Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní program: Biochemie
Studijní obor: Biochemie

Mgr. Adéla Jílková

Strukturní a funkční analýza katepsinu B1 z krevničky *Schistosoma mansoni*

Structural and functional analysis of cathepsin B1 from the blood fluke, *Schistosoma mansoni*

Disertační práce

Školitel: RNDr. Michael Mareš, CSc.

Praha, 2014

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 7.7.2014

Podpis

Poděkování

Ráda bych zde poděkovala všem, kteří přispěli k realizaci této práce. Zejména děkuji svému školiteli Michalovi Marešovi za jeho cenné rady, vstřícnost, ochotu a čas, který mi věnoval, Martinovi Hornovi za praktické rady a věčný optimismus a Irče Pražákové za technickou pomoc v laboratoři. Rovněž děkuji i ostatním členům naší biochemické skupiny za všestrannou pomoc a vytvoření příjemného prostředí.

Ze servisních pracovišť bych chtěla poděkovat zejména Pavlíně Řezáčové za vždy laskavý přístup a veškerou pomoc v oblasti krystalografie, Zdeňkovi Voburkovi za sekvenování proteinů a Mirce Blechové za syntézu peptidů.

Díky patří i spolupracovníkům z týmu výpočetní chemie, zejména Martinovi Lepšíkovi, Jindrovi Fanfrlíkovi a Jiřímu Vondráškovi.

Rovněž děkuji všem kolegům ze spolupracujícího pracoviště UCSF v San Francisku, kde jsem mohla strávit několik krásných a přínosných měsíců a naučit se spoustu nového, zejména Jimovi McKerrowovi, Conorovi Caffreymu a Brianovi Suzukimu.

Na závěr bych chtěla poděkovat mé rodině a přátelům za velkou podporu a trpělivost v průběhu celého studia.

Obsah

Seznam zkratek

AMC	7-amino-4-methylkumarin
CA074	L-3-*trans*-(propylkarbamoyl)oxiran-2-karbonyl)-L-isoleucyl-L-prolin
CCP4	"Collaborative Computational Project No. 4", software
E-64	N-[N-(L-3-trans-karboxyoxirin-2-karbonyl)-L-leucin]-agmatin
FRET	"Fluroescence Resonance Energy Transfer"
GAG	glykosaminoglykan
$\Delta G'_{cov}$	rozdíl volné energie mezi kovalentním a nekovalentním komplexem
HB	"heparin-binding", vázající heparin
IC_{50}	koncentrace inhibitoru potřebná k dosažení 50% inhibice enzymu
K11777	N-methyl-piperazin-Phe-homoPhe-vinylsulfon-fenyl
k_{2nd}	rychlostní konstanta 2. řádu
k_{cat}	číslo přeměny
k_{cat}/K_m	katalytická účinnost enzymu
kDa	kilodalton
K_i	inhibiční konstanta
K_m	Michaelisova konstanta
PDB	"Protein Data Bank", databáze proteinových struktur
PISA	"Proteins, Interfaces, Structures and Assemblies"
RNAi	RNA interference
RP-HPLC	"Reverse Phase High Performance Liquid Chromatography"
$score_{cov}$	kovalentní skóre
$score_{nc}$	nekovalentní skóre
SDS-PAGE	elektroforéza na polyakrylamidovém gelu v přítomnosti SDS
SmAE	asparaginylendopeptidasa ze *Schistosoma mansoni*
SmCB1	katepsin B1 ze *Schistosoma mansoni*

SmCC	katepsin C ze *Schistosoma mansoni*
SmCD	katepsin D ze *Schistosoma mansoni*
SmCL	katepsin L ze *Schistosoma mansoni*
SmLAP	leucylaminopeptidasa ze *Schistosoma mansoni*
SP	sulfatovaný polysacharid
UCSF	"University of California San Francisco"
VS	vinylsulfon
Z	benzyloxykarbonyl

Abstrakt (ČJ)

Schistosomóza je závažné infekční onemocnění postihující více než 200 milionů lidí v oblasti tropů a subtropů. Původcem je parazit krevnička, který žije v cévách člověka a živiny získává degradací hemoglobinu z krve hostitele působením trávicích proteas. V současnosti je k dispozici pouze jediný lék proti schistosomóze a hrozba vzniku rezistence vyvolává potřebu vývoje nových terapeutik. Katepsin B1 (SmCB1) je klíčovou trávicí proteasou krevničky střevní (*Schistosoma mansoni*) a představuje cílovou molekulu pro terapeutický zásah. Tato práce podává komplexní charakterizaci SmCB1 zaměřenou na vztah struktury a aktivity a na inhibiční regulaci s využitím šesti vyřešených krystalografických struktur molekulárních forem a komplexů SmCB1.

SmCB1 je syntetizován ve formě neaktivního zymogenu, ve kterém N-koncový propeptid působí jako přirozený intramolekulární inhibitor blokující aktivní místo. Detailní biochemickou a strukturní analýzou byl identifikován nový unikátní mechanismus procesu aktivace zymogenu, při kterém je propeptid odštěpen, a byla popsána regulace tohoto procesu pomocí glykosaminoglykanů. Studium proteolytické aktivity SmCB1 prokázalo, že se jedná o enzym působící jako endopeptidasa i exopeptidasa a představuje tak účinný nástroj pro trávení hemoglobinu hostitele. Významná část práce se zabývala vyhledáním účinných molekul pro inhibici SmCB1 a to s využitím dvou přístupů. Zaprvé byly navrženy a syntetizovány reverzibilní peptidové inhibitory odvozené ze struktury propeptidu, které jsou funkční *in vitro* v mikromolárních koncentracích. Zadruhé byly identifikovány peptidomimetické inhibitory s vinylsulfonovou reaktivní skupinou, které jsou funkční *in vitro* v nanomolárních koncentracích. Mechanismus jejich interakce byl detailně studován pomocí krystalových struktur komplexů s SmCB1 a s využitím metod výpočetní chemie. Účinnost vinylsulfonových inhibitorů byla prokázána *ex vivo* při supresi živých parazitů.

Závěrem, tato práce pomohla definovat SmCB1 jako cílovou molekulu pro supresi *S. mansoni* a přinesla nové typy inhibitorů pro potenciální vývoj léčiv proti schistosomóze.

Abstract (EN)

Schistosomiasis is a serious infectious disease that afflicts over 200 million people in tropical and subtropical regions. It is caused by *Schistosoma* blood flukes that live in human blood vessels and obtain nutrients from host hemoglobin, which is degraded by digestive proteases. Current therapy relies on a single drug and concern over resistance necessitates new drug development. In *Schistosoma mansoni*, cathepsin B1 (SmCB1) is a critical digestive protease that is a target molecule for therapeutic interventions. This thesis provides a comprehensive characterization of SmCB1 focused on structure-activity relationships and inhibitory regulation based on six crystal structures solved for SmCB1 molecular forms and complexes.

SmCB1 is biosynthesized as an inactive zymogen in which the N-terminal propeptide operates as a natural intra-molecular inhibitor by blocking the active site. Detailed biochemical and structural analyses have identified a new and, so far, unique mechanism of SmCB1 zymogen activation through which the propeptide is proteolytically removed and the regulatory role of glycosaminoglycans in this process has been described. A study of SmCB1 proteolytic activity has revealed that the enzyme acts in two modes, as endopeptidase and exopeptidase, which makes it an efficient tool for host hemoglobin digestion. A major part of the thesis focused on the identification of new molecules for SmCB1 inhibition using two different approaches. First, reversible peptide inhibitors derived from the structure of the SmCB1 propeptide were designed and synthesized that are effective *in vitro* in the micromolar concentration range. Second, peptidomimetic inhibitors with a vinyl sulfone reactive group were identified that are effective *in vitro* in the nanomolar concentration range. The mechanism of the interaction between SmCB1 and the vinyl sulfone inhibitors was studied in detail using crystallographic structures of SmCB1 in complexes with the inhibitors and by computational chemistry methods. Finally, the *ex vivo* efficiency of the vinyl sulfone inhibitors was demonstrated by the suppression of live schistosoma parasites in culture.

To conclude, this thesis has defined SmCB1 as a target molecule for the suppression of *S. mansoni* and has identified new types of inhibitors for the development of potential anti-schistosomal drugs.

1. Úvod

Schistosomóza je chronické onemocnění vyvolané parazity rodu krevnička (*Schistosoma*). Touto nemocí je infikováno více než 200 milionů lidí v tropických a subtropických oblastech světa a představuje závažný zdravotní i sociálně ekonomický problém [1-4]. Tři nejvýznamnější druhy krevničky ohrožující člověka jsou *Schistosoma mansoni* (krevnička střevní), *S. japonicum* (krevnička jaterní) a *S. haematobium* (krevnička močová). K infekci člověka dochází při kontaktu s kontaminovanou vodou, ve které se nacházejí infekční larvy krevničky schopné penetrace nepoškozenou kůží člověka. Přes tkáně se dostávají do krevního oběhu, kde rostou a vyvíjejí se v dospělé krevničky, které pak žijí v cévách obklopujících střevo nebo urogenitální trakt. Samička zde klade denně až tisíce vajíček, která jsou hlavním patogenním agens schistosomózy [5]. Imunitní odpověď na vajíčka způsobuje záněty, vznik granulomů, vede k poškození orgánů a při infekci bývá zvýšeno riziko vzniku rakoviny [6].

Proti schistosomóze v současnosti neexistuje žádná očkovací látka a pro léčbu se používá téměř výhradně lék praziquantel [7]. S ohledem na to, že praziquantel je jediný účinný lék a nepřetržitě se používá již více než 20 let, stále narůstá pravděpodobnost vzniku rezistence a potřeba vývoje nových alternativních léčiv [8-12]. Tento výzkum se v současnosti zaměřuje hlavně na porozumění biochemie parazita a jeho molekulárních drah a na hledání cílových molekul, zásadních a unikátních pro parazita, jejichž blokování by bylo pro parazita letální.

Proteasy hrají velmi důležitou úlohu ve všech vývojových stádiích krevničky, zejména při interakci s hostitelem. Účastní se procesů invaze do hostitele, migrace tkáněmi, degradace krevních proteinů v procesu trávení, překonání imunitního systému hostitele a aktivace a modulace zánětu [13]. Dospělé krevničky žijící v cévách člověka získávají živiny pro růst a vývoj odbouráváním krevních proteinů. Proteolytické trávení je pro přežití parazita zásadní a je předmětem intenzivního vědeckého výzkumu. Na degradaci hemoglobinu se podílí kaskáda trávicích enzymů z tříd cysteinových, serinových, aspartátových proteas a metaloproteas. Ve střevě krevničky *S. mansoni* byly identifikovány proteasy katepsin B1, katepsin C, katepsiny L1, L2 a L3, katepsin D, asparaginylendopeptidasa a leucylaminopeptidasa [14-21].

Katepsin B1 ze *S. mansoni* (SmCB1), kterým se zabývá tato práce, je nejvíce zastoupená proteasa ve střevě krevničky a hraje v degradaci hemoglobinu klíčovou roli [14;22]. Tento

enzym působí komplexně jako endopeptidasa i exopeptidasa [14]. SmCB1 je cílová molekula pro léčbu schistosomózy, jak bylo prokázáno *in vivo* na myším modelu pomocí inhibitoru K11777 blokujícího aktivitu SmCB1 [23]. Účinné inhibitory SmCB1 tak představují perspektivní molekuly pro vývoj chemoterapeutik proti schistosomóze. Důkladná strukturně funkční charakterizace SmCB1 předkládaná v této práci je nezbytná pro racionální design těchto nových inhibičních molekul a pro pochopení biochemických mechanismů aktivity a regulace SmCB1.

