified in the second reaction into the pNIC-CTHF vector by *restriction free* (RF) cloning. In this reaction the pNIC-CTHF-hCdc45 (expressing the *wild-type* hCdc45 protein) was used as the template, while the construct amplified in the second PCR reaction was used as a mega-primer. The amplification product was extensively digested with DpnI restriction enzyme (NEB) to eliminate the *wild-type* pNIC-CTHF-hCdc45 plasmid used as the template. A portion of this mixture was then used to transform *E. coli* DH5α cells (Invitrogen), which were then spread on plates containing kanamycine and incubated over night at 37 °C. Bacterial colonies grown on the plates were screened to identify recombinant clones harbouring the DNA fragment of interest.

By using this method plasmid constructs codifying for three mutants derivatives of the protein were cloned:

- hCdc45_A1_LP_1 (a. a. from 138 to 182 substituted with STGSSGTSGG)
- hCdc45-L1(a. a. from 113 to 181 substituted with STGSSGTSGG)
- hCdc45-L2 (a. a. from 254 to 351 substituted with STGSSGTSGG).

Plasmid constructs for the expression of truncated forms of Cdc45 were produced by using two sequential PCRs (Fig. 12).

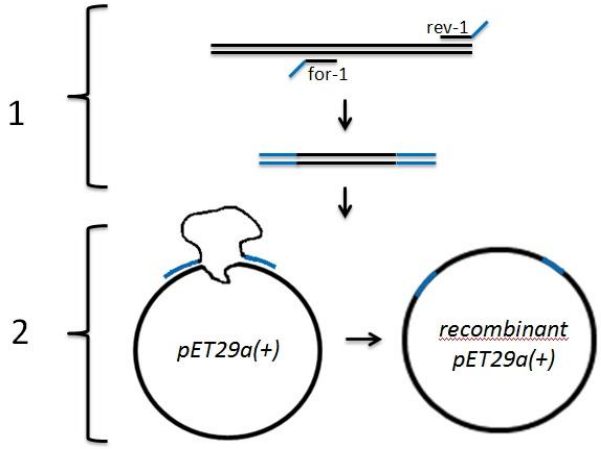

Figure 12 – Schematic representation of the two PCR steps used to construct plasmids expressing truncated forms of Cdc45.

In the first PCR the region of interest of the Cdc45 CDS was amplified by using two different primers. Half-part of each primer was complementary to the Cdc45 CDS while the other part was complementary to the pET29a(+) (Novagen) sequence present in the chosen insertion site. In this first PCR the pNIC-CTHF-hCdc45 plasmid was used as the template. In the second PCR the amplification product from the previous reaction was used as a mega-primer and the empty pET29a(+) was used as template to get a *restriction free* cloning of the sequence of interest into this vector. The amplification product was extensively digested with the DpnI restriction enzyme (NEB) to eliminate the empty pET29a(+) vector used as the template. A portion of this mixture was then used to transform *E. coli* DH5α cells (Invitrogen), which were then spread on plates containing kanamycine and incubated over night at 37 °C. *E. coli* colonies grown on the plates were screened to identify recombinant clones.

By using this method plasmid constructs codifying for two different truncated forms of the protein were cloned:

- hCdc45-A13/1 (a. a. 1-258)
- hCdc45-C-ter (a. a. 394-566)

Cloning, expression and purification of the recombinant *Xenopus* Cdc45 protein

An oligonucleotide sequence codifying for the *Xenopus* Cdc45 protein with 6xHis and 3xFLAG tags at the C-terminus had been previously cloned into a pCR2.1-TOPO vector (Invitrogen) in our laboratory.

I was able to sub-clone this sequence into a pET29a(+) bacterial expression vector (Novagen) by cutting both the pCR2.1-TOPO-xCdc45 construct and the empty pET29a(+) vector with BamHI and NdeI restriction enzymes (Roche). The digestion was carried out in buffer B (Roche). DNA fragments of interest were purified by gel extraction and employed for a ligation reaction using the T4 DNA ligase (Roche).

Half of the ligase reaction mixture was then used to transform *E. coli* DH5α cells (Invitrogen) then spread on kanamycine plates and incubated over night at 37 °C. *E. coli* colonies were screened to get positive clones.

Expression of the xCdc45-His-FLAG protein was carried out in (DE3)-R3-Rosetta cells (SGC, Oxford). Bacteria were grown up to an optical density of 2 and then the expression of the protein was induced by adding 0.1 mM IPTG to the cell medium. After the induction, cells were kept over night at 18 °C on shaking, then harvested by centrifugation for 15 minutes at 6,000 rpm in a F12-6x500Y rotor (Sorvall – RC 6+ centrifuge).

Cell pellet from 3 liters of culture was re-suspended in 50 ml of Lysis Buffer (50 mM Tris/HCl pH 7.5, 200 mM NaCl, 10% (v/v) glycerol, 2 mM β-mercaptoethanol) supplemented with *EDTA-free protease inhibitor* cocktail (Roche). Cells were disrupted by 2 cycles of French Press and the crude extract obtained was treated with 50 units/ml Benzonase nuclease (Novagen) for 1h at 4°C on shaking. Whole cell extract was then clarified by centrifugation at 40,000 rpm for 40 minutes in a LE-80 BECKMAN ultracentrifuge.

The soluble extract obtained was incubated with 3 ml of *Ni-NTA agarose* resin (Qiagen) for 1 h at 4 °C on shaking. This mixture was then loaded on an affinity chromatography column and the unbound material let flow through. Resin was washed with 200 ml of Lysis Buffer and then 5 elution steps were carried out respectively with:

- 10 ml of Lysis Buffer plus 10 mM imidazole
- 10 ml of Lysis Buffer plus 25 mM imidazole
- 10 ml of Lysis Buffer plus 50 mM imidazole
- 10 ml of Lysis Buffer plus 100 mM imidazole
- 10 ml of Lysis Buffer plus 250 mM imidazole

The 50 mM imidazole elution fraction, which contained the peak of the protein eluted from the resin, was loaded on a MonoQ HR 16/10 column (GE Healthcare) previously equilibrated with Buffer A (25 mM Tris/HCl pH 7.5, 200 mM NaCl, 10% (v/v) glycerol, 2 mM β-mercaptoethanol). The elution was carried out with a linear salt gradient from 200 mM to 1M NaCl. 1-ml fractions were collected and analysed by SDS-PAGE to check the presence of the protein of interest.

Fractions containing xCdc45-His-FLAG were pooled and loaded on a Superdex 200 16/60 column (GE Healthcare) for a size-exclusion chromatography step. The column had been previously equilibrated in Buffer D (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol). After SDS-PAGE analysis, fractions containing the protein of interest were pooled and concentrated with a *Vivaspin 2* device (30k MWCO – GE Healthcare) to reach a final concentration of 1 mg/ml for xCdc45-His-FLAG protein. The result from the purification protocol is shown in Fig. 13.

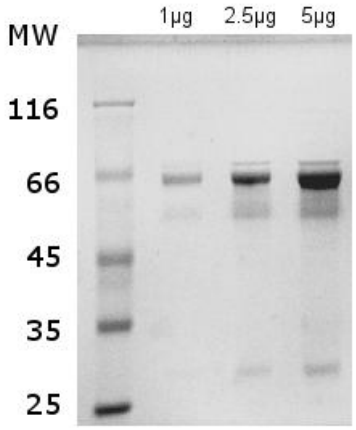

Figure 13 – Purification of xCdc45-His-FLAG. The indicated amounts of the purified recombinant protein are shown by Coomassie staining after SDS-PAGE.

***E*lectrophoretic *m*obility *s*hift *a*ssays (EMSAs)**

The synthetic oligonucleotide used as single-stranded DNA had the following sequence: 5'-TCTACCTGGACGACCGGGTATATAGGGCCCTATATATAGGGCCAG CAGGTCCATCA-3'. A complementary synthetic oligonucleotide used to prepare the blunt DNA duplex had the following sequence: 5'-TGATGGACCTGCTGGCCCTATAT ATAGGGCCCTATATACCCGGTCGTCCAGGTAGA-3'. The first oligonucleotide was labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP and annealed to a 2-fold molar excess of the cold complementary strand to prepare the double-stranded DNA ligand. For the DNA mobility shift assays, 10-µl mixtures were prepared that contained 100 fmol of ^{32}P-labeled DNA in 20 mM Tris-HCl, pH 7.5, and the indicated amounts of protein (0.5 – 5 µg). Following incubation for 20 min at 27 °C, complexes were separated by electrophoresis through 5% polyacrylamide/bis gels (19:1) in 0.5X TBE (1X TBE: 89 mM Tris Base, 89 mM Boric Acid, 2 mM EDTA, pH 8.3). Gels were dried down and analyzed by phosphorimaging. Experiments were performed in triplicate, and the results were averaged. The error bars on the graphs are the standard error. For the super-shift analysis assays were carried out after incubation of hCdc45-His-FLAG-DNA complexes with anti-FLAG monoclonal antibodies (Abcam). For these experiments 10-µl mixtures were prepared that contained 50 fmol of ^{32}P-labeled oligonucleotide in 20mM Tris-HCl, pH 7.5, and 5 µg of hCdc45-His-FLAG. Following incubation for 20 min at 27 °C, anti-FLAG antibody was added (0.5, 1, and 2 µg in 2 µl of the following buffer: 10 mM sodium phosphate, pH 7.4, 150mM NaCl, 50% (v/v) glycerol; an equal volume of buffer (2 µl) was added into the samples where antibody was omitted, and the incubation was continued for additional 30 min.

The mixtures were subjected to electrophoresis, as previously described, and the gels were analyzed by phosphorimaging.

EMSAs on a radio-labeled 20-mer synthetic single-stranded RNA (100 fmoles/assay) were carried out using the following single-stranded RNA sequence: 5'-GAUCGGGAGAUCUCAGACCA-3'. The protocol used for the assay is the same used for single-stranded DNA.

Analytical gel filtration and DNA-binding activity of hCdc45

An aliquot of purified hCdc45 (1.9 mg) was loaded onto an analytical gel-filtration column (Bio-Sil SEC-250, Bio-Rad). The column was developed with 50mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM β-mercaptoethanol, 5% (v/v) glycerol at a flow rate of 1.0 ml/min. Fractions (200 μl) were collected, and aliquots (5 μl) were run through an 8% (w/v) polyacrylamide-SDS gel. The gel-filtration column was calibrated using the following markers: tyroglobulin, bovine γ-globulin, chicken ovalbumin, equin myoglobin, and vitamin B-12. The DNA binding activity of each peak fraction (aliquots of 1 μl) was analyzed by EMSAs using radio-labelled single-stranded DNA as a probe, as previously described.

Pyrophosphatase activity assays

Pyrophosphatase (PPase) activity was determined following the method described by Park *et al.* [92].

PPase activity was assessed for the hCdc45 protein in presence of different metal ions and at different pH conditions. Increasing amounts of recombinant hCdc45 protein (from 100 ng to 5 μg) were added to reaction mixtures (volume: 300 μl) having the following composition: 50 mM Tris–HCl, pH 6.5 (or pH 7.5, or pH 8.5), 5 mM $MgCl_2$ (or $MnCl_2$, or $ZnCl_2$, or $CoCl_2$), 1 mM inorganic pyrophosphate. The reaction was started by the addition of the recombinant protein and kept at 37 °C for 15 min. Inorganic phosphate present in the mixture was detected by adding 700 μl of *phosphate detection reagent* (1.5% ascorbic acid (w/v), 1% ammonium molybdate, 0.83 M Sulfuric Acid). After inorganic phosphate production a deep blue colour is developed due to the presence of a phosphomolybdate complex and its absorbance can be measured at 820 nm. The amount of inorganic phosphate can then be quantified using a phosphate standard curve obtained from serial dilutions of a NaH_2PO_4 standard solution.

DNA exonuclease activity assays

Assays were carried out to check for 5' to 3' exonuclease activity. hCdc45 recombinant protein was incubated in presence of an oligonucleotide (having the following sequence 5'-GCCGTGATCACCAATGCAGATTGACGAACCTTTGCCCAC GT-3') that was labelled with a Cy3 fluorescent molecule (PRIMM). Exonuclease activity was assessed for the hCdc45 protein in presence of different metal ions and at different pH conditions. Different amounts of recombinant hCdc45 protein (from

100 ng to 1 μg) were added to reaction mixtures (volume: 20 μl) with the following composition: 50 mM Tris–HCl, pH 6.5 (or pH 7.5, or pH 8.5), 5 mM $MgCl_2$ (or $MnCl_2$, or $ZnCl_2$, or $CoCl_2$), 150 mM NaCl, 20 nM DNA substrate. The reaction was started by the addition of the recombinant protein and kept at 37 °C for 30 min. Reactions were terminated by the addition of 5X Stop Solution (0.5% SDS, 40 mM EDTA, 0.5 mg/ml proteinase K, 20% glycerol). The eventual degradation of the substrate was assessed by loading the reaction mix on a 8% polyacrylamide gel containing 0.1% SDS that run in 0.5X TBE buffer. Fluorescence was detected by using a VersaDoc instrument (Bio Rad).

Production of recombinant baculoviruses

The pFastBac derivatives harbouring the following human CDSs were provided by Dr Lori Frappier (University of Toronto, Toronto, Canada): MCM2 (6xHis tag), MCM2 (STREP tag), MCM3, MCM4, MCM5, MCM6, MCM6 (HA+6xHis tag), MCM7 (FLAG+6xHis tag).

A baculovirus expressing the human Cdc45 protein fused to a 6xHis tag had been previously produced in the laboratory.

A baculovirus expressing the MCM3 protein fused to a STREP tag and four baculoviruses expressing the four single subunits of the GINS complex (Psf1, Psf2, Psf3 and Sld5) were produced using the *Gateway® System* with the *pENTR™ Directional TOPO® Cloning* kit (Invitrogen). The CDSs of interest were amplified by using two different primers. The forward primer for each reaction was designed to start with a specific sequence (CACC) that is needed for the next topoisomerase reaction step. The forward primer used for cloning the MCM3 CDS contained the sequence codifying for the STREP tag (5'-ATGGCTAGCTGGAGTCATCCT CAATTCGAAAAA-3'). The amplified sequences were cloned into linearised pENTR vector by means of the *Directional TOPO® Cloning* system (Invitrogen). An aliquot of the *TOPO® Cloning* reaction was then used to transform *One Shot® Competent E. coli* cells (Invitrogen), which were then spread on plates containing kanamycine and incubated over night at 37 °C. *E. coli* colonies grown on plates were screened to identify recombinant clones. The recombinant pENTR plasmids were purified from *E. coli* cells and used in the subsequent *Gateway® LR Recombination Reaction* (Invitrogen). In this reaction the cloned sequences are transferred by the *LR Clonase™ II Enzyme* (Invitrogen) into the baculovirus linear DNA exploiting the *attL* and *attR* sites present in both the pENTR vector and in the baculovirus linear DNA. An aliquot of this reaction was then used to transfect *Sf9* cells (Novagen) by using *Cellfectin® II Reagent* (Invitrogen) and *Grace's Insect Medium Unsupplemented* (GIBCO). After the transfection step, *Sf9* cells were grown in *Sf-900™ II SFM* medium (GIBCO) supplemented with 10% FBS (*Fetal Bovine Serum*; Euroclone) and gentamicin (50 μg/ml; GIBCO). The viral stocks obtained from the transfection reactions were used for other two rounds of *Sf9* cells infection to get high-titre viral stocks. These latters viral stocks were subjected to the *plaque assay* according to Invitrogen manual (BaculoDirect™ Baculovirus Expression System) to get a single viral population expressing the proteins of interest. Eventually the protein expression was assayed for each clone, after 72-h infection of *Sf9* cells by means of western blot analysis.

Production of multi-gene baculoviruses with the *MultiBac^{Turbo}* system

Single baculoviruses co-expressing in insect cells more than one protein were produced using the *MultiBac^{Turbo}* system. More precisely, one baculovirus expressing FLAG-tagged MCM3, MCM5 and MCM7 was produced (named Multi-MCM357). A second baculovirus was produced for the simultaneous expression of MCM2, MCM4 and MCM6 (named Multi-MCM246). A third baculovirus was engineered for the co-expression of Cdc45, Psf1, Psf2, Psf3 and Sld5, this latter fused to a GST-tag (named Multi-GINS/Cdc45).

First of all, the single CDSs of interest were cloned into different plasmids useful for the next step of *in vitro* recombination. The method used for these cloning steps was the same *restriction free* cloning procedure described before for the production of the Cdc45 truncation mutants (Fig. 12). All the plasmid vectors were purchased from ATG biosynthetics GmbH (Merzhausen, Germany). Here below the plasmid vectors are reported into which each CDS of interest has been cloned:

- MCM7 → pACEBac1
- MCM5 → pIDC
- FLAG-MCM3 → pIDK
- MCM4 → pACEBac1
- MCM2 → pIDC
- MCM6 → pIDK
- GST-Sld5 + Psf2 → pIDK (sequentially cloned *via* the homing endonuclease/BstXI)
- Cdc45 → pIDC
- Psf1 + Psf3 → pACEBac1 (sequentially cloned *via* the homing endonuclease/BstXI).

The method of the homing endonuclease/BstXI to clone multiple sequences into the same plasmid vector is based on couples of enzymes (CeuI/BstXI for pACEBac1 and SceI/BstXI for pIDK). Digestion of plasmid DNA with CeuI and SceI releases sticky ends that are complementary to the ones obtained by digestion of the BstXI site. Thus a specific sequence can be cut from these vectors by using the couple CeuI/BstXI or SceI/BstXI and ligated into a vector previously digested only with BstXI enzyme. Cloning a CeuI/BstXI- or SceI/BstXI-cleaved DNA fragment into a BstXI linearized plasmid does not reproduce the BstXI recognition site at the CeuI- or SceI-digested end, so that the BstXI site at the 3'-end of the inserted sequence can be used again for an additional cloning round.

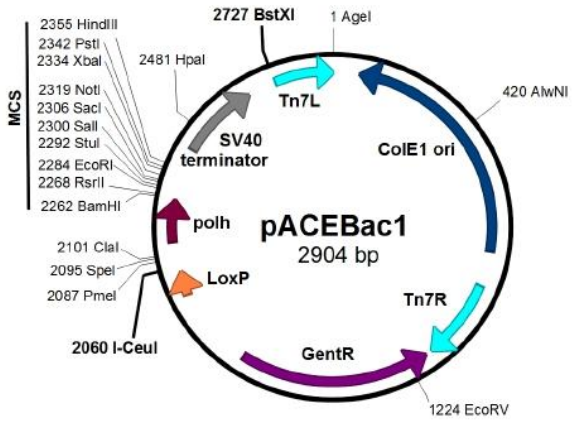

Figure 14 - Circle map representation of the pACEBac1 vector.

The pACEBac1 is indicated by the *ATG biosynthetics* manual as *acceptor vector* since it can accept incoming cloning cassettes from the other vectors indicated as *donor vectors* (pIDC, pIDK) during a reaction catalysed by *Cre recombinase* enzyme, which works on the *Lox P* sites present on both the acceptor and the donor vectors (Fig. 14 and 15). With this method the pIDK-FLAG-MCM3 and the pIDC-MCM5 plasmid constructs were fused with pACEBac1-MCM7, the pIDC-MCM2 and the pIDK-MCM6 constructs were fused with pACEBac1-MCM4 and the pIDC-Cdc45 and the pIDK-GST-Sld5/Psf2 constructs were fused with the pACEBac1-Psf1/Psf3 plasmid. Detailed protocols can be found on the manual provided by the *ATG biosynthetics* (*ACEMBL Expression System Series – MultiBacTurbo - Multi-Protein Expression in Insect Cells*) which is downloadable from the website of the company (https://www.atg-biosynthetics.com).

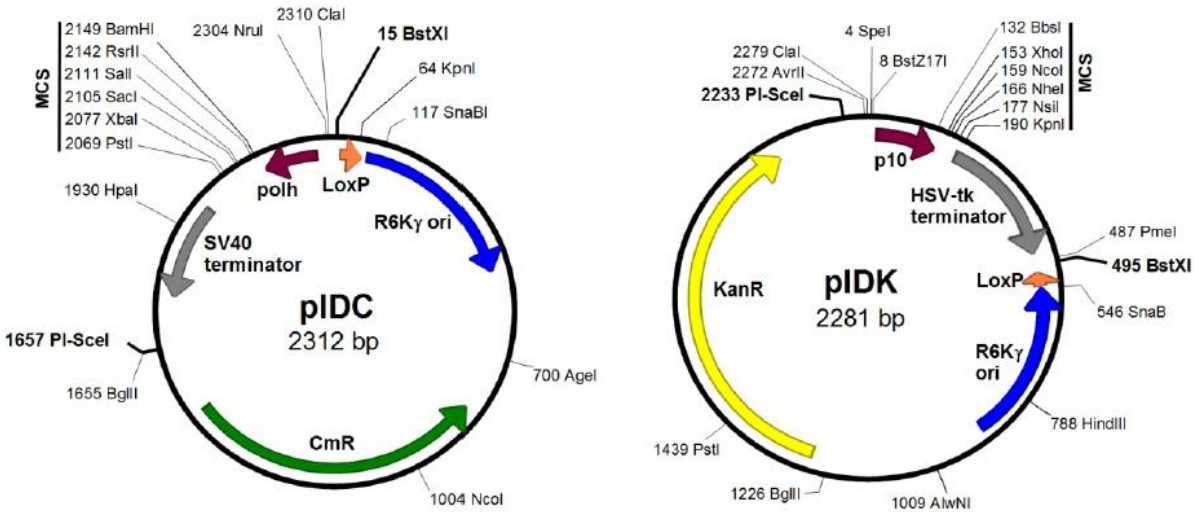

Figure 15 - Circle map representation of the pIDC and pIDK vectors.

For each recombination reaction the final construct comes to have the pACEBac1 as base structure. Thus the *Tn7L* and *Tn7R* sequences present in the pACEBac1

sequence can be exploited to get the transposition of the cloned sequences into the baculoviral genome contained in the *DH10MultiBac*^{Turbo} cells (*ATG biosynthetics*). To this purpose *DH10MultiBac*^{Turbo} cells were transformed with the *in vitro* recombination reaction mixture. *DH10MultiBac*^{Turbo} cells also contain a helper plasmid, which encodes the transposase required for the transposition reaction. Transformed cells were spread on plates and incubated over night at 37 °C. At this step blue-white selection is used to select colonies containing recombinant baculoviral genome and resistance to several antibiotics allows to select cells, which contains markers derived not only from the baculoviral genome but also from the recombined plasmid constructs (e.g. pACEBac1, pIDC, pIDK). Positive clones were identified and the baculoviral DNA was purified from the *E. coli* cells. This DNA preparation was then used to transfect *Sf9* cells (Novagen) by using *Cellfectin® II Reagent* (Invitrogen) and *Grace's Insect Medium Unsupplemented* (GIBCO). After the transfection step, *Sf9* cells were grown in *Sf-900™ II SFM* medium (GIBCO) supplemented with 10% FBS (Euroclone) and gentamicin (50 µg/ml; GIBCO). The viral stocks obtained from the transfection reactions were used for other two rounds of *Sf9* cells infection to get high-titre viral stocks to be used for protein production. Expression of the recombinant proteins after insect cells infection was assayed by western blot analysis with specific antibody.

Production of the recombinant CMG complex

Three main different protocols were set up to reconstitute a recombinant CMG complex in insect cells infected with recombinant baculoviruses.

Sf9 cells were grown in *Sf-900™ II SFM* medium (GIBCO) supplemented with 10% FBS (Euroclone) and gentamicin (50 µg/ml; GIBCO) at 27 °C.

In the first protocol 400 x 10^6 Sf9 cells (20 x 10^6 cells in each of 20 dishes (150X25mm)) were infected with eleven baculoviruses codifying for the single subunits of the CMG complex. More precisely, the viruses codifying for the following proteins were employed:

- MCM2 (6xHis-tag at the C-terminus)
- MCM3 (No tag)
- MCM4 (No tag)
- MCM5 (No tag
- MCM6 (6xHis + HA-tag at the N-terminus)
- MCM7 (6xHis + 3xFLAG-tag at the N-terminus)
- Cdc45 (6xHis-tag at the C-terminus)
- Psf1 (No tag)
- Psf2 (No tag)
- Psf3 (No tag)

- Sld5 (No tag)

The FLAG-tag present on MCM7 was exploited in the next affinity purification step.

250 µl of each baculoviral stock was used to infect a single dish of cells for a total amount of 5 ml employed for each virus in the whole infection round. The infections were kept at 27° C. After 60 hours infected cells were collected, washed with PBS and kept at -80° C or directly employed to purify the recombinant CMG complex. The cell pellet from the previous step was re-suspended in a volume of 4 ml of Lysis Buffer (40 mM Hepes-NaOH, pH 7.5, 100 mM sodium acetate, 1 mM DTT (Dithiothreitol), 20 mM magnesium acetate, 1 mM EDTA, 0.01% NP-40) with the addition of protease inhibitors (*Complete EDTA-free protease inhibitor cocktail*; Roche) and phosphatase inhibitors (*PhosSTOP phosphatase inibitor cocktail*; Roche). Cells were disrupted by 4 rounds of sonication with an amplitude of 5 microns (Soniprep MSE 150 sonicator; AL.BRA.) for 15 s each round. Crude extracts were clarified by centrifugation at 40,000 rpm for 40 min in a LE-80 BECKMAN ultracentrifuge.

Soluble extract was then incubated with *ANTI-FLAG M2 affinity gel* resin (2 ml; SIGMA) for 1 h at 4 °C on shaking. This mixture was loaded on a chromatographic column and the unbound material was let flowing through. The resin was then washed with 100 ml of Lysis Buffer. Then, 5 elution steps were carried out by using each time 1 ml of Lysis Buffer supplemented with 200 µg/ml FLAG peptide (SIGMA). In each elution step the mixture (resin + elution buffer) was incubated for 15 min at 4° C on shaking.

In the second protocol used in order to get a recombinant CMG complex a STREP-tagged version of MCM2 or MCM3 were alternatively employed for the production of a complex. In this case 400 x 10^6 Sf9 cells were infected with 11 baculoviruses codifying for the subsequent recombinant proteins:

- MCM2 (STREP-tag at the C-terminus) / MCM2 (6xHis-tag at the C-terminus)
- MCM3 (No tag) / MCM3 (STREP-tag at the N-terminus)
- MCM4 (No tag)
- MCM5 (No tag
- MCM6 (6xHis + HA-tag at the N-terminus)
- MCM7 (6xHis + 3xFLAG-tag at the N-terminus)
- Cdc45 (6xHis-tag at the C-terminus)
- Psf1 (No tag)
- Psf2 (No tag)
- Psf3 (No tag)
- Sld5 (No tag)

Each of 20 dishes containing 20 x 10^6 Sf9 cells was infected with 20 µl of baculovirus codifying for MCM2 (STREP-tag at the C-terminus) or alternatively a baculovirus

codifying for MCM3 (STREP-tag at the N-terminus). Then 120 µl of each baculoviruses codifying for all the other MCM proteins and 240 µl of the viruses codifying for Cdc45 and the proteins forming the GINS complex were added. We tried to favor the formation of a stable CMG complex rather than of other sub-complexes by balancing in this way the amount of viruses used in the infection.

After 72 hours infected cells were collected, washed with PBS and kept at -80° C or directly employed for the purification of the recombinant CMG complex.

The cell pellet from the previous step was re-suspended in a volume (4 ml) of Lysis Buffer (40 mM Hepes-NaOH, pH 7.5, 200 mM sodium acetate, 1 mM DTT, 10 mM magnesium acetate, 0.05% NP-40) with the addition of protease inhibitors (*Complete EDTA-free protease inhibitor cocktail*; Roche) and phosphatase inhibitors (*PhosSTOP phosphatase inibitor cocktail*; Roche). Cells were disrupted by 4 rounds of sonication with an amplitude of 5 microns (Soniprep MSE 150 sonicator; AL.BRA.) for 15 s each round. Crude extracts were clarified by centrifugation at 40,000 rpm for 40 min in a LE-80 BECKMAN ultracentrifuge.

Soluble extract was then incubated with *Strep-Tactin Sepharose resin* (2 ml; IBA) for 1 h at 4°C on shaking. This mix was loaded on a chromatographic column and the unbound material was let flowing through. The resin was then washed with Lysis Buffer (120 ml) and 5 elution steps were carried out by using each time 1 ml of Lysis Buffer supplemented with 2.5 mM Desthiobiotin (IBA). In each step, the mixture (resin + elution buffer) was incubated for 10 min at 4° C on shaking.

The last purification protocol was set up after the construction of the three multi-gene baculoviruses. In this case 400 x 10^6 Sf9 cells were infected with the following three baculoviruses:

- **Multi-MCM246** → codifying for MCM2 (no tag), MCM4 (no tag) and MCM6 (no tag).
- **Multi-MCM357** → codifying for MCM3 (FLAG-tag at N-terminus), MCM5 (no tag) and MCM7 (no tag)
- **Multi-GINS/Cdc45** → codifying for Cdc45 (no tag), Psf1 (no tag), Psf2 (no tag), Psf3 (no tag) and Sld5 (GST-tag at the N-terminus).

700 µl of each viral stock was used to infect a single dish containing 20 x 10^6 Sf9 cells (x 20 dishes).

After 60 hours infected cells were collected, washed with PBS and kept at -80° C or directly employed for the purification of the recombinant CMG complex.

The cell pellet from the previous infection step was re-suspended in 4 ml of Lysis Buffer (20 mM Hepes-NaOH, pH 7.5, 5 mM KCl, 1.5 mM $MgCl_2$) with the addition of protease inhibitors (*Complete EDTA-free protease inhibitor cocktail*; Roche). Cells in suspension were kept on ice and disrupted by using a Dounce homogenizer (30-40 strokes).

Crude cell extract was then adjusted to 0.3 M potassium acetate and centrifuged at 40,000 rpm for 40 min in a LE-80 BECKMAN ultracentrifuge.

Soluble extract was then incubated with *anti-FLAG M2 affinity gel* resin (1.5 ml; SIGMA) for 1 h at 4 °C on shaking. This mix was loaded on a chromatographic column and the unbound material was let flowing through. The resin was then washed with a volume (100 ml) of Wash Buffer (20 mM Hepes-NaOH, pH 7.5, 0.3 M potassium acetate, 1 mM DTT, 1 mM EDTA, 0.01% NP40, 10% glycerol). 5 elution steps were carried out by using for each step a volume (1 ml) of Elution Buffer (20 mM Hepes-NaOH, pH 7.5, 150 mM potassium acetate, 1 mM DTT, 1 mM EDTA, 0.01% NP-40, 10% glycerol) supplemented with 200 µg/ml 3xFLAG peptide (SIGMA). In each step the mixture (resin + Elution Buffer) was incubated with the resin for 15 min at 4° C on shaking.

Preparation of extracts from *Xenopus laevis* eggs

Female frogs (*Xenopus laevis*) were injected with 380-400 µl of 2500 U/ml of human chorionic gonadotropin (SIGMA) to stimulate ovulation. About 16 hours after the injection, eggs laid by the frogs were collected. A selection was carried based on the coloration of the egg cytoplasm: eggs that are completely white or spotted were removed, while eggs were kept that showed a typical bipartite coloration (with the vegetal pole more clear respect to the brown/black animal pole).

For each set of eggs coming from a single frog, after the selection, the water in excess was removed and eggs were treated with 15 ml of De-jellying Buffer (20 mM Tris/HCl, pH 8.5, 110 mM NaCl, 5 mM DTT) at R/T (room temperature). Eggs were occasionally shaken during this treatment to facilitate the removal of the jelly layer from the eggs. Then, the De-jellying Buffer was replaced with fresh buffer for a treatment which did not last over 5 min. The proceeding of the *dejellying* process can be observed because eggs start to touch closely with each other. The de-jellied eggs were washed three times with abundant 1/4MMR (1/4MMR buffer was got by diluting 20 folds a 5X MMR buffer stock: 100 mM Hepes/NaOH, pH 7.5, 2 M NaCl, 10 mM KCl, 5 mM $MgSO_4$, 10 mM $CaCl_2$, 0.5 mM EDTA).

Unfertilized eggs are stuck at the metaphase of meiosis II. Eggs need to be activated and complete the meiotic cycle to get interphase extracts. To this purpose 2 µl of 10 mM A23187 calcium ionophore solution (SIGMA) were added to a final volume of 20 ml 1/4MMR buffer containing the eggs. Activation of the eggs can be checked by following the typical contraction of the pigment in the animal pole. Moreover, due to this contraction, the animal pole starts to appear darker and the separation between the two poles starts to be sharper. This activation process is supposed not to last for more than 5 min. Again, during this activation step, eggs that do not display the typical bipartite coloration can be highlighted and removed.

After the activation step, eggs were quickly rinsed 3 times with abundant 1/4MMR buffer and, after the final wash, the excess buffer was removed as much as possible. Starting from this point eggs were kept on ice. Eggs were then washed several times with ice cold S-Buffer (50 mM Hepes/KOH, pH 7.5, 50 mM KCl, 2.5 mM $MgCl_2$, 250 mM sucrose) supplemented with 2 mM β-mercaptoethanol and 15 µg/ml leupeptin (SIGMA) as protease inhibitor. Eggs were eventually transferred into 2-ml tubes by using a plastic Pasteur pipet to avoid disturbing cells.

Tubes were centrifuged for 10 s at 6,000 rpm in a refrigerated micro-centrifuge to compact eggs and remove the excess of buffer. Then, tubes were centrifuged for 10 min at 13,200 rpm to crush eggs and to get a first rough separation. At this step extracts appear to be divided into 3 main layers: an upper cap of fats and yolk, a central more clear cytoplasmatic extract and a bottom pigments layer (Fig. 16).

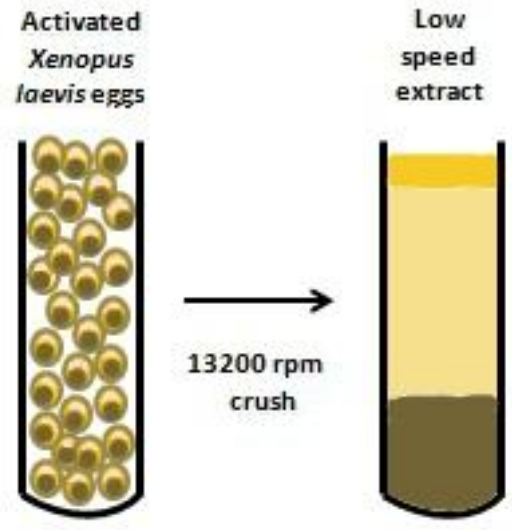

Figure 16 – Schematic representation of the result got from the first crushing centrifugation

A small hole in the fat layer was made by using a cut yellow tip (for P200 pipettes) and the central cytoplasmatic extract was collected into 15-ml tubes avoiding the upper fat layer and the pigments layer at the bottom. This extract was then supplemented with 40 µg/ml cytochalasin B (SIGMA) to avoid actin polymerization and then further centrifuged for 15 min at 70,000 rpm at 4 °C in an Optima Max-XP ultracentrifuge (Beckman) by using a TLA 100.3 rotor. This step of centrifugation provides a further clarification of the extract. Residual lipids will be stratified at the top, while mitochondria and residual pigments will be pushed at the bottom. Once again the central portion of the stratified extract is collected by using a cut yellow tip, transferred into new tubes and kept on ice. This final collected extract is supplemented with 15 mM creatin phosphate (SIGMA) and 150 µg/ml creatine phosphokinase (SIGMA), which will work as an ATP-regeneration system during the replication assays, and 10 µg/ml cycloheximide (SIGMA) to block protein synthesis. Moreover, glycerol to a final concentration of 3% (v/v) was added, before freezing the extracts in liquid nitrogen.

Preparation of Xenopus *laevis* de-membranated sperm nuclei

De-membranated sperm nuclei from male frogs were prepared following the method described by Blow and Laskey [91].

Quality control of the *Xenopus laevis* egg extracts

After each preparation egg extracts were checked for the competence to provide *in vitro* DNA replication.

Aliquots (volume 20 μl) of the freshly prepared extract were supplemented with Cy3-dCTP (GE Healthcare) to a final concentration of 0.5 nM. To this cold-kept extract de-membranated sperm nuclei were added to a final concentration of 3000 nuclei/μl of extract. After the addition of the sperm nuclei DNA replication is started by keeping the extract at 23 °C. After 30 min (and after additional 30 min), an aliquot (volume: 3 μl) of this mixture was picked up, spotted on a microscope slide and fixed by the addition of 3 μl of Stop Buffer (15 mM PIPES, pH 7.2, 15 mM NaCl, 80 mM KCl, 10% formalin; 2 μg/ml Hoechst, 50% glycerol). De-condensation of the chromatin coming from the sperm nuclei and incorporation of Cy3-dCTP was checked by fluorescence microscopy to evaluate whether the extract was competent to provide *in vitro* DNA replication (Fig. 17).

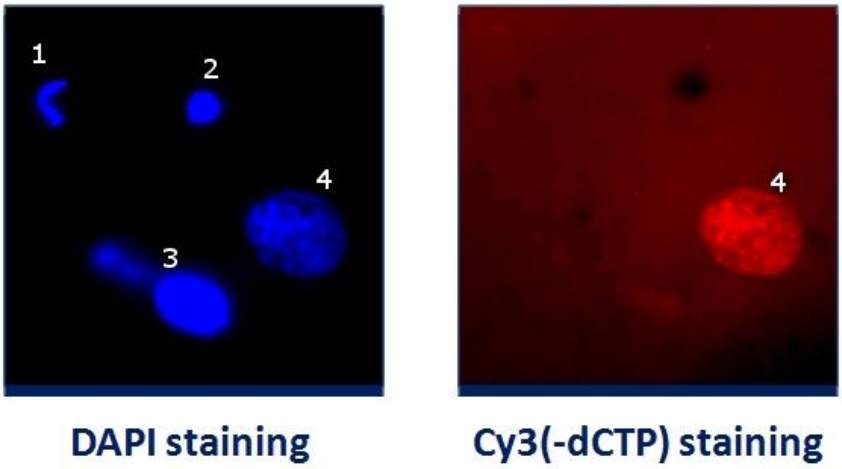

Figure 17 - Quality control of *Xenopus laevis* egg extracts after preparation. On the *left* DAPI-staining of four different chromatin structures is shown for the sperm nuclei in the extracts. The structure reported as # 1 represents a sperm nucleus whose chromatin is still fully condensed. The structure reported as # 4 represents a sperm nucleus that has fully de-condensed its chromatin forming a chimeric nuclear structure. With # 2 and # 3 two intermediate structures between the two previous states are indicated. On the *right* the same field detected by Cy3-staining shows that dCTP is incorporated only into the structure # 4, where chromatin is fully de-condensed and DNA replication can occur.

Extracts which could give the replication of the 90% of the nuclei visualized at the microscope were considered to be fully competent and used for DNA replication assays.

Immuno-depletion of the endogenous Cdc45 factor from *X. laevis* egg extracts

Rabbit polyclonal antibodies directed against the Cdc45 frog factor (α-xCdc45) used in this study were provided by Dr V. Costanzo (Cancer Research, UK).

To immuno-deplete xCdc45 protein from 200 μl of egg extract, 80 μl of Protein A Sepharose Fast Flow beads (GE Healthcare) were employed. Beads were washed 3

times with PBS (*P*hosphate-*B*uffered *S*aline, LONZA) and then incubated with 80 µl of α-xCdc45 for 1 h at 4 °C. Then, beads were rinsed 3 times with 1 ml of PBS to wash away the unbound antibodies. Beads were then divided into 4 aliquots in different tubes which were used to get 4 sequential rounds of depletion on the same extract. 200 µl of egg extract were incubated with the first aliquot of conjugated beads and left at 4 °C for 1h on a slow rotating wheel. After 1 h the samples were centrifuged for 30 s at 3000 rpm in a refrigerated micro-centrifuge. The overlying extract was then collected and incubated with the second aliquot of beads for 45 min. The same procedure was repeated for the third and the fourth incubation that lasted 30 min. After the last round of immuno-depletion beads were collected by centrifugation for 30 s at 3000 rpm and the overlying collected extract was ready to be used for our experiments.

The efficacy of the immuno-depletion was checked each time by western blot analysis on the straight depleted extract.

To get negative controls each time the same amount of extract was used for a mock immuno-depletion, according to the same protocol but using beads conjugated with rabbit pre-immune serum.

DNA replication assays

To check DNA replication activity with the *Xenopus laevis* egg extract system aliquots of extract (20 µl) were used for a single reaction. This extract aliquot was supplemented with 0.1 µl of [α-^{32}P]dCTP (3000 Ci/mmol; Perkin Elmer) and 60000 sperm nuclei while kept on ice.

xCdc45 recombinant protein (or hCdc45 recombinant protein or hCdc45_A1_LP1 mutant, as indicated) was added to the reaction mixture at a final concentration of 54 ng/µl. In control reactions an equal amount of protein storage buffer was used (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol).

DNA replication assays were started by keeping the reactions at 23 °C. After 1 or 2 h the reactions were stopped by the addition of 100 µl of DNA Stop Solution (80 mM Tris/HCl pH 8, 0.13% Phosphoric Acid, 10% Ficoll, 5% SDS, 0.2% BPB (Bromophenol Blue), 8 mM EDTA) supplemented with 2 mg/ml proteinase K (Roche). Stopped reactions were kept for 1 h at 50 °C or, alternatively, over night at 37 °C.

Aliquots of each samples (30 µl) were loaded on an agarose gel (0.8% agarose in 1X TBE). The gel was run for 90 min at a constant voltage of 150 V in 1X TBE buffer.

After the run the gel was incubated in a 10% TCA (Trichloroacetic Acid) bath for 10-15 min to precipitate the DNA. Then, the agarose gel was dried by means of a 3-mm paper/tissue paper sandwich and the radioactivity revealed by phosphorimaging.

RESULTS

Cdc45 shows sequence similarity to archaeal proteins belonging to the DHH family of phosphoesterases

As described in the Introduction section, Cdc45 is a fundamental factor for the initiation and the elongation phases of the DNA replication mechanism [2,36]. It is part of the CMG complex, the DNA replicative helicase, together with the MCM2-7 and the GINS complexes [37,38].

Most of the factors which form the DNA replication machinery in Eukarya are present as homologous proteins also in the archaeal system. Archaea are prokaryotic organisms that for many aspects of their cellular physiology present similarity with the eukaryotic organisms.

Indeed, when we started focusing our attention on the CMG complex a counterpart was known for the MCM proteins in the archaeal system. In fact in most archaeal species a single MCM protein was identified, which is homologous to the eukaryotic MCM proteins, and can form homo-hexamers in its physiological state [72]. As for the GINS complex, two proteins are usually present in the archaeal species which are homologous to the eukaryotic ones and they can form tetrameric assemblies (i.e. dimers of dimers). One protein is named GINS15 because it is similar to eukaryotic Psf1 and Sld5 factors, while the other one is named GINS23 because it is related to the Psf2 and Psf3 subunits [73].

Thus, looking at this general framework, it was pretty strange that there was no counterpart known in the archaeal system of the Cdc45 replication factor and to assess this point we started a bio-informatic analysis of the human Cdc45 factor.

First of all, a database search was carried out using the human Cdc45 aminoacidic sequence as a query by means of the Position-Specific Interactive BLAST algorithm, using default parameters. The human Cdc45 protein was found to have a weak but significant degree of similarity with two archaeal sequences. The first protein was annotated as “phosphoesterase domain-containing protein” from the archaeon *Candidatus korarchaeum cryptofilum* OPF8 (E-value: $2x10^{-4}$). The second protein was indicated as “putative single-stranded DNA-specific exonuclease RecJ” from *Methanocella paludicola* SANAE (E-value: $3x10^{-4}$). Both proteins belong to the DHH family of phosphoesterases.

These two sequences had been only recently added to the databases, explaining why the similarity with the eukaryotic Cdc45 proteins had not been previously identified.

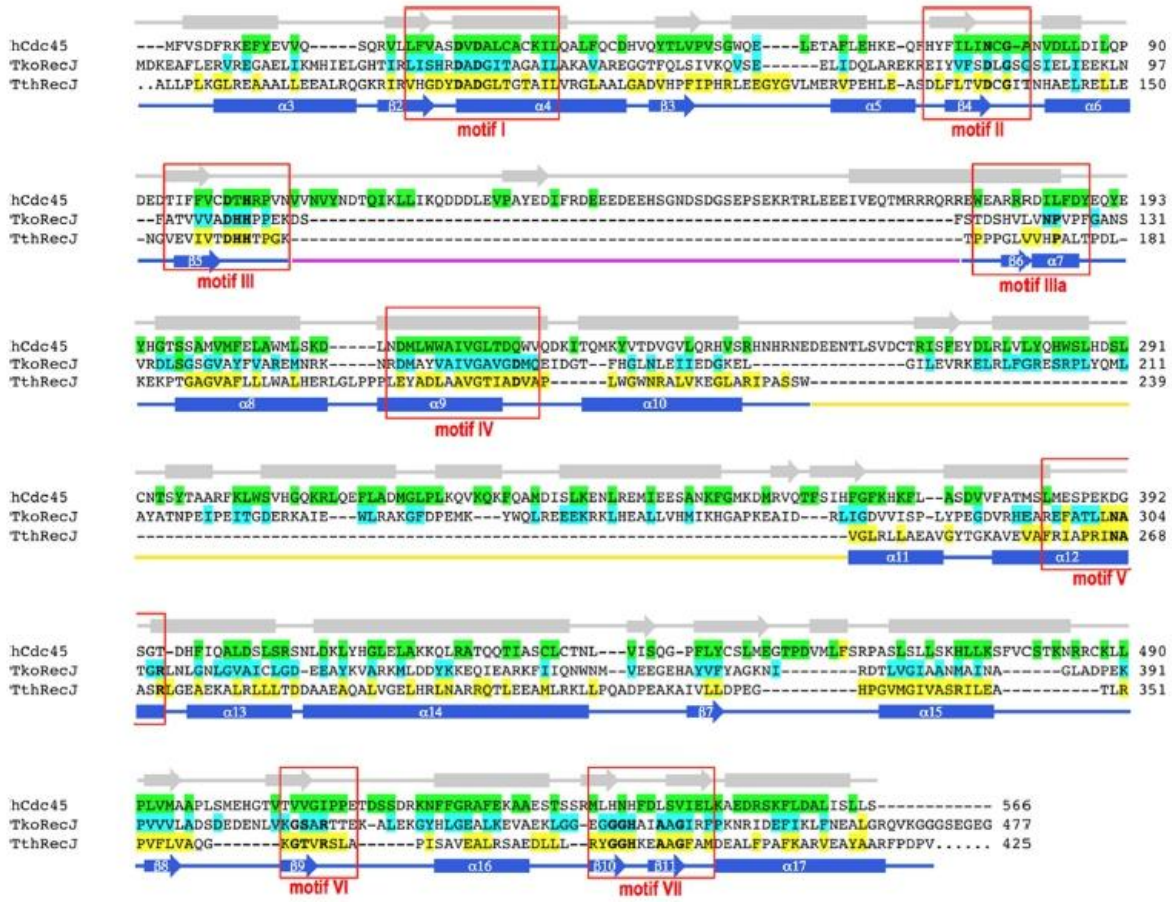

Figure 18 - Sequence alignment between hCdc45, RecJ^{Cdc45} from *T. kodakaraensis*, and RecJ from *T. thermophilus*. The alignment presented here is based on an extended multiple alignment using 15 eukaryotic Cdc45 sequences, 16 archaeal sequences, and 10 bacterial RecJ sequences, selected from evolutionary diverse organisms. Only the RecJ core (residues 50–425, comprising domains I and II) has been used in the alignment. Residues that are conserved in more than 70% of the eukaryotic, archaeal, and bacterial sequences are highlighted in *green*, *cyan*, and *yellow*, respectively. The following groups of amino acid residues were considered similar: Asp/Glu, Lys/Arg, Phe/Tyr, Ser/Thr, Gly/Ala, and Val/Leu/Ile/Met. The position of the secondary structural elements in the *T. thermophilus* RecJ crystal structure (PDB code: 2ZXP) is indicated at the bottom, whereas the predicted secondary structure for human Cdc45 is shown at the top. Secondary structure elements are named according to the TthRecJ nomenclature [76]. The position of the characteristic RecJ motifs is shown by *red boxes*, with the residues conserved highlighted in bold. The alignment was carried out using the multiple sequence alignment program MUSCLE [78] and manually modified to take into account the structural constraints. Up to motif IV the similarity is strong enough to be detected based on sequence alone, whereas the second half of the alignment relies on the threading data, which identify similarity patterns in the absence of high sequence homology, as exemplified by the conservation of the patterns of hydrophobic residues and the excellent match between RecJ secondary structure elements and the prediction for Cdc45. An insertion unique to eukaryotic Cdc45 orthologues and containing many charged amino acid residues is shown in *magenta*. The putative helical insertion present in both archaeal and eukaryotic proteins is shown in *yellow*.

Although the sequence similarity detected by BLAST involved only the first 130-140 residues, multiple alignments between Cdc45 and putative archaeal orthologues show that some similarity can be noticed throughout the entire sequence (Fig. 18).

Because archaeal species possess multiple RecJ-like proteins, we have selected those displaying the closest match to the *Candidatus korarchaeum cryptofilum* and

Methanocella paludicola homologues. To distinguish the archaeal RecJ/DHH proteins that may be putative Cdc45 orthologues from other members of the DHH family, I will use the notation $RecJ^{Cdc45}$. However, most Archaea seem to possess two close paralogues of the putative Cdc45 counterpart.

A study, focused on detecting the archaeal counterpart of *E. coli* RecJ protein, identified two highly similar proteins from *Methanocaldococcus jannaschii* (MJ0977 and MJ0831). Although both were able to partially complement a RecJ mutant phenotype in *E. coli*, single-stranded DNA nuclease activity could be observed only for the MJ0977 protein [79]. Two highly similar $RecJ^{Cdc45}$ proteins (AF0699 and AF0735) were also found associated with a replication protein network in *Archaeoglobus fulgidus* [80]. More recently, a RecJ-like single-stranded 5'-3' DNA exonuclease from *Thermococcus kodakaraensis* was found to physically associate with GINS, and this association was found to stimulate its exonuclease activity [81].

A putative RecJ homologue has been reported to co-purify with MCM and GINS from cellular extracts of *Sulfolobus solfataricus* (SSO0295 [82]). This sequence is rather divergent from most of the archaeal proteins mentioned above and lacks some of the characteristic motifs; in particular, some of the conserved residues putatively responsible for the nuclease activity are absent. An open reading frame highly similar to SSO0295 is present in most *Sulfolobales* genomes. Divergent sequences are also detected in other Crenarchaeota, such as *Aeropyrum pernix*.

Both the eukaryotic Cdc45 proteins and the archaeal $RecJ^{Cdc45}$ sequences match only to the RecJ catalytic core (domains I and II, as defined in the crystal structure of *Thermus thermophilus* RecJ [75], while lacking the 50-residues N-terminal extension as well as domains III and IV. These additional elements form a closed-ring structure that is predicted to encircle single-stranded DNA.

Cdc45 sequences only partially retain the motifs typical of the DHH family that are involved in metal binding. Whereas in motif I both aspartate residues coordinating the metal are conserved (^{26}DVD), the DHH motif III, which gives the name to the family, is mutated to ^{99}DTH. Although a number of residues belonging to motifs II and IV are conserved suggesting that the protein fold is similar in those regions, the key residues for metal binding and catalysis are not conserved, with the aspartate of motif II mutated to asparagine, and the aspartate of motif IV to glutamine (Fig. 18). In contrast, most of the archaeal $RecJ^{Cdc45}$ comprise all of the motifs that are conserved in the bacterial RecJ exonucleases, with the exception of the putative orthologues from *Sulfolobales*, whose sequence is rather divergent and lacks most of the RecJ canonical residues.

Threading algorithms confirm the presence of a RecJ/DHH-like core fold in hCdc45

A variety of threading/fold recognition algorithms were used to verify the similarity among hCdc45 and DHH family members. The Protein Fold Recognition Server Phyre [83] identifies a three-dimensional similarity between hCdc45 and a number of DHH protein structures such as the RecJ exonuclease from *T. thermophilus* (PDB code: 1IR6), as well as various inorganic pyrophosphatases (PDB codes: 1WPN, 1I74, and 2HAW), but the similarity was restricted to the first 110 amino acids.

When a multiple sequence alignment, including a number of Cdc45 sequences, was used as input for the HHPRED server [84], a similarity (involving the first 315 residues) was detected to an inorganic pyrophosphatase from *Bacillus subtilis* belonging to the DHH family (PDB code: 1WPN).

The profile-profile alignment and fold-recognition server FFAS03 [85] was also used both with the full-length hCdc45 sequence and a number of fragments corresponding to putative domains. A score below the recommended threshold (-9) was obtained when the first 140 amino acid residues of hCdc45 were used as input, matching the N-terminal domain of a number of manganese-dependent inorganic pyrophosphatases from the DHH family, with the *B. subtilis* inorganic pyrophosphatase giving the best agreement (-10.5, PDB code: 1K23). Using as input longer hCdc45 fragments still provides a match with the DHH proteins, although the scores get progressively higher, indicating a lower degree of confidence; using the full-length protein it is possible to detect some similarity to 1K23.

The sequence-structure homology recognition server FUGUE [86] unambiguously identifies both the *T. thermophilus* RecJ (1IR6) as well as the *Streptococcus gordonii* inorganic pyrophosphatase (1K20) structures as the two best hits along the entire sequence with a degree of confidence higher than 99%. When using as input the sequences of a variety of archaeal RecJ^{Cdc45} proteins, all the servers predicted a strong structural similarity to the bacterial RecJ proteins, along the entire sequence, as expected from the significant sequence homology and the conservation of the characteristic motifs.

A central region in both Cdc45 (residues 252–363 in hCdc45) and archaeal RecJ^{Cdc45} (residues 188–275 in the sequence from *T. kodakaraensis*) proteins, appears as a long insertion into the RecJ/DHH core. This region is reasonably well conserved between eukaryotic and archaeal proteins. Secondary structure predictions suggest a helical fold, and threading algorithms tend to find matches with helical bundle proteins, such as acyl-CoA-binding proteins and helix-turn-helix transcription factors. An additional insertion is unique to the eukaryotic Cdc45 proteins (residues 108–175 in the human sequence), and the central part includes a large number of charged residues (aspartate, glutamate, arginine, and lysine), thus suggesting a partially structured region (Fig. 18). Based on the results of both the sequence analysis and the threading algorithms, we have produced a sequence alignment that summarizes the putative relationships among bacterial RecJ single-stranded DNA exonucleases, archaeal RecJ^{Cdc45} and the eukaryotic Cdc45 (Fig. 18).

In the first half of the molecule (up to motif IV) the similarity is strong enough to be detected based on sequence alone, whereas for the second half the alignment relies on the threading data, which identify similarity patterns in the absence of high sequence conservation. Consistent with the threading results, Fig. 18 shows the conservation of patterns of hydrophobic residues and an excellent match between the experimentally determined α-helices and β-strands of RecJ and the predicted Cdc45 secondary structure elements. Although the correspondence is good overall, there are a few uncertainties in the central region of the proteins. For example it is difficult to establish unambiguously whether the first Cdc45 insertion occurs after β5 (as depicted in Fig. 18) or after α7. In the same way the second insertion (common to eukaryotic and archaeal proteins) may slide from the current position in Fig. 18 (after α10) to an alternative position after α11. However the match is very convincing in the

first half of domain I as well as in domain II, including the long helix connecting the two domains (α14), suggesting that the relationship extends throughout the entire sequence.

While we were finalizing our report, a bioinformatic analysis [87] was published suggesting a similarity between Cdc45 and RecJ proteins, limited to the first 100 amino acid residues. Our report is consistent with that one [87] but further extends the analysis by showing that the sequence and structural similarity covers the entire length of the protein.

Biochemical characterization of the recombinant hCdc45

We produced in bacterial cells hCdc45 as a fusion protein with a C-terminal His_6-FLAG tag, using the pNIC-CTHF vector [88]. The tag could be cleaved with the TEV protease and the protein purified to homogeneity (Fig. 19).

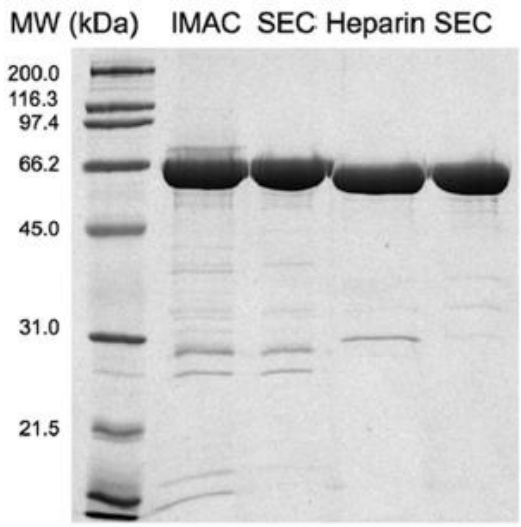

Figure 19 - Purification of hCdc45. SDS-PAGE analysis of samples throughout the purification protocol (see detailed description under "Experimental Procedures"), starting from the protein obtained after the first step of Ni-affinity purification (IMAC (immobilized metal-ion affinity chromatography)), followed by size-exclusion chromatography (SEC); the protein after cleavage of the His_6-FLAG tag using TEV protease and purification over a heparin column (Heparin) to a final round of SEC.

Although many of the residues that in RecJ/DHH proteins are involved in Mn^{2+} binding and catalysis are not conserved in hCdc45, we examined the possibility that the few remaining aspartate and histidine residues (namely D26, D28, D99, and H101) may possibly coordinate one metal ion. We have therefore used both inductively coupled plasma/atomic emission spectroscopy and atomic absorption spectroscopy to test whether the purified protein contains manganese, magnesium, or zinc, but we failed to confirm the presence of any of these metals.

Dr Roberto Udisti (University of Florence) worked for determination of metal ions in protein samples by inductively coupled plasma-atomic emission spectroscopy, while Dr Spartaco De Gennaro and Dr Achille Palma (Metapontum Agrobios) carried out

the inductively coupled plasma-MS analysis. Dr Marco Trifuoggi (University of Naples) collaborated for the atomic absorption spectroscopy analysis.

We also carried out activity assays to check whether hCdc45 displays either pyrophosphatase or exonuclease activity. Consistently with the absence of metal ions and some of the putative catalytic residues, we were unable to detect any of the above enzymatic activities.

Both its role in DNA replication and the putative similarity with a single-stranded DNA exonuclease, such as RecJ, suggested that Cdc45 could be a DNA-binding protein.

We used EMSAs to evaluate the DNA-binding properties of hCdc45 (Fig. 20).

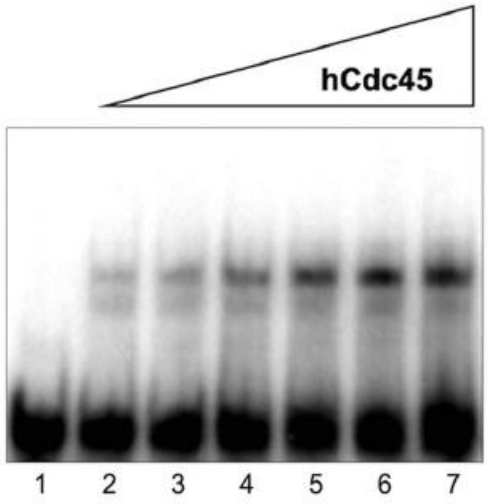

Figure 20 - DNA-binding activity of hCdc45. Example of an EMSA on single-stranded DNA is shown. The assay was carried out with increasing concentrations of hCdc45 (0.5, 1, 2, 3, 4, and 5 μg of protein were present in the mixtures loaded into the *lanes* from *2* to *7*). A radio-labeled 56-mer DNA was used as a ligand (see Experimental Procedures section for details). A control mixture without protein was run on *lane 1*.

The purified recombinant protein binds single-stranded synthetic oligonucleotides with a weak but detectable affinity, comparable with the one observed for the *Drosophila* GINS complex [12] or for the full-length *T. thermophilus* RecJ [75].

No increase in affinity was observed when a fork-containing DNA molecule was used as a ligand in the EMSAs, and negligible binding was detected to short blunt double-stranded DNA (Fig. 21).