2. Literární přehled

2.1. Katepsin B v kontextu cysteinových proteas

2.1.1. Proteasy, katepsiny a jejich dělení

Proteasy (proteolytické enzymy, peptidasy) jsou enzymy katalyzující hydrolýzu peptidové vazby proteinů a peptidů. Vyskytují se ve všech živých organismech a jsou zapojeny v mnoha důležitých biologických a fyziologických procesech, jako je například lysosomální degradace proteinů, krevní srážení, trávení či apoptóza [24;25]. Podle uspořádání aktivního místa a katalytického mechanismu se proteasy dělí do tříd na cysteinové, serinové, aspartátové, threoninové, glutamátové a metaloproteasy. Podle způsobu štěpení substrátu (peptidového řetězce) se rozlišují endopeptidasy, štěpící uvnitř řetězce, a exopeptidasy, odštěpující jednotlivé aminokyseliny (nebo skupiny aminokyselin) z jeho N nebo C konce.

Katepsiny byly původně považovány za savčí proteasy zajišťující výhradně nespecifické intracelulární odbourávání proteinů uvnitř lysosomů. Dnes je známo, že hrají důležitou úlohu i při specifické aktivaci a regulaci řady proteinů v mnoha fyziologických procesech. Kromě lysosomů byly navíc lokalizovány i v jiných kompartmentech buňky (cytoplasma, buněčné jádro) nebo sekretované z buňky do extracelulární matrix [26-28]. Proteasy katepsinového typu byly rovněž identifikovány u jiných organismů než u savců. Například u parazitů (klíšťat, helmintů, a prvoků) se účastní interakcí s hostitelem, u herbivorního hmyzu interakce s rostlinou a u rostlin se podílejí na obranyschopnosti proti patogenům. Katepsiny dnes představují velmi různorodou skupinu zahrnující enzymy ze tří proteasových tříd včetně cysteinových (katepsin B, C, L, K, S, F, H, O, V, X, W), serinových (katepsin A a G) i aspartátových proteas (katepsin D a E), které jsou klasifikovány podle savčích modelových katepsinů [25].

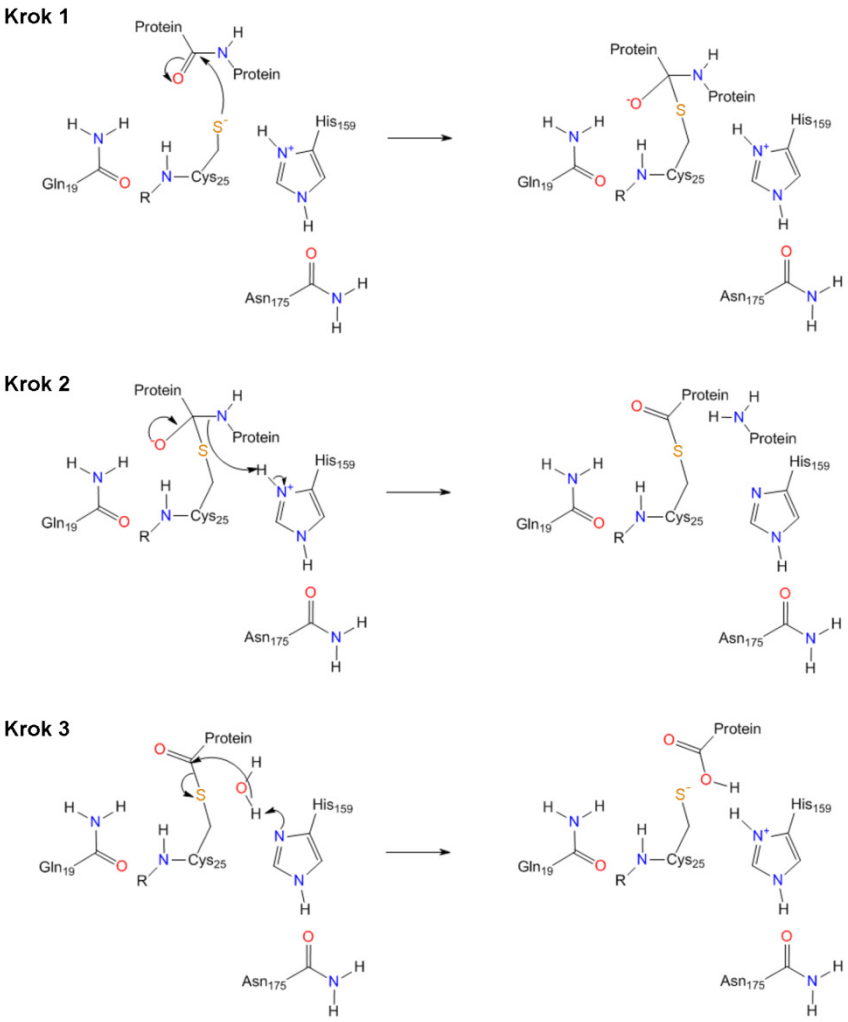
2.1.2. Cysteinové katepsiny

2.1.2.1. Funkce a struktura cysteinových katepsinů

Cysteinové proteasy vyžadují pro svou hydrolytickou aktivitu přítomnost cysteinu s nukleofilní thiolovou skupinou v aktivním místě. Podle homologie jejich primárních struktur se v databázi MEROPS cysteinové proteasy dělí do zhruba 70 rodin, z nichž ***C1 rodina papainu*** je nejpočetnější a nejlépe charakterizovaná [29]. Do této rodiny patří kromě modelové proteasy papainu i příbuzné enzymy rostlin (např. bromelain, chymopapain), virů, prvoků, bakterií, hub a živočichů [25]. Řadí se sem i velká skupina

cysteinových katepsinů, kterou tvoří zejména savčí katepsiny, ale i homologické enzymy katepsinového typu přítomné u řady obratlovců, bezobratlých i rostlin. U savců bylo nalezeno 11 zástupců cysteinových katepsinů (katepsin B, C, F, H, K, L, O, S, V, X a W), obecně jsou rozšířeny téměř ve všech tkáních a buňkách a podílejí se kromě lysosomální degradace proteinů například na procesech aktivace a prezentace antigenů, remodelace extracelulární matrix a aktivace hormonů (přehledně v [30]). Jejich chybná regulace může vyvolat patologické stavy jako rakovina, ateroskleróza, osteoporóza či revmatická artritida, jsou proto častým cílem vývoje nových terapeutik [31-33]. Medicinálně významné jsou rovněž parazitární cysteinové katepsiny, které zajišťují základní biologické funkce parazitů a jsou vhodnými cílovými molekulami pro vývoj nových antiparazitik. Výzkum se zaměřuje například na vývoj inhibitorů proteas typu katepsinu L parazitů *Trypanosoma cruzi* (cruzain) a *Plasmodium falciparum* (falcipain) jako potenciálních léčiv proti Chagasově chorobě a malárii [34;35].

Katalytický mechanismus cysteinových katepsinů je stejný jako pro ostatní cysteinové proteasy rodiny papainu a je popsán na obr. 1 na str. 9. Aktivní místo je tvořeno dvěma vysoce konzervativními aminokyselinovými zbytky Cys25 a His159 (číslování papainu). Další aminokyselinové zbytky, které napomáhají při mechanismu katalýzy, jsou Asn175 a Gln19. V základním stavu je thiolová skupina Cys25 deprotonovaná a vytváří thiolát-imidazolový iontový pár s histidinem His159. Iontový pár je stabilizovaný zbytkem Asn175. Hydrolýzu substrátu začíná nukleofilní atak thiolové skupiny Cys25 na karbonylovou skupinu štěpené amidové vazby za tvorby tetrahedrálního intermediátu stabilizovaného tzv. oxyanionovou dírou obsahující Gln19 (krok 1, obr. 1, str. 9). Následuje acylace enzymu, štěpení peptidové vazby a uvolnění prvního produktu, aminu, který deprotonuje His159 (krok 2, obr. 1, str. 9). His159 následně deprotonuje molekulu vody, která atakuje karbonylový uhlík acyl-enzymu za tvorby druhého tetrahedrálního intermediátu. Následuje deacylace enzymu (uvolnění druhého produktu) a regenerace enzymu do základního stavu (krok 3, obr. 1, str. 9).

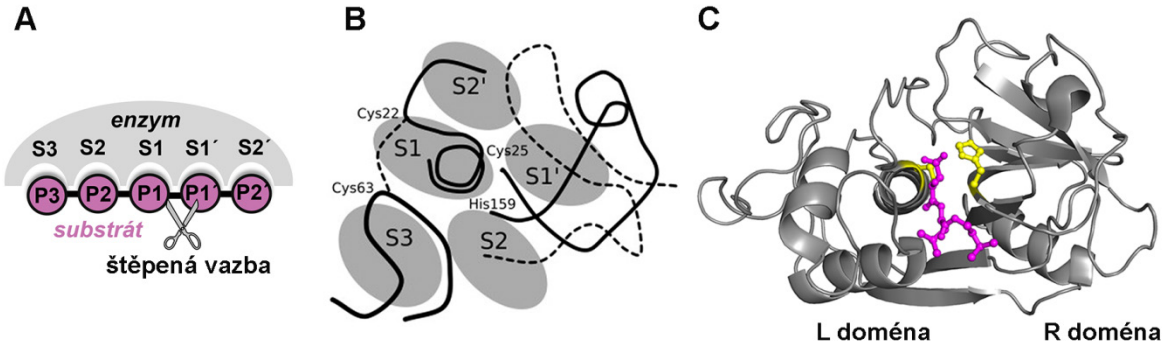


Obr. 1. ***Katalytický mechanismus cysteinových proteas.*** *Detaily jsou uvedeny v textu. Číslování aminokyselinových zbytků je dle papainu. Převzato a upraveno podle [36].*

Cysteinové katepsiny jsou ***biosyntetizovány*** jako prekurzory (preproenzymy) složené ze signální sekvence, propeptidu a zralého enzymu [37]. Signální sekvence zajišťuje transport biosyntetizovaného proteinu přes membránu endoplasmatického retikula a poté je odštěpena. Propeptid napomáhá správnému sbalení enzymu a řízenému transportu do cílových kompartmentů (především lysosomů, endosomů), přičemž zůstává navázán do aktivního místa zymogenu (proenzymu) a brání tak nežádoucí proteolytické aktivitě enzymu [38]. Ze zymogenu je propeptid proteolyticky odštěpen během aktivace buď pomocí jiné proteasy [39] nebo autokatalyticky [40;41]. Stavba propeptidu je jedním z důležitých rysů, kterými se odlišují katepsiny typu B, typu L a typu F [42;43]. Katepsiny typu L mají delší propeptid (více než 100 aminokyselin) a obsahují charakteristickou sekvenci ERFNIN. Katepsiny typu F se vyznačují sekvenčním motivem ERFNAQ.

Katepsiny typu B nemají žádný z uvedených sekvenčních motivů a délka propeptidu je obvykle kolem 60 aminokyselin.

Známé ***krystalové struktury*** cysteinových katepsinů vykazují značnou vzájemnou podobnost a obsahují charakteristické strukturní znaky modelové proteasy papainu [44]. Zralý aktivní enzym je protein o molekulové hmotnosti kolem 25-35 kDa, výjimkou je katepsin C, který tvoří tetramer o molekulové hmotnosti 200 kDa [45]. Molekula je tvořena levou a pravou doménou, mezi nimiž je uložena dutina aktivního místa s katalytickými zbytky Cys25 a His159 (číslování papainu). Aktivní místo obsahuje pět relativně dobře definovaných vazebných podmíst, značených S3, S2, S1, S1´ a S2´, do kterých se váží odpovídající aminokyselinové zbytky substrátu P3, P2, P1, P1´ a P2´ (obr. 2) [46]. Proteasy rodiny papainu mají poměrně širokou ***substrátovou specificitu***, pro vazbu substrátu je určující zejména vazebné podmísto S2, ve kterém jsou preferovány velké hydrofobní aminokyseliny. Substrátová specificita pro jednotlivé enzymy byla detailně studována a je dostupná v databázích MEROPS (http://merops.sanger.ac.uk) nebo BRENDA (http://www.brenda-enzymes.org). Mezi katepsinové proteasy patří hlavně endopeptidasy, exopeptidasová aktivita některých proteas (katepsin H, B, C, X) je důsledkem specifických strukturních elementů, které částečně blokují některá vazebná podmísta [45;47;48].