Addition of Mg^{2+}, Mn^{2+}, or Zn^{2+} ions into the buffer, as well as variation of pH in the range 5.5-8.5, were found to have no effect on the hCdc45 DNA-binding activity. Preliminary experiments indicated that the affinity of hCdc45 for single-stranded RNA is weaker with respect to single-stranded DNA (Fig. 22).

The protein shows a preference for an oligo(dG) in comparison to oligo(dC), oligo(dA), and oligo(dT) (Fig. 23).

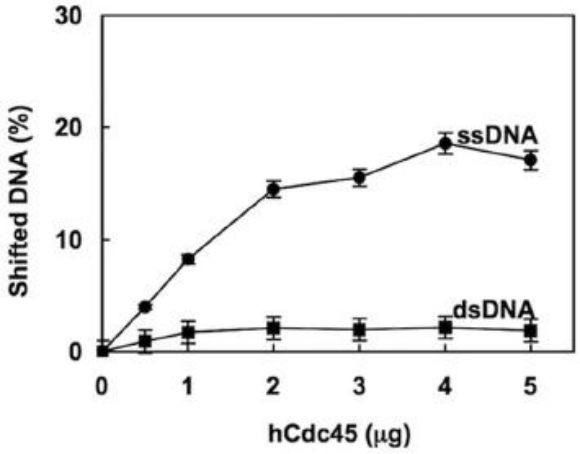

Figure 21 - Single-stranded *versus* double-stranded DNA binding. Shifted DNA (either in single- (●) or double-stranded (▄) form) is reported *versus* the amount of hCdc45. Experiments were performed in triplicate, and the results are averaged. *Curves* represent best fits to the data points. The *error bars* on the *graphs* are the ±S.E.

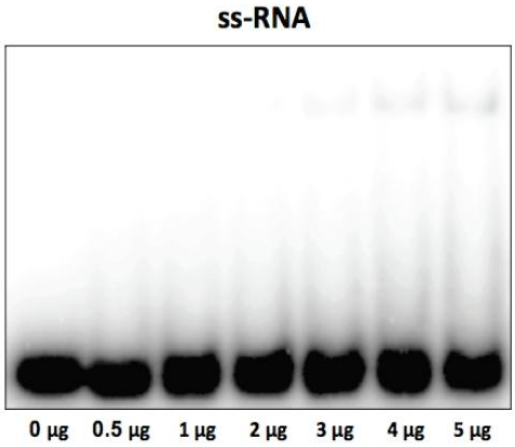

Figure 22 - Binding of hCdc45 on single-stranded RNA. EMSAs on a radio-labeled 20-mer synthetic single-stranded RNA (100 fmoles/assay) were carried out using the indicated amounts of hCdc45.

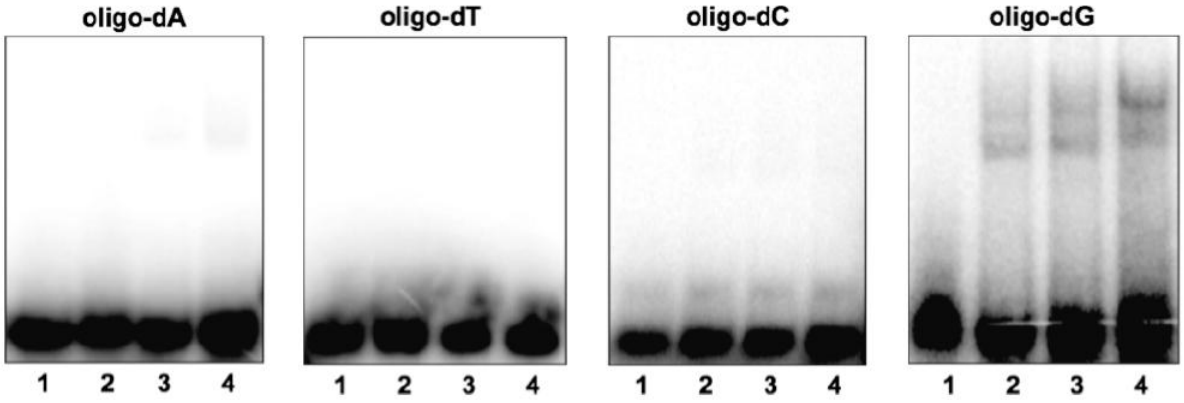

Figure 23 - DNA-binding activity of hCdc45 on various ligands. EMSAs on radio-labeled 20-mer synthetic oligo-dA, oligo-dT, oligo-dC and oligo-dG (100 fmoles/assay, as described above) were carried out using 0,1, 2.5 and 5 μg of hCdc45 (lanes from 1 to 4 in each gel, respectively).

To demonstrate that the weak DNA-binding activity is indeed due to hCdc45, and not to trace amounts of a contaminating protein, we analyzed the DNA-binding activity of the protein fractions following size-exclusion chromatography. The DNA-binding activity profile precisely co-migrates with the protein peak (Fig. 24) suggesting that the ability to bind DNA is a truly intrinsic feature of hCdc45. Furthermore, we analyzed whether the protein-DNA complex could be super-shifted by a specific antibody. For this experiment, we used a FLAG-tagged version of hCdc45 (purified according to the same protocol as the untagged protein with the omission of the TEV cleavage step) and a monoclonal anti-FLAG antibody. This analysis (Fig. 25) revealed that the anti-FLAG antibody was able to super-shift the protein-DNA complex.

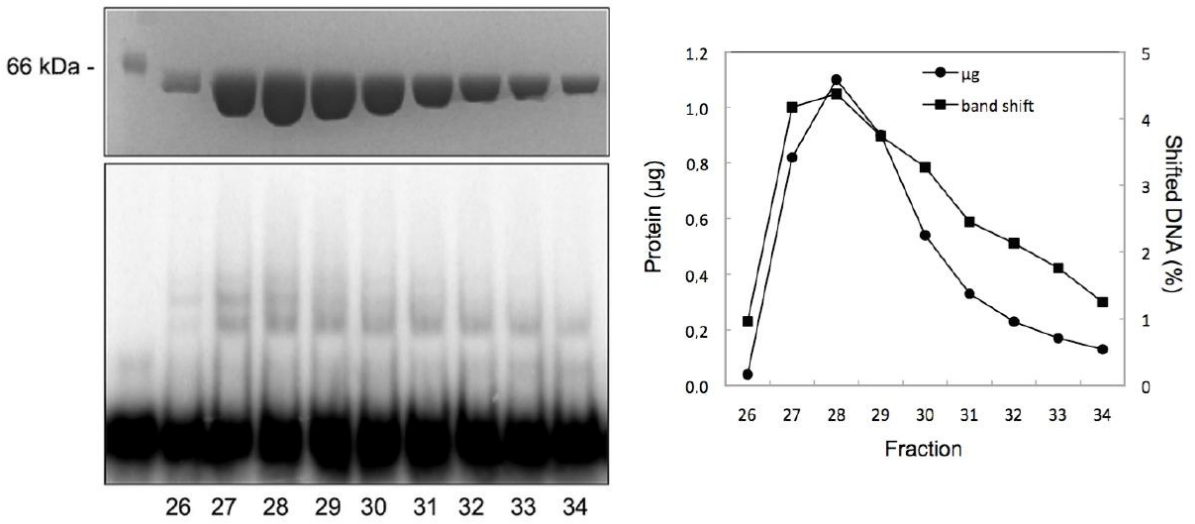

Figure 24 - Gel filtration analysis of hCdc45 and EMSAs of the corresponding peak fractions. Gel-filtration chromatography of purified hCdc45 was performed using a Bio-Sil SEC-250 column (Bio Rad) as described under “Experimental Procedures.” Peak fractions were analyzed by SDS-PAGE (5µl/fraction) and used in a gel shift experiment (1 µl/fraction). On the right a plot of the shifted radio-labelled DNA versus each fraction is shown.

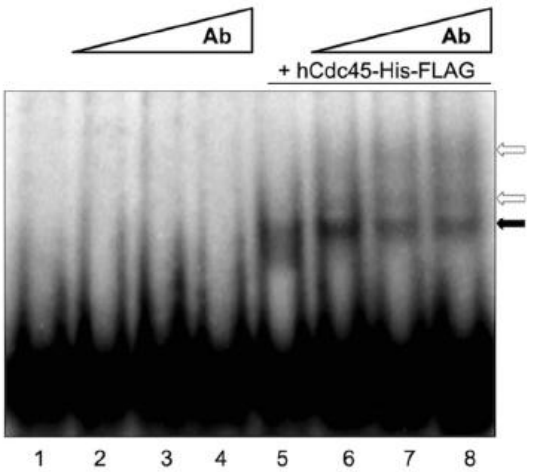

Figure 25 - EMSAs with hCdc45-His-FLAG in the presence of anti-FLAG antibody. The assays were carried out by adding increasing amounts of a monoclonal anti-FLAG antibody (0.5, 1, and 2 µg, *lanes 2* and *6*, *3* and *7*, and *4* and *8*, respectively) into mixtures containing the single-stranded DNA probe with hCdc45-His-FLAG (*lanes 6–8*; 5µg of protein) or without the recombinant protein (*lanes 2–4*; see text for details). A *black arrow* indicates the Cdc45-DNA complex, whereas the *white arrows* identify the ternary complexes with the anti-FLAG antibody.

SAXS analyses of the recombinant hCdc45 protein are consistent with the three-dimensional structure of the RecJ core

SAXS data were collected from highly purified samples of recombinant hCdc45 and corrected for the scattering from the buffer. *Ab initio* modeling was performed using the program DAMMIN [89], with 44 runs being averaged using the program DAMAVER [90]. The model obtained can be described as a compact core with two lateral extensions (Fig. 26).

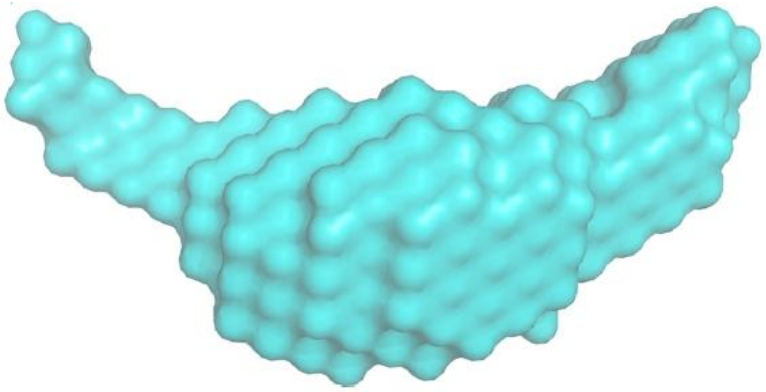

Figure 26 - Small angle x-ray scattering data. Final model reconstructed from the scattering curve for hCdc45.

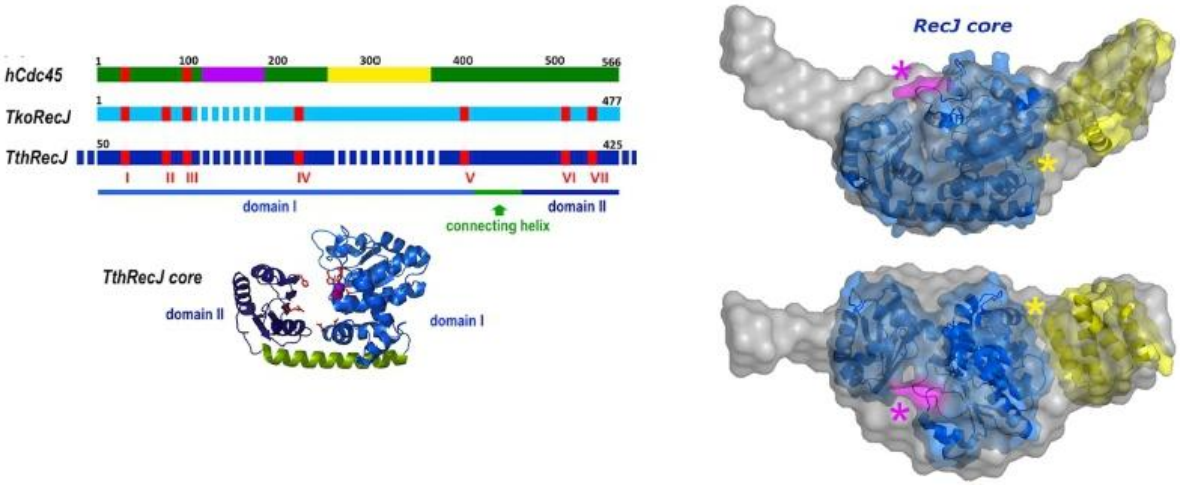

Figure 27 - SAXS data are consistent with a RecJ-like fold. On the left, a *schematic diagram* summarizing the result of the bioinformatic analysis and showing the relationship between eukaryotic Cdc45 proteins, archaeal RecJCdc45, and bacterial RecJ single-stranded DNA exonucleases is shown. The RecJ motifs are shown in *red*, the Cdc45 charged insertion is in *magenta*, and the helical archaeal/eukaryotic insertion is in *yellow*. The crystal structure of the core of *T. thermophilus* RecJ (PDB code: 2ZXP, domains I and II) is shown below the diagram, with domain I in *blue*, domain II in *dark blue*, and the connecting helix in *green*. The conserved residues in the seven RecJ motifs are shown in *red*. The *ab initio* calculated SAXS model for hCdc45 (depicted as *gray light spheres*) is superimposed to the crystal structure of the core of *T. thermophilus* RecJ, in *blue*. Highlighted in *magenta* and indicated by a *magenta asterisk* is the putative position of the insertion, which is unique to the eukaryotic Cdc45 orthologues; highlighted in *yellow* and indicated by a *yellow asterisk* is the position of the helical bundle insertion that is common to both archaeal and eukaryotic proteins (see Fig. 18). As an example, the helical domain of the acyl-CoAbinding protein (PDB code: 2FDQ) has been fitted to the map, consistently with the results of the threading algorithms. The two views are roughly related by a 90° rotation around a *horizontal axis*.

One of the two lateral extensions is larger and better defined and can be assigned to the putative helical insertion domain that is common to both the eukaryotic Cdc45 and the archaeal RecJ^{Cdc45} proteins (residues 188–275 in hCdc45), as the putative insertion loop within the RecJ core is positioned in a manner compatible with this interpretation (Figs. 18 and 27). As an example, we choose to fit to this region of the map the helical bundle from the acyl-CoA-binding protein (PDB code: 2FDQ) as suggested from the results of the threading analysis. The second insertion has a more elongated shape and can be allocated to the partially unstructured insertion that is unique to the Cdc45 sequences (residues 108–175 in hCdc45, Figs. 18 and 27). As discussed above, there is some uncertainty in the exact positions of the two long insertions with respect to the RecJ core, and alternative insertion points can be suggested. However, in both cases the alternative loops are still positioned in a manner compatible with the interpretation of the SAXS data.

Production of the human recombinant CMG complex

It is very useful to reconstitute a recombinant CMG complex to study *in vitro* the role played by Cdc45 together with GINS in the activation of MCM2-7 DNA helicase activity.

We carried out several attempts to produce a stable recombinant CMG complex by co-infecting *Sf9* cells with 11 baculoviruses codifying for the single subunits of the complex. Different constructs expressing the same factors with different tags were employed and different strategies were utilized for the first affinity chromatography step (e.g. FLAG-, STREP-affinity chromatography). Using these strategies we were usually able to identify all the 11 proteins, which were actually expressed in the extract got from infected cells, as indicated by western blot analyses. Nevertheless, the 11 proteins were never found together in the fractions eluted from the affinity chromatography steps, indicating that in our preparations at least the most of these factors did not associate to form a stable CMG complex. We figured out that the main problem with the production of a stable complex could be due to the fact that only a small percentage of the produced proteins associate to form a complex. Most of the produced material remains as single proteins or associates to form sub-complexes different from the CMG.

To address this point we tried to modulate the amount of the different viruses used in the co-infections, most of the time using an higher amount of the viruses codifying for Cdc45 and GINS proteins respect to the ones expressing the 6 MCM proteins, to push towards the formation of the CMG complex. However, all these attempts were not successful.

Finally, we decided to use the MultiBac^{Turbo} system, a novel technology that allows the construction of multi-gene baculoviruses. We constructed the following multi-gene baculoviruses: Multi-MCM246, Multi-MCM357 and Multi-GINS/Cdc45 (see *Experimental Procedures*). The rational at the basis of this choice was the idea that co-infecting cells with only 3 baculoviruses could provide a more coordinated expression of the 11 proteins, bringing to the association of a stable complex. Also the cloning of these constructs was designed to reduce the formation of sub-complexes among the MCM proteins, which could compete with the formation of the

MCM2-7 complex. Indeed MCM subunits known to be involved in the formation of sub-complexes were cloned into different constructs.

By using this method (see *Experimental Procedures*) we were able to pull-down all the 11 proteins likely composing a stable CMG complex with a FLAG-affinity chromatography step. Nevertheless, we got very low amounts of protein complex which were not sufficient so far to undergo a subsequent purification step or any kind of analytical assay.

A western blot analysis after a typical purification round is shown in Fig. 28.

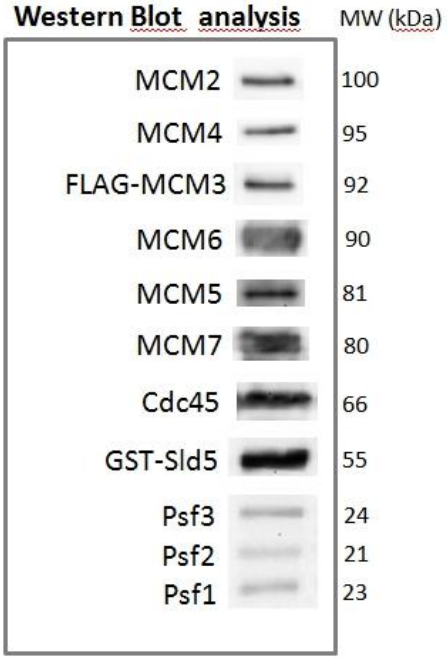

Figure 28 – Western Blot analysis of the CMG complex components in the FLAG specific eluted fraction. The presence of all the components of the CMG complex was checked by western blot analysis. Note that Psf1 has an altered migration on SDS-PAGE despite its expected molecular mass.

The N-terminal portion of hCdc45 is not involved in the DNA binding activity

We demonstrated that hCdc45 protein retains single-stranded DNA binding activity. Structural motifs that are usually present in DNA binding proteins have not been identified in hCdc45. We postulated that the weak single-stranded DNA binding activity shown by this protein could be due to the presence of several positively charged amino acids in the C-terminal region of the protein.

Thus, to assess which portion of the protein was actually involved in the binding of the single-stranded DNA molecules, we produced two different truncated forms of the protein. The first construct, hCdc45-A13/1, comprises the first 258 amino acids of the wild type protein, while the second one, hCdc45-C-ter codify for the region from the residue 394 to the end of the *wild type* protein.

Unfortunately the C-terminal truncation form is not stable in solution when produced as a recombinant protein. Thus we were not able to carry out any DNA-binding assays with it. On the other hand, the hCdc45-A13/1 mutant was produced and found

to be very stable as recombinant protein. We checked this construct for single-stranded DNA binding activity but we did not observe any kind of binding in EMSAs.

The hCdc45 L1 aminoacidic insertion loop does not seem to be needed for DNA replication

The *Xenopus laevis* egg extract experimental system had been employed since years for the study of factors involved in the DNA replication process. Indeed it is a very powerful system by which DNA replication process can be monitored *in vitro*.

In the extracts obtained from the *X. laevis* eggs all the replicative factors are present at high concentration and they can support duplication of an exogenous DNA molecule *in vitro*. When de-membranated sperm nuclei from the male frogs are added to the extracts, their tight chromatin structure starts to uncondense and a chimeric nuclear structure is formed which can provide the replication of the chromatin coming from the sperm nuclei (Fig. 29).

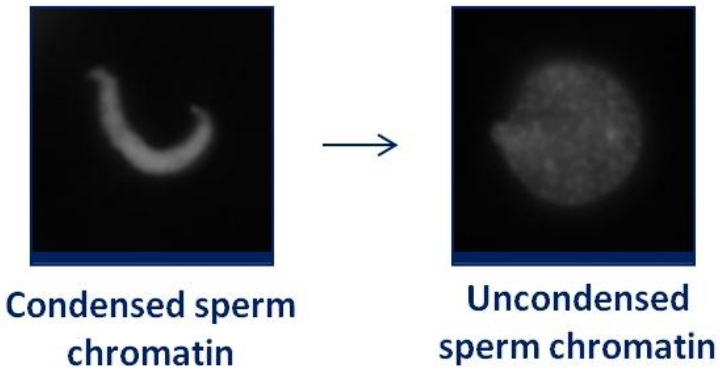

Figure 29 – Decondensation of a demembranated sperm nucleus in the *X. laevis* egg extract.

I was able to set up an immuno-depletion protocol of the *Xenopus* endogenous Cdc45 factor from the extracts by using specific polyclonal antibodies. When I carried out DNA replication assays I could observe a strong decrease in the DNA replication activity (followed by incorporation of radio-labeled nucleotides – *see Experimental Procedures*) if Cdc45 had been depleted from the extract, compared to the activity shown by a mock-treated extract (immuno-depletion carried out with a rabbit pre-immune serum). Moreover, when the recombinant *Xenopus* Cdc45 protein was added back to the specifically immuno-depleted extract a rescue of the DNA replication activity could be observed (Fig. 30).

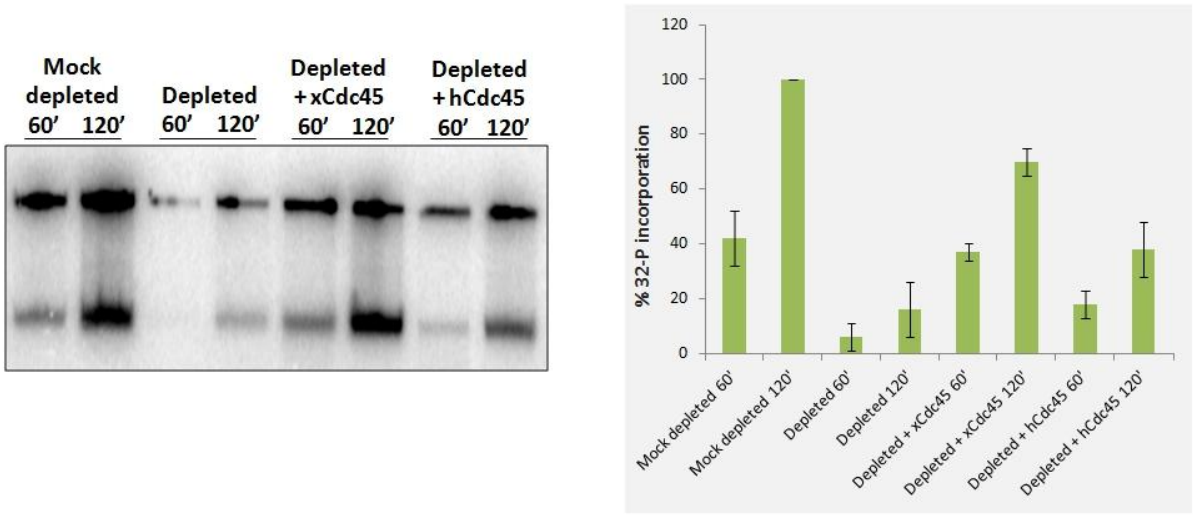

Figure 30 – DNA replication assays with frog egg extracts after the depletion of the endogenous Cdc45 factor. On the *left* a typical result obtained with phopshorimaging detection of the gels is shown. On the *right* a graphic shows a quantification of the results obtained in three different experiments.

Also a rescue in the DNA replication activity could be observed even when the human Cdc45 recombinant factor was used to supplement the extracts after the immuno-depletion of the endogenous frog factor. Even if the level of rescue is about the half respect to the one observed for the *Xenopus* Cdc45 factor, the human protein was shown to work in the *Xenopus* system to give DNA replication. This result is consistent with the finding that the primary structure of the frog and human Cdc45 share more than 80% identity.

Our bio-informatic analysis suggested that two long aminoacidic insertions were introduced during evolution into the eukaryotic Cdc45 protein (Fig. 18), one of which is shared with the archaeal $RecJ^{Cdc45}$ proteins. These two insertions likely form the two lateral extensions which characterize the structure predicted for the human Cdc45 protein (Fig. 26). In order to understand whether these two insertions play any role during DNA replication in unperturbed conditions, we produced three mutant derivatives of human Cdc45. In the mutant referred to as hCdc45_A1_LP1 about 70% of the first aminoacidic insertion (the one present only in the eukaryotic Cdc45 proteins) was deleted. In the mutant named hCdc45-LP1 this whole sequence was removed. In the derivative named hCdc45-LP2 the second aminoacidic insertion (the one present also in the archaeal counterparts) was fully deleted. Unfortunately the mutants hCdc45-LP1 and hCdc45-LP2, even if soluble, were found to form multimeric protein aggregates in solution as indicated by size exclusion chromatography analyses. Therefore, these Cdc45 mutant forms could not be used for the analysis with the frog egg extract DNA replication system.

Thus, only the hCdc45_A1_LP1 mutant protein could be employed to carry out DNA replication assays with the *Xenopus* egg extract system. Our results show that this mutant can rescue, as well as the *wild type* protein, the DNA replication activity after immuno-depletion of the endogenous factor from the extracts (Fig. 31).

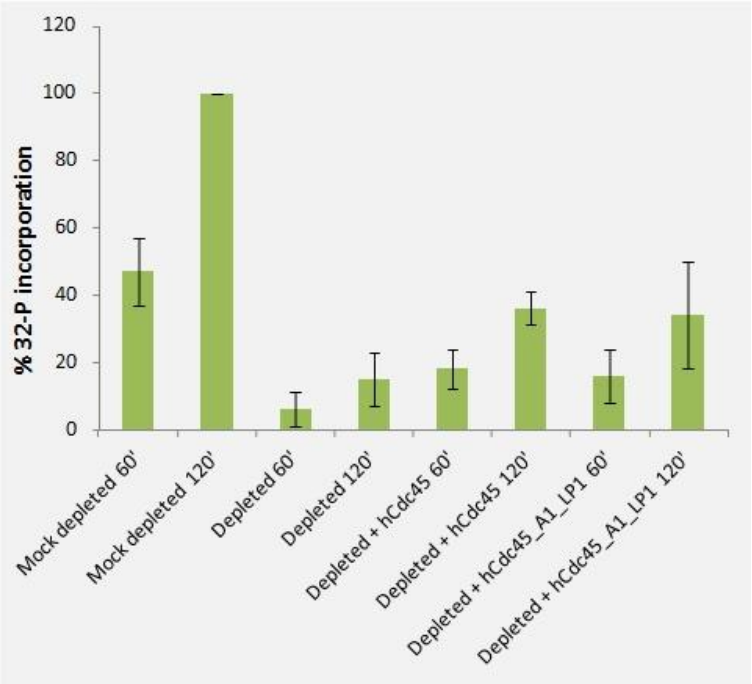

Figure 31 - DNA replication assays carried out with the hCdc45_A1_LP1 mutant. The ability of hCdc45_A1_LP1 to rescue the DNA replication activity after depletion of the frog Cdc45 factor was assessed and compared to the one of the hCdc45 *wild type* protein.

After the immuno-depletion of the endogenous factor from the extract the system is stressed and this explains why usually a good rescue in the DNA replication activity can reach at most the 70% of the activity shown by the mock-depleted extract. Moreover, even if the human protein is very similar to the frog counterpart of course there is a gap between the functionality of the two proteins in the *Xenopus* system. Nevertheless, a rescue in the DNA replication activity after the depletion of an endogenous replicative factor can only represent a specific result. As demonstrated since years this kind of experimental system can unlikely provide false positive results.

DISCUSSION

The analyses that have been carried out on the human Cdc45 replicative factor during this work represent a pioneering investigation about the biochemistry of this protein. The results of these biochemical and structural analyses suggest that Cdc45 is evolutionarily linked with the DHH proteins family and the hypothesis that Cdc45 is derived from an ancestral pyrophosphatase cannot be completely ruled out.

It has been long suggested that a pyrophosphatase activity may be associated to the DNA replication machinery [93]. Pyrophosphate hydrolysis could enhance the catalytic activity of replicative DNA polymerases, because removal of the pyrophosphate, a product of dNTPs incorporation, would drive forward the polymerization reaction. However, an important clue about the function and the evolutionary origin of Cdc45 comes from the finding that some of its putative archaeal homologues display a much closer similarity to the DHH subfamily 1 that includes the RecJ single-stranded DNA exonucleases. Moreover, the *ab initio* model obtained from SAXS data is more consistent with a compact RecJ core, rather than the more "open" conformation of inorganic pyrophosphatases.

A number of RecJ-like proteins have been found associated with the DNA replication machinery in a variety of archaeal organisms, such as *S. solfataricus* [82], *A. fulgidus* [80], *T. kodakaraensis* [81]. In particular, the *T. kodakaraensis* $RecJ^{Cdc45}$ homologue has been shown to possess a 5' to 3' exonuclease activity that is highly stimulated by physical interaction with the GINS15 subunit [81]. It has been proposed that this enzyme could participate in maturation of the Okazaki fragments in a pathway that is redundant with the FEN1- and Dna2-dependent pathways (two factors involved in this mechanism respectively in humans and yeast). Out of the multiple RecJ-like proteins present in archaeal organisms, it is plausible that some are associated with the DNA replication machinery and have a direct role in genome duplication; whereas others are "true" RecJ homologues, being involved in DNA repair/recombination transactions.

Based on our biochemical and bioinformatic analyses we suggest that Cdc45 originated from an ancestral 5' to 3' exonuclease loosing during evolution its catalytic activity and only retaining the ability to bind single-stranded DNA. The proteins from *Sulfolobales* may have followed a similar evolutionary path, as demonstrated by the absence of key catalytic residues and their tight association with GINS15. Association with the replication fork and GINS, in particular, may have led Cdc45 to loose its enzymatic activity and to acquire the function of "chaperoning" the lagging strand that is sterically excluded from the central channel of the MCM2–7 complex, as suggested from the cryo-electron microscopy of the CMG complex. In the CMG context Cdc45 (together with GINS) contributes to the formation of a tracking channel for the lagging strand [66].

Taking into account this model, the two lateral extensions showed in the hCdc45 predicted structure (Fig. 26) could be likely involved in contacting GINS and MCM2-7 on the two sides with the central core of the protein looking at the center of the gap formed by this association (Fig. 6). The finding that the hCdc45_A1_LP1 mutant, in which about 70% of the left arm of the protein has been deleted (Fig. 26), can still work to give *in vitro* DNA replication activity, could indicate that the contact surface

between Cdc45 and the other components of the CMG complex is very restricted, at least on one of the two sides. The single-stranded DNA binding activity seems to be related to the C-terminal portion of the protein which would be folded in the central core of the protein as can be hypothesized by the result got from the super-imposition of the *T. thermophilus* RecJ core crystal structure with the predicted structure for the human protein.

It would have been really useful to study the Cdc45 factor in association with GINS and MCM2-7 in the formation of a stable recombinant complex. So far, we are able to get the co-purification of all the 11 proteins forming the complex but the amount got is definitely low. About this, the inability to carry out large suspension cell culture was likely another important issue. Indeed if it is expected that the yield of the whole CMG complex is low, it is actually needed to start with a massive protein production to get amounts of reconstituted complex enough to carry out biochemichal analyses. Anyway, the basis for the production of the recombinant complex has been setted up and in the future the protocol can be optimized to get relevant amounts of the CMG complex. This would allow us to analyse and characterize the role played as an auxiliary factor of the MCM2-7 helicase using this reconstituted *in vitro* system.

Cdc45 has been shown to be a proliferation-associated antigen [67,68]. Its loading onto chromatin coincides with the origin firing and it is actually used by cells to switch on the DNA replication process by activating the replicative forks. In cancerous cells the normal mechanism by which the association of Cdc45 with the replicative forks is tightly regulated is out of control. These cells show high level of expression of the Cdc45 protein, which is actually needed to keep the DNA replication process on, while in normal cells, in most of the adult normal organs there is no expression at all for this factor [68]. Novel anti-cancer therapeutics can be hypothesized for the future, in which Cdc45 can be used as a molecular target to get a very specific effect only on tumoral cells. By inhibiting the association of Cdc45 with the other components of the CMG complex, or interfering with its DNA binding activity (if this is a requirement for the DNA replication activity) the DNA helicase function at the fork can be stopped and the DNA replication process can be switched off in cells that have escaped the normal control mechanisms.

REFERENCES

1. Dutta, A., Bell. S, P, (1997) Annu. Rev. Cell. Dev. Biol. 13:293-332.
2. Bell, S. P., Dutta, A. (2002) Annu. Rev. Cell. Dev. Biol. 71:333-374.
3. Machida, Y. J., Hamlin, J. L. , Dutta, A.(2005) Cell 123:13-24.
4. Blow, J. J., Dutta, A. (2005) Nat. Rev. Mol. Cell. Biol. 6:476-86.
5. Wohlschlegel, J. A., Dwyer, B. T., Dhar, S. K., Cvetic, C., Walter, J. C., Dutta, A., (2000) Science 290:2309-2312.
6. Nasheuer, H. P., Pospiech, H., Syvaoja, J., (2006) Genome Dynamics & Stability, Vol.1, D.H. Lankenau (ed.), Springer, p.27
7. Masai, H., Araki, K. I. (2002) J. Cell. Physiol. 190:287-296.
8. Sclafani, R. A. (2000) J. Cell. Sci. 113:2111-2117.
9. Jares, P., Donaldson, A., Blow, J. J. (2000) EMBO Rep. 1:319-322.
10. Doonan, J.H., Kitsios, G., (2009) Mol. Biotechnol. 42:14-29.
11. Moyer, S. E., Lewis, P. W., Botchan, M. R. (2006) Proc. Natl. Acad. Sci. USA. 103:10236-10241.
12. Ilves, I., Petojevic, T., Pasavento, J. J., Botchan, M. R. (2010) Mol. Cell 37:247-258.
13. Lehman, I.R., Kaguni, L.S., (1989) J. Biol. Chem. 15:4265-4268.
14. Braithwaite, D.K., Ito, J. (1993) Nucleic Acid Res. 21:787-802.
15. Gibson, S.I., Surosky, R.T., Tye, B.K. (1990) Mol. Cell. Biol. 10:5707-20.
16. Passmore, S., Maine, G.T., Elble, R., Christ, C., Tye, B.K. (1988) J. Mol. Biol. 204:593-606.
17. Forsburg, S.L. (2004) Microbiol. Mol. Biol. Rev. 68:109-31.
18. Madine, M.A., Khoo, C.Y., Mills, A.D., Laskey, R.A. (1995) Nature 375:421-4.
19. Tye, B.K. (1999) Annu. Rev. Biochem. 68:649-86.
20. Diffley, J.F., Labib, K. (2002) J. Cell. Sci. 115:869-72.
21. Sakwe, A.M., Nguyen, T., Athanasopoulos, V., Shire, K., Frappier, L. (2007) Mol. Cell. Biol. 27:3044-55.
22. Hubscher, U., Maga, G., Spadari, S. (2002) Annu. Rev. Biochem. 71:133-63.

23. Nasheuer, H.P., Smith, R., Bauerschmidt, C., Grosse, F., Weisshart, K. (2002) Prog. Nucleic Acid. Res. Mol. Biol. 72:41-94.
24. Schwacha, A., Bell, S.P. (2001) Mol. Cell. 8:1093-104.
25. Kang, Y.H., Galal, W.C., Farina, A., Tappin, I., Hurwitz, J. (2012) Proc. Natl. Acad. Sci U S A 109:6042-7.
26. Ishimi, Y. (1997) J. Biol. Chem. 272:24508-13.
27. Lee, J.K., Hurwitz, J. (2011) Prot. Natl. Acad. Sci. USA 98:54-9.
28. Labib, K., Tercero, J.A., Diffley, J.F. (2000) Science 288:1643-7.
29. Labib, K., Kearsey, S.E., Diffley, J.F. (2001) Mol. Cell. Biol. 12:3658-67.
30. Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A., Labib, K. (2005) Genes Dev. 19:1905-19.
31. Pacek, M., Tutter, A.V., Kubota, Y., Takisawa, H., Walter, J.C. (2006) Mol. Cell. 21:581-7.
32. Blow, J.J. (2001) Embo J. 20:3293-7.
33. Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., Labib, K. (2006) Nat. Cell. Biol. 8:358-66.
34. Moyer, S.E., Lewis, P.W., Botchan, M.R. (2006) Proc. Natl. Acad. Sci. USA 103:10236-41.
35. Moir, D., Stewart, S.E., Osmond, B.C., Botstein, D. (1982) Genetics 4:547-63.
36. Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J., Grunstein, M. (2002) Mol. Cell. 10:1223-33.
37. Bauershmidt, C., Pollok, S., Kremmer, E., Nasheuer, H.P., Grosse, F. (2007) Genes Cells 12:745-58.
38. Kneissl, M., Putter, V., Szalay, A.A., Grummt, F. (2003) J.Mol. Biol. 327:111-28.
39. Loebel, D., Huikeshoven, H., Cotterill, S. (2000) Nucleic Acid Res. 28:3897-903.
40. Zou, L., Stillman, B. (2000) Mol. Cell. Biol. 20:3795-806.
41. Katou, Y., Kanoh, Y., Bando, M., Noguchi, H., Tanaka, H., Ashikari, T., Sugimoto, K., Shirahige, K. (2003) Nature 424:1078-83.

42. Saha, P., Thome, K.C., Yamaguchi, R., Hou, Z., Weremowicz, S., Dutta, A. (1998) J. Biol. Chem. 273:18205-9.
43. Schmidt, U., Wollmann, Y., Franke, C., Grosse, F., Saluz, H.P., Hanel, F. (2007) Biochem. J. 409:169-77.
44. Wohlschlegel, J.A., Dhar, S.K., Prokhorova, T.A., Dutta, A., Walter, J.C. (2002) Mol. Cell. 9:233-40.
45. Zhu, W., Ukomadu, C., Jha, S., Senga, T., Dhar, S.K., Wohlschlegel, J.A., Nutt, L.K., Kornbluth, S., Dutta, A. (2007) Genes Dev. 21:2288-99.
46. Hartwell, L.H., Weinert, T.A. (1989) Science 246:629-34.
47. Weinert, T., Hartwell, L.H. (1989) J. Cell. Sci. Suppl. 12:145-8.
48. Kastan, M.B., Bartek, J. (2004) Nature 432:316-23.
49. Hartwell, L.H., Kastan, M.B. (1994) Science 266:1821-8.
50. Luciani, M.G., Oehlmann, M., Blow, J.J. (2004) J. Cell. Sci. 117:6019-30.
51. Liu, P., Barkley, L.R., Day, T., Bi, X., Slater, D.M., Alexandrow, M.G., Nasheuer, H.P., Vaziri, C. (2006) J. Biol. Chem. 281:30631-44.
52. Jazayeri, A., Falck, J., Lukas, C., Bartek, J., Smith, G.C., Lukas, J., Jackson, S.P. (2006) *Nat. Cell Biol*. 8:37-45.
53. Liu, Q., Guntuku, S., Cui, X.S., Matsuoka, S., Cortez, D., Tamai, K., *et al.* (2000) Genes Dev. 14:1448-1459.
54. Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, V., Ziv, Y., Vanagaite, L., *et al.* (1995) Science 268:1749-1753.
55. Guo, Z., Kumagai, A., Wang, S.X., Dunphy, W.G. (2000) Genes Dev. 14:2745-2756.
56. Hekmat-Nejad, M., You, Z., Yee, M., Newport, J.W., Cimprich, K.A. (2000) Curr. Biol. 10:1565-1573.
57. Falck, J. Petrini, J.H., Williams, B. R., Lukas, J., Bartek, J. (2002) Nature Genet. 30:290-294.
58. Wong., P.G., Winter., S.L., Zaika., E., Cao, T.V., Oguz, U., Koomen, J.M., Hamlin, J.L., Alexandrow, M.G. (2011) Plose One 6:e17533.

59. Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., Araki, H. (2003) Genes Dev. 17:1153-65.

60. Kubota, Y., Takase, Y., Komori, Y., Hashimoto, Y., Arata, T., Kamimura, Y., Araki, H., Takisawa, H. (2003) Genes Dev. 17:1141-52.

61. Makarova, K.S., Wolf, Y.I., Mekhedov, S.L., Mirkin, B.G., Koonin, E.V. (2005) Nucleic Acids Res. 33:4626-38.

62. Boskovic, J., Coloma, J., Aparicio, T., Zhou, M., Robisnon, C.V., Mendez, J., Montoya, G. (2007) EMBO Rep. 8:678-84.

63. Chang, Y.P., Wang, G., Bermudez, V., Hurwitz, J., Chen, X.S. (2007) Proc. Natl. Acad. Sci. USA 104:12685-90.

64. Choi, J.M., Lim, H.S., Kim, J.J., Song, O.K., Cho, Y. (2007) Genes Dev. 21:1316-21.

65. Kamada, K., Kubota, Y., Arata, T., Shindo, Y., Hanaoka, F. (2007) Nat. Struct. Mol. Biol. 14:388-96.

66. Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M.R., Berger, J.M. (2011) Nat. Struct. Mol. Biol. 18:471-7.0

67. Pollok1, S., Bauerschmidt, C., Sanger, J., Nasheuer, H.P., Grosse, F. (2007) FEBS J. 274:3669-84.

68. Tomita, Y., Imai, K., Senju, S., Irie, A., Inoue, M., Hayashida, Y., Shiraishi, K., Mori, T. *et al.* (2011) Cancer Sci. 102:697-705.

69. Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., Araki, H. (2001) Curr. Biol. 21:2055-63.

70. Krastanova, I., Sannino, V., Amenitsch, H., Gileadi, O., Pisani, F.M., Onesti, S. (2012) J. Biol. Chem. 287:4121-8.

71. Maine, G.T., Sinha, P., Tye, B.K. (1984) Genetics 106:365-85.

72. Slaymaker, I.M., Chen, X.S. (2012) Subcell. Biochem. 62:89-111.

73. Kamada, K. (2012) Subcell. Biochem. 62:135-56.

74. Aravind, L., Koonin, E.V. (1998) Trends Biochem. Sci. 23:17-9.

75. Yamagata, A., Kakuta, Y., Masui, R., Fukuyama, K. (2002) Proc. Natl. Acad. Sci. U.S.A. 99:5908-5912.

76. Wakamatsu, T., Kitamura, Y., Kotera, Y., Nakagawa, N., Kuramitsu, S., Masui, R. (2010) J. Biol. Chem. 285:9762-9769.
77. Fabrichniy, I. P., Lehtiö, L., Salminen, A., Zyryanov, A. B., Baykov, A. A., Lahti, R., Goldman, A. (2004) Biochemistry 43:14403-14411.
78. Edgar, R.C. (2004) Nucleic Acids Res. 32:1792-1797.
79. Rajman, L. A., Lovett, S. T. (2000) J. Bact. 182:607-612.
80. Motz, M., Kober, I., Girardot, C., Loeser, E., Bauer, U., Albers, M., Moeckel, G., Minch, E., Voss, H., Kilger, C., Koegl, M. (2002) J. Biol. Chem. 277: 16179-16188.
81. Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., Hurwitz J., Reeve, J. N., and Kelman, Z. (2011) Nucleic Acids Res. 39:6114-6123.
82. Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V., Bell, S. D. (2006) EMBO rep. 7:539-545.
83. Kelley, L. A., Sternberg, M. J. (2009) Nature Protocols 4:363-371.
84. Söding, J., Biegert, A., Lupas, A. N. (2005) Nucleic Acids Res. 33:W244-W248.
85. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., Godzik, A. (2005) Nucleic Acids Res. 33:W284-288.
86. Shi, J., Blundell, T. L., Mizuguchi, K. (2001) J. Mol. Biol. 310:243-257.
87. Sanchez-Pulido, L., Ponting, C. P. (2011) Bioinformatics 27:1885-1888.
88. Savitsky, P., Bray, J., Cooper, C. D., Marsden, B. D., Mahajan, P., Burgess-Brown, N. A., Gileadi, O. (2010) J. Struct. Biol. 172:3-13.
89. Svergun, D. I. (1999) Biophys. J. 76:2879–2886.
90. Volkov, V. V., Svergun, D. I. (2003) J. Appl. Crystallogr. 36:860–864.
91. Blow, J.J., Laskey, R.A. (1986) Cell. 47:577-87.
92. Park, S.Y., Lee, B., Park, K.S., Chong, Y., Yoon, M.Y., Jeon, S.J., Kim, D.E. (2010) Appl. Microbiol. Biotechnol. 85:807-12.
93. Aravind, L., Koonin, E. V. (1998) Nucleic Acids Res. 26:3746–3752.

LIST OF SCIENTIFIC PUBLICATIONS

1. Krastanova I.*, Sannino V.*, Amenitsch H., Gileadi O., Pisani F.M.#, Onesti S.#. Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. J Biol Chem 2012; 287:4121-4128. (*Joint First Authors; #Corresponding Authors).

2. Pennacchio A., Sannino V., Sorrentino G., Rossi M., Raia C.A., Esposito L. Biochemical and structural characterization of recombinant short-chain NAD(H)-dependent dehydrogenase/reductase from Sulfolobus acidocaldarius highly enantioselective on diaryl diketone benzil. Appl Microbiol Biotechnol 2012 (DOI 10.1007/s00253-012-4273-z).

3. Di Perna R., Aria V., De Falco M., Sannino V., Okorokov A., Pisani F.M., De Felice M. The physical interaction of Mcm10 with Cdc45 modulates their DNA binding properties. Biochem. J. 2013, *under revision*.

MEETINGS COMMUNICATIONS

1. De Falco M., De Felice M., Aria V., Sannino V., Rossi M., Pisani F.M.. Identification and biochemical characterization of interactors of the Mini-Chromosome Maintenance (MCM) complex, the replicative DNA helicase at the human replication fork. Convegno del Dipartimento Scienze della Vita, 11th – 12th October 2010, C.N.R, Roma.

2. Aria V., De Felice M., De Falco M., Sannino V., You Z., Masai H., Pisani F.M. Interaction of the human Tim1/Tipin complex with MCM2-7. "1st workshop R3: DNA Replication,Recombination and Repair", 30th June – 2nd July 2010, Conservatorio Santa Chiara, San Miniato, Pisa (IT).

3. Krastanova I., Sannino V., Amenitsch H., Gileadi O., Pisani F.M., Onesti S. Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. "Eukaryotic DNA replication & genome maintenance", 6th -10th September 2011, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY).

4. Aria V., De Felice M., Sannino V., De Falco M., Hubscher U., Syvaoja J., You Z., Masai H., Pisani F.M. Physical and functional interaction of the human Timeless and Tipin proteins with DNA polymerase ε and the MCM complex. "Eukaryotic DNA replication & genome maintenance", 6th -10th September 2011, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY)

5. Sannino V., Krastanova I., Onesti S., Costanzo V., Rossi M., Pisani F.M. Structural and functional insights into the DNA replication factor Cdc45. "PhD Student Conference – MRC Laboratory for Molecular Cell Biology & Institute of Protein Biochemistry", 4th -6th November 2012, Vico Equense, Napoli (IT).

FOREIGN EXPERIENCE

3 months stage (March to May 2012) at the Cancer Research UK, London Research Institute – Clare Hall Site in the laboratory of Dr. Vincenzo Costanzo.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 6, pp. 4121–4128, February 3, 2012

Structural and Functional Insights into the DNA Replication Factor Cdc45 Reveal an Evolutionary Relationship to the DHH Family of Phosphoesterases*[S]

Received for publication, July 24, 2011, and in revised form, November 18, 2011 Published, JBC Papers in Press, December 6, 2011, DOI 10.1074/jbc.M111.285395

Ivet Krastanova[‡1], **Vincenzo Sannino**[§1], **Heinz Amenitsch**[¶], **Opher Gileadi**[‖], **Francesca M. Pisani**[§2], **and Silvia Onesti**[‡3]

From the [‡]*Structural Biology Laboratory, Sincrotrone Trieste, Trieste 34149, Italy, the* [§]*Istituto di Biochimica delle Proteine, Consiglio Nazionale delle Ricerche, Napoli 80131, Italy, the* [¶]*Institut of Biophysics and Nanosystems Research, Austrian Academy of Sciences, Schmiedlstrasse 6, Graz 8042, Austria, and the* [‖]*Structural Genomics Consortium, Oxford OX3 7DQ, United Kingdom*

Background: Although Cdc45 is a key replication factor, there are no biochemical or structural studies on the isolated protein.
Results: We report the first purification and biochemical characterization of human Cdc45, as well as the first structural data on the isolated Cdc45 by small angle x-ray scattering.
Conclusion: Cdc45 is related to the RecJ/DHH family of phosphoesterases and binds single-stranded DNA.
Significance: The similarity has important evolutionary implications.

Cdc45 is an essential protein conserved in all eukaryotes and is involved both in the initiation of DNA replication and the progression of the replication fork. With GINS, Cdc45 is an essential cofactor of the Mcm2–7 replicative helicase complex. Despite its importance, no detailed information is available on either the structure or the biochemistry of the protein. Intriguingly, whereas homologues of both GINS and Mcm proteins have been described in Archaea, no counterpart for Cdc45 is known. Herein we report a bioinformatic analysis that shows a weak but significant relationship among eukaryotic Cdc45 proteins and a large family of phosphoesterases that has been described as the DHH family, including inorganic pyrophosphatases and RecJ ssDNA exonucleases. These enzymes catalyze the hydrolysis of phosphodiester bonds via a mechanism involving two Mn^{2+} ions. Only a subset of the amino acids that coordinates Mn^{2+} is conserved in Cdc45. We report biochemical and structural data on the recombinant human Cdc45 protein, consistent with the proposed DHH family affiliation. Like the RecJ exonucleases, the human Cdc45 protein is able to bind single-stranded, but not double-stranded DNA. Small angle x-ray scattering data are consistent with a model compatible with the crystallographic structure of the RecJ/DHH family members.

Cdc45 is an essential factor required for the establishment (1–3) and progression (4–7) of the DNA replication fork in eukaryotic cells. As many other DNA replication factors, Cdc45 is more abundant in proliferating cells, whereas it is almost absent from long term quiescent, terminally differentiated, and senescent cells (8). Several genetic studies, two-hybrid screens, and co-immunoprecipitation analyses revealed that Cdc45 interacts with a number of other replication factors, including the Mcm2–7 complex, GINS, MCM10, replication protein A, DNA polymerases α, δ, and ϵ, the origin recognition complex subunit 2, and TopBP1 (for a review see Ref. 9). In a variety of eukaryotic organisms Cdc45 has been found to stably associate with Mcm2–7 and GINS to form a complex (the CMG),[4] which is believed to act as the DNA helicase at the replication fork (7, 10–13). This hypothesis has been reinforced by the demonstration that the *Drosophila melanogaster* CMG complex can be reconstituted by co-producing its protein components in baculovirus-infected insect cells and is found to possess a robust DNA-unwinding activity *in vitro*, whereas the Mcm2–7 complex alone is almost completely unable to unwind duplex DNA (14). This analysis suggests that Cdc45 and GINS are helicase auxiliary factors whose association with the Mcm hetero-hexameric ring is absolutely required to reconstitute an active complex.

The critical biological function played by Cdc45 is underscored by its ubiquitous distribution and high degree of sequence conservation from yeast to man. Nevertheless, not much is known on the exact role of Cdc45 either in the pre-initiation complex or the CMG helicase, due to the lack of biochemical studies on the isolated protein. The analysis of the primary structure of Cdc45 has failed to reveal the presence of any significant similarity to known protein family or any characteristic sequence motif.

Due to the complexity of the eukaryotic DNA replication machinery, most of the information on the structure and biochemistry of replication proteins has been inferred from the study of the simpler archaeal system. In fact, Archaea possess a simplified version of the Mcm (15, 16) and GINS (17–19) complexes, but so far no archaeal homologue of Cdc45 has been identified.

* The work was supported by a Friuli Venezia Giulia grant (DAGEAST to I. K.) and by the Associazione Italiana per la Ricerca sul Cancro (Grant IG10646 to S. O. and Grant IG9087 to F. M. P.).
[S] This article contains supplemental Figs. S1–S3 and Table S1.
[1] Both authors contributed equally to this work.
[2] To whom correspondence may be addressed. Tel.: 39-081-613-2292; Fax: 39-081-613-2277; E-mail: fm.pisani@ibp.cnr.it.
[3] To whom correspondence may be addressed. Tel.: 39-040-375-8451; Fax: 30-40-938-0902; E-mail: silvia.onesti@elettra.trieste.it.

[4] The abbreviations used are: CMG, Cdc45/Mcm2–7/GINS complex; TEV, tobacco etch virus; SAXS, small angle x-ray scattering.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The nucleotide sequence encoding the human Cdc45 (hCdc45) was amplified using the Platinum Pfx DNA polymerase (Invitrogen), using a cDNA clone from the mammalian Gene Collection (IMAGE ID 2964592) as a template, and the following primers: 5′-TTAAGAAGGAGATATACTATGTTCGTGTCCGATTTCCGCAAAG-3′ and 5′-GATTGGAAGTAGAGGTTCTCTGCGGACAGGAGGGAAATAAGTGCG-3′.

The amplified fragment was then Ligation independent-cloned into the pNIC-CTHF vector (Structural Genomics Consortium, Oxford (20)) for protein expression as a fusion with a C-terminal His_6-FLAG tag cleavable with tobacco etch virus (TEV) protease.

Expression of hCdc45 was carried out in *Escherichia coli* BL21 (DE3)-R3-Rosetta cells (SGC, Oxford) grown overnight at 25 °C in TB broth, following 0.1 mM isopropyl 1-thio-β-D-galactopyranoside induction. A frozen cell pellet (corresponding to 1 liter of culture) was resuspended in 100 ml of buffer A (50 mM sodium phosphate buffer, pH 7.4, 1 M NaCl, 10% (v/v) glycerol, 2 mM β-mercaptoethanol) containing EDTA-free protease inhibitor mixture (Roche Applied Science), 50 units/ml benzonase nuclease (Novagen), and 1 mg/ml lysozyme (Sigma-Aldrich) and incubated for 2 h at 4 °C. Cells were further disrupted by sonication (Bandelin Sonopuls HD3200 ultrasonic homogenizer). The insoluble material was removed by centrifugation (17,000 rpm for 1 h at 4 °C, Beckman Allegra 64R), and the supernatant was loaded onto a 5-ml HiTrap chelating column (GE Healthcare) previously charged with Ni^{2+} ($NiSO_4$) and equilibrated with buffer A. An imidazole step gradient (25 mM, 250 mM, and 500 mM) was applied, and protein elution was achieved using buffer A plus 250 mM imidazole. Fractions containing the hCdc45 were pooled and subsequently loaded onto a Superdex 200 16/60 size-exclusion column (GE Healthcare) equilibrated with buffer B (50 mM sodium phosphate, pH 7.0, 150 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol). The eluted protein was incubated with TEV protease for overnight tag cleavage at 4 °C. 12% SDS-PAGE was carried out to check that the cleavage was complete. The hCdc45-TEV mix was loaded onto a 5-ml Heparin column (GE Healthcare), previously equilibrated with buffer B. The cleaved hCdc45 was eluted from the column applying a linear (150 mM to 1 M) NaCl gradient. Finally the protein was subjected to size-exclusion chromatography in buffer D (20 mM Tris-HCl, pH 7.9, 150 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol).

EMSAs—The synthetic oligonucleotide used as single-stranded DNA had the following sequence: 5′-TCTACCTGGACGACCGGGTATATAGGGCCCTATATATAGGGCCAGCAGGTCCATCA-3′. A complementary synthetic oligonucleotide used to prepare the blunt DNA duplex had the following sequence: 5′-TGATGGACCTGCTGGCCCTATATATAGGGCCCTATATACCCGGTCGTCCAGGTAGA-3′. The first oligonucleotide was labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP and annealed to a 2-fold molar excess of the cold complementary strand to prepare the double-stranded DNA ligand. For the DNA mobility shift assays, 10-μl mixtures were prepared that contained 100 fmol of ^{32}P-labeled DNA in 20 mM Tris-HCl, pH 7.5, and the indicated amounts of protein (0.5–5 μg). Following incubation for 20 min at 27 °C, complexes were separated by electrophoresis through 5% polyacrylamide/bis gels (19:1) in 0.5× TBE (1× TBE (89 mM Tris Base, 89 mM Boric Acid, 2 mM EDTA, pH 8.3)). Gels were dried down and analyzed by phosphorimaging. Experiments were performed in triplicate, and the results were averaged. The *error bars* on the *graphs* are the standard error.

Electrophoretic mobility shift assays (EMSAs) were carried out after incubation of hCdc45-His-FLAG-DNA complexes with anti-FLAG monoclonal antibodies (Abcam). For these experiments 10-μl mixtures were prepared that contained 50 fmol of ^{32}P-labeled oligonucleotide in 20 mM Tris-HCl, pH 7.5, and 5 μg of hCdc45-His-FLAG. Following incubation for 20 min at 27 °C, anti-FLAG antibody was added (0.5, 1, and 2 μg in 2 μl of the following buffer: 10 mM sodium phosphate, pH 7.4, 150 mM NaCl, 50% glycerol; an equal volume of buffer (2 μl) was added into the samples where the antibody was omitted, and the incubation was continued for additional 30 min. The mixtures were subjected to electrophoresis, as previously described, and the gels were analyzed by phosphorimaging.

Analytical Gel Filtration and DNA-binding Activity of hCdc45—An aliquot of purified hCdc45 (1.9 mg) was loaded onto an analytical gel-filtration column (Bio-Sil SEC-250, Bio-Rad). The column was developed with 50 mM Tris-HCl. pH 8.0, 150 mM NaCl, 2 mM β-mercaptoethanol, 5% (v/v) glycerol at a flow rate of 1.0 ml/min. Fractions (200 μl) were collected, and aliquots (5 μl) were run through an 8% (w/v) polyacrylamide-SDS gel. The gel-filtration column was calibrated using the following markers: tyroglobulin, bovine γ-globulin, chicken ovalbumin, equin myoglobin, and vitamin B-12. The DNA-binding activity of each peak fraction (aliquots of 1 μl) was analyzed by EMSAs using radiolabeled ssDNA as a probe, as previously described.

SAXS—The synchrotron scattering data were collected at the Austrian small angle x-ray scattering (SAXS) beamline of the electron storage ring ELETTRA (21) at a wavelength λ of 1.54 Å. A Pilatus 100K (Tectris, Baden, Switzerland) was used as a detector, and a sample distance of 0.75 m was set to resolve the momentum transfer, q ($q = 4\pi \sin(\theta)/\lambda$, with θ as half scattering angle), in the range from 0.17 to 4.5 nm^{-1}. Samples of recombinant hCdc45 at a concentration of 0.44, 0.90 mg/ml, and 1.85 mg/ml in buffer D were used. All samples were kept and measured in a 1.5-mm glass capillary (Glass, Berlin, Germany) at 8 °C. The three 30-s exposures have been averaged, because the comparison of the first and last pattern did not show any effect of radiation damage.

The primary data reduction was performed using IGOR Pro (Wavemetrics, Lake Oswego, OR), which was also used to estimate the molecular mass of the solutes by comparing them to the scattering of water. The reduced data were treated further with the ATSAS program collection (22). The indirect Fourier transformation was calculated by using the program GNOM (23), which determines the distance distribution function, the radius of gyration R_g, the forward scattering $I(0)$, as well as the maximal dimension D_{max} of the proteins. The results are summarized in supplemental Table S1, in which additionally the

molecular mass of each sample has been calculated. As the sample c (1.85 mg/ml) showed some evidence of aggregation, as seen by the SAXS pattern change at $q < 0.4$ nm^{-1} and the difference in the relative molecular mass, the low q-data of sample b (0.9 mg/ml) has been merged with the high q-data of the sample c (1.85 mg/ml; supplemental Fig. S1). The *ab initio* shape of the protein was reconstructed using DAMMIF (24) with the combined scattering curve *bc*. The initial volume was a sphere with a D_{max} of 12.5 nm consisting of 5594 individual spheres (diameter 0.32 nm). The best 44 models have been averaged to obtain the final model with the program DAMAVER (25).