Obr. 2. ***Krystalová struktura a popis aktivního místa cysteinových proteas rodiny papainu.*** *U cysteinových proteas rodiny papainu je v aktivním místě definováno pět vazebných podmíst.* ***A,*** *Schematické znázornění vazby jednotlivých aminokyselin peptidového substrátu (P3, P2, P1, P1´, P2´, fialově) do odpovídajících vazebných podmíst enzymu (S3, S2, S1, S1´, S2´, šedě). Převzato a upraveno podle [13].* ***B,*** *Schematické znázornění aktivního místa cysteinových proteas rodiny papainu s prostorovým rozmístěním vazebných podmíst. Definování podmíst vychází z krystalových struktur proteas v komplexu s peptidovými inhibitory navázanými do aktivního místa. Zobrazeny jsou katalytické zbytky Cys25 a His159 a dále Cys22 a Cys63, které tvoří disulfidický můstek (číslování papainu). Převzato z [37].* ***C,*** *Prostorová struktura papainu (šedě) s molekulou inhibitoru E-64 (fialově) kovalentně navázanou v aktivním místě (PDB kód: 1PPP). Struktura je rozdělena na levou a pravou doménu (L a R), uprostřed které se nachází aktivní místo s katalytickými zbytky Cys25 a His159 (žlutě). Připraveno pomocí programu Pymol.*

2.1.2.2. Regulace aktivity cysteinových katepsinů

Aktivita cysteinových katepsinů musí být přísně regulována, aby nedocházelo k nekontrolované proteolýze, která může vyvolat řady patologií. První regulace nastává už na ***úrovni genové exprese*** a dále ***řízeným transportem*** do cílových kompartmentů buňky [28]. Může také docházet k alternativnímu sestřihu mRNA transkriptů, což ovlivňuje rychlost proteinové exprese a vede k nežádoucí extracelulární sekreci či změně cílového kompartmentu [49;50].

Dalším regulačním mechanismem je proces ***aktivace zymogenů*** na aktivní enzymy. Cysteinové katepsiny jsou syntetizovány jako neaktivní zymogeny (prokatepsiny), ve kterých N-koncový propeptid funguje jako přirozený intramolekulární inhibitor a blokuje aktivní místo. Orientace propeptidu v aktivním místě je opačná než při vazbě substrátu, čímž se zamezuje hydrolýze peptidové vazby [38]. Proteolytické odstranění propeptidu při aktivaci probíhá buď autokatalyticky, nebo vyžaduje přítomnost jiného enzymu. Tuto aktivaci mohou ovlivňovat i parametry prostředí jako pH a redoxní potenciál. Propeptid po odštěpení disociuje, je rozbalen a degradován proteasami [51]. Již během aktivace může být propeptid štěpen na několika různých místech, což snižuje jeho inhibiční funkci [41;52]. Inhibiční schopnost a selektivita propeptidu pro danou proteasu z něho činí nástroj pro vývoj nových inhibitorů [53]. *In vitro* může uvolněný propeptid i po odštěpení působit jako inhibitor aktivovaného enzymu [54], přičemž byla prokázána částečná selektivita interakce propeptidu s rodičovským enzymem [55;56].

Aktivita katepsinů je regulována pomocí ***pH.*** Většina lidských katepsinů (s výjimkou katepsinu S [57]) má pH optimum v mírně kyselé oblasti a plná aktivita je tak zajištěna hlavně v prostředí lysosomů [58;59]. V neutrálním pH jsou lysosomální katepsiny obvykle rychle a ireversibilně degradovány, což představuje kontrolní mechanismu pro potlačení jejich nežádoucí proteolytické aktivity při proniknutí do cytosolu [59]. Naproti tomu mnoho parazitárních cysteinových katepsinů má pH optimum v neutrální oblasti pH, což odpovídá jejich roli v extracelulárních procesech jako je např. invaze do buněk a tkání, encystace a obrana před imunitním systémem hostitele [13;60].

Novější studie ukazují, že aktivita i stabilita cysteinových proteas může být modulovaná ***glykosaminoglykany (GAG)***, což jsou sulfatované polysacharidy se silným negativním nábojem přirozeně se vyskytující v extracelulární matrix a na povrchu buněk. GAG zvyšují stabilitu katepsinu B v neutrálním pH [61], modulují kolagenolytickou aktivitu katepsinu K

[62] a působí jako alosterické modulátory na trypanosomální katepsin L, brucipain [63]. GAG rovněž urychlují autoaktivaci několika cysteinových katepsinů [64-67]

Největší možnost selektivní regulace poskytují různé typy inhibitorů. Cysteinové katepsiny jsou inhibovány zejména přirozenými ***makromolekulárními proteinovými inhibitory*** z rodiny tyropinů, cystatinů, stefinů, kininogenů, chagasinů, clitocypinů, macrocypinů, Kunitzova typu a některými serpiny [68-71]. Mezi nízkomolekulární inhibitory izolované z přírodních zdrojů patří leupeptin (Ac-Leu-Leu-Arg-H) nebo epoxysukcinát E-64 (trans-epoxysukcinyl-L-leucylamido-4-guanidinobutan).

V ***in vitro pokusech*** je E-64 široce používaný kovalentní ireverzibilní inhibitor cysteinových proteas a jeho objevení zahájilo vývoj nových ***syntetických derivátů.*** Byl z něho odvozen specifický inhibitor pro proteasy typu katepsinu B, CA074, a další jeho analoga [72]. Další syntetické deriváty peptidů (tzv. peptidomimetika), které se používají pro inhibici papainových proteas *in vitro* jsou jednak inhibitory s reverzibilním účinkem (peptidyl aldehydy, methylketony, nitrily) a dále inhibitory s ireverzibilním působením (diazomethany, halomethylketony, acyloxymethylketony, epoxysukcinylové deriváty a vinyl sulfony) [37]. Tyto inhibitory se liší reaktivní skupinou, reakčním mechanismem a specifitou vůči různým proteasám [73]. Chemicky reaktivní inhibitory jsou využívány i pro konstrukci značených proteomických prób tzv. "***activity-based probes***" pro detekci a vizualizaci cysteinových katepsinů. Reaktivní část "activity-based probes" na bázi inhibitoru se kovalentně váže do aktivního místa enzymu a je propojena s detekční značkou (např. fluorofor, biotin, radioizotop), která umožňuje následnou detekci. "Activity-based probes" jsou vysoce selektivní a slouží k vizualizaci aktivních cysteinových katepsinů v komplexních biologických materiálech jako jsou tkáňové homogenáty, buňky nebo i celé organismy [74].

2.1.3. Katepsin B: biosyntéza, aktivita a inhibice

Katepsin B (EC 3.4.22.1) je hojně se vyskytující cysteinová proteasa rodiny papainu. U ***savců*** je exprimován téměř ve všech tkáních a orgánech. V lysosomech byl nalezen ve velmi vysokých koncentracích (až 1 mM) [75]. Kromě nespecifické lysosomální degradace proteinů se podílí i na proteolytické úpravě (processingu) jiných enzymů [76], antigenů [77] či prekurzorů hormonů [78]. Velký zájem je o výzkum katepsinu B hlavně v souvislosti s různými patologickými stavy jako je revmatoidní artritida [79], osteoartróza [80] a různé druhy rakoviny, kdy je většinou spojován se stádii progrese a invaze tumorů

[81-83] a je považován za prognostický marker [84;85]. V nádorové tkáni je zvýšena exprese katepsinu B, je sekretován z buněk [86;87] a podílí se na degradaci extracelulární matrix [88]. Katepsin B uvolněný z lysosomů do cytosolu může spouštět apoptosu [89;90].

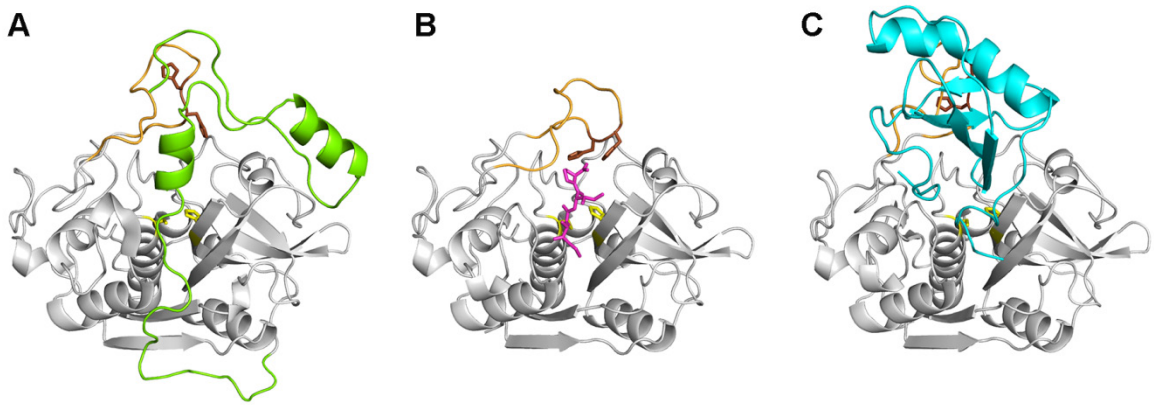
U ***rostlin*** se katepsin B účastní obranných mechanismů spojených s invazí patogenů a procesu stárnutí rostlin [91;92]. Proteasy typu katepsinu B jsou také využívány řadou ***parazitů*** pro zajištění základních metabolických funkcí (trávení potravy a příjem živin) a interakcí mezi hostitelem a parazitem (invaze do hostitele, migrace tkáněmi, obrana před hostitelským imunitním systémem) [93]. Katepsin B využívají např. parazitičtí prvoci (*Trypanosoma* [94], *Leishmania* [95], *Toxoplasma* [96]), motolice (*Schistosoma* [20], *Trichobilharzia* [97], *Fasciola* [98;99]), hlístice (*Necator* [100], *Ancylostoma* [101]) či členovci (*Ixodes* [102]). Tato práce se zabývá katepsinem B1 z krevničky *S. mansoni*, označovaným SmCB1, který hraje klíčovou roli v trávení parazita [21].