RESULTS AND DISCUSSION

Cdc45 Shows Sequence Similarity to Archaeal Proteins Belonging to the DHH Family of Phosphoesterases—We carried out a bioinformatic analysis on Cdc45. Database searches with the Position-Specific Interactive BLAST algorithm, using the human Cdc45 sequence as search model and default parameters, identified weak but significant similarity to two archaeal sequences on the second iteration run. The proteins were annotated as "phosphoesterase domain-containing protein" from *Candidatus korarchaeum cryptofilum OPF8* (E-value: 2×10^{-4}) and "putative single-stranded DNA-specific exonuclease RecJ" from *Methanocella paludicola SANAE* (E-value: 3×10^{-4}). They both belong to the DHH family of phosphoesterases that was first described by Aravind and Koonin (26). Both sequences have been recently added to the databases, explaining why the similarity had not been previously identified. Although the sequence similarity detected by BLAST involved only the first 130–140 residues, multiple alignments between Cdc45 and putative archaeal orthologues show that some similarity can be noticed throughout the entire sequence (Fig. 1).

The DHH family is defined by a number of conserved sequence motifs and can be split into two distinct clusters: subfamily 1 (whose prototype is the *Escherichia coli* RecJ protein), is so far represented only in Archaea and Bacteria; subfamily 2 (which includes the *Drosophila melanogaster* Prune protein and the *Saccharomyces cerevisiae* exopolyphosphatase PPX1), is also found in Eukarya. Crystallographic analysis of some DHH proteins (27–29) has revealed the presence of a catalytic core formed by two domains: in the N-terminal domain invariant residues (aspartic acid and histidine residues) coordinate two metals ions (typically manganese ions), suggesting a two-metal mediated hydrolysis reaction.

Because Archaea possess multiple RecJ-like proteins, we have selected the sequences displaying the closest match to the *Candidatus* and *Methanocella* homologues. To distinguish the archaeal RecJ/DHH proteins that may be putative Cdc45 orthologues from other members of the DHH family, we will use the notation RecJ^{Cdc45}. However, most Archaea seem to possess two close paralogues of the putative Cdc45 counterpart. A study, focused on detecting the archaeal equivalent of *E. coli* RecJ, identifies two highly similar proteins from *Methanocaldococcus jannaschii* (MJ0977 and MJ0831). Although both were able to partially complement a *RecJ* mutant phenotype in *E. coli*, ssDNA nuclease activity could only be observed for the MJ0977 protein (30). Two highly similar RecJ^{Cdc45} proteins (AF0699 and AF0735) were also found associated with a replication protein network in *Archaeoglobus fulgidus* (31). More recently, a RecJ-like single-stranded 5′-3′ DNA exonuclease from *Thermococcus kodakaraensis* was found to physically associate with GINS, and this association was found to stimulate its exonuclease activity (32).

A putative RecJ homologue has been reported to co-purify with Mcm and GINS from cellular extracts of *Sulfolobus solfataricus* (SSO0295 (18)). This sequence is rather divergent from most of the archaeal proteins mentioned above and lacks some of the motifs; in particular, some of the conserved residues putatively responsible for the nuclease activity are absent. An open reading frame highly similar to SSO0295 is present in most *Sulfolobales* genomes. Divergent sequences are also detected in other Crenarchaeota, such as *Aeropyrum pernix*.

Both the eukaryotic Cdc45 proteins and the archaeal RecJ^{Cdc45} sequences match only to the RecJ catalytic core (domains I and II, as defined in the crystal structure of *Thermus thermophilus* RecJ (27), while lacking the 50-residues N-terminal extension as well as domains III and IV. These additional elements form a closed-ring structure that is predicted to encircle ssDNA.

Cdc45 sequences only partially retain the motifs typical of the DHH family that are involved in metal binding. Whereas in motif 1 both aspartate residues coordinating the metal are conserved (^{26}**D**V**D**), the DHH motif 3, which gives the name to the family, is mutated to ^{99}**D**T**H**. Although a number of residues belonging to motifs 2 and 4 are conserved suggesting that the protein fold is similar in those regions, the key residues for metal binding and catalysis are not conserved, with the aspartate of motif 2 mutated to asparagine, and the aspartate of motif 4 to a glutamine (Fig. 1). In contrast, most of the archaeal RecJ^{Cdc45} comprise all of the motifs that are conserved in the bacterial RecJ exonucleases, with the exception of the putative orthologues from *Sulfolobales*, whose sequence is rather divergent and lacks most of the RecJ canonical residues.

Threading Algorithms Confirm the Presence of a RecJ/DHH-like Core Fold in hCdc45—A variety of threading/fold recognition algorithms were used to verify the similarity among hCdc45 and DHH family members. The Protein Fold Recognition Server Phyre (33) identifies a three-dimensional similarity between hCdc45 and a number of DHH protein structures such as the RecJ exonuclease from *T. thermophilus* (PDB code: 1IR6), as well as various inorganic pyrophosphatases (PDB codes: 1WPN, 1I74, and 2HAW), but the similarity was restricted to the first 110 amino acids. When a multiple sequence alignment, including a number of Cdc45 sequences, was used as input for the HHPRED server (34), a similarity (involving the first 315 residues) was detected to an inorganic pyrophosphatase from *Bacillus subtilis* belonging to the DHH family (PDB code: 1WPN). The profile-profile alignment and fold-recognition server FFAS03 (35) was also used both with the full-length hCdc45 sequence and a number of fragments corresponding to putative domains. A score below the recommended threshold (−9) was obtained when the first 140 amino acid residues of hCdc45 were used as input, matching the N-terminal domain of a number of manganese-dependent inorganic pyrophosphatases from the DHH family, with the *B. subtilis* inorganic pyro-

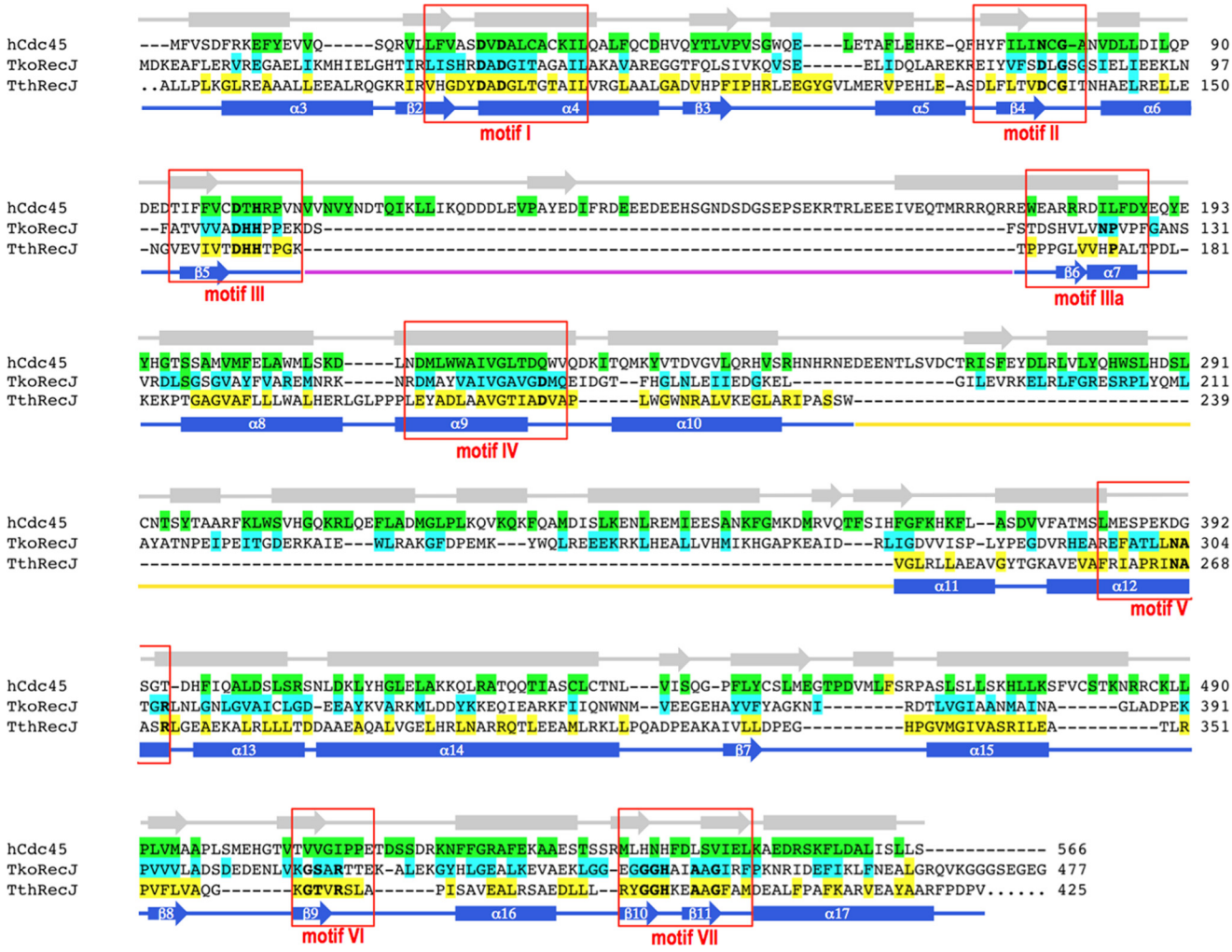

FIGURE 1. **Sequence alignment between hCdc45, RecJ^{Cdc45} from *T. kodakaraensis*, and RecJ from *T. thermophilus*.** The alignment presented here is based on an extended multiple alignment using 15 eukaryotic Cdc45 sequences, 16 archaeal sequences, and 10 bacterial RecJ sequences, selected from evolutionary diverse organisms. Only the RecJ core (residues 50–425, comprising domains I and II) has been used in the alignment. Residues that are conserved in more than 70% of the eukaryotic, archaeal, and bacterial sequences are highlighted in *green*, *cyan*, and *yellow*, respectively. The following groups of amino acid residues were considered similar: Asp/Glu, Lys/Arg, Phe/Tyr, Ser/Thr, Gly/Ala, and Val/Leu/Ile/Met. The position of the secondary structural elements in the *T. thermophilus* RecJ crystal structure (PDB code: 2ZXP) is indicated at the *bottom*, whereas the predicted secondary structure for human Cdc45 is shown at the *top*. Secondary structure elements are named according to the *Tth*RecJ nomenclature (28). The position of the characteristic RecJ motifs is shown by *red boxes*, with the residues conserved highlighted in *bold*. The alignment was carried out using the multiple sequence alignment program MUSCLE (41) and manually modified to take into account the structural constraints, and the results of the -fold recognition/threading algorithms. Up to motif IV the similarity is strong enough to be detected based on sequence alone, whereas the second half of the alignment relies on the threading data, which identify similarity patterns in the absence of high sequence homology, as exemplified by the conservation of the patterns of hydrophobic residues and the excellent match between RecJ secondary structure elements and the prediction for Cdc45. An insertion unique to eukaryotic Cdc45 orthologues and containing many charged amino acid residues is shown in *magenta*. The putative helical insertion present in both archaeal and eukaryotic proteins is shown in *yellow*.

phosphatase giving the best agreement (−10.5, PDB code: 1K23). Using as input longer hCdc45 fragments still provides a match with the DHH proteins, although the scores get progressively higher, indicating a lower degree of confidence; using the full-length protein it is possible to detect some similarity to 1K23. The sequence-structure homology recognition server FUGUE (36) unambiguously identifies both the *T. thermophilus* RecJ (1IR6) as well as the *Streptococcus gordonii* inorganic pyrophosphatase (1K20) structures as the two best hits along the entire sequence with a degree of confidence higher than 99%. When using as input the sequences of a variety of archaeal RecJ^{Cdc45} proteins, all the servers predicted a strong structural similarity to the bacterial RecJ proteins, along the entire sequence, as expected from the significant sequence homology and the conservation of the characteristic motifs.

A central region in both Cdc45 (residues 252–363 in hCdc45) and archaeal RecJ^{Cdc45} (residues 188–275 in the sequence from *T. kodakaraensis*) proteins, appears as a long insertion into the RecJ/DHH core. This region is reasonably well conserved between eukaryotic and archaeal proteins. Secondary structure predictions suggest a helical fold, and threading algorithms tend to find matches with helical bundle proteins, such as acyl-CoA-binding proteins and helix-turn-helix transcription factors. An additional insertion is unique to the eukaryotic Cdc45 proteins (residues 108–175 in the human sequence), and the central part includes a large number of charged residues (aspartate, glutamate, arginine, and lysine), thus suggesting a partially structured region (Fig. 1).

Based on the results of both the sequence analysis and the threading algorithms, we have produced a sequence alignment

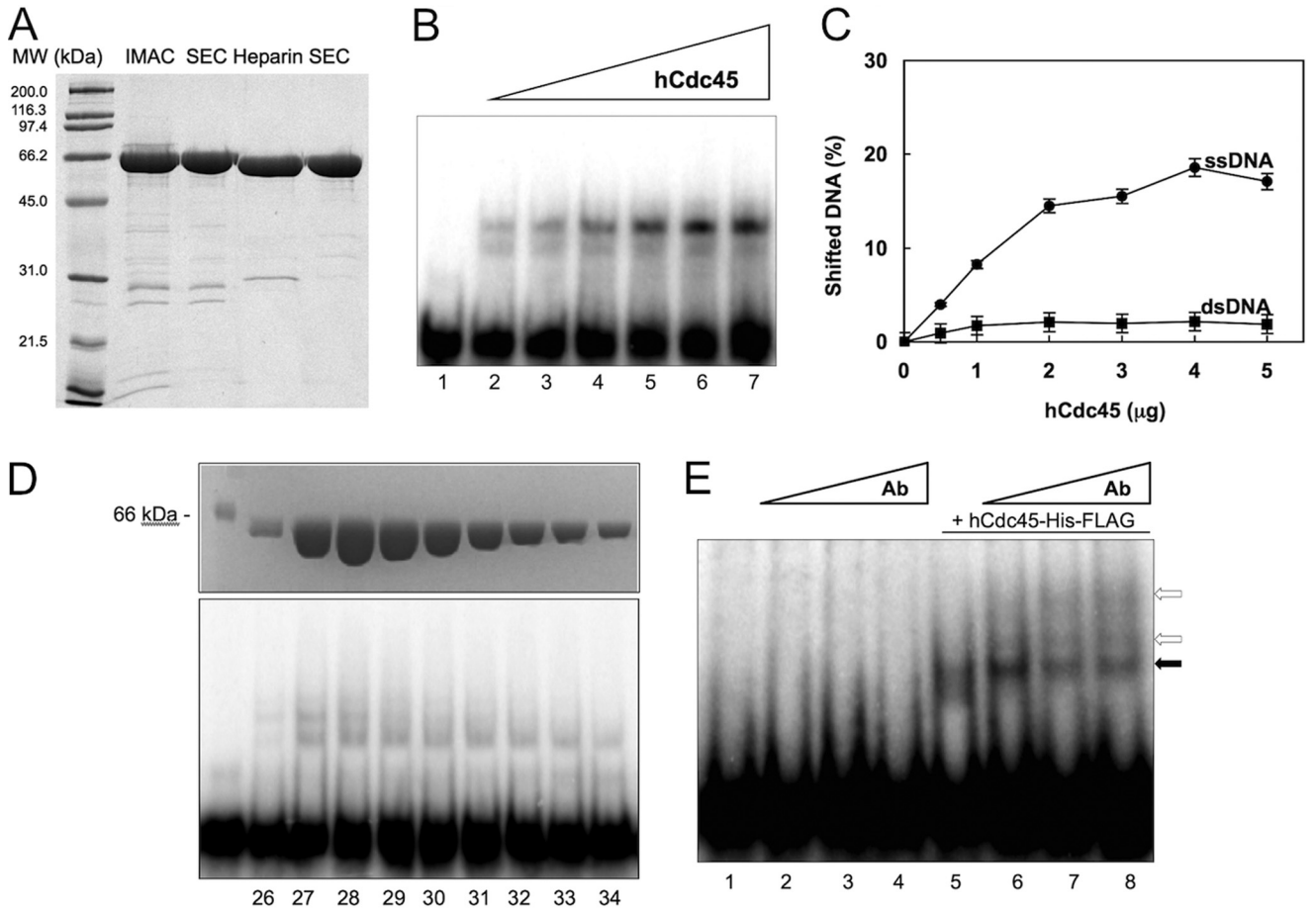

FIGURE 2. **Biochemical characterization of recombinant human Cdc45.** *A*, purification of hCdc45. SDS-PAGE analysis of samples throughout the purification protocol (see detailed description under "Experimental Procedures"), starting from the protein obtained after the first step of Ni-affinity purification (IMAC (immobilized metal-ion affinity chromatography)), followed by size-exclusion chromatography (*SEC*); the protein after cleavage of the His_6-FLAG tag using TEV protease and purification over a heparin column (*Heparin*) to a final round of SEC. *B*, DNA-binding activity of hCdc45. Example of an EMSA on single-stranded DNA is shown. The assay was carried out with increasing concentrations of hCdc45 (0.5, 1, 2, 3, 4, and 5 μg of protein were present in the mixtures loaded into the *lanes* from *2* to *7*). A radiolabeled 56-mer DNA was used as a ligand (see the text for details). A control mixture without protein was run on *lane 1*. *C*, single-stranded *versus* double-stranded DNA binding. Shifted DNA (either in single- (●) or double-stranded (■) form) is reported *versus* the amount of hCdc45. Experiments were performed in triplicate, and the results are averaged. *Curves* represent best fits to the data points. The *error bars* on the *graphs* are the ±S.E. *D*, gel-filtration analysis of hCdc45 and EMSAs of the corresponding peak fractions. Gel-filtration chromatography of purified hCdc45 was performed using a Bio-Sil SEC-250 column (Bio-Rad) as described under "Experimental Procedures." Peak fractions were analyzed by SDS-PAGE (5 μl/fraction) and used in a gel shift experiment (1 μl/fraction). *E*, EMSA with hCdc45-His-FLAG in the presence of anti-FLAG antibody. The assays were carried out by adding increasing amounts of a monoclonal anti-FLAG antibody (0.5, 1, and 2 μg, *lanes 2* and *6*, *3* and *7*, and *4* and *8*, respectively) into mixtures containing the single-stranded DNA probe with hCdc45-His-FLAG (*lanes 6–8*; 5 μg of protein) or without the recombinant protein (*lanes 2–4*; see text for details). A *black arrow* indicates the Cdc45-DNA complex, whereas the *white arrows* identify the ternary complexes with the anti-FLAG antibody.

that summarizes the putative relationships among bacterial RecJ ssDNA exonucleases, archaeal $RecJ^{Cdc45}$ and the eukaryotic Cdc45 (Fig. 1). In the first half of the molecule (up to motif IV) the similarity is strong enough to be detected based on sequence alone, whereas for the second half the alignment relies on the threading data, which identify similarity patterns in the absence of high sequence conservation. Consistent with the threading results, Fig. 1 shows the conservation of patterns of hydrophobic residues and an excellent match between the experimentally determined α-helices and β-strands of RecJ and the predicted Cdc45 secondary structure elements.

Although the correspondence is good overall, there are a few uncertainties in the central region of the proteins. For example it is difficult to establish unambiguously whether the first Cdc45 insertion occurs after β5 (as depicted in Fig. 1) or after α7. In the same way the second insertion (common to eukaryotic and archaeal proteins) may slide from the current position in Fig. 1 (after α10) to an alternative position after α11. However the match is very convincing in the first half of domain I as well as in domain II, including the long helix connecting the two domains (α14), suggesting that the relationship extends throughout the entire sequence.

While we were finalizing this report, a bioinformatic analysis (37) was published suggesting a similarity between Cdc45 and RecJ proteins, limited to the first 100 amino acid residues. Our report is consistent with that one (37) but further extends the analysis by showing that the sequence and structural similarity covers the entire length of the protein.

Biochemical Characterization of the Recombinant hCdc45—We produced in bacterial cells hCdc45 as a fusion protein with a C-terminal His_6-FLAG tag, using the pNIC-CTHF vector (20). The tag could be cleaved with the TEV protease and the protein purified to homogeneity (Fig. 2*A*).

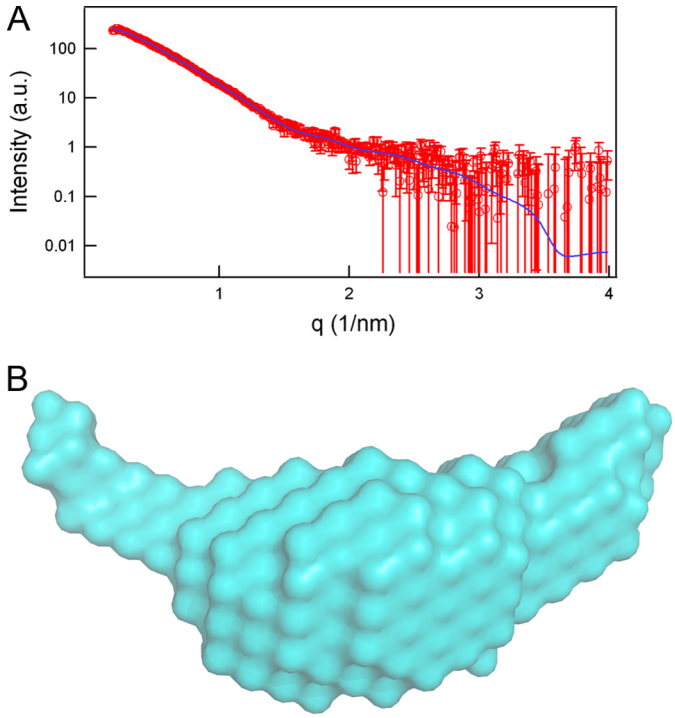

FIGURE 3. **Small angle x-ray scattering data.** *A*, the experimental SAXS profile (log intensity as a function of the momentum transfer) of hCdc45 (*red points*) is compared with the theoretical scattering curves calculated from the *ab initio* model (*blue line*). *B*, final model reconstructed from the scattering curve.

Although many of the residues that in RecJ/DHH proteins are involved in Mn^{2+} binding and catalysis are not conserved in hCdc45, we examined the possibility that the few remaining aspartate and histidine residues (namely Asp-26, Asp-28, Asp-99, and His-101) may possibly coordinate one metal ion. We have therefore used both inductively coupled plasma/atomic emission spectroscopy and atomic absorption spectroscopy to test whether the purified protein contains manganese, magnesium, or zinc, but we failed to confirm the presence of any of these metals (data not shown).

We also carried out activity assays to check whether hCdc45 displays either pyrophosphatase or exonuclease activity. Consistently with the absence of metal ions and some of the putative catalytic residues, we were unable to detect any of the above enzymatic activities (data not shown).

Both its role in DNA replication and the putative similarity with a single-stranded DNA exonuclease, such as RecJ, suggested that Cdc45 could be a DNA-binding protein. We used EMSAs to evaluate the DNA-binding properties of hCdc45 (Fig. 2*B*). The purified recombinant protein binds single-stranded synthetic oligonucleotides with a weak but detectable affinity, comparable with the one observed for the *Drosophila* GINS complex (14) or for the full-length *T. thermophilus* RecJ (27). No increase in affinity was observed when a fork-containing DNA molecule was used as a ligand in the EMSAs, and negligible binding was detected to short blunt double-stranded DNA (Fig. 3*C*). Addition of Mg^{2+}, Mn^{2+}, or Zn^{2+} ions into the buffer, as well as variation of pH in the range 5.5–8.5, were found to have no effect on the hCdc45 DNA-binding activity (data not shown).

Preliminary experiments indicated that the affinity of hCdc45 for ssRNA is weaker with respect to ssDNA. The protein shows a preference for an oligo(dG) in comparison to oligo(dC), oligo(dA), and oligo(dT) (supplemental Fig. S2).

To demonstrate that the weak DNA-binding activity is indeed due to hCdc45, and not to trace amounts of a contaminating protein, we analyzed the DNA-binding activity of the protein fractions following size-exclusion chromatography. The DNA-binding activity profile precisely co-migrates with the protein peak (Fig. 2*D* and supplemental Fig. S3) suggesting that the ability to bind DNA is a truly intrinsic feature of hCdc45. Furthermore, we analyzed whether the protein-DNA complex could be super-shifted by a specific antibody. For this experiment, we used a FLAG-tagged version of hCdc45 (purified according to the same protocol as the untagged protein with the omission of the TEV cleavage step) and a monoclonal anti-FLAG antibody. This analysis (Fig. 2*E*) revealed that the anti-FLAG antibody was able to super-shift the protein-DNA complex.

The ssDNA-binding properties of hCdc45 are consistent with the role of the protein as inferred from the single particle electron microscopy structure of the *Drosophila* CMG complex (38). In the CMG context Cdc45 (together with GINS) contributes to the formation of a tracking channel for the lagging strand. DNA binding is therefore not a property of the isolated factor but is achieved synergically with Mcm2–7 and GINS.

SAXS Data from Recombinant hCdc45 Are Consistent with the Three-dimensional Structure of the RecJ Core—SAXS data were collected from highly purified samples of recombinant hCdc45 and corrected for the scattering from the buffer. *Ab initio* modeling was performed using the program DAMMIN (39), with 44 runs being averaged using the program DAMAVER (25). The model obtained can be described as a compact core with two lateral extensions (Fig. 3). We can fit the RecJ core (encompassing domains I and II, residues 50–421) from the *T. thermophilus* RecJ crystallographic structure (PDB code: 2ZXP) in the central part of the envelope (Fig. 4). One of the two lateral extensions is larger and better defined and can be assigned to the putative helical insertion domain that is common to both the eukaryotic Cdc45 and the archaeal $RecJ^{Cdc45}$ proteins (residues 188–275 in hCdc45), as the putative insertion loop within the RecJ core is positioned in a manner compatible with this interpretation (Figs. 1 and 4). As an example, we choose to fit to this region of the map the helical bundle from the acyl-CoA-binding protein (PDB code: 2FDQ) as suggested from the results of the threading analysis. The second insertion has a more elongated shape and can be allocated to the partially unstructured insertion that is unique to the Cdc45 sequences (residues 108–175 in hCdc45, Figs. 1 and 4).

As discussed above, there is some uncertainty in the exact positions of the two long insertions with respect to the RecJ core, and alternative insertion points can be suggested. However, in both cases the alternative loops are still positioned in a manner compatible with the interpretation of the SAXS data.

Evolutionary Implications—The results of these biochemical and structural analyses suggest an evolutionary link between Cdc45 and DHH proteins. The hypothesis that Cdc45 is derived from an ancestral pyrophosphatase cannot be completely ruled

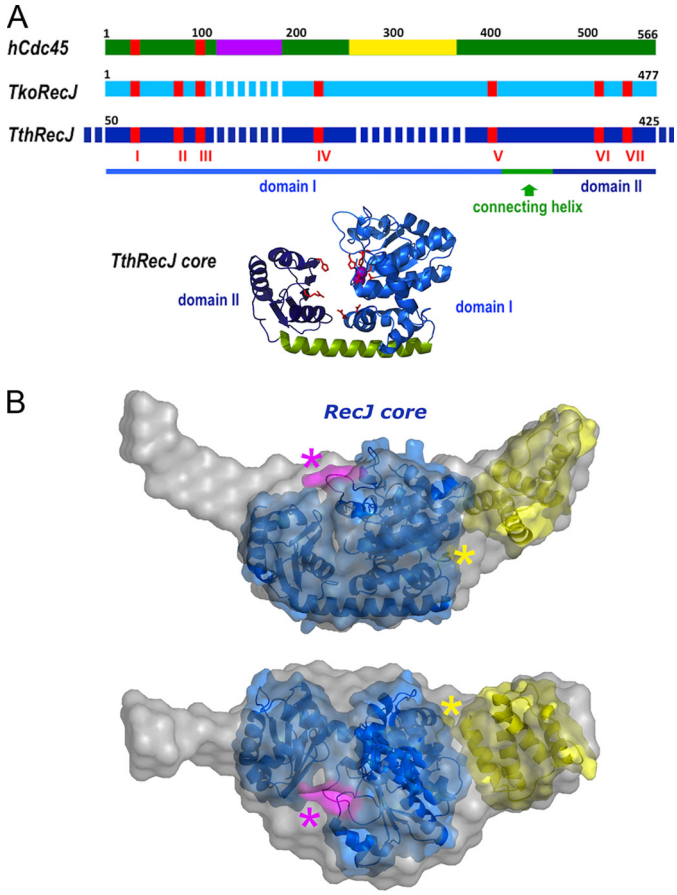

FIGURE 4. **SAXS data are consistent with a RecJ-like fold.** *A*, a *schematic diagram* summarizing the result of the bioinformatic analysis and showing the relationship between eukaryotic Cdc45 proteins, archaeal RecJ^{Cdc45}, and bacterial RecJ single-stranded DNA exonucleases. The RecJ motifs are shown in *red*, the Cdc45 charged insertion is in *magenta*, and the helical archaeal/eukaryotic insertion is in *yellow*. The crystal structure of the core of *T. thermophilus* RecJ (PDB code: 2ZXP, domains I and II) is shown below the diagram, with domain I in *blue*, domain II in *dark blue*, and the connecting helix in *green*. The conserved residues in the seven RecJ motifs are shown in *red*. *B*, the *ab initio* calculated SAXS model for hCdc45 (depicted as *gray light spheres*) is superimposed to the crystal structure of the core of *T. thermophilus* RecJ, in *blue*. Highlighted in *magenta* and indicated by a *magenta asterisk* is the putative position of the insertion, which is unique to the eukaryotic Cdc45 orthologues; highlighted in *yellow* and indicated by a *yellow asterisk* is the position of the helical bundle insertion that is common to both archaeal and eukaryotic proteins (see Fig. 1). As an example, the helical domain of the acyl-CoA-binding protein (PDB code: 2FDQ) has been fitted to the map, consistently with the results of the threading algorithms. The two views are roughly related by a 90° rotation around a *horizontal axis*.

out. It has been long suggested that a pyrophosphatase activity may be associated to the replication machinery (40). Pyrophosphate hydrolysis could enhance the catalytic activity of replicative DNA polymerases, because removal of the pyrophosphate, a product of dNTPs incorporation, would drive forward the polymerization reaction.