Základní informace o struktuře a aktivaci katepsinu B byly získány pro savčí enzymy. Stejně jako ostatní proteasy rodiny papainu je katepsin B ***syntetizován*** jako preproenzym. Propeptid lokalizovaný na N konci proenzymu se váže po celé délce aktivního místa, čímž blokuje aktivitu proteasy a působí jako intramolekulární inhibitor. U lidského prokatepsinu B je propeptid tvořen 62 aminokyselinami [103]. Po odštěpení je propeptid účinný a značně selektivní inhibitor zralého katepsinu B [55]. Komplex enzymu s propeptidem je stabilní v neutrálním pH [54], ale disociační konstanta komplexu se zvyšuje v kyselé oblasti pH. K aktivaci proteolytickým odštěpením propeptidu dochází autokatalyticky v kyselém prostředí lysosomů, ale není vyloučeno ani působení jiné proteasy [40;104].

pH optimum ***in vitro autoaktivace*** savčích katepsinů B je kolem 4.5 [65;105] a jedná se zřejmě o kombinaci unimolekulárního a bimolekulárního procesu [106]. Předpokládá se, že při nižším pH, kdy je interakce mezi propeptidem a enzymem zeslabena [55;107], dojde k částečnému uvolnění propeptidu z aktivního místa, čímž vzniká prokatepsin B se stopovou katalytickou aktivitou. Tato konformační změna je unimolekulární proces. Další krok je bimolekulární, kdy částečně aktivní prokatepsin B může odštěpit propeptid druhé molekuly proenzymu za vzniku zralého katepsinu B. Ten je pak schopen aktivovat další molekuly proenzymu [106]. Přesný popis tohoto schématu je v současné době stále značně hypotetický, zatím nebyla jednoznačně dořešena řada aspektů včetně existence a úlohy velmi nestabilních intermediátů během procesu aktivace. Proces autoaktivace lidského katepsinu B *in vitro* je usnadněn v přítomnosti glykosaminoglykanů, které indukují konformační změny a konvertují prokatepsin B na lepší substrát pro aktivní katepsin B

[64]. Aktivace pomocí glykosaminoglykanů může probíhat i při vyšším pH (kolem 6.5) [65].

Zralý katepsin B z různých biologických druhů je monomer o molekulové hmotnosti 24-27 kDa. U savčích katepsinů B bývá působením legumainu vyštěpen dipeptid v pozici 48-49 aktivního enzymu [108] a molekula je pak tvořena dvěma řetězci propojenými disulfidickými můstky. ***Struktura*** katepsinu B vykazuje klasické znaky proteas rodiny papainu, rozdíl je v přítomnosti unikátní smyčky tzv. "occluding loop" v levé doméně proteasy, tvořené inzercí asi 20 aminokyselin, která se u jiných proteas této rodiny nevyskytuje (obr. 3) [47]. V současnosti je v PDB databázi uloženo 26 rentgenových struktur savčích katepsinů B (lidských, hovězích a krysích), z toho 4 jsou struktury proenzymu a 22 struktur aktivního enzymu nebo jeho komplexu s nízkomolekulárními nebo přirozenými proteinovými inhibitory. Rovněž byla vyřešena struktura proenzymu a 2 struktury aktivovaného katepsinu B parazita *Trypanosoma brucei*. Tato dizertační práce přispívá 6 novými strukturami katepsinu B z krevničky *Schistosoma mansoni* k pochopení inhibičního a aktivačního mechanismu.



Obr. 3. ***Krystalová struktura lidského katepsinu B.*** *Část katepsinu B odpovídající zralému enzymu je zobrazena šedým stužkovým modelem. Zvýrazněny jsou smyčka "occluding loop" (oranžově) s dvěma histidinovými zbytky (hnědě), katalytické zbytky Cys29 a His199 (žlutě). Všechny molekuly jsou ve stejné orientaci; pozice smyčky "occluding loop" se mění při vazbě různých strukturních elementů do aktivního místa.* ***A,*** *Struktura proenzymu s propeptidem navázaným do aktivního místa (zeleně); PDB kód: 3PBH.* ***B,*** *Struktura aktivního enzymu s kovalentně navázaným malým syntetickým inhibitorem CA030 (růžově); PDB kód: 1CSB.* ***C,*** *Struktura zralého katepsinu B v komplexu s proteinovým inhibitorem stefinem A (tyrkysově); PDB kód: 3K9M. Obrázky byly připraveny pomocí programu Pymol.*

Z hlediska ***aktivity*** jsou savčí katepsiny B komplexními proteasami, protože mohou zastávat funkci jak endopeptidasy, tak exopeptidasy [58]. Exopeptidasová aktivita je umožněna smyčkou "occluding loop", která interaguje s aktivním místem a stericky

blokuje vazbu substrátu za S2´ podmístem, čímž zaručuje peptidyldipeptidasovou aktivitu (odštěpování dipeptidů z C-konce substrátu). Dva histidinové zbytky přítomné na smyčce jsou důležité pro její funkci, His110 vytváří interakce, které stabilizují pozici smyčky, a His 111 pomáhá ukotvení C-koncového karboxylu P2´ substituentu substrátu (obr. 3, str. 14). Smyčka je flexibilní v závislosti na pH, což je rovněž regulováno His110 a His111 [107;109]. Při nízkém pH jsou tyto zbytky protonované a smyčka se váže do aktivního místa. Při vyšším pH deprotonace histidinů způsobuje pohyb smyčky mimo aktivní místo a katepsin B může vázat substrát po celé délce aktivního místa a štěpit v endopeptidasovém módu [47].

Savčí katepsiny B vykazují hydrolytickou aktivitu i stabilitu v slabě kyselém prostředí, ***pH optimum*** je 5.0-6.0 (v závislosti na substrátu) [58;110]. Endopeptidasová aktivita s malými syntetickými substráty byla detekována i při vyšších pH (maximální hodnoty mezi 7.5 až 8) [111], ale nad pH 7 je katepsin B nestabilní a ireverzibilně degraduje [58;112]. Katepsin B jako cysteinová proteasa vyžaduje pro svou maximální aktivitu redukční prostředí (např. v aktivitních testech přítomnost dithiotreitolu či cysteinu spolu s EDTA), jinak dochází k oxidaci katalytického cysteinu a inaktivaci enzymu [113].

Substrátová specifita savčích katepsinů B je určena zejména S2 podmístem, kde stejně jako řada ostatních proteas rodiny papainu preferuje velké hydrofobní aminokyseliny [114], ale na rozdíl od ostatních zde akceptuje i arginin [115]. Selektivita ostatních vazebných podmíst není tak striktní, nicméně byla dokázána jistá preference pro pozitivně nabité a alifatické zbytky na pozici P1 [116-118], aromatické a hydrofobní zbytky na pozici P3 [119], aromatické a větší alifatické zbytky v pozici P1´ [120] a aromatické zbytky v pozici P2´ [121]. *In vitro* jsou pro katepsin B běžně používané malé syntetické substráty Z-Phe-Arg-AMC a více selektivní Z-Arg-Arg-AMC. Katepsin B je *in vitro* schopen hydrolyzovat složky extracelulární matrix laminin, kolagen typu IV, fibronectin a tenascin C [122-124].

Specifickými ireversibilními ***inhibitory*** pro katepsin B jsou ***CA074*** (N-(L-3-trans-karboxyoxiran-2-karbonyl)-L-isoleucyl-L-prolin) a jeho deriváty, které byly vytvořeny modifikací neselektivního inhibitoru cysteinových proteas E-64 s epoxidovou reaktivní skupinou [72]. Krystalové struktury katepsinu B s CA074 a deriváty [125;126] potvrdily, že karboxylová skupina prolinu interaguje s histidiny His 110 a His111 na smyčce "occluding loop" stejně jako v případě vazby substrátu. Katepsin B je inhibován i velkými proteinovými inhibitory cysteinových proteas jako jsou např. cystatiny, stefiny a

chagasiny, ale afinita těchto inhibitorů je menší než u jiných cysteinových proteas díky sterické zábraně aktivního místa smyčkou "occluding loop". Krystalové struktury komplexů katepsinu B s chagasinem [127] a stefinem A [128] ukazují, že vazba větších inhibitorů je možná díky konformační flexibilitě smyčky "occluding loop" (obr. 3, str. 14).

2.2. Parazitické krevničky a onemocnění schistosomóza

2.2.1. Taxonomické zařazení krevniček

Rod krevnička (*Schistosoma*) se ***taxonomicky řadí*** do kmene ploštěnců (Platihelminthes), třídy motolic (Trematoda), podtřídy dvourodí (Digenea), řádu Strigeatoidea, čeledi krevničkovití (*Schistosomatidae*). Dvourodí z třídy motolic jsou jednou z nejrozšířenějších skupin parazitů ohrožující všechny třídy obratlovců a zasahující řadu jejich tkání. Mnoho druhů vyvolává ekonomické problémy způsobené infekcí domestikovaných zvířat, jiné jsou vysoce významné z hlediska humánní medicíny. Jedná se o endoparazity se složitým alespoň dvouhostitelským životním cyklem, kdy prvním hostitelem je vždy měkkýš [129].

Krevnička bývá řazena společně např. s tasemnicemi (Cestoda) a hlísticemi (Nematoda) do skupiny ***helmintů*** (Helminthes). Helminti netvoří zoologický taxon; název se užívá pro velmi různorodou skupinu organismů, dříve označovanou jako parazitičtí červi. Helminti sdružují fylogeneticky a ontogeneticky variabilní a nepříbuzné skupiny organismů, které jsou významnými původci lidských parazitóz a zoonóz, tzv. helmintóz [130].

Je známo nejméně 5 druhů krevniček, které mohou parazitovat na lidech *Schistosoma mansoni* (krevnička střevní), *S. haematobium* (krevnička močová), *S. japonicum* (krevnička jaterní), *S. mekongi* a *S. intercalatum*, z nichž nejběžnější jsou první tři zmiňované druhy [5]. Krevničky rodu *Schistosoma* obývají tropické a subtropické oblasti světa (tab. 1, str. 17 a obr. 6, str. 20). *S. mansoni* se vyskytuje v Africe, na Středním východě, některých Karibských ostrovech, v Brazílii, Venezuele a v Surinamu, *S. japonicum* je rozšířena v Číně, Indonésii a na Filipínách a *S. haematobium* se vyskytuje v Africe a na Středním východě. V našich zeměpisných šířkách můžeme nalézt některé druhy z čeledi krevničkovití parazitující na ptácích (rod *Trichobilharzia*).

Tabulka 1
Přehled medicínsky nejvýznamnějších druhů parazita krevničky způsobujících lidskou schistosomózu. Upraveno podle [131] a [5].

Druh krevničky	Geografická distribuce	Lokalizace dospělé krevničky v lidském hostiteli	Způsob vylučování vajíček	Mezihostitel (vodní plž)
Urinální schistosomóza				
S. haematobium	Afrika, Střední východ	vezikulární cévy	močí	rod *Bulinus*
Intestinální schistosomóza				
S. mansoni	Afrika, Střední východ, Karibské ostrovy, Brazílie, Venezuela, Surinam	cévy mezenteria	stolicí	rod *Biomphalaria*
S. intercalatum	Deštné pralesy centrální Afriky	cévy mezenteria	stolicí	rod *Bulinus*
S. japonicum	Čína, Indonésie, Filipíny	cévy mezenteria	stolicí	rod *Oncomelania*
S. mekongi	Některé části Kambodže a Laosu	cévy mezenteria	stolicí	rod *Tricula*