However, an important clue about the function and the evolutionary origin of Cdc45 comes from the finding that some of its putative archaeal homologues display a much closer similarity to the DHH subfamily 1 that includes the RecJ single-stranded DNA exonucleases. Moreover, the *ab initio* model obtained from SAXS data is more consistent with a compact RecJ core, rather than the more "open" conformation of inorganic pyrophosphatases. A number of RecJ-like proteins have been found associated with the DNA replication machinery in a variety of archaeal organisms, such as *S. solfataricus* (18), *A. fulgidus* (31), *T. kodakaraensis* (32). In particular, the *T. kodakaraensis* RecJ^{Cdc45} homologue has been shown to possess a 5′-3′ exonuclease activity that is highly stimulated by physical interaction with the GINS15 subunit (32). It has been proposed that this enzyme could participate in maturation of the Okazaki fragments in a pathway that is redundant with the FEN1- and Dna2-dependent pathways. Out of the multiple RecJ-like proteins present in archaeal organisms, it is plausible that some are associated with the DNA replication machinery and have a direct role in genome duplication; whereas others are "true" RecJ homologues, being involved in DNA repair/recombination transactions.

Based on our biochemical and bioinformatic analyses we suggest that Cdc45 originated from an ancestral 5′-3′ exonuclease loosing during evolution its catalytic activity and only retaining the ability to bind single-stranded DNA. The proteins from *Sulfolobales* may have followed a similar evolutionary path, as demonstrated by the absence of key catalytic residues and their tight association with GINS15. Association with the replication fork and GINS, in particular, may have led Cdc45 to loose its enzymatic activity and to acquire the function of "chaperoning" the lagging strand that is sterically excluded from the central channel of the MCM2–7 complex, as suggested from the cryoelectron microscopy of the CMG complex (38).

Acknowledgments—We thank Roberto Udisti (University of Florence) for determination of metal ions in protein samples by inductively coupled plasma-atomic emission spectroscopy, Spartaco De Gennaro and Achille Palma (Metapontum Agrobios) for inductively coupled plasma-MS analysis, and Marco Trifuoggi (University of Naples) for atomic absorption spectroscopy analysis. We are grateful to Alessandro Costa and James Berger (University of California at Berkeley) for sharing results before publication.

REFERENCES

1. Aparicio, O. M., Weinstein, D. M., and Bell, S. P. (1997) Components and dynamics of DNA replication complexes in *S. cerevisiae*. Redistribution of MCM proteins and Cdc45p during S phase. *Cell* **91,** 59–69
2. Owens, J. C., Detweiler, C. S., and Li, J. J. (1997) CDC45 is required in conjunction with CDC7/DBF4 to trigger the initiation of DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 12521–12526
3. Uchiyama, M., Arai, K., and Masai, H. (2001) Sna41goa1, a novel mutation causing G_1/S arrest in fission yeast, is defective in a CDC45 homolog and interacts genetically with polα. *Mol. Genet. Genomics* **265,** 1039–1049
4. Mimura, S., and Takisawa, H. (1998) *Xenopus* Cdc45-dependent loading of DNA polymerase α onto chromatin under the control of S-phase Cdk. *EMBO J.* **17,** 5699–5707
5. Pacek, M., and Walter, J. C. (2004) A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication. *EMBO J.* **23,** 3667–3676
6. Tercero, J. A., Labib, K., and Diffley, J. F. (2000) DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. *EMBO J.* **19,** 2082–2093
7. Bauerschmidt, C., Pollok, S., Kremmer, E., Nasheuer, H. P., and Grosse, F. (2007) Interactions of human Cdc45 with the Mcm2–7 complex, the GINS complex, and DNA polymerases δ and ε during S phase. *Genes Cells* **12,** 745–758
8. Pollok, S., Bauerschmidt, C., Sänger, J., Nasheuer, H. P., and Grosse, F.

(2007) Human Cdc45 is a proliferation-associated antigen. *FEBS J.* **274,** 3669–3684
9. Pospiech, H., Grosse, F., and Pisani, F. M. (2010) The initiation step of eukaryotic DNA replication. *Subcell. Biochem.* **50,** 79–104
10. Zou, L., and Stillman, B. (2000) Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. *Mol. Cell Biol.* **20,** 3086–3096
11. Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R. D., and Labib, K. (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat. Cell Biol.* **8,** 358–366
12. Pacek, M., Tutter, A. V., Kubota, Y., Takisawa, H., and Walter, J. C. (2006) Localization of MCM2–7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol. Cell* **21,** 581–587
13. Moyer, S. E., Lewis, P. W., and Botchan, M. R. (2006) Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 10236–10241
14. Ilves, I., Petojevic, T., Pesavento, J. J., and Botchan, M. R. (2010) Activation of the MCM2–7 helicase by association with Cdc45 and GINS proteins. *Mol. Cell* **37,** 247–258
15. Sakakibara, N., Kelman, L. M., and Kelman, Z. (2009) Unwinding the structure and function of the archaeal MCM helicase. *Mol. Microbiol.* **72,** 286–296
16. Costa, A., and Onesti, S. (2009) Structural biology of MCM helicases. *Crit. Rev. Biochem. Mol. Biol.* **44,** 326–342
17. Makarova, K. S., Wolf, Y. I., Mekhedov, S. L., Mirkin, B. G., and Koonin, E. V. (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33,** 4626–4638
18. Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V., and Bell, S. D. (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO rep.* **7,** 539–545
19. Yoshimochi, T., Fujikane, R., Kawanami, M., Matsunaga, F., and Ishino, Y. (2008) The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J. Biol. Chem.* **283,** 1601–1609
20. Savitsky, P., Bray, J., Cooper, C. D., Marsden, B. D., Mahajan, P., Burgess-Brown, N. A., and Gileadi, O. (2010) High-throughput production of human proteins for crystallization. The SGC experience. *J. Struct. Biol.* **172,** 3–13
21. Amenitsch, H., Bernstorff, S., Kriechbaum, M., Lombardo, D., Mio, H., Rappolt, M., and Laggner, P. (1997) *J. Appl. Crystallogr.* **30,** 872–876
22. Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006) *J. Appl. Crystallogr.* **39,** 277–286
23. Svergun, D. I. (1992) *J. Appl. Crystallogr.* **25,** 495–503
24. Franke, D., and Svergun, D. I. (2009) *J. Appl. Crystallogr.* **42,** 342–346
25. Volkov, V. V., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* **36,** 860–864
26. Aravind, L., and Koonin, E. V. (1998) A novel family of predicted phosphoesterases includes *Drosophila* prune protein and bacterial RecJ exonuclease. *Trends Biochem. Sci.* **23,** 17–19
27. Yamagata, A., Kakuta, Y., Masui, R., and Fukuyama, K. (2002) The crystal structure of exonuclease RecJ bound to Mn^{2+} ion suggests how its characteristic motifs are involved in exonuclease activity. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 5908–5912
28. Wakamatsu, T., Kitamura, Y., Kotera, Y., Nakagawa, N., Kuramitsu, S., and Masui, R. (2010) Structure of RecJ exonuclease defines its specificity for single-stranded DNA. *J. Biol. Chem.* **285,** 9762–9769
29. Fabrichniy, I. P., Lehtiö, L., Salminen, A., Zyryanov, A. B., Baykov, A. A., Lahti, R., and Goldman, A. (2004) Structural studies of metal ions in family II pyrophosphatases. The requirement for a Janus ion. *Biochemistry* **43,** 14403–14411
30. Rajman, L. A., and Lovett, S. T. (2000) A thermostable single-strand DNase from *Methanococcus jannaschii* related to the RecJ recombination and repair exonuclease from *Escherichia coli*. *J. Bact.* **182,** 607–612
31. Motz, M., Kober, I., Girardot, C., Loeser, E., Bauer, U., Albers, M., Moeckel, G., Minch, E., Voss, H., Kilger, C., and Koegl, M. (2002) Elucidation of an archaeal replication protein network to generate enhanced PCR enzymes. *J. Biol. Chem.* **277,** 16179–16188
32. Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., Hurwitz J., Reeve, J. N., and Kelman, Z. (2011) A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res.* **39,** 6114–6123
33. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web. A case study using the Phyre server. *Nature Protocols* **4,** 363–371
34. Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33,** W244–W248
35. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005) FFAS03. A server for profile–profile sequence alignments. *Nucleic Acids Res.* **33,** W284–288
36. Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) FUGUE. Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310,** 243–257
37. Sanchez-Pulido, L., and Ponting, C. P. (2011) Cdc45. The missing RecJ ortholog in eukaryotes? *Bioinformatics* **27,** 1885–1888
38. Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M. R., and Berger, J. M. (2011) The structural basis for MCM2–7 helicase activation by GINS and Cdc45. *Nat. Struct. Mol. Biol.* **18,** 471–477
39. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* **76,** 2879–2886
40. Aravind, L., and Koonin, E. V. (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res.* **26,** 3746–3752
41. Edgar, R. C. (2004) MUSCLE. Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797

Appl Microbiol Biotechnol
DOI 10.1007/s00253-012-4273-z

Biochemical and structural characterization of recombinant short-chain NAD(H)-dependent dehydrogenase/reductase from *Sulfolobus acidocaldarius* highly enantioselective on diaryl diketone benzil

Angela Pennacchio · Vincenzo Sannino · Giosuè Sorrentino · Mosè Rossi · Carlo A. Raia · Luciana Esposito

Received: 16 May 2012 / Revised: 27 June 2012 / Accepted: 28 June 2012

Abstract The gene encoding a novel alcohol dehydrogenase that belongs to the short-chain dehydrogenases/reductases superfamily was identified in the aerobic thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius* strain DSM 639. The *saadh2* gene was heterologously overexpressed in *Escherichia coli*, and the resulting protein (SaADH2) was purified to homogeneity and both biochemically and structurally characterized. The crystal structure of the SaADH2 NADH-bound form reveals that the enzyme is a tetramer consisting of identical 27,024-Da subunits, each composed of 255 amino acids. The enzyme has remarkable thermophilicity and thermal stability, displaying activity at temperatures up to 80 °C and a 30-min half-inactivation temperature of ~88 °C. It also shows good tolerance to common organic solvents and a strict requirement for NAD(H) as the coenzyme. SaADH2 displays a preference for the reduction of alicyclic, bicyclic and aromatic ketones and α-ketoesters, but is poorly active on aliphatic, cyclic and aromatic alcohols, showing no activity on aldehydes. Interestingly, the enzyme catalyses the asymmetric reduction of benzil to (*R*)-benzoin with both excellent conversion (98 %) and optical purity (98 %) by way of an efficient in situ NADH-recycling system involving a second thermophilic ADH. The crystal structure of the binary complex SaADH2–NADH, determined at 1.75 Å resolution, reveals details of the active site providing hints on the structural basis of the enzyme enantioselectivity.

Keywords Archaea · *Sulfolobus acidocaldarius* · Short-chain dehydrogenases/reductases · Crystal structure · Bioreduction · Benzil

Electronic supplementary material The online version of this article (doi:10.1007/s00253-012-4273-z) contains supplementary material, which is available to authorized users.

A. Pennacchio (✉) · V. Sannino · M. Rossi · C. A. Raia
Istituto di Biochimica delle Proteine, CNR,
Via P. Castellino 111,
80131 Naples, Italy
e-mail: a.pennacchio@ibp.cnr.it

G. Sorrentino · L. Esposito (✉)
Istituto di Biostrutture e Bioimmagini, CNR,
Via Mezzocannone 16,
80134 Naples, Italy
e-mail: luciana.esposito@unina.it

Introduction

Dehydrogenases/reductases are found throughout across a wide range of organisms where they are involved in a broad spectrum of metabolic functions (Jörnvall 2008), and a system of short-, medium- and long-chain dehydrogenase/reductases has been recently identified based on molecular size, sequence motifs, mechanistic features and structural comparisons (Kavanagh et al. 2008; Persson et al. 2009). Many studies have been addressed to characterize alcohol dehydrogenases (ADHs) from thermophiles and hyperthermophiles, mainly to understand their evolution and structure/function/stability relationship (Radianingtyas and Wright 2003) and develop their biotechnological potential in the synthesis of chiral alcohols (Jones and Beck 1976; Keinan et al. 1986; Hummel 1999; Kroutil et al. 2004; Müller et al. 2005; Goldberg et al. 2007). Recently, ADHs displaying distinct substrate specificity, good efficiency and high enantioselectivity have been described, such as the NADP-dependent (*R*)-specific ADH from *Lactobacillus brevis* (LB-RADH) (Schlieben et al. 2005), the NAD-

Published online: 18 July 2012

dependent ADH from *Leifsonia* sp. strain S749 (LSADH) (Inoue et al. 2006), and the NADP-dependent carbonyl reductase from *Candida parapsilosis* (Nie et al. 2007). These enzymes originate from mesophilic microorganisms and belong to the short-chain dehydrogenases/reductases (SDRs) superfamily (Kavanagh et al. 2008) which is characterized by ~250 residue subunits, a Gly-motif in the coenzyme-binding regions, and a catalytic tetrad formed by the highly conserved residues Asn, Ser, Tyr and Lys (Filling et al. 2002; Schlieben et al. 2005; Persson et al. 2009). Representative examples of ADHs from thermophilic microorganisms are medium-chain enzymes, including *Thermoanaerobacter brockii* ADH (Korkhin et al. 1998), the ADH from *Bacillus stearothermophilus* strain LLD-R (BsADH) (Ceccarelli et al. 2004) and two archaeal enzymes, the ADH from *Aeropyrum pernix* (Guy et al. 2003) and *Sulfolobus solfataricus* (Giordano et al. 2005; Friest et al. 2010). However, two archaeal short-chain ADHs have been identified in *Pyrococcus furiosus*, an NADP(H)-dependent SDR (van der Oost et al. 2001; Machielsen et al. 2008), and an NAD(H)-preferring ADH, that belongs to the aldo-keto reductase superfamily (Machielsen et al. 2006; Zhu et al. 2006, 2009). Furthermore, two short-chain NAD(H)-dependent ADHs, TtADH and SaADH, identified in *Thermus thermophilus* HB27 and *Sulfolobus acidocaldarius*, respectively, have been recently purified and characterized in our laboratory (Pennacchio et al. 2008, 2010a, 2010b, 2011).

With the aim to find novel dehydrogenase/reductases that are both stable and NAD^+ dependent, we focused our attention on the genomes of thermophilic organisms containing genes encoding putative ADHs belonging to the SDR superfamily and applied the criteria used for TtADH and SaADH to select enzymes with a high probability of being NAD^+ dependent (Pennacchio et al. 2008). An open reading frame coding for a protein belonging to the SDR superfamily with relatively high sequence identity to that of most representative ADHs (LB-RADH, LSADH, TtADH and SaADH) was found in the genome of *S. acidocaldarius*, an aerobic thermoacidophilic crenarcheon which grows optimally at 80 °C and pH 2 (Chen et al. 2005). The amino acid sequence revealed the presence of a glutamic acid residue at position 39. Noteworthy, an aspartate residue at the homologous position plays a critical role in determining the preference of SDRs for NAD(H) as shown for the LB-RADH mutant G37D (Schlieben et al. 2005) and by the evidence that the SDRs LSADH (Inoue et al. 2006), TtADH and SaADH (Pennacchio et al. 2008, 2010a, b) are strictly NAD(H)-dependent and have an aspartate residue at the same position within the sequence (Fig. 1). Due to Glu and Asp chemical similarity, the putative dehydrogenase/reductase identified in *S. acidocaldarius* (SaADH2) was expected to display a preference for NAD(H) rather than NADP(H). This preference as well as an intrinsic thermostability are the features that make an oxidoreductase more attractive from an application perspective (Kroutil et al. 2004; Zhu et al. 2006; Huisman et al. 2010).

This report describes cloning, heterologous expression and structural characterization of the *S. acidocaldarius saadh2* gene, which encodes the SDR SaADH2. The purified enzyme was characterized in terms of substrate specificity, kinetics and stability as well as enantioselectivity. SaADH2 was found to be highly efficient and enantioselective in reducing the diaryl diketone benzil to (*R*)-benzoin.

Materials and methods

Chemicals

$NAD(P)^+$ and NAD(P)H were obtained from AppliChem (Darmstadt, Germany). Alcohols, aldehydes, ketones and keto esters were obtained from Sigma-Aldrich. 1-Butyl-3-methylimidazolium tetrafluoroborate (BMIMBF4) was a kind gift from Professor S. Cacchi. Recombinant BsADH was prepared as previously described (Fiorentino et al. 1998). Other chemicals were A grade substances from Applichem. Solutions of NADH and NAD^+ were prepared as previously reported (Raia et al. 2001). All solutions were made up with MilliQ water.

Amplification and cloning of the *saadh2* gene

Chromosomal DNA was extracted by caesium chloride purification as described by Sambrook et al. (1989). Ethidium bromide and caesium chloride were removed by repeated extraction with isoamyl alcohol and extensive dialysis against 10 mM Tris–HCl pH 8.0 and 1 mM EDTA, respectively. DNA concentration was determined spectrophotometrically at 260 nm, and the molecular mass checked by electrophoresis on 0.8 % agarose gel in 90 mM Tris–borate pH 8.0 and 20 mM EDTA, using DNA molecular size markers. The *saadh2* gene was amplified by polymerase chain reaction (PCR) using oligonucleotide primers based on the *saadh2* sequence of *S. acidocaldarius* strain DSM 639 (GenBank accession no. YP_255871.1). The following oligonucleotides were used: (5′-TGCATAGTAGCATATG TCATATCAGAGTTTG-3′) as the forward primer (the *Nde I* restriction site is underlined in the sequence) and the oligonucleotide (5′-AGGAATTCACATTACAGTACAG TTAAACCACC-3′) as the reverse primer. This latter oligonucleotide introduces a translational stop following the last codon of the ADH gene, followed by an *EcoRI* restriction site, which is underlined in the sequence shown. The PCR product was digested with the appropriate restriction enzymes and cloned into the expression vector pET29a (Novagen, Madison, Wisconsin, USA) to create the

```
SaADH2          ---MSYQSLKNKVVIVTGAGSGIGRAIAKKFALNDSIVVAVELLEDRLNQIVQELRGMGK 57
SaADH           MDIDRLFSVKGMNAVVLGASSGIGKAIAEMFSEMGGKVVLSDIDEEGLKRLSDSLRSRGH 60
LB-RADH         ----MSNRLDGKVAIITGGTLGIGLAIATKFVEEGAKVMITGRHSDVGEK-AAKSVGTPD 55
LSADH           ---MAQYDVADRSAIVTGGGSGIGRAVALTLAASGAAVLVTDLNEEHAQAVVAEIEAAGG 57
TtADH           ---MGLF--AGKGVLVTGGARGIGRAIAQAFAREGALVALCDLRPEGKE--VAEAIGGAF 53
                          .  .:: *.  *** *:*  :   .. *   *    :  :    .  .

SaADH2          EVLGVKADVSKKKDVEEFVRRTFETYSRIDVLCNNAGIMDGVTPVAEVSDELWERVLAVN 117
SaADH           EVNHMKCDITDLNQVKKLVNFSLSVYGNVDALYVTPSIN-VRKSIENYTYEDFEKVINVN 119
LB-RADH         QIQFFQHDSSDEDGWTKLFDATEKAFGPVSTLVNNAGIA-VNKSVEETTTAEWRKLLAVN 114
LSADH           KAAALAGDVTDPAFGEASVA-GANALAPLKIAVNNAGIGGEAATVGDYSLDSWRTVIEVN 116
TtADH           FQVDLEDERERVRFVEE----AAYALGRVDVLVNNAAIAAPGSAL-TVRLPEWRRVLEVN 108
                    .  :                . . :.    ...*     .:       :. :: **

SaADH2          LYSAFYSSRAVIPIMLKQGKGV-IVNTASIAGIRGGFAGAPYTVAKHGLIGLTRSIAAHY 176
SaADH           LKGNFMVVKEFLSVMKNNKGGGSVVLFSSIRGTVVEPGQSVYAMTKAGIIQLAKVAAAEY 179
LB-RADH         LDGVFFGTRLGIQRMKNKGLGASIINMSSIEGFVGDPSLGAYNASKGAVRIMSKSAALDC 174
LSADH           LNAVFYGMQPQLKAMAANGGGA-IVNMASILGSVGFANSSAYVTAKHALLGLTQNAALEY 175
TtADH           LTAPMHLSALAAREMRKVGGGA-IVNVASVQGLFAEQENAAYNASKGGLVNLTRSLALDL 167
                * . :         *     *  ::  :*: *       . *  :* .:  ::: * .

SaADH2          GDQ--GIRAVAVLPGTVKTN---IGLGSSKPSELGMRTLTKLMSLSSRLAEPEDIANVIV 231
SaADH           GKY--NIRVNVIAPGVVDTP-----LTRQIKSDPEWFKAYTEKTILKRWATPEEIANVAL 232
LB-RADH         ALKDYDVRVNTVHPGYIKTP-----LVDDLPGAEEAMSQRTKTPMG-HIGEPNDIAYICV 228
LSADH           AAD--KVRVVAVGPGFIRTP--LVEANLSAD---ALAFLEGKHALG-RLGEPEEVASLVA 227
TtADH           APL--RIRVNAVAPGAIATEAVLEAIALSPDPERTRRDWEDLHALR-RLGKPEEVAEAVL 224
                .     :*. .: ** : *         .              .:  : . *:::*

SaADH2          FLASDEASFVNGDAVVVDGGLTVL-------- 255
SaADH           FLAMPASSYITGTVIYVDGGWTAIDGRYDPKV 264
LB-RADH         YLASNESKFATGSEFVVDGGYTAQ-------- 252
LSADH           FLASDAASFITGSYHLVDGGYTAQ-------- 251
TtADH           FLASEKASFITGAILPVDGGMTASFMMAGRPV 256
                :**   :.: .*    **** *.
```

Fig. 1 Multiple-sequence alignment of the *S. acidocaldarius* ADH (SaADH2) and ADHs belonging to the SDR family, including *S. acidocaldarius* ADH (SaADH; NCBI accession no. YP_256716.1), *L. brevis* ADH (LB-RADH; PDB code 1ZK4), *Leifsonia* sp. strain S749 ADH (LSADH; NCBI accession no. BAD99642) and *T. thermophilus* ADH (TtADH; PDB code 2D1Y). The sequences were aligned using the ClustalW2 program. *Grey shading* indicates residues highly conserved in the SDR family. The four members of the catalytic tetrad are indicated by a *black background*. The following positions are indicated by *bold type*: the glycine-rich consensus sequence and the sequence motif Dhx[cp] that (in all SDRs) have a structural role in coenzyme binding (Kallberg et al. 2002). The *star* indicates the major determinant of the coenzyme specificity. The LB-RADH G37D mutant shows preference for NAD^+ over $NADP^+$ (Schlieben et al. 2005)

recombinant pET29a–saADH plasmid. The insert was sequenced in order to verify that mutations had not been introduced during PCR.

Expression and purification of recombinant SaADH2

Recombinant protein was expressed in *Escherichia coli* BL21(DE3) cells (Novagen) transformed with the corresponding expression vector. Cultures were grown at 37 °C in 2 L of LB medium containing 30 μg ml^{-1} kanamycin. When the A_{600} of the culture reached 1.4, protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside to a concentration of 1.0 mM. The bacterial culture was incubated at 37 °C for a further 24 h. Cells were harvested by centrifugation, and the pellet stored at −20 °C until use. The cells obtained from 2 L of culture were suspended in 20 mM Tris–HCl buffer (pH 7.5) containing 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and were lysed using a French pressure cell (Aminco Co., Silver Spring, MD) at 2,000 psi (1 psi=6.9 kPa). The lysate was centrifuged, and the supernatant incubated in the presence of DNase I (50 μg/ml of solution) and 5 mM $MgCl_2$ for 30 min at 37 °C, followed by protamine sulphate (1 mg/ml of solution) at 4 °C for 30 min. The nucleic acid fragments were removed by centrifugation, and the supernatant incubated at 70 °C for 15 min. The host protein precipitate was removed by centrifugation. The supernatant was dialysed overnight at 4 °C against 20 mM Tris–HCl, pH 8.4 (buffer

A) containing 1 mM PMSF. The dialysed solution was applied to a DEAE-Sepharose Fast Flow (1.6×12 cm) column equilibrated in buffer A at 120 ml h^{-1}. The active pool did not bind to the column matrix. The flowthrough fractions were combined and dialysed against buffer A, concentrated 5-fold with a 30,000 MWCO centrifugal filter device (Millipore), and applied to a Sephadex G-75 (1.6×30 cm) column equilibrated in buffer A containing 0.15 M NaCl. The active pool was dialysed against buffer A and concentrated to obtain 2.0 mg protein·ml^{-1} as described previously. SaADH2 was stored at −20 °C in buffer A containing 50 % glycerol, without loss of activity following several months of storage. SDS-PAGE was carried out according to the Laemmli method (1970). The subunit molecular mass was determined by electrospray ionization mass spectrometry (ESI-MS) with a QSTAR Elite instrument (Applied Biosystems, USA). The protein concentration was determined with a Bio-Rad protein assay kit using BSA as a standard.

Sucrose density gradient centrifugation

One hundred microlitres of a 1 mg/ml purified SaADH2 solution was layered on top of a 10-ml 5–20 % preformed sucrose gradient in 50 mM Tris–HCl, pH 8.4. The standards were also prepared by layering 100 μl of 1 mg/ml chymotripsinogen (25 kDa), BSA (67 kDa), *T. thermophilus* ADH (a 108-kDa tetramer) and *S. solfataricus* ADH (a 150-kDa tetramer) on top of another identical sucrose gradient. By using a Beckman SW 41 Ti rotor, the tubes containing the samples were spun at 37,000 rpm for 18 h at 4 °C in a Beckman LE-80 ultracentrifuge. Immediately after centrifugation, gradients were collected in 250-μl fractions and assayed for absorption at 280 nm to determine the position of the marker protein. The mass of SaADH2 was estimated from the position of the peak relative to those of the markers.

Enzyme assay

SaADH2 activity was assayed spectrophotometrically at 65 °C by measuring the change in absorbance of NADH at 340 nm using a Cary 1E spectrophotometer equipped with a Peltier effect-controlled temperature cuvette holder. The standard assay for the reduction reaction was performed by adding 5−25 μg of the enzyme to 1 ml of preheated assay mixture containing 20 mM ethyl 3-methyl-2-oxobutyrate (EMO) and 0.2 mM NADH in 37.5 mM sodium phosphate, pH 5.0. The standard assay for the oxidation reaction was performed using a mixture containing 18 mM cycloheptanol and 3 mM NAD^+ in 50 mM glycine/NaOH, pH 10.5. Screening of the substrates was performed using 1 ml of assay mixture containing either 10 mM alcohol and 3 mM NAD^+ in 50 mM glycine/NaOH, pH 10.5, or 10 mM carbonyl compound and 0.1 mM NADH in 37.5 mM sodium phosphate, pH 5.0. One unit of SaADH2 represented 1 μmol of NADH produced or utilized per minute at 65 °C, on the basis of an absorption coefficient of 6.22 mM^{-1} cm^{-1} for NADH at 340 nm.

Effect of pH on activity

The optimum pH value for the reduction and oxidation reactions was determined at 65 °C under the conditions used for EMO and cycloheptanol, respectively, except that different buffer were used. The concentration of each buffer solution was diluted to get the similar conductivity (~1.1 mS). The pH was controlled in each assay mixture at 65 °C.

Kinetics

The SaADH2 kinetic parameters were calculated from measurements determined in duplicate or triplicate and by analysing the kinetic results using the program GraFit (Leatherbarrow 2004). The turnover value (k_{cat}) for SaADH2 was calculated on the basis of a molecular mass of 27 kDa, assuming that the four subunits are catalytically active.

Thermophilicity and thermal stability

SaADH2 was assayed in a temperature range of 25–90 °C using standard assay conditions and 22 μg protein·ml^{-1} of assay mixture. The stability at various temperatures was studied by incubating 0.2 mg ml^{-1} protein samples in 50 mM Tris–HCl, pH 9.0, at temperatures between 25 and 95 °C for 30 min. Each sample was then centrifuged at 5 °C, and the residual activity assayed as described above.

The effect of chelating agents on enzyme stability was studied by measuring the activities before and after exhaustive dialysis of the enzyme against buffer A, containing 1 mM EDTA, and then against buffer A alone. An aliquot of the dialysed enzyme was then incubated at 70 °C in the absence and presence of 1 mM EDTA, and the activity assayed at different times.

Effects of compounds on enzyme activity

The effects of salts, metal ions and chelating agents on SaADH2 activity were investigated by assaying the enzyme in the presence of an appropriate amount of each compound in the standard assay mixture used for the oxidation reaction.

The effects of organic solvents were investigated by measuring the activity in enzyme samples (0.2 mg ml^{-1} in 100 mM sodium phosphate, pH 7.0) immediately after the addition of organic solvents at different concentrations and

after incubation for 6 and 24 h at 50 °C. The percentage activity for each sample was calculated by comparison with the value measured prior to incubation. The volume of the solution in a tightly capped test tube did not change during incubation.

Enantioselectivity

The enantioselectivity of SaADH2 was determined by examining the reduction of acetophenone, the bicyclic ketones benzyl and 2,2′-dichlorobenzil using an NADH regeneration system consisting of BsADH and a substrate alcohol (see below). The reaction mixture contained 2 mM NAD^+, 5 mM carbonyl compound, 4 to 18 %*v*/*v* alcohol, 250 μg SaADH2 and 50 μg of BsADH in 1 ml of 100 mM sodium phosphate, pH 7.0. The reactions were carried out at 50 °C for 24 h in a temperature-controlled water bath. Upon termination of the reaction, the reaction mixture was extracted twice with ethyl acetate. The percentage of conversion and enantiomeric purity of the product were determined on the basis of the peak areas of ketone substrates and alcohol products by HPLC, on a Chiralcel OD-H column (Daicel Chemical Industries, Ltd., Osaka, Japan). The absolute configuration of product alcohols was identified by comparing the chiral HPLC data with the standard samples. Products were analyzed with isocratic elution, under the following conditions: hexane/2-propanol (9:1) (mobile phase), flow rate of 1 ml min^{-1}, detection at 210 nm. Retention times were the following: 5.6, 11.9 and 16.7 for benzil, (*S*)- and (*R*)-benzoin, respectively; 6.98, 12.73 and 15.13, for 2,2′-dichlorobenzil, (*S*)- and (*R*)-1,2-bis(2-chlorophenyl)-2-hydroxyethanone, respectively and 5.27, 6.2 and 7.5 for acetophenone, (*R*)- and (*S*)-1-phenylethanol, respectively. The absolute stereochemistry of 1,2-bis(2-chlorophenyl)-2-hydroxyethanone enantiomers was assigned by analogy to the values of (*S*)- and (*R*)-benzoin.

Structural characterization

The purified protein was concentrated to 8–10 mg ml^{-1} in 20 mM Tris/HCl buffer at pH 8.4. The protein solution was added with NADH to obtain a 8.5-mg ml^{-1} protein solution containing 2.2 mM NADH in the above buffer. Crystallization was performed by hanging-drop vapour diffusion method at 296 K. The best crystals were grown by mixing 1 μl of protein solution with the same volume of a solution containing 2.0 M ammonium sulphate in 0.1 M HEPES buffer (pH 7.5). The crystals belong to P21 space group with unit cell parameters a=83.80 Å, b=60.48 Å, c=107.1 Å and β=99.8°. Crystals were transferred to a stabilizing solution containing 20 % glycerol as cryoprotectant and X-ray data were then collected in-house at 100 K using a Rigaku Micromax 007 HF generator producing Cu Kα radiation and equipped with a Saturn944 CCD detector. Intensity data up to 1.75 Å resolution were processed and scaled using program HKL2000 (HKL Research). Data collection statistics are shown in Table 1.

The crystal structure was solved by MR methods by using as a search model the coordinates of a putative glucose/ribitol dehydrogenase from *Clostridium thermocellum* (PDB code 2HQ1), which showed the best sequence alignment with SaADH2 (39 % sequence identity). Model building was first performed with ARP/WARP package using the warpNtrace automated procedure. Later model building and refinement was performed by the program CNS version 1.1. The final structure shows R and R-free factors of 16.8 and 19.1 %, respectively. Refinement statistics are summarized in Table 1. Fold similarities searches were performed with the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server).

Results

Expression and protein purification

Analysis of the *S. acidocaldarius* genome (Chen et al. 2005) for genes encoding short-chain ADHs resulted in identification of a putative oxidoreductase gene. The sequence of the 27,024-Da protein, named SaADH2, showed the highest level of identity to four typical SDRs, LB-RADH (29 % identity), LSADH (34 %), TtADH (35 %) and SaADH (30 %; Fig. 1). The *saadh2* gene was successfully expressed in *E. coli* cells, yielding an active enzyme accounting for

Table 1 Data collection and refinement statistics

Data collection	
Space group	P21
Cell dimensions (Å;°)	a=83.80, b=60.48, c=107.1; β=99.78
Resolution range (Å)	40.0–1.75 (1.81–1.75)
No. of unique reflections	106,431 (10,388)
Average redundancy	4.2 (2.8)
Completeness	99.7 (98.0)
$<I/\sigma(I)>$	19.8 (3.6)
Rmerge	0.065 (0.36)
Refinement	
No. of reflections used (with $\|F\| > 0$)	101,786
No. of reflections working/test set	96,639/5,147
Resolution range (Å)	40.0–1.75
R-factor/R-free (%)	16.8/19.1
No. of protein atoms	7,512
No. of water/glycerol/sulphate/NADH	687/5/1/4
RMS deviation on bond distances (Å)	0.007
RMS deviation on bond angles (°)	1.4

about 9 % of the total protein content of the cell extract (Supplementary material Table S1). Host protein precipitation at 70 °C was found to be the most effective purification step. An overall purification of 5.4-fold was achieved from crude-cell-free extracts with an overall yield of 48 %. SDS-PAGE of the purified protein showed a single band corresponding to a molecular mass of ~29 kDa (data not shown).

The quaternary structure of the enzyme was investigated by sucrose density gradient centrifugation. SaADH2 sedimented as one peak nearly overlapping that of TtADH (Supplementary Material Fig. S1). The plot of molecular mass of the markers vs fraction number (inset Fig. S1) allowed to determine an apparent molecular mass of 109± 10 kDa for SaADH2. The molecular mass of the subunit determined by ESI-MS analysis proved to be 27,024.0 Da (average mass), in agreement with the theoretical value of the sequence.

Optimal pH

The pH dependence of SaADH2 in the reduction and oxidation reaction was analysed (Fig. S2). The SaADH2 activity was found to be closely dependent on pH in the reduction reaction, displaying a peak of maximum activity at around pH 5.0. The oxidation reaction showed a less marked dependence on pH, displaying a peak with a maximum at around pH 10.0.

Thermophilicity and thermal stability

The effect of temperature on SaADH2 activity is shown in Fig. 2. The reaction rate increases up to 78 °C and then decreases rapidly due to thermal inactivation. This optimal temperature value is similar to that of TtADH (73 °C) and SaADH (75 °C) (Pennacchio et al. 2008, 2010a, b), and lower than that of *P. furiosus* aldo-keto reductase (100 °C) (Machielsen et al. 2006). The thermal stability of SaADH2 was determined by measuring the residual enzymatic activity after 30 min of incubation over a temperature range from 25 to 95 °C (Fig. 2). SaADH2 was shown to be quite stable up to a temperature of ~75 °C, above which its activity decreased abruptly, resulting in a $T_{1/2}$ value (the temperature of 50 % inactivation) of ~88 °C.

Coenzyme and substrate specificity

The enzyme showed no activity with NADP(H) and full activity with NAD(H).

The specificity of SaADH2 for various alcohols, aldehydes and ketones was examined (Table 2). The enzyme showed a poor activity on a discrete number of aliphatic linear and branched alcohols, such as 2-propyn-1-ol, 4-methyl-1-pentanol, the *S* enantiomers of 2-butanol and 2-pentanol as well as on aliphatic cyclic and bicyclic alcohols, except for isoborneol and cycloheptanol which rank first and second, respectively, among the tested alcohols. Benzyl alcohol and substituted benzyl alcohols were found to be poor substrates. Among the aromatic secondary alcohols tested SaADH2 showed a low activity on (*S*)-1-phenylethanol and no activity with the *R* enantiomer, whereas displayed similar poor activities towards the (*S*)- and (*R*)- forms of α-(trifluoromethyl)-benzyl alcohol, 1-(2-naphthyl)ethanol and methyl and ethyl mandelates. Moreover, the enzyme showed poor activity on (±)-1-phenyl-1-propanol and its *ortho*-chloro derivative, and *para*-halogenated 1-phenylethanols. However, SaADH2 showed a relatively high activity with 1-indanol and α-tetralol, but a poor activity on the β-hydroxy ester ethyl (*R*)-4-chloro-3-hydroxybutyrate.

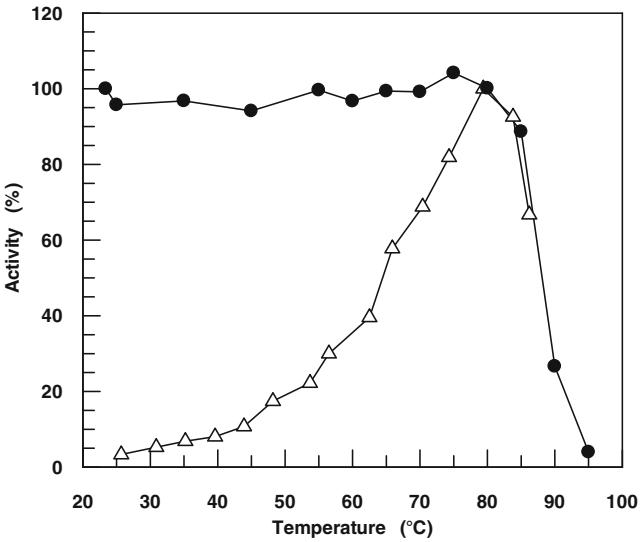

Fig. 2 Effect of temperature on activity and stability of SaADH2 monitored by dehydrogenase activity. The assays at the increasing temperature values (*triangles*) were carried out as described in Materials and methods, using cycloheptanol as the substrate. The thermal stability (*circles*) was studied by incubating 0.2 mg ml^{-1} protein samples in 50 mM Tris–HCl, pH 9.0 for 30 min at the indicated temperatures. Activity measurements were carried out under the conditions of the standard assay using cycloheptanol as the substrate. The assay temperature was 65 °C. The percentage of residual activity was obtained by the ratio to the activity without heating

The enzyme was not active on aliphatic and aromatic aldehydes, and on aliphatic linear, and branched ketones (data not shown). However, it was active on aliphatic cyclic and bicyclic ketones such as cyclohexanone, methyl-substituted cyclohexanones and decalone (Table 2). Two aryl diketones, 1-phenyl-1,2-propanedione and benzil were good substrates of SaADH2 which also showed a relatively high reduction rate with 2,2-dichloro- and 2,2,2-trifluoroacetophenone and penta-substituted fluoroacetophenone (Table 2). The electronic factor accounts for the relatively high activity measured with the

Table 2 Substrate specificity of SaADH2 in the oxidation and reduction reactions[a]

Substrate	Relative activity (%)[b]	Substrate	Relative activity (%)[b]
Alcohols			
2-Propyn-1-ol	5	1-(2-Chlorophenyl)-1-propanol[c]	2
4-Methyl-1-pentanol	2	(±)-1-Indanol[d]	26
(*S*)-2-Butanol	3	(±)-α-Tetralol[c]	51
(*R*)-2-Butanol	0	(±)-1-(2-Naphthyl)ethanol	2
2,3-Butanediol	2	(*S*)-1-(2-Naphthyl)ethanol	3
(±)-2-Pentanol	2	(*R*)-1-(2-Naphthyl)ethanol	2
(*S*)-2-Pentanol	7	Benzoin	0
(*R*)-2-Pentanol	0	*trans*-1,3-Diphenyl-2-propen-1-ol	5
3-Pentanol	1	Ethyl (*R*)-4-chloro-3-hydroxybutyrate	2
2-Hexanol[c]	2	Methyl (*S*)-(−)-mandelate	4
2-Heptanol[c]	2	Methyl (*R*)-(+)-mandelate	3
6-Methyl-5-hepten-2-ol[c]	3	Ethyl (*S*)-(−)-mandelate	2
1-Octanol	1	Ethyl (*R*)-(+)-mandelate	2
Geraniol	2	Ketones[c]	
Cyclopentanol	2	Cyclohexanone	13
Cyclohexanol	6	2-Methylcyclohexanone	33
3-Methylcyclohexanol[c]	10	3-Methylcyclohexanone	60
Cycloheptanol	55	4-Methylcyclohexanone	41
Cyclohexylmethanol[c]	2	Cycloheptanone	0
2-Cyclohexylethanol	1	1-Decalone	85
Chrysanthemyl alcohol	1	(±)-Camphor	0
cis-Decahydro-1-naphthol	3	Acetophenone	0
Isoborneol	100	2,2-Dichloroacetophenone	30
Benzyl alcohol	7	2,2,2-Trifluoroacetophenone	35
2-Methoxybenzyl alcohol	5	2′,3′,4′,5′,6′-Pentafluoroacetophenone	45
3-Methoxybenzyl alcohol	6	1-Phenyl-1,2-propanedione	100
4-Methoxybenzyl alcohol	3	1-Indanone[d]	0
4-Bromobenzyl alcohol[c]	9	α-Tetralone	0
(*R*,*S*)-1-Phenylethanol[c]	2	Benzil	62
(*S*)-(−)-1-Phenylethanol	6	2,2′-Dichlorobenzil	0
(*R*)-(+)-1-Phenylethanol	0	Chalcone	0
(*R*,*S*)-α-(Trifluoromethyl)benzyl alcohol	2	Keto esters[c]	
(*S*)-α-(Trifluoromethyl)benzyl alcohol	2	Ethyl pyruvate	16
(*R*)-α-(Trifluoromethyl)benzyl alcohol	2	Ethyl 3-methyl-2-oxobutyrate	100
1-(4-Fluorophenyl)ethanol[c]	5	Ethyl 4-chloroacetoacetate	0
1-(4-Chlorophenyl)ethanol[c]	2	Methyl benzoylformate	8
(±)-1-Phenyl-1-propanol[c]	2	Methyl *o*-chlorobenzoylformate	5
trans-Cinnamyl alcohol[c]	2	Ethyl benzoylformate	23

[a] The activity was measured at 65 °C as described in Materials and methods. The concentration of each substrate was 5 mM

[b] The percent values refer to cycloheptanol for alcohols, benzil for ketones and to ethyl 3-methyl-2-oxobutyrate for keto esters

[c] The substrates were dissolved in 100 % 2-propanol

[d] The substrate was dissolved in 100 % acetonitrile

halogenated acetophenones, as compared to the apparent zero activity observed with acetophenone. The electron withdrawing character of fluorine (or chlorine) favours hydride transfer, inductively decreasing electron density at the acceptor carbon C1. On the other hand, the corresponding ketones of cycloheptanol and isoborneol, cycloheptanone and camphor did not show any apparent activity due to the deactivating effect exerted by the

electron donating alkyl groups on the acceptor carbon C1. SaADH2 was also active on aliphatic and aryl α-keto esters but not on β-ketoesters (Table 2).

Kinetic studies

The kinetic parameters of SaADH2 determined for the most active substrates are shown in Table 3. Based on the specificity constant (k_{cat}/K_m), isoborneol is the best substrate in the oxidation reaction; compared to the alicyclic cycloheptanol and aromatic bicyclic alcohols, it displays a higher affinity to the active site probably due to its alicyclic bridged structure. However, the enzyme shows a 6-fold greater preference for (*S*)- than (*R*)-1-indanol and a 23-fold greater preference for (*S*)- than (*R*)-α-tetralol. In the reduction reaction, 2,2-dichloroacetophenone was preferred 22-fold more than 2,2,2-trifluoroacetophenone and 2′,3′,4′,5′,6′-pentafluoroacetophenone, and 1.6- and 6-fold more than the two diketones, benzil and 1-phenyl-1,2-propanedione, respectively, due to its higher affinity. Ethyl 3-methyl-2-oxobutyrate shows the highest turnover among the carbonyl compound tested, although it binds to the catalytic site with relatively low affinity. Moreover, the specificity constant value is 6-fold higher for NADH than NAD^+.

Effects of various compounds

The effects of salts, ions and reagents on SaADH2 activity were studied by adding each compound to the standard assay mixture. The enzyme activity in the presence of 1 mM of the Li^+, Na^+, K^+, Ca^{++}, Mg^{++} and Mn^{++} chlorides was 115, 113, 102, 101, 101 and 110 %, respectively, and in the presence of 1 mM of the sulphate of heavy metal ions such as Fe^{++}, Co^{++}, Ni^{++}, Cu^{++} and Zn^{++} was 93, 108, 104, 102 and 107 %, correspondingly, when compared to the enzyme activity measured in the absence of additional metal ions. The presence of 5 % ionic liquid (BMIMBF4) inactivated by 50 % the enzyme presumably due to a competition of the BF_4^- ion with the coenzyme phosphate moiety for the anion-binding site of the enzyme.

The addition of 4 mM iodoacetate and 1 mM Hg^{++} had no significant influence on the enzyme activity, which resulted in 101 and 113 % of the control activity, respectively, suggesting that the only Cys residue per monomer, C90, has no functional role. Even the metal-chelating agents did not affect enzyme activity. The enzyme activity in the presence of either 1 mM *o*-phenanthroline, or 10 mM EDTA was 98 and 87 %, respectively, suggesting that either the protein does not require metals for its activity or the chelating molecule was not able to remove the metal under the assay conditions. Furthermore, the enzyme showed no loss in activity following exhaustive dialysis against EDTA. The EDTA-dialysed enzyme turned out to be quite stable at 70 °C for 5 h, both in the absence and the presence of EDTA (data not shown).

Stability in organic solvents

The effects of common organic solvents, such as acetonitrile, DMSO, 1,4-dioxane and ethyl acetate on SaADH2 were investigated at 50 °C, at two different time points and 18 % concentration (Fig. 3). SaADH2 activated after 6 and 24 h incubation in aqueous buffer (120 and 105 % the initial values, respectively), and inactivated by 15 and 30 % in the presence of 18 % acetonitrile following incubation for 6 and 24 h,

Table 3 Steady-state kinetic constants of SaADH2

Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
Cyclohepthanol	7.0	5.1	1.37
Isoborneol	16.6	0.81	20.5
(±)-1-Indanol	6.2	8.7	0.71
(*R*)-Indanol	2.7	15.3	0.18
(*S*)-Indanol	10.6	10.0	1.06
α-Tetralol	9.6	8.5	1.13
(*R*)-α-Tetralol	1.15	9.2	0.13
(*S*)-α-Tetralol	6.0	2.0	3.0
NAD^+	19.3	0.18	107
Benzil	1.6	0.43	3.72
1-Phenyl-1,2-propanedione	5.3	5.0	1.06
2,2-Dichloroacetophenone	0.65	0.11	5.91
2,2,2-Trifluoroacetophenone	1.7	6.3	0.27
2′,3′,4′,5′,6′-Pentafluoroacetophenone	3.2	11.9	0.27
Ethyl 3-methyl-2-oxobutyrate	26	16.6	1.57
Ethyl benzoylformate	1.9	4.2	0.45
NADH	26.2	0.04	655

The activity was measured at 65 °C as described in Materials and methods. Kinetic constants for NAD^+ and NADH were determined with 15 mM isoborneol and 30 mM ethyl 3-methyl-2-oxobutyrate, respectively

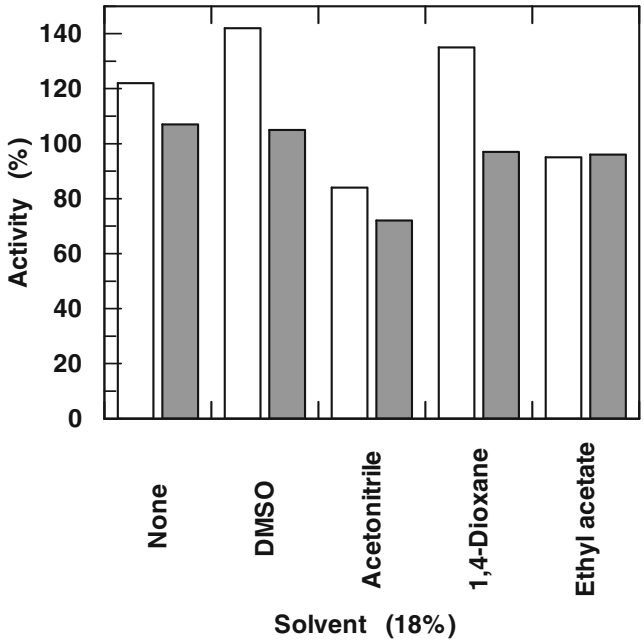

Fig. 3 Effects of various solvents on SaADH2. Samples of enzyme (0.20 mg/ml) were incubated at 50 °C in the absence and presence of the organic solvents at 18 % concentrations, and the assays were performed after 6 h (*white bars*) and 24 h (*grey bars*). The activity assays were performed at 50 °C as described in Materials and methods using cycloheptanol as the substrate. The data obtained in the absence and presence of organic solvents are expressed as percentage of activity relative to the value determined prior to incubation

respectively. A slightly reduced activity (~95 %) remained after 6 and 24 h incubation in aqueous solution containing 18 % ethyl acetate. Interestingly, significant increases in enzyme activity occurred after 6 h incubation in the presence of 18 % DMSO (>140%) and 1,4-dioxane (135 %). After 24 h incubation in the presence of these two solvents, the enzyme activity remained nearly unchanged with respect to the initial value.

Enantioselectivity

The enantioselectivity of SaADH2 was tested on benzil using an NADH recycling system consisting of thermophilic NAD (H)-dependent BsADH (Fig. 4). The latter enzyme is mainly active on aliphatic and aromatic primary and secondary alcohols and aldehydes (Guagliardi et al. 1996), but not on aliphatic and aromatic ketones, nor on the carbonyl substrates of SaADH2 and corresponding alcohols (data not shown). Since 2-propanol is not a substrate of SaADH2, it may be a suitable substrate for BsADH in NADH recycling, as well as being used as a co-solvent. Bioconversions were carried out for 24 h using 5 mM benzil as substrate and 2-propanol at three different concentrations. Sodium phosphate buffer, pH 7.0 and 50 °C was chosen as a compromise between cofactor stability and catalytic activity of the two ADHs at suboptimal pH (Pennacchio et al. 2008, 2011). Chiral HPLC analysis of the extracts obtained from the bioconversions showed that benzil

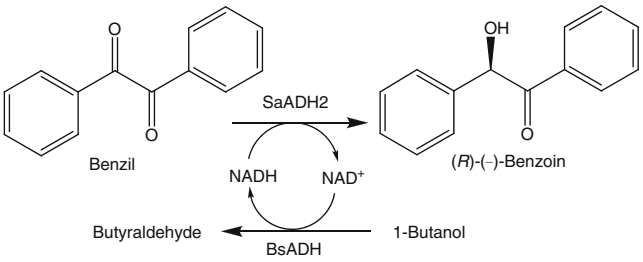

Fig. 4 Coenzyme recycling in the production of chiral diaryl alcohol with SaADH2 utilizing *B. stearothermophilus* ADH (BsADH) and 1-butanol

was reduced by the archaeal enzyme to the (*R*)-enantiomer of benzoin with a level of conversion of 35, 100 and 99 % and an enantiomeric excess (*ee*) of 85, 82 and 80 % using 4, 14 and 18 % (*v*/*v*) 2-propanol as ancillary substrate, respectively.

In addition to 2-propanol (k_{cat}/K_m=9 s^{-1} mM^{-1}), BsADH oxidizes other alcohols with even greater efficiency such as ethanol (k_{cat}/K_m=64 s^{-1} mM^{-1}), 1-propanol (286 s^{-1} mM^{-1}), 1-butanol (437 s^{-1} mM^{-1}), 1-pentanol (64 s^{-1} mM^{-1}) and 1-hexanol (64 s^{-1} mM^{-1}; Raia, unpublished data). We therefore tested these alcohols as alternative hydride source for NADH recycling and also to improve the solubility of the substrate in the aqueous phase. Moreover, these alcohols and the respective aldehydes are not substrates of SaADH2. Thus, only the cofactor is the co-substrate of SaADH2 and BsADH.

Figure 5 summarizes the results of the bioconversions carried out using different alcohol substrates at 18 % concentrations. As for the case of 2-propanol, benzil was reduced to the (*R*)-alcohol with excellent conversion (>99 %) but modest enantioselectivity using ethanol and 1-propanol. However, in the presence of 1-butanol, 1-pentanol and 1-hexanol excellent optical purity (98–99 % *ee*) and levels of conversion decreasing from 98 to 93 % and 66 %, respectively, were obtained. This suggests that ethanol and linear/branched propanol acted as good substrates and solvents for the bioconversion system, but negatively affected the prochiral selectivity of SaADH2 that was instead enhanced by longer chain-linear alcohols (four to six carbons). The decrease in conversion rate with 1-pentanol and 1-hexanol could be due to different effects: (1) substrate excess inhibition, (2) BsADH deactivation and (3) reduced substrate solubility.

The SaADH2 enantioselectivity was further examined with other aromatic ketones using the system BsADH/1-butanol. Table 4 shows that SaADH2 preferably reduced benzil to (*R*)-benzoin with an *ee* of 98 % and 98 % conversion after 24 h of reaction, and that the selectivity was reduced when 2-propanol instead of 1-butanol was used. However, 2,2′-dichlorobenzil was reduced to (*R*)-1,2-bis(2-chlorophenyl)-2-hydroxyethanone with lower conversion (33 %) and *ee* of 91 %, despite the apparent inactivity under different assay conditions (Table 2). Presumably, the presence of chlorine, a bulky

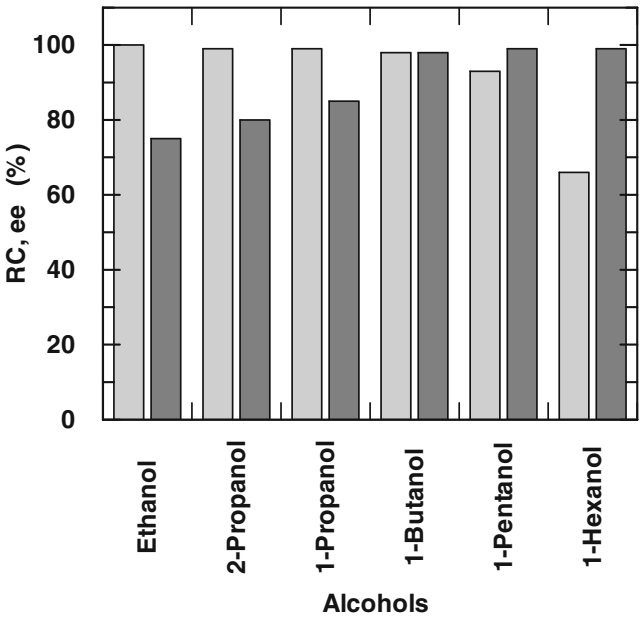

Fig. 5 Reduction of benzil catalyzed by SaADH2. Biotransformations were carried out at 50 °C with different alcohols at 18 % (*v*/*v*) concentrations. The reactions were stopped after 24 h by addition of ethyl acetate. The dried extracts were analysed by chiral HPLC to determine the relative conversion (*RC*, *grey bars*) and the enantiomeric excess (*ee*, *dark grey bars*). Error limit, 2 % of the stated values. The alcohols are plotted according to increasing log *P* (*P*=partition coefficient) values; from *left to right*: log *P*=−0.24, 0.07, 0.28, 0.8, 1.3 and 1.8. Log *P* values were from Laane et al. (1987)

atom with electron-withdrawing properties, disturbs the fitting of the diketone molecule in the substrate-binding pocket as well as the interactions that stabilize the transition state. Acetophenone was reduced to (*S*)-1-phenylethanol with a 4 % conversion and an *ee* of 92 %, while no conversion was observed for camphor, the ketone corresponding to the highly preferred alcohol isoborneol (Table 3). Kinetic tests showed that neither racemic benzoin nor (*R*)-benzoin was oxidized in the presence of NAD^+ as well as both were not reduced in the presence of NADH. Accordingly, the HPLC analysis indicated that no further reduction of benzoin occurred since no trace of hydrobenzoin could be found within the detection limits (data not shown).

Structural characterization

The determined crystal structure reveals a tetrameric organization similar to other members of the short-chain dehydrogenase/reductase (SDR) family. Although SaADH2 has low sequence similarities with other SDR proteins, it shares common structural features with them (Kallberg et al. 2002; Kavanagh et al 2008). The overall structure is typical of SDRs that is an α/β fold with characteristic Rossmann-fold motifs for the binding of the dinucleotide cofactor at the N-terminal region of the chain. Indeed, the structure confirmed that SaADH2 is a NAD-dependent enzyme with a cofactor-binding domain composed of a seven-stranded parallel sheet flanked by helices on each side.

Overall tertiary and quaternary structure

The structural comparison of SaADH2 with other SDRs demonstrated that SaADH2 exhibits a high degree of overall structural homology. Indeed, when aligning the monomeric structures by the DALI server the first 100 best superimpositions showed RMS deviations on C^{α} atoms ranging from 1.5 to 2.5 Å (with a minimum of 222/255 aligned residues). Some structural neighbours are superimposed onto a single chain of SaADH2 in Fig. S3. Major conformational differences are found on the molecular surface and at the terminal region of the polypeptide chain.

Analysis of the crystal packing (using PISA server:http://www.ebi.ac.uk/pdbe/prot_int/pistart.html), indicates that the tetramer is the biologically active form. The tetramer shows 222 point group symmetry and approximate dimensions of 80×70×55 Å (Fig. S4). The largest dimeric interface is the one involving AB or CD assemblies (~1,630 $Å^2$). The central elements in this interface are the long α6 helices (residues 156–179, Fig. S5) facing each other from adjacent subunits. Other segments involved in this interface are α4 and α5 helices (Fig. S5). The second large interface involving AC or BD assemblies (~1,250 $Å^2$) mainly comprises the C-terminal regions (residues 216–255) of the two subunits. The very terminal segment (residue 253–255) also

Table 4 Asymmetric reduction of carbonyl compounds by SaADH2

Substrate	Product	Conversion (%)	ee (%)	*R*/*S*
Benzil	Benzoin	98 (99)	98 (80)	*R*
2,2′-Dichlorobenzil	1,2-Bis(2-chlorophenyl)-2-hydroxyethanone	33	91	*R*
Acetophenone	1-Phenylethanol	4	92	*S*

Reactions were performed at 50 °C for 24 h as described in Materials and methods, using 18 % 1-butanol as ancillary substrate. Conversion and enantiomeric excess were determined by chiral HPLC analysis. Configuration of product alcohol was determined by comparing the retention time with that of standard samples as described in Material and methods. Values in parentheses refer to data obtained with 18 % 2-propanol as ancillary substrate

contributes to the smallest interface involving AD or BC assemblies (~445 Å^2); indeed, this segment forms inter-subunit interactions with residues belonging to 150–153 segment. It is worth noting that residues from the 209–218 region of the structure also mediate inter-subunit interactions in this interface.

Cofactor-binding site

A highly defined electron density was observed for the whole NAD(H) molecule thus allowing unambiguous identification of the positions of the cofactor in all the four molecules of the asymmetric unit (Fig. S6). The cofactor-binding site is located at the C-terminal end of the central parallel β-sheet.

The NAD(H) cofactor specificity is explained by the presence of Glu39 whose side chain oxygen atoms form hydrogen bonds with two oxygen atoms of the adenosine ribose of NAD(H).

The adenine ring of NAD binds in a hydrophobic pocket on the enzyme surface formed by the side chains of Val116, Leu40, Val66, Ala93 and Gly15. Hydrogen bonds are formed by N6 of adenine and the Asp65 side chain and by N1 and the peptide nitrogen of Val66. The segment 14-TGAGSGIG-21 corresponds to the glycine-rich sequence (TGxxx[AG]xG) present in NAD-binding site of "classical" SDR (Kavanagh et al 2008). This region interacts with the adenine ribose and the diphosphate moiety of NAD(H). The phosphate groups also interact with Thr193-Ala194. The 2′- and 3′-hydroxyls of the second ribose, attached to the nicotinamide ring, form a bifurcated hydrogen bond with Lys162. The 3′-hydroxyl is also hydrogen bonded to the active site Tyr158. The *syn* face of the nicotinamide ring displays vdW contacts with the hydrophobic side-chains of Ile20, Pro188 and Val191. The amide part of the nicotinamide portion is anchored to the protein by hydrogen bonds to Thr193 and Val191.