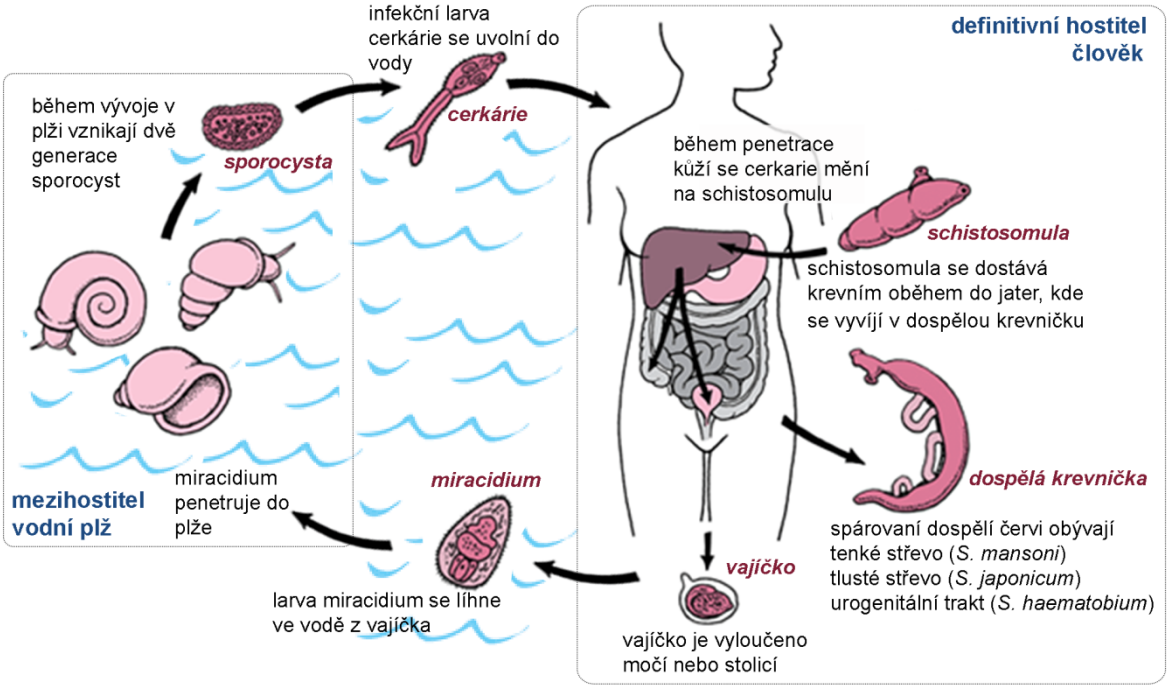
2.2.2. Morfologie a životní cyklus krevniček

Životní cyklus krevničky probíhá ve dvou hostitelích (obr. 4, str. 18) a jednotlivá vývojová stadia prodělávají výrazné morfologické a fyziologické změny (obr. 5, str. 19). Definitivním hostitelem je člověk, ve kterém probíhá pohlavní rozmnožování a tvorba vajíček. Líhnutí vajíček je vázáno na vodní prostředí a nepohlavní rozmnožování parazita probíhá v mezihostiteli, kterým je tropický sladkovodní plž.

K infekci člověka dochází při kontaktu lidské kůže s kontaminovanou vodou, ve které se nacházejí infekční stadia krevničky, cerkárie (obr. 5A, str. 19). Cerkárie má oválné tělo (300 x 70 μm) a charakteristický rozdvojený ocásek [132]. Vyhledává hostitele na základě látek obsažených v kůži hostitele (arginin, kyselina linolová, ceramidy) a pomocí sekretů vyloučených penetračními žlázami (enzymů a mucinů) je schopna přichycení a narušení nepoškozené kůže člověka [133;134]. Při penetraci kůží ztrácí cerkárie ocásek a mění se ve stádium zvané schistosomula (obr. 5B, str. 19). Po průniku tkáněmi se dostávají schistosomuly do krevního řečiště a krevním oběhem do plic, do srdce a do portální žíly, kde získávají živiny z portální krve a dochází k jejich růstu a vývoji v dospělé krevničky. Za 3-4 týdny po infekci se dospělí červi spárují (obr. 5C, str. 19) a sameček přenáší samičku do cév mezenteria obklopujících střeva (*S. mansoni* a *S. japonicum*) nebo do cév urogenitálního traktu (*S. haematobium*), kde samička klade vajíčka. Za den může samička

vyprodukovat stovky až tisíce vajíček (obr. 5D, str. 19) [5]. Vajíčka pronikají přes stěnu cévy, tkáně a mukózu střeva nebo močového měchýře a jsou vyloučena močí (*S. haematobium*) nebo stolicí (*S. mansoni, S. japonicum*). Některá vajíčka se mohou zachytit ve tkáních a jsou potom hlavním patogenním agens schistosomózy.

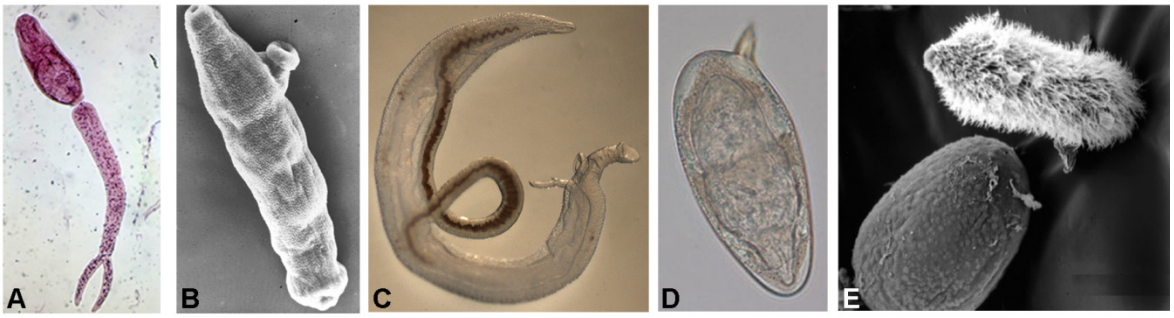
Při kontaktu s vodou se vajíčko vylíhne a uvolní se obrvená larva miracidium (obr. 5E, str. 19). Miracidium vyhledá na základě chemických a světelných stimulů specifického hostitelského plže (tab. 1, str. 17), penetruje jeho kůží, ztrácí obrvenou epidermis a vyvíjí se v mateřskou sporocystu, která dává vznik dceřiným sporocystám. Ty migrují do trávicích žláz plže, kde nepohlavním rozmnožováním vznikají tisíce nových sporocyst a v určité fázi dochází k produkci larev cerkárií. Cerkárie opouští tělo plže asi 4 týdny od infekce a její uvolnění je stimulováno světlem [5]. Nepohlavní rozmnožování umožňuje namnožení parazita a zvyšuje jeho šanci na infikování definitivního hostitele [7]. Jeden plž infikovaný jedním miracidiem může produkovat tisíce cerkárií každý den [5].



Obr. 4. ***Životní cyklus parazita krevničky.*** *Detailní popis cyklu je v textu. Převzato a upraveno podle [135].*

Morfologie dospělé krevničky. Dospělé krevničky jsou drobní červi válcovitého tvaru o délce 10-26 mm a šířce 0.5-1 mm žijící v krevních cévách člověka. Jsou to gonochoristé s výrazným pohlavním dimorfismem. Samička je delší a štíhlejší a je permanentně uložena

v podélné rýze (canalis gyneacophorus) na břišní straně samečka, kde probíhá kopulace (obr. 5C). Povrch parazita je pokrytý vrstvou tzv. tegumentu, který chrání před imunitním systémem hostitele a podílí se na příjmu potravy. Většina potravy je přijímána ústním otvorem obklopeným ústní přísavkou pro uchycení na stěnu cévy, poblíž které se nachází ještě druhá břišní přísavka (acetabulum) [129]. Průměrná doba života dospělé krevničky je 3-5 let, ale někdy mohou žít v cévách člověka až 30 let [5].



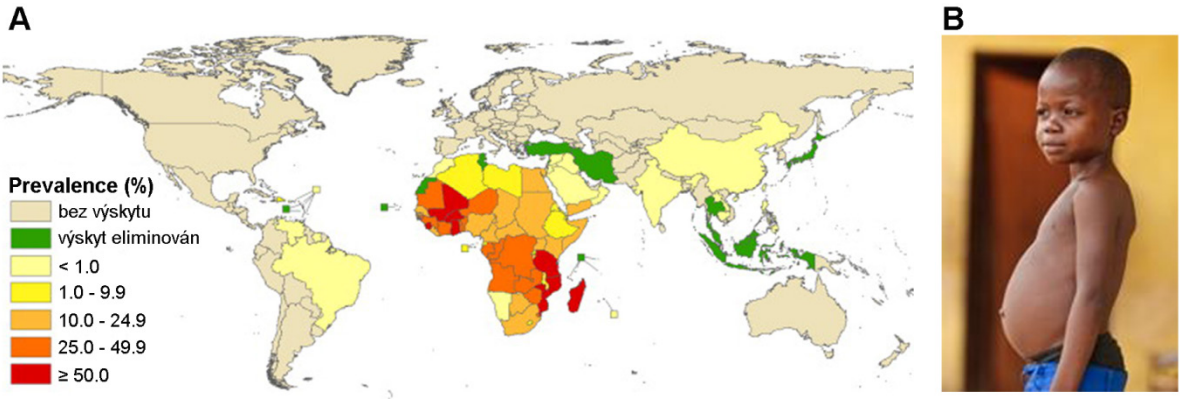
Obr. 5. ***Vývojová stadia parazita krevničky. A,*** *Infekční larva cerkárie.* ***B,*** *Larva schistosomula.* ***C,*** *Pár dospělých krevniček, které se vyskytují v cévách člověka. Samička je drobnější a je uložena v canalis gyneacophorous na břišní straně samečka. Dobře patrný je tmavý pigment hemozoin ve střevě samičky, který vzniká degradací hemu. Poskytl Dr. C. Caffrey.* ***D,*** *Vajíčko krevničky S. mansoni s charakteristickým laterálním trnem.* ***E,*** *Vylíhnutá obrvená larva miracidium. Převzato z [136;137;138].*

2.2.3. Schistosomóza: symptomatika a patogeneze

Schistosomóza je tropická nemoc vyvolaná parazitickými krevničkami rodu *Schistosoma*. Někdy je podle svého objevitele Theodora Bilharze nazývána též bilharzióza [139]. Je to po malárii druhé nejzávažnější parazitární onemocnění. Podle světových statistik je na světě nakaženo přes 200 milionu lidí a více než 600 milionů je v ohrožení nákazou [1-4]. Schistosomóza je endemická v 78 zemích světa, přičemž přibližně 97% všech nakažených se vyskytuje v oblastech Afriky (obr. 6A, str. 20) [2;140]. Onemocnění je rozšířené v tropických a subtropických oblastech, zejména v chudých částech bez přístupu k bezpečné pitné vodě a s nedostačujícími hygienickými podmínkami. Zakládání nových zavlažovacích ploch a přehrad a globální oteplování mohou vést k rozšíření schistosomózy do zatím neendemických oblastí [4;141].

Schistosomóza se řadí mezi tzv. opomíjené tropické nemoci ("Neglected Tropical Diseases" [142]), skupinu chronických infekcí způsobujících postižení a následně i zhoršení sociálně ekonomických podmínek především v extrémně chudých venkovských

oblastech nebo znevýhodněných městských populacích [143-145]. Dostupnost léčby a potlačení nemoci stále není dostatečné a počet infikovaných lidí není trvale redukován [3].



Obr. 6. ***Schistosomóza. A,*** *Mapa zachycující rozšíření onemocnění. Převzato a upraveno z [140].* ***B,*** *Nakažený schistosomózou se symptomem ascites - hromaděním tekutiny v břišní dutině. Převzato z [146].*

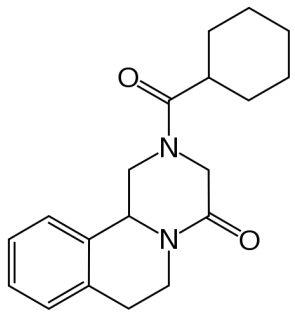
Patogeneze u onemocnění schistosomóza je způsobena vajíčky a lidskou imunitní reakcí na ně spíše než samotnými parazity přítomnými v cévách člověka. Průběh nemoci se dá rozdělit na tři fáze, které souvisí s vývojem parazita: migrační, akutní a chronickou. Migrační fáze zahrnuje období od penetrace cerkárie do vývoje v dospělé schistosomy a produkce vajíček. Tato fáze je často bez příznaků, může se objevit vyrážka v místě průniku cerkárie. Akutní fáze (někdy zvaná Katayamská horečka) nastává 4-10 týdnů po počátku infekce, kdy schistosomy začínají produkovat vajíčka. Symptomy jsou vyvolané imunitní odpovědí na antigeny produkované vajíčky a zahrnují horečku, nevolnost, únavu, bolest hlavy, bolest svalů. Typická je eosinofilie a zvětšená játra a slezina. Vajíčka, která mají charakteristický trn, opouštějí tělo s močí (*S. haematobium*) nebo stolicí (*S. japonicum*, *S. mansoni*) a mohou způsobovat hematurii nebo krvácivý průjem. Chronická fáze bývá opět často bez příznaků, ale v těle se začínají projevovat výrazné patologie. Vajíčka, která jsou zachycena ve tkáních různých orgánů, vyvolávají zánětlivé imunitní reakce, což vede k tvorbě granulomů. Po opakovaných zánětech se množství granulomů zvětší natolik, že orgány kalcifikují a dochází k poškození jejich funkce [129]. U intestinální schistosomózy (způsobené *S. mansoni* a *S. japonicum*) je zasaženo hlavně tlusté střevo a játra, dochází k portální fibróze a může vznikat ascites (hromadění tekutiny v břišní dutině, obr. 6B) a nádory. U urogenitální schistosomózy (způsobené *S. haematobium*) je zasažen urogenitální

trakt a onemocnění může být spojeno se vznikem rakoviny a neplodnosti [6]. Některá vajíčka se mohou dostat krevním oběhem i do dalších orgánů, kde působí patologie.

Diagnóza schistosomózy se většinou stanovuje na základě mikroskopické identifikace vajíček v moči nebo stolici. V oblastech postižených schistosomózou je důležitá ***prevence***, která spočívá v kontrole kvality vody, v celkovém zlepšení životních a hygienických podmínek, preventivní léčbě dostupnými léky a kontrole a likvidaci mezihostitelských vodních plžů [140].

2.2.4. Současná léčba schistosomózy a nové směry vývoje léků

Proti schistosomóze v současnosti neexistuje žádná očkovací látka, pro léčbu se dnes používá téměř výhradně ***praziquantel*** (obr. 7). Praziquantel je účinný proti všem druhům krevniček, je dobře snášen s minimálními vedlejšími účinky, nízkou toxicitou a nízkou cenou (cca 0,2 USD na léčbu jednoho dítěte) [7]. Praziquantel způsobuje tetanické stahy svaloviny dospělých červů, což vede k jejich uvolnění od stěny cévy a migraci do jater. Navíc dochází k poškození a vakuolizaci tegumentu a k expozici antigenů na povrch parazita [7;147]. Přesný mechanismus účinku praziquantelu není známý, předpokládá se souvislost interakce praziquantelu s Ca^{2+} kanály krevniček [148]. Hlavní nevýhodou praziquantelu je, že působí pouze na dospělé krevničky, ale nemá žádný vliv na schistosomuly a vajíčka. V průběhu prvních 21 dní infekce tedy není účinný, léčba bývá kombinovaná s kortikosteroidy, které potlačí symptomy akutní fáze. Je možná kombinace praziquantelu s lékem artemisininem, což je antimalarikum, které je účinné proti nedospělým červům krevničky [149]. Tato kombinace však vede ke snížení efektu praziquantelu na redukci vajíček [150] a zároveň se nedoporučuje plošné užívání arteminisinu v endemických oblastech kvůli možnému vývoji rezistence v parazitické zimničce (*Plasmodium*) způsobující malárii [5].



Obr. 7. Strukturní vzorec účinné látky v léčivu praziquantel.

S ohledem na to, že praziquantel je jediný účinný lék a nepřetržitě se používá už více než 20 let, stále narůstají obavy z vývoje rezistence [8]. Případy nízké citlivosti vůči praziquantelu již byly zaznamenány [9-11] a v laboratorních podmínkách na myších se rezistenci podařilo vyvolat [12]. Proto se už nyní zaměřuje pozornost na hledání nových léků či vakcín.

Vývoji vakcín je věnováno velké úsilí, hlavně proto, že infekci je v počátečních stadiích velmi těžké odhalit vzhledem k častému nesymptomatickému průběhu a než je započata léčba, může již dojít ke vzniku vážných patologií. Několik potenciálních antigenů bylo navrženo, ale indukovaly méně než 50% protekci v myším modelu. Zatím nejúspěšnější kandidát na vakcínu je rekombinantní glutathion-S-transferasa Sh28GST (bilhvax, Eurogenentech, Belgie) pro urinární schistosomózu. Vakcína je nyní ve třetí fázi klinických studií [151]. Vývoj vakcín zůstává stále velkou výzvou [152;153], protože dospělé krevničky jsou na život v krevním řečišti člověka velmi dobře adaptované (přežívají zde až 30 let) a mají dobře vyvinuté strategie pro únik lidskému imunitnímu systému [154].

Stále se tedy zdá být snazší ***vývoj nových chemoterapeutik***, který se snaží ubírat novým směrem. Dodnes používané léky pocházejí z výzkumných programů založených na testování různých látek bez znalosti cílových molekul nebo mechanismu působení dané látky [8]. Nyní se výzkum zaměřuje hlavně na porozumění biochemie parazita a jeho molekulárních drah a na hledání cílových molekul, zásadních a unikátních pro parazita, jejichž zablokování by pro něho bylo letální. Do této oblasti spadají také inhibitory trávicích proteas krevniček jako je katepsin B1 [23]. Pro zjišťování účinnosti nových inhibitorů nebo pro RNAi pokusy je v současnosti k dispozici účinná metoda *in vitro* testování na larvách schistosomula uchovávaných v kultivačních médiích a vyhodnocování změny fenotypu. Tato metoda může být dnes již částečně automatizovaná a umožňuje efektivní testování různých látek [155].

2.3. Proteolytický systém krevničky *Schistosoma mansoni*

Proteolytické enzymy hrají v životě parazitů velmi důležitou úlohu. Krevnička je využívá ve všech svých vývojových stadiích a to zejména při interakci s hostitelem. Mají klíčovou roli v procesech invaze do hostitele, migrace tkáněmi, degradace krevních proteinů při trávení, překonání imunitního systému hostitele a aktivace a modulace zánětu [13].

2.3.1. Genom a degradom S. mansoni

V roce 1994 bylo zahájeno ***sekvenování*** jaderného genomu krevničky *S. mansoni* s cílem lépe pochopit biologii a fyziologii krevničky a vyhledávat geny pro cílený vývoj racionálních léků či vakcín. Sekvenování bylo dokončeno v roce 2009 [156], transkripční a sekvenační data byla dále v roce 2012 upravena a rozšířena na údaje pro různá vývojová stadia (cerkárie, schistosomula, dospělá krevnička) [157] a jsou nyní k dispozici ve veřejných databázích SchistoDB (www.schistodb.net) a GeneDB (http://www.genedb.org/Homepage/Smansoni). Genom krevničky je asi desetkrát větší než genom parazitických prvoků, je tvořen 360 Mb a obsahuje 10852 genů kódujících proteiny, 81% genomu se podařilo přiřadit na chromozomy [157]. Pro *S. mansoni* je rovněž osekvenovaný mitochondriální genom [158], který má význam hlavně pro mezidruhové porovnání a epidemiologické studie [159;160].

V genomu *S. mansoni* bylo identifikováno 335 sekvencí proteas, což činí 2.5% genomu. Byly v něm nalezeny sekvence zástupců pěti tříd proteas (aspartátové, cysteinové, serinové, threoninové a metaloproteasy). Procentuální zastoupení jednotlivých tříd je značně podobné jako v genomu člověka. Nejvýraznějším rozdílem je u krevničky menší zastoupení enzymů rodiny S1 chymotrypsinu, které se u člověka podílí na vysoce komplexních a regulovaných proteolytických kaskádách jako je např. srážení krve (16 u krevničky oproti 135 u člověka) [161]. Oproti tomu má krevnička více členů cysteinových proteas rodiny C1, které se účastní především trávení krevničky. Bylo identifikováno 15 zástupců této rodiny, z čehož 8 má konzervativní doménu katepsinu B [162]. V genomu bylo rovněž nalezeno 34 proteasových inhibitorů včetně cystatinů - proteinových inhibitorů cysteinových proteas [156].

2.3.2. Trávicí systém a trávicí proteasy S. mansoni

Dospívající larvy schistosomula a dospělé krevničky žijí v cévách člověka a získávají živiny nezbytné pro růst, vývoj, rozmnožování a produkci vajíček rozkladem krevních proteinů obsažených v červených krvinkách a krevním séru. Ve střevě krevničky jsou krevní proteiny (zejména hemoglobin) degradovány trávicími proteasami [20].

Trávicí soustava krevničky začíná ústním otvorem obklopeným silnou ústní přísavkou, kterou se krevnička uchycuje na stěny cévy. Pokračuje párové střevo, které se spojuje asi v polovině délky červa a pokračuje jako jedna trubice. Trávicí trakt je slepý a nestrávené zbytky potravy jsou průběžně vyvrhovány zpět do krve hostitele [5]. Odhaduje se, že

dospělý sameček je schopen přijmout desítky tisíc erytrocytů za hodinu, samička, která má větší energetické výdaje spojené s produkcí vajíček, může přijmout až stovky tisíc erytrocytů za hodinu [163]. Část potravy je zřejmě rovněž přijímána povrchem těla.

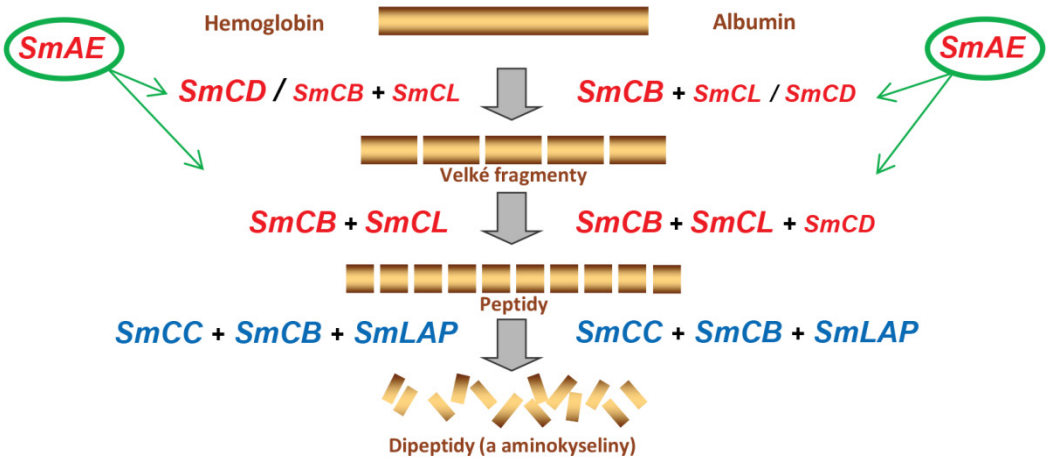
Rozklad erytrocytů začíná v jícnu (esophagus), mechanismus hemolýzy není přesně známý. ***Hemoglobin*** uvolněný z červených krvinek se dostává do střeva, kde je degradován kaskádou endopeptidas a exopeptidas, nejprve na oligopeptidy, a dále na dipeptidy a jednotlivé aminokyseliny, které jsou využity parazitem. Ve střevě krevniček byly identifikovány pomocí imunochemických a biochemických metod cysteinové proteasy rodiny papainu katepsin B1 (SmCB1), katepsin L1, L2 a L3 (SmCL1-3) a dipeptidylpeptidasa I (katepsin C, SmCC), cysteinová proteasa klanu CD asparaginylendopeptidasa (legumain, SmAE), aspartátová proteasa katepsin D (SmCD) a metaloproteasa leucylaminopeptidasa (SmLAP) [14-21]. Až na SmLAP jsou všechny tyto enzymy syntetizovány jako neaktivní zymogeny, které se aktivují na zralé aktivní enzymy buď autoaktivací v kyselém pH nebo působením jiných proteas.