Substrate-binding site

The SaADH2 contains the conserved catalytic residues, Ser145, Tyr158 and Lys162, located in the active-site cleft with a spatial arrangement that is common to that found in other SDRs. SaADH2 also shows the conserved Asn residue (Asn117) which has been suggested as a key residue in a catalytic mechanism based on a tetrad instead of a triad (Filling et al. 2002). Its side chain atoms form hydrogen bonds with main chain atoms of Ile95.

In many SDRs the substrate-binding site is a deep cleft with a floor created by the NADH molecule and the right-side and left-side walls formed by the substrate-binding region (residues 193-221 in SaADH2 numbering) and the loop between β5 and β6 strands (Fig. 6).

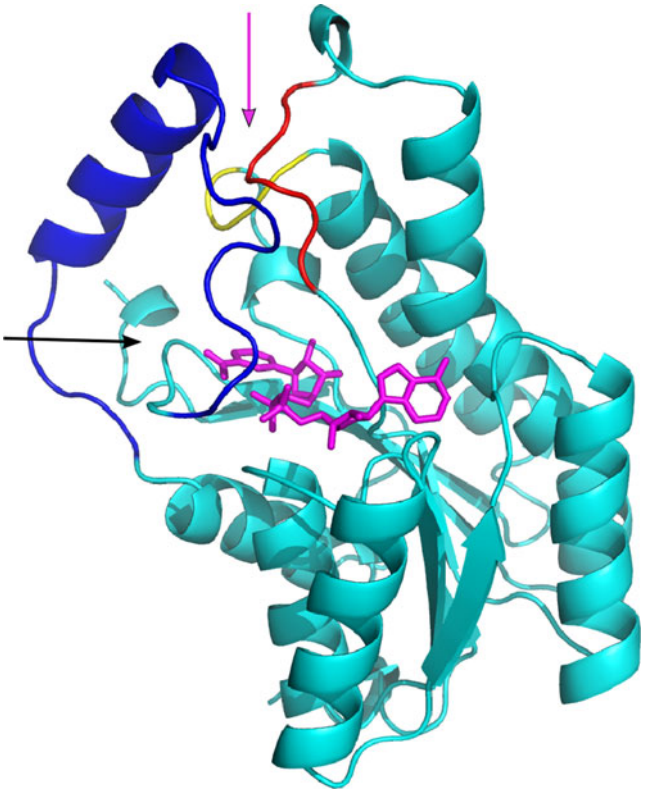

Fig. 6 Overall tertiary structure of one of the subunits of SaADH2. Bound NADH is shown in stick representation (*magenta*). The substrate-binding loop, the stretches 152–155 and 96–101 are highlighted in *blue*, *yellow* and *red*, respectively. The *arrows* point toward the substrate-binding cleft; the directions of approach are shown in *magenta* and *black* (see the text for details)

In three out of four subunits of the asymmetric unit there is a glycerol molecule occupying the position that is approximately suited for the binding of substrate. Indeed, the glycerol O2 and O3 oxygen atoms are hydrogen bonded to the oxygen of the NAD carboxamide and to the oxygen of Tyr158 side chain, respectively. Glycerol was used as cryo-protectant agent and therefore diffused through the water channels of the crystals up to the binding site. The lacking of glycerol binding in one of the subunits (chain B) of the tetramer is due to crystal packing which restricts the access to NADH in subunit B. Indeed, a nearby symmetric molecule (#B) inserts its 50–57 region into the substrate-binding region (residues 193–221) that covers the access to the NADH site. On the other hand, the other subunits of the tetramer do not show direct interactions with adjacent symmetric molecules through the substrate-binding loop.

It is worth noting that, despite the different crystal environments experienced by the four molecules of the asymmetric unit, the substrate-binding loops adopt similar conformations in the four subunits of the tetramer with pairwise RMSDs calculated on C^{α} atoms lying in the range 0.36–0.68 Å.

Furthermore, the structure of the loop (in blue in Fig. 6) represents the most significant difference between SaADH2 and the other SDR structures aligned with it. Indeed, as can be seen in Fig. S3, this loop (in cyan) is remarkably different from the other structures. In most of SDR structures the substrate-binding loop contains two segments of α-helix on opposite sides of the loop. In addition, part of the loop is often disordered thus revealing a considerable mobility. In SaADH2 the loop shows the following features: (1) it is well defined in the electron density; (2) it contains only one stretch of α-helix lacking the one usually present at the C-terminal end and (3) it partially obstructs the access to the substrate-binding cleft indicated with a magenta arrow in Fig. 6. Indeed, a detailed comparison with other SDR aligned structures reveals that three stretches (193–221, 96–101 and 152–155) of SaADH2 structure adopt a different conformation by getting closer to the NADH nicotinamide ring which becomes inaccessible to solvent/substrate from the top of the molecule (magenta arrow in Fig. 6). The cofactor ring is, however, still accessible to ligands through a side cavity (black arrow in Fig. 6) lined by residues Met96, Ser145, Ile146, Ala147, Phe153, Ala154, Tyr158, Gly189, Thr190, Ile195, Gly196, Leu197, Leu210 and Met214 (Fig. S7). The mouth of the cavity is composed of atoms from residues 190, 195–197, 210 and 214, as confirmed by CastP server (http://sts.bioengr.uic.edu/castp/calculation.php).

The benzil molecule is an interesting substrate of the enzyme that reduces it to *R*-benzoin in the presence of NADH with excellent conversion (>99 %). Several attempts were performed to crystallize a ternary complex of the enzyme bound to a benzil molecule, but they were unsuccessful. To envision how the enzyme binds the substrate and to probe the stereoselectivity of the hydride transfer we have manually modelled the benzil molecule in the active site.

The starting structure for the alpha-diketone modelling was the crystal structure present in the small molecule structure database (Allen 2002). This structure shows a conformation of the molecule with the two aromatic rings forming a dihedral angle of 114° around the diketo bond. A single benzil carbonyl group is bonded to two substituents: a smaller phenyl group and a larger carbophenyl group. In order to obtain a (*R*)-alcohol, the hydride has to be added to the *Si* face of the carbonyl ketone and this means that the smaller group is located opposite to the bulky carboxamide group of NADH.

A rigid docking of the benzil compound performed by the patchdock server identifies an orientation in the substrate site which is consistent with the stereoselectivity of the reduction reaction leading to (*R*)-benzoin. However, it has to be noted that in the best docked conformational ensemble of benzil, the carbonyl C atom, to be attacked by hydride, is still rather far from the hydrogen at the NADH C4 position (>5.5 Å). In addition, the oxygen atom of the carbonyl is not hydrogen bonded to the catalytic Tyr158 side chain (4.7 Å apart) as expected in case of a fully productive state of the enzyme–substrate complex.

Due to the steric hindrance of the benzil molecule an efficient accommodation in the substrate site would require a structural rearrangement of residues from the three segments (193–221, 96–101, and 152–155) which line the site and/or a change in the conformation of the diketone. A better packing of benzil in the site can be indeed obtained when the torsional angle around the carbon–carbon bond linking the two carbonyl moieties is reduced to values around 80–90°. Although approximate, this manually docked substrate position identifies a small pocket deep into the active site (Fig. 7) which is lined by Ser145, Leu146, Ala147, Tyr158, Phe153, Ala154 and Gly189. This pocket can accommodate the small phenyl group of benzil. Indeed, this phenyl ring would display mostly favourable contacts with Phe153, Ile146 and Ala147; the phenyl ring lies with its ring between two main-chain oxygen atoms from residues Phe153 and Glu189. On the other hand, the large carbophenyl group of benzil would point toward the channel opened to the solvent and lined by Met96, Thr190, Ile195, Gly196, Leu197, Leu210 and Met214. However, this bulky carbophenyl group would display steric clashes with the NADH and Thr190 when the intercarbonyl dihedral angle is about 80°. Some bad contacts can be relieved by a deviation of the phenyl group from the plane defined by the carbonyl moiety as well as by small rearrangements of the carboxyamide group of the nicotinamide ring (Fig. 7).

Conformational rearrangements with respect to the crystal structure of benzil can be accessible at a reasonable energetic cost since both experimental and theoretical studies have indicated a significant flexibility of the molecular structure even depending on the environmental conditions (Pawelka et al. 2001; Lopes et al. 2004).

The presence of the deep pocket in the active site, well suited for an aromatic ring, also explains why, in the case of acetophenone and its derivatives, the alcohol obtained by the reduction reaction is the (*S*)-enantiomer.

Discussion

In search of novel dehydrogenase/reductases that are both stable and NAD^+ dependent, a short-chain dehydrogenase (SDR) from *S. acidocaldarius* was identified and overexpressed. The enzyme was successfully purified and functionally/structurally characterized.

The enzyme is endowed with a high thermal stability showing a $T_{1/2}$ value of ~88 °C. It is rather stable in common organic solvents also exhibiting an increased activity in presence of DMSO and 1,4-dioxane. The activation

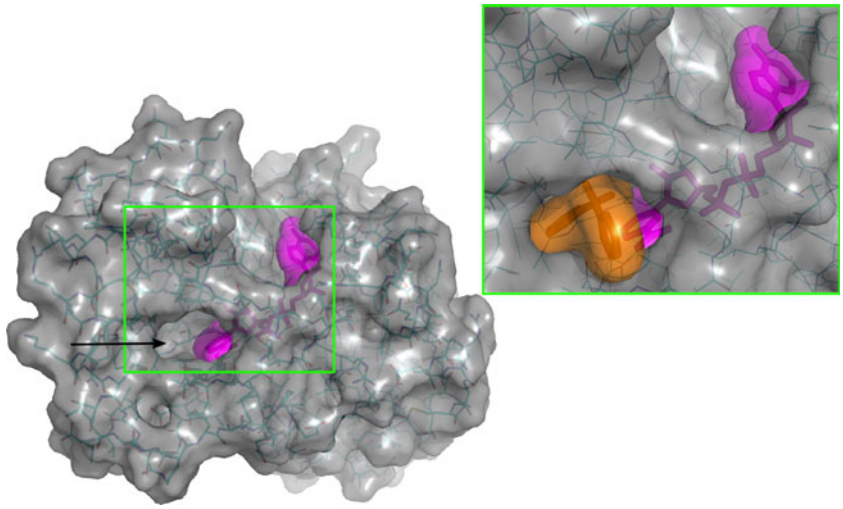

Fig. 7 Surface representation of a single chain of SaADH2. The NADH molecule is highlighted in *magenta*, as both stick and surface representations. The *black arrow* indicates the channel by which the nicotinamide ring is accessible. The *green box* is a close-up view of the binding site with the benzil molecule (in *orange*) manually docked into it. The figure was prepared using the PyMOL software (www.pymol.org)

of thermophilic enzymes by loosening of their rigid structure in the presence of protein perturbants is a well-documented phenomenon (Fontana et al. 1998; Liang et al. 2004). The two organic solvents discussed above (Fig. 3) may induce a conformational change in the protein molecule to a more relaxed and flexible state that is optimal for activity.

The analysis of the effects of ions reveals that SaADH2 does not require metals for its activity or structural stabilization; this is typical of non-metal SDR enzymes, although the well-known LB-RADH shows strong Mg^{++} dependency (Niefind et al. 2003).

Kinetic data show that SaADH2 is a strictly NAD(H)-dependent oxidoreductase with discrete substrate specificity. SaADH2 is more similar to the archaeal thermophilic TtADH and SaADH (Pennacchio et al. 2008; 2010a, b), than to the two representative SDR mesophilic ADHs, i.e. LB-RADH (Schlieben et al. 2005) and LSADH (Inoue et al. 2006), which are active on a variety of aliphatic as well as aromatic alcohols, ketones, diketones and keto esters. Indeed, the SaADH2 catalytic activity against various putative substrates was checked. Poor activity was observed on aliphatic linear and branched alcohols as well as on most of cyclic and bicyclic alcohols with the exception of cycloheptanol and isoborneol (Tables 2 and 3). Significant activity was found on alicyclic alcohols fused to an aromatic ring such as tetralol and indanol substrates. The kinetic data show that for most of chiral substrates SaADH2 is (*S*)-stereospecific and that the physiological direction of the catalytic reaction is reduction rather than oxidation (Table 3). The enzyme was active on aliphatic cyclic and bicyclic ketones such as cyclohexanone, methyl-substituted cyclohexanones and decalone (Table 2). It also showed a good reduction rate on halogenated acetophenone derivatives (Table 2) probably due to the electron withdrawing character of halogens which favours hydride transfer by decreasing electron density at the carbonyl carbon.

Interestingly, two aryl diketones, 1-phenyl-1,2-propanedione and benzil are good substrates of SaADH2. Although the natural substrates and function of SaADH2 remain unknown, the catalyzed reduction of the benzil α-diketone to (*R*)-benzoin with high yield and optical purity makes SaADH2 an interesting and biotechnologically relevant enzyme. Bioconversion studies have shown that SaADH2 can be conveniently coupled with a thermophilic bacillar ADH (BsADH) resulting in an efficient regeneration of NADH. Moreover, it is noteworthy that the SaADH2 enantioselectivity changes from moderate to excellent on going from ethanol to 1-hexanol (Fig. 5), i.e. that enantioselectivity increases as the hydrophobicity of the medium increases. The finely tuned solvent dependence of the prochiral selectivity shown by SaADH2 is remarkable and agrees with the general observation that enzyme selectivity (enantiomeric, prochiral and regiomeric) can change, or even reverse, from one solvent to another (Carrea and Riva 2000; Klibanov 2001). Overall, the solvent enantioselectivity study emphasizes the versatility and high efficiency of the BsADH/alcohol substrate system in recycling the reduced cofactor.

α-Hydroxy ketones are highly valuable building blocks for many applications for the fine chemistry sector as well as pharmaceuticals (Hoyos et al. 2010). It is noteworthy that optically pure (*R*)-benzoin is useful as urease inhibitor (Tanaka et al. 2004), and to prepare amino alcohols used as important chiral synthons (Aoyagi et al. 2000), or to synthesize different heterocycles (Wildemann et al. 2003). Many examples of stereoselective reduction of several benzils to the corresponding benzoins using whole microbial cells have been recently reviewed (Hoyos et al. 2010). The microorganisms *Bacillus cereus*, *Pichia glucozyma*, *Aspergillus oryzae*, *Fusarium roseum*, *Rhizopus oryzae* (at pH 4.5–5.0) and *B. cereus* Tm-r01 were reported to give (*S*)-benzoin while the microorganisms *R. oryzae* (at pH 6.8–8.5) and *Xanthomonas oryzae* were reported to transform benzil to (*R*)-benzoin with different yields and optical purity

(Hoyos et al. 2010 and references therein). However, recombinant *B. cereus* NADPH-dependent benzil reductase belonging to SDR superfamily which reduces benzil to 97 % optically pure (*S*)-benzoin in vitro was characterized (Maruyama et al. 2002). To our knowledge, this study represents the first biochemical characterization of an NAD(H)-dependent archaeal short-chain ADH reducing the diketone compound benzil to (*R*)-benzoin with high yield and optical purity.

The functional characterization of SaADH2 was complemented by the determination of the crystal structure of SaADH2–NADH complex at 1.75 Å resolution.

The determined crystal structure reveals a tetrameric organization that corroborates the results obtained by sucrose density gradient centrifugation. The structural comparison of SaADH2 with other SDRs demonstrated that SaADH2 exhibits a high degree of overall structural homology despite the low sequence similarities. The largest differences with other SDR structures are on the surface and at the C-terminal region of the polypeptide chain which contributes to the shaping of the active site.

The reductase activity measurements showed that the protein is not active with NADP(H), having a definite preference for NAD(H) cofactor. This cofactor specificity is consistent with structural results. Indeed, the side chain oxygen atoms of Glu39 form hydrogen bonds with two oxygen atoms of the adenosine ribose of NAD(H), thus hampering the binding of NADP(H). Therefore, the Glu residue plays the same role of the corresponding Asp residue found in many SDR structures.

The active site of SaADH2 structure contains the conserved catalytic residues, Ser145, Tyr158 and Lys162, which adopt position and orientation similar to those found in other SDRs (Schlieben et al. 2005). Besides the catalytic triad, also Asn117 is structurally conserved; this residue has been suggested to play a role in maintaining the active-site architecture and the building up of a proton-relay system in SDR structures (Filling et al. 2002; Schlieben et al. 2005).

A comparison with other structures of the SDR superfamily reveals that the substrate-binding loop (residues 193–221) is the most different region of the subunit structure (Fig. S3). This is not unexpected since the conformation, the length and the composition of the substrate-binding loop are rather variable in SDR proteins thus conferring different substrate specificity. Besides this loop, also the segment 152–155 shows a different conformation by getting closer to the NADH nicotinamide ring (Fig. 6). Interestingly, this region contains a Phe residue (Phe153) which points deep into the active site and can play a relevant role for substrate specificity. The region 96–101 is also closer to the nicotinamide ring thus contributing to the reduced accessibility of the site where hydride transfer has to occur (Fig. 6). As a consequence, the manual modelling of the bulky benzil molecule in the active site, guided by the position of glycerol molecule bound and by preliminary results of rigid docking, cannot avoid the presence of unfavourable contacts. However, it can be suggested that minor changes in the three segments of the protein structure discussed above as well as in the conformation of the benzil α-diketone may well accommodate this substrate in the site (Fig. 7).

The elucidation of the ternary complex structure will ultimately provide details of the conformational changes in both the enzyme and the substrate necessary to obtain a fully productive state which guarantees high enantioselectivity. Nonetheless, the modelling explains the enantioselectivity of the benzil reduction. The (*R*)-benzoin is obtained since the hydride transfer occurs on the *Si*-face of the carbonyl ketone attacked. There is a deep pocket in the active site which can accommodate the small phenyl group of benzil (Fig. 7). Indeed, this phenyl ring would display mostly favourable contacts with Phe153, Ile146 and Ala147. On the other hand, the large carbophenyl group of benzil would point toward the channel opened to the solvent (Fig. 7) and lined with hydrophobic residues (Fig. S7).

Furthermore, the modelling of benzil provides clues on the different stereoselective course of the reduction reaction for other substrates such as acetophenone and its derivatives. Indeed, the deep pocket in the active site, well suited for an aromatic ring, can efficiently accommodate the phenyl group of the acetophenone whereas the methyl group can point toward the channel; this orientation of the substrate in the site leads to the (*S*)-1-phenylethanol product (Table 4).

In conclusion, a new short-chain dehydrogenase from *S. acidocaldarius* was identified and overexpressed. The purified enzyme was shown to possess remarkable thermal resistance as well as high enantioselectivity on the aryl diketone benzil molecule. It shows many advantages with regard to its preparative application: ease of purification, preservability of the biocatalyst and absolute preference for NAD(H). It is also amenable for coupling with a thermophilic bacillar ADH (BsADH) in an efficient and sustainable bioconversion process based on coenzyme recycling, where both enzymes have only the coenzyme as co-substrate. The remarkable resistance of SaADH2 to organic solvents proves the sturdiness of this biocatalyst and suggests exploratory investigations on conversions of poorly water-soluble prochiral substrates. In addition, the availability of the crystal structure of the enzyme provides the opportunity to rationalize the stereospecificity and stereoselectivity of the enzyme and to design proper mutations to broaden the scope of possible substrates.

Acknowledgments This work was funded by FIRB (Fondo per gli Investimenti della Ricerca di Base) RBNE034XSW and by the ASI project MoMa n. 1/014/06/0.

Atomic coordinates as well as structure factors have been deposited within the PDB under the accession code 4FN4.

References

Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Crystallogr B 58:380–388

Aoyagi Y, Agata N, Shibata N, Horiguchi M, Williams RM (2000) Lipase TL-mediated kinetic resolution of benzoin: facile synthesis of (1R,2S)-erythro-2-amino-1,2-diphenylethanol. Tetrahedron Lett 41:10159–10162

Carrea G, Riva S (2000) Properties and synthetic applications of enzymes in organic solvents. Angew Chem Int Ed Engl 39:2226–2254

Ceccarelli C, Liang ZX, Strickler M, Prehna G, Goldstein BM, Klinman JP, Bahnson BJ (2004) Crystal structure and amide H/D exchange of binary complexes of alcohol dehydrogenase from *Bacillus stearothermophilus*: insight into thermostability and cofactor binding. Biochemistry 43:5266–5277

Chen L, Brügger K, Skovgaard M, Redder P, She Q, Torarinsson E, Greve B, Awayez M, Zibat A, Klenk HP, Garrett RA (2005) The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota*. J Bacteriol 187:4992–4999

Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, Oppermann U (2002) Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. J Biol Chem 277:25677–25684

Fiorentino G, Cannio R, Rossi M, Bartolucci S (1998) Decreasing the stability and changing the substrate specificity of the *Bacillus stearothermophilus* alcohol dehydrogenase by single amino acid replacements. Protein Eng 11:925–930

Fontana A, De Filippis V, Polverino de Laureto P, Scaramella E, Zambonin M (1998) Rigidity of thermophilic enzymes. In: Ballestreros A, Plou FJ, Iborra JL, Halling PJ (eds) Stability and stabilization in biocatalysis, vol 15. Elsevier Sciences, Amsterdam, pp 277–294

Friest JA, Maezato Y, Broussy S, Blum P, Berkowitz DB (2010) Use of a robust dehydrogenase from an archael hyperthermophile in asymmetric catalysis–dynamic reductive kinetic resolution entry into (S)-profens. J Am Chem Soc 132:5930–5931

Giordano A, Febbraio F, Russo C, Rossi M, Raia CA (2005) Evidence for co-operativity in coenzyme binding to tetrameric *Sulfolobus solfataricus* alcohol dehydrogenase and its structural basis: fluorescence, kinetic and structural studies of the wild-type enzyme and non-co-operative N249Y mutant. Biochem J 388:657–667

Goldberg K, Schroer K, Lütz S, Liese A (2007) Biocatalytic ketone reduction—a powerful tool for the production of chiral alcohols—part I: processes with isolated enzymes. Appl Microbiol Biotechnol 76:237–248

Guagliardi A, Martino M, Iaccarino I, De Rosa M, Rossi M, Bartolucci S (1996) Purification and characterization of the alcohol dehydrogenase from a novel strain of *Bacillus stearothermophilus* growing at 70 °C. Int J Biochem Cell Biol 28:239–246

Guy JE, Isupov MN, Littlechild JA (2003) The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*. J Mol Biol 331:1041–1051

Hoyos P, Sinisterra JV, Molinari F, Alcántara AR, Domínguez de María P (2010) Biocatalytic strategies for the asymmetric synthesis of alpha-hydroxy ketones. Acc Chem Res 43:288–299

Huisman GW, Liang J, Krebber A (2010) Practical chiral alcohol manufacture using ketoreductases. Curr Opin Chem Biol 14:122–129

Hummel W (1999) Large-scale applications of NAD(P)-dependent oxidoreductases: recent developments. Trends Biotechnol 17:487–492

Inoue K, Makino Y, Dairi T, Itoh N (2006) Gene cloning and expression of *Leifsonia* alcohol dehydrogenase (LSADH) involved in asymmetric hydrogen–transfer bioreduction to produce (*R*)-form chiral alcohols. Biosci Biotechnol Biochem 70:418–426

Jones JB, Beck JF (1976) Applications of biochemical systems in organic chemistry. In: Jones JB, Sih CJ, Perlman D (eds) Techniques of chemistry series, part I, vol 10. Wiley, New York, pp 248–401

Jörnvall H (2008) Medium- and short-chain dehydrogenase/reductase gene and protein families: MDR and SDR gene and protein superfamilies. Cell Mol Life Sci 65:3873–3878

Kallberg Y, Oppermann U, Jörnvall H, Persson B (2002) Short-chain dehydrogenases/reductases (SDRs). Eur J Biochem 269:4409–4417

Kavanagh KL, Jörnvall H, Persson B, Oppermann U (2008) The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. Cell Mol Life Sci 65:3895–3906

Keinan E, Hafeli EK, Seth KK, Lamed R (1986) Thermostable enzymes in organic synthesis. 2. Asymmetric reduction of ketones with alcohol dehydrogenase from *Thermoanaerobium brockii*. J Am Chem Soc 108:162–169

Klibanov AM (2001) Improving enzymes by using them in organic solvents. Nature 409(6817):241–246

Korkhin Y, Kalb(Gilboa) AJ, Peretz M, Bogin O, Burstein Y, Frolow F (1998) NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*. J Mol Biol 278:967–981

Kroutil W, Mang H, Edegger K, Faber K (2004) Recent advances in the biocatalytic reduction of ketones and oxidation of sec-alcohols. Curr Opin Chem Biol 8:120–126

Laane C, Boeren S, Hilhorst R, Veeger C (1987) Optimization of biocatalysis in organic media. In: Laane C, Tramper J, Lilly MD (eds) Biocatalysis in organic media, vol 29. Elsevier, Amsterdam, pp 65–84

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

Leatherbarrow RJ (2004) *GraFit* version 5.0.11. Erithacus Software Ltd, Horley

Liang ZX, Lee T, Resing KA, Ahn NG, Klinman JP (2004) Thermal-activated protein mobility and its correlation with catalysis in thermophilic alcohol dehydrogenase. Proc Natl Acad Sci U S A 101:9556–9561

Lopes S, Gòmez-Zavaglia A, Lapinski L, Chattopadhyay N, Fausto R (2004) Matrix-isolation FTIR spectroscopy of benzil: probing the flexibility of the C–C torsional coordinate. J Phys Chem A 108:8256–8263

Machielsen R, Uria AR, Kengen SW, van der Oost J (2006) Production and characterization of a thermostable alcohol dehydrogenase that belongs to the aldo-keto reductase superfamily. Appl Environ Microbiol 72:233–238

Machielsen R, Leferink NG, Hendriks A, Brouns SJ, Hennemann HG, Daussmann T, van der Oost J (2008) Laboratory evolution of *Pyrococcus furiosus* alcohol dehydrogenase to improve the production of (2S,5S)-hexanediol at moderate temperatures. Extremophiles 12:587–594

Maruyama R, Nishizawa M, Itoi Y, Ito S, Inoue M (2002) The enzymes with benzil reductase activity conserved from bacteria to mammals. J Biotechnol 94:157–169

Müller M, Wolberg M, Schubert T, Hummel W (2005) Enzyme-catalyzed regio- and enantioselective ketone reductions. Adv Biochem Eng Biotechnol 92:261–287

Nie Y, Xu Y, Mu XQ, Wang HY, Yang M, Xiao R (2007) Purification, characterization, gene cloning, and expression of a novel alcohol dehydrogenase with anti-prelog stereospecificity from *Candida parapsilosis*. Appl Environ Microbiol 73:3759–3764

Niefind K, Müller J, Riebel B, Hummel W, Schomburg D (2003) The crystal structure of R-specific alcohol dehydrogenase from *Lactobacillus brevis* suggests the structural basis of its metal dependency. J Mol Biol 327:317–328

Pawelka Z, Koll A, Zeegers-Huyskens Th (2001) Solvent effect on conformation of benzil. J Mol Struct 597:57–66

Pennacchio A, Pucci B, Secundo F, La Cara F, Rossi M, Raia CA (2008) Purification and characterization of a novel recombinant highly enantioselective, short-chain NAD(H)-dependent alcohol dehydrogenase from *Thermus thermophilus*. Appl Environ Microbiol 74:3949–3958

Pennacchio A, Giordano A, Pucci B, Rossi M, Raia CA (2010a) Biochemical characterization of a recombinant short-chain NAD (H)-dependent dehydrogenase/reductase from *Sulfolobus acidocaldarius*. Extremophiles 14:193–204

Pennacchio A, Giordano A, Esposito L, Langella E, Rossi M, Raia CA (2010b) Insight into the stereospecificity of short-chain *Thermus thermophilus* alcohol dehydrogenase showing pro-S hydride transfer and prelog enantioselectivity. Protein Pept Lett 17:437–443

Pennacchio A, Giordano A, Rossi M, Raia CA (2011) Asymmetric reduction of α-keto esters with *Thermus thermophilus* NADH-dependent carbonyl reductase using glucose dehydrogenase and alcohol dehydrogenase for cofactor regeneration. Eur J Org Chem 23:4361–4366

Persson B, Kallberg Y, Bray JE, Bruford E, Dellaporta SL, Favia AD, Duarte RG, Jörnvall H, Kavanagh KL, Kedishvili N, Kisiela M, Maser E, Mindnich R, Orchard S, Penning TM, Thornton JM, Adamski J, Oppermann U (2009) The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative. Chem Biol Interact 178:94–98

Radianingtyas H, Wright PC (2003) Alcohol dehydrogenases from thermophilic and hyperthermophilic archaea and bacteria. FEMS Microbiol Rev 27:593–616

Raia CA, Giordano A, Rossi M (2001) Alcohol dehydrogenase from *Sulfolobus solfataricus*. Methods Enzymol 331:176–195

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, New York

Schlieben NH, Niefind K, Muller J, Riebel B, Hummel W, Schomburg D (2005) Atomic resolution structures of R-specific alcohol dehydrogenase from *Lactobacillus brevis* provide the structural bases of its substrate and cosubstrate specificity. J Mol Biol 349:801–813

Tanaka T, Kawase M, Tani S (2004) Alpha-hydroxyketones as inhibitors of urease. Bioorg Med Chem 12:501–505

van der Oost J, Voorhorst WG, Kengen SW, Geerling AC, Wittenhorst V, Gueguen Y, de Vos WM (2001) Genetic and biochemical characterization of a short-chain alcohol dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Eur J Biochem 268:3062–3068

Wildemann H, Dünkelmann P, Müller M, Schmidt B (2003) A short olefin metathesis-based route to enantiomerically pure arylated dihydropyrans and alpha, beta-unsaturated delta-valero lactones. J Org Chem 68:799–804

Zhu D, Malik HT, Hua L (2006) Asymmetric ketone reduction by a hyperthermophilic alcohol dehydrogenase. The substrate specificity, enantioselectivity and tolerance of organic solvents. Tetrahedron-Asymmetry 17:3010–3014

Zhu D, Hyatt BA, Hua L (2009) Enzymatic hydrogen transfer reduction of α-chloro aromatic ketones catalyzed by a hyperthermophilic alcohol dehydrogenase. J Mol Catal B: Enzym 56:272–276