Role jednotlivých trávicích enzymů (SmCB1, SmCL, SmAE a SmCD) v kaskádě štěpení hemoglobinu a sérového albuminu byla zkoumána pomocí specifických inhibitorů jednotlivých tříd proteas v kombinaci s RNAi [21]. Působení cysteinových a aspartátových proteas na hlavní proteinové substráty je kooperativní a je schematicky znázorněno na obr. 8 na str. 25. Hemoglobin je v první fázi efektivně štěpen pomocí SmCD, s přispěním SmCB1 a SmCL. Cysteinové endopeptidasy (SmCB1, SmCL) dále degradují produkty vzniklé počátečním štěpením. Štěpení dokončuje exopeptidasová aktivita SmCB1 a dále exopeptidasy SmCC a SmLAP. Při štěpení albuminu se v počáteční fázi uplatňuje nejvíc SmCB1 s částečnou redundancí SmCD. Dále jsou fragmenty albuminu štěpeny pomocí endopeptidas SmCD, SmCB1 a SmCL. Štěpení je opět dokončeno pomocí exopeptidas jako u hemoglobinu. Role SmAE v celém procesu je zřejmě v aktivaci SmCB1 a/nebo ve specifické limitované proteolýze substrátu s přímou synergií s katepsiny.

Otázkou stále zůstává, kde přesně k procesu trávení dochází. Uvnitř střeva bylo stanoveno pH 6.0-6.4 [14], ale pH optimum degradace hemoglobinu a albuminu je nižší, kolem 4.0 a v pH 6.0 je degradováno jen asi 40% obou proteinů [21]. To naznačuje možnost trávení v diskrétních kompartmentech v gastrodermis nebo v luminálních "kapsách" uvnitř střeva, které zajišťují kyselé prostředí pro trávení krevních proteinů a kde byly trávicí proteasy lokalizovány [14;21].

Proteolytická degradace hemoglobinu způsobuje uvolnění ***hemu***, prostetické skupiny, který je pro parazita toxický. Krevnička, stejně jako jiní hematofágní parazité, detoxifikuje hem přeměnou na tmavě hnědý krystalický pigment hemozoin, který se akumuluje ve střevě krevničky a je průběžně vyvrhován [7;164].

Podobná ***proteolytická kaskáda*** degradace hemoglobinu zahrnující cysteinové a aspartátové proteasy jako u *S. mansoni* byla nalezena i u jiných hematofágních parazitů jako jsou klíšťata (Artropoda), měchovci (Nematoda) nebo zimničky (Protozoa), ale výrazně se liší od trávení obratlovců, kterého se účastní zejména serinové proteasy [13;165;166]. Identifikace klíčových enzymů, které zabezpečují degradaci hostitelských proteinů a tím i přežití parazita, je důležitá jak pro pochopení biologie parazita, tak pro výběr vhodných cílových molekul pro racionální a biospecifický design nových léků proti schistosomóze.



Obr. 8. ***Schéma degradace krevních proteinů pomocí trávicích proteas krevničky S. mansoni.*** *Dva hlavní proteiny hemoglobin (vlevo) a albumin (vpravo) jsou pomocí trávicích proteas postupně štěpeny na velké fragmenty, menší peptidy a absorbovatelné dipeptidy/aminokyseliny. Modře jsou zvýrazněny exopeptidasy, červeně endopeptidasy. Proteasa, která má větší roli v daném kroku, je zvýrazněna větším písmem. Asparaginylendopeptidasa se účastní aktivace ostatních proteas. Upraveno podle [21]. Vysvětlivky: SmCD, katepsin D, SmCB, katepsin B, SmCL, katepsin L, SmCC, katepsin C, SmLAP, leucylaminopeptidasa, SmAE, asparaginylendopeptidasa (legumain).*

2.3.3. Katepsin B1 ze S. mansoni (SmCB1)

2.3.3.1. Funkce a lokalizace SmCB1

Katepsin B1 (SmCB1) je nejvíce zastoupená cysteinová proteasa ve střevě dospělé krevničky *S. mansoni* i v celotělním homogenátu [14;22]. Pomocí imunologických a biochemických metod byl lokalizován v lumen střev a v buňkách gastrodermis vystýlající

střevní dutinu [14;167;168]. Je na něj soustředěna značná pozornost, protože je považován za klíčovou proteasu v degradaci lidského hemoglobinu [21]. Z hlediska transkripční analýzy je SmCB1 exprimován zejména dospělými krevničkami a larvami schistosomula [169]. SmCB1 vykazuje vysokou antigenní schopnost v infikovaných lidech a myších a je tak potenciální serodiagnostický marker [170-173].

Existují ***2 geny*** pro SmCB1, označované SmCB1.1 a SmCB1.2 (UniProt: Q8MNY2, Q8MNY1). První uveřejněná nukleotidová sekvence UniProt: P25792 [174] byla pravděpodobně získaná ze směsi obou těchto isoforem [14]. Kromě SmCB1 byl u krevniček nalezen homologický katepsin B2, který byl lokalizován v tegumentu a parenchymových buňkách dospělých krevniček [175] a na trávení krevničky se nepodílí.

2.3.3.2. Sekvenční a biochemická charakterizace SmCB1

Sekvence SmCB1 zymogenu vykazuje 48-57% homologii se zymogeny savčích katepsinů B a prokazuje přítomnost smyčky "occluding loop" s konzervativními histidinovými zbytky, která u proteas typu katepsinu B zajišťuje exopeptidasovou aktivitu [121;176]. Propeptid SmCB1 je dlouhý 71 aminokyselin a neobsahuje sekvenční motiv ERFNIN a ERFNAQ, což řadí SmCB1 mezi katepsiny typu B [42;43]. Na rozdíl od lidského katepsinu B nebyl SmCB1 nikdy nalezen v krevničkách jako dvouřetězcová molekula. V sekvenci SmCB1 se vyskytuje tzv. "hemoglobinasový motiv" (YWLIAN*SWxxDWGE; x, nekonzervativní aminokyselina, N*, katalytický zbytek Asn290), který je přítomen v proteasách hematofágních parazitů podílejících se na degradaci hemoglobinu a mohl by ovlivňovat interakci se substrátem [177].

Molekulová hmotnost aktivního SmCB1 vypočtená z aminokyselinové sekvence je kolem 28 kDa. V sekvenci SmCB1 se vyskytují dvě potenciální N-glykosylační místa a výsledná glykosylace zvyšuje molekulovou hmotnost na 31 kDa, podle které byl SmCB1 dříve nazýván Sm31 [178]. Pokud jde o prostorovou strukturu, byl doposud k dispozici pouze homologní model SmCB1 [179] a první krystalové struktury přinášejí publikace č. 2 a 4 této dizertační práce.

Přirozený SmCB1 byl izolován z celotělního homogenátu dospělých červů pomocí afinitní chromatografie s použitím inhibitoru Gly-Phe-Gly-semikarbazon [175]. Rekombinantní SmCB1 se podařilo připravit expresí v hmyzích buňkách [180], v kvasinkách *Saccharomyces cerevisiae* [181] a *Pichia pastoris* [14] a dále pomocí *in vitro* transkripce a translace ("cell-free" systém) [178]. Většina systémů poskytovala pouze malé

výtěžky. Nejúspěšnější byla metoda exprese v *P. pastoris*, kterou bylo získáno dostatečné množství enzymu pro další charakterizaci [14] a je nadále běžně používána [182;183].

Přirozený SmCB1 izolovaný z celotělního homogenátu dospělých krevniček *S. mansoni* umožnil základní biochemickou charakterizaci. Byly stanoveny enzymologické konstanty se syntetickými substráty pro katepsin B (Z-Phe-Arg-AMC a Z-Arg-Arg-AMC), identifikována interakce s inhibitorem CA074 savčích katepsinů B a prokázána částečná degradace hemoglobinu a sérového albuminu [184]. ***Rekombinantní SmCB1*** připravený expresí v *P. pastoris* hydrolyzuje syntetické substráty s podobnou účinností jako přirozený SmCB1 a byla pro něj prokázána endopeptidasová i exopeptidasová aktivita [14]. Endopeptidasová aktivita SmCB1 byla podrobněji zkoumána pomocí substrátové knihovny z hlediska specifity v S1-S4 vazebných podmístech a nalezena substrátová specifita SmCB1 byla podobná jako pro lidský katepsin B [182].

2.3.3.3. Biosyntéza a aktivace SmCB1

SmCB1 je ***syntetizován*** jako neaktivní proenzym (zymogen) obsahující 323 aminokyselin. Prvních 71 aminokyselin tvoří propeptid, který je odštěpen v průběhu aktivace, čímž vzniká zralý aktivní enzym. Proces aktivace nebyl zatím u SmCB1 detailně popsán. *In vitro* aktivace zymogenu SmCB1 byla úspěšně provedena pomocí asparaginylendopeptidasy ze *S. mansoni* (SmAE) [14]; starší pokusy s aktivací pepsinem nevedly k plně funkčnímu produktu [181]. Trans-aktivace pomocí SmAE poskytla aktivní zralý enzym, který svými vlastnostmi odpovídal nativnímu SmCB1. Při proteolýze katalyzované SmAE dochází k odštěpení 69 aminokyselin propeptidu, tj. dvě aminokyseliny před štěpným místem odpovídajícím N-konci nativního enzymu. Tyto 2 aminokyseliny mohou být dále odstraněny pomocí katepsinu C (SmCC) s odpovídající dipeptidylpeptidasovou aktivitou [14]. V dosud publikovaných pracích se nepodařilo autoaktivovat rekombinantní zymogen SmCB1 na zralý aktivní enzym a byly připraveny pouze neaktivní intermediátní formy SmCB1 [14;181].

2.3.3.4. SmCB1 jako cílová molekula pro léčbu schistosomózy

V několika publikacích byly podány důkazy o zásadní roli SmCB1 pro životaschopnost krevničky *S. mansoni*, což z něho činí potenciální cílovou molekulu pro vývoj nových léčiv proti schistosomóze. SmCB1 je nejvíce zastoupená proteasa ve střevě krevničky, která *in vitro* účinně degraduje hemoglobin [14;21]. Potlačení exprese SmCB1 v dospělých krevničkách *S. mansoni* pomocí RNAi experimentů vyvolalo u parazitů pokles schopnosti

štěpit krevní proteiny a bylo docíleno zpomalení růstu krevniček [21;185]. Hlavní důkaz o SmCB1 jako cílové molekule poskytla práce z r. 2007 provedená v laboratoři Dr. Caffreyho na University of California San Francisco [23], ve které byl testován inhibitor K11777 na myším modelu schistosomózy. K11777 je inhibitor vinylsulfonového typu navržený proti cruzainu, cysteinové protease rodiny papainu z parazita *Trypanosoma cruzi* způsobujícího Chagasovu chorobu [34;186]. Účinnost tohoto inhibitoru byla testována na myších infikovaných cerkáriemi *S. mansoni*. Bylo dokázáno, že opakované podávání tohoto inhibitoru infikovaným myším snižuje patologické stavy způsobené krevničkou; redukován byl počet parazitů, produkce vajíček i postižení jater a sleziny. Jako hlavní molekulární cíl inhibitoru K11777 byla pomocí aktivitní próby identifikována trávicí proteasa SmCB1, jejíž aktivita byla po aplikaci inhibitoru potlačena [23].

3. Cíle práce

Disertační práce se zabývá trávicí proteasou katepsinem B1 (SmCB1) z krevničky střevní (*Schistosoma mansoni*), způsobující závažné parazitární onemocnění schistosomózu. Tato proteasa má klíčovou roli při trávení krevních proteinů hostitele a představuje cílovou molekulu pro vývoj nových léčiv proti schistosomóze. Práce je zaměřena na strukturně funkční charakterizaci SmCB1, popis mechanismu aktivace zymogenu a nalezení účinných inhibitorů SmCB1 jako potenciálních chemoterapeutik.

Dílčí cíle této práce jsou:

1. Příprava SmCB1 rekombinantní expresí a jeho purifikace.
2. Určení krystalové struktury SmCB1.
3. Analýza substrátové specifity SmCB1 pomocí syntetických i přirozených substrátů.
4. Studium inhibice SmCB1 pomocí vinylsulfonových (VS) inhibitorů:
   a) Určení inhibiční specifity SmCB1 pomocí sady syntetických VS inhibitorů.
   b) Strukturní popis kritických interakcí mezi SmCB1 a VS inhibitory.
   c) Zjištění biologické účinnosti VS inhibitorů na larvách *S. mansoni*.
   d) Enzymologická analýza inhibičního mechanismu VS inhibitorů pro vývoj výpočetní metody skórování účinnosti inhibice.
5. Příprava syntetických peptidových inhibitorů SmCB1 odvozených ze struktury propeptidu.
6. Biochemická a strukturní analýza mechanismu aktivace zymogenu SmCB1.

4. Materiál a metodika

Tato kapitola podává obecný přehled o základní metodice, vybavení a použitém materiálu; detailní informace jsou uvedeny v jednotlivých přiložených publikacích.

4.1. Materiál a laboratorní vybavení

Většina dat byla získána s využitím vybavení a přístrojů v laboratořích Ústavu organické chemie a biochemie AV ČR (ÚOCHB AV ČR). Práce vznikla ve spolupráci s pracovištěm University of California, San Francisco (UCSF, Dr. C. Caffrey), které poskytlo sadu vinylsulfonových inhibitorů a kde byly provedeny biologické experimenty s larvami krevničky. Analýzy metodami výpočetní chemie byly prováděny ve spolupracujících laboratořích Dr. Lepšíka a Dr. Vondráška ÚOCHB AV ČR. Difrakční data pro řešení krystalografických struktur byla získána na synchrotronu na pracovištích (1) Helmholtz-Zentrum Berlin, Bessy II electron storage ring, Berlin-Adlershof, Německo a (2) Structural Biology Center, Advanced Photon Source, Argonne National Laboratory, Argonne, U.S.A. Řešení krystalových struktur probíhalo ve spolupráci s laboratoří Dr. Řezáčové na ÚOCHB AV ČR.

Celkově bylo připraveno 5 konstruktů zymogenu SmCB1 v plasmidu pPICZα pro expresi v kvasinkách *Pichia pastoris*: (1) přirozený $SmCB1^{WT}$, (2) deglykosylovaný $SmCB1^{DG}$ se dvěma mutacemi v sekvenčních motivech pro N-glykosylaci (Thr168→Ala, Thr283→Ala), (3) SmCB1 obsahující mutace v motivu vázajícím heparin: $SmCB1^{HepA}$ (Arg57→Ala, Arg59→Ala) a $SmCB1^{HepN}$ (Arg57→Asn, Arg59→Asn) a (4) SmCB1 s mutací katalytického cysteinu v aktivním místě (Cys100→Ser100). Syntetické peptidy a FRET substráty byly připraveny na ÚOCHB AV ČR.

4.2. Metodika

Základní metody použité v přiložených publikacích jsou následující:

Metody molekulární biologie:

Klonování do plasmidu pPICZα, cílená mutageneze SmCB1 konstruktů, transformace buněk *P. pastoris* (kmen X-33), rekombinantní exprese v *P. pastoris*.

Biochemické metody:

Elektroforetická separace proteinů pomocí SDS-PAGE a detekce proteinů na membráně metodou Western blot, vizualizace proteas pomocí proteomických aktivitních prób, chromatografická purifikace proteinů pomocí FPLC, separace peptidů pomocí RP-HPLC,

analýza proteinů a peptidů metodami termofluor, cirkulární dichroismus, hmotová spektrometrie a N-koncové sekvenování.

Enzymologické metody:

Měření aktivity enzymů pomocí syntetických fluorogenních a FRET substrátů, stanovení kinetických parametrů SmCB1 pro substráty (k_{cat}, K_m, k_{cat}/K_m) a inhibitory (K_i, IC_{50}, k_{2nd}).

Krystalografické metody:

Krystalizace proteinů metodou visící kapky, řešení struktur proteinů metodou molekulárního nahrazení, analýza krystalografických struktur pomocí programů CCP4, Pymol, PISA.

Biologické metody:

Test *ex vivo* účinnosti inhibitorů na larvách *S. mansoni* (schistosomula) v kultivačním médiu a vyhodnocování indukovaných změn fenotypů.

5. Výsledky

Výsledky této disertační práce jsou shrnuty v celkem čtyřech publikovaných pracích (kapitola 5.1, 5.2, 5.3, 5.5) a v jednom rukopise (kapitola 5.4) připraveném k podání do impaktovaného časopisu. Všechny publikace uvedené v této dizertační práci se zabývají katepsinem B1(SmCB1) z krevničky střevní (*Schistosoma mansoni*), hlavní trávicí proteasou podílející se na procesu degradace krevních proteinů hostitele. Publikace na sebe volně navazují a vzájemně se doplňují. V této části jsou předloženy uvedené publikace, každé předchází krátký souhrn, který rekapituluje hlavní výsledky.

Seznam publikací

Publikace č. 1: Mapping the Pro-Peptide of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Modulation of Inhibition by Heparin and Design of Mimetic Inhibitors

Martin Horn, **Adéla Jílková**, Jiří Vondrášek, Lucie Marešová, Conor R. Caffrey, Michael Mareš

ACS Chem Biol, 6 (6), 609-617 (2011)

Publikace č. 2: Structural Basis for Inhibition of Cathepsin B Drug Target from the Human Blood Fluke, *Schistosoma mansoni*

Adéla Jílková, Pavlína Řezáčová, Martin Lepšík, Martin Horn, Jana Váchová, Jindřich Fanfrlík, Jiří Brynda, James H. McKerrow, Conor R. Caffrey, Michael Mareš

J Biol Chem, 286 (41), 35770-35781 (2011)

Publikace č. 3: Quantum Mechanics-Based Scoring Rationalizes the Irreversible Inactivation of Parasitic *Schistosoma mansoni* Cysteine Peptidase by Vinyl Sulfone Inhibitors

Jindřich Fanfrlík, Pathik S Brahmkshatriya, Jan Řezáč, **Adéla Jílková**, Martin Horn, Michael Mareš, Pavel Hobza, Martin Lepšík

J Phys Chem B 117(48), 14973-14982 (2013)

Publikace č. 4: Activation Route of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Structural Map with a Glycosaminoglycan Switch

Adéla Jílková, Martin Horn, Pavlína Řezáčová, Lucie Marešová, Pavla Fajtová, Jiří Brynda, Jiří Vondrášek, Conor R. Caffrey, Michael Mareš

Rukopis připraven k podání do časopisu *Structure*

Publikace č. 5: Katepsinové proteasy v patologii

Martin Horn, **Adéla Jílková**, Michael Mareš

Chem listy, 4, 358-363 (2014)

5.1. Publikace č. 1: Mapping the Pro-Peptide of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Modulation of Inhibition by Heparin and Design of Mimetic Inhibitors

5.1.1. Souhrn

Jedním z hlavních cílů disertační práce je nalezení nových účinných inhibitorů SmCB1 jako potenciálních léčiv proti schistosomóze. Tato publikace se zaměřila na vývoj inhibitorů odvozených ze struktury propeptidu SmCB1. Propeptid je N-koncovou částí zymogenu (proenzymu) cysteinových katepsinů a působí jako přirozený intramolekulární inhibitor tím, že se váže do aktivního místa zymogenu a blokuje tak jeho aktivitu. Po proteolytickém odštěpení propeptidu během procesu aktivace vzniká plně aktivní enzym. Propeptid se i po odštěpení může vázat zpět do aktivního místa a inhibovat enzym, ze kterého byl vyštěpen [54]. Strukturní principy inhibice propeptidem se dají využít při konstrukci malých peptidomimetických inhibitorů, jak bylo prokázáno v případě lidského katepsinu L [53]. Tato publikace se zabývá (1) identifikací inhibičních oblastí v propeptidu pomocí jeho syntetických fragmentů navržených s využitím homologního modelu zymogenu SmCB1 a (2) racionálním vývojem nových inhibitorů odvozených z nalezeného přirozeného inhibičního motivu.

Analýza 3D homologního modelu zymogenu SmCB1 ukázala, že oproti propeptidům savčích katepsinů B se v propeptidu SmCB1 nachází unikátní α-helix, který obsahuje sekvenční motiv XBBXBBX (kde B je bazická aminokyselina a X je hydropatická aminokyselina) známý schopností vázat heparin ("HB motiv" [187]). Porovnáním se sekvencemi propeptidu katepsinů B z jiných organismů bylo zjištěno, že tento motiv je přítomný pouze u katepsinů B z parazitických motolic, které se podílejí na trávení krve hostitele. Identifikovaný "HB motiv" se může účastnit intermolekulárních interakcí SmCB1 s glykosaminoglykany (GAG). Příklady interakcí mezi savčími cysteinovými katepsiny a GAG, které způsobují modulaci aktivity nebo usnadňují aktivaci zymogenů, byly již popsány [61;62;64;66;67].

Pomocí 3D modelu SmCB1 byla navržena a syntetizována sada peptidových fragmentů pokrývající celou délku propeptidu s ohledem na sekundární struktury. Pro tyto peptidy byly určeny hodnoty inhibičních konstant K_i s rekombinantním SmCB1. V propeptidu byly identifikovány dvě inhibiční oblasti, jedna interagující s aktivním místem SmCB1 a druhá, která obsahuje "HB motiv", ta byla však účinná pouze v přítomnosti heparinu.

Inhibiční oblast interagující s aktivním místem byla dále použita jako pentapeptidový templát, z něhož byly pomocí 3D modelu SmCB1 a metody dokování navrženy deriváty

s vyšší inhibiční účinností. Nejlepší z cca 50 derivátů dosahovaly mikromolárních hodnot K_i s SmCB1. Druhá inhibiční oblast obsahuje "HB motiv" a působí inhibičně pouze v přítomnosti heparinu. Tento efekt může být způsoben duální interakcí heparinu jednak se syntetickým "HB motivem" a jednak s SmCB1. Tato hypotéza byla podložena analýzou uvedených interakcí pomocí afinitní chromatografie, metodou termofluor a určením módu inhibice. Funkce "HB motivu" je dále detailně zkoumána pomocí bodových mutací v publikaci č. 4.

Můj podíl na práci spočíval v produkci SmCB1 k enzymologickým pokusům, což zahrnovalo vývoj metody pro aktivaci SmCB1 na zralý enzym a jeho chromatografickou purifikaci. Dále jsem prováděla analýzu interakcí heparinu s enzymem SmCB1.

5.1.2. Publikace č. 1

Horn, M., **Jílková**, **A.**, Vondrášek, J., Marešová, L., Caffrey, C. R. a Mareš, M.

Mapping the Pro-Peptide of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Modulation of Inhibition by Heparin and Design of Mimetic Inhibitors

ACS Chemical Biology, 6 (6), 609-617 (2011)

ACS chemical biology

ARTICLES

pubs.acs.org/acschemicalbiology

Mapping the Pro-Peptide of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Modulation of Inhibition by Heparin and Design of Mimetic Inhibitors

Martin Horn,[†] Adéla Jílková,[†,‡] Jiří Vondrášek,[†] Lucie Marešová,[†] Conor R. Caffrey,[§] and Michael Mareš[†,*]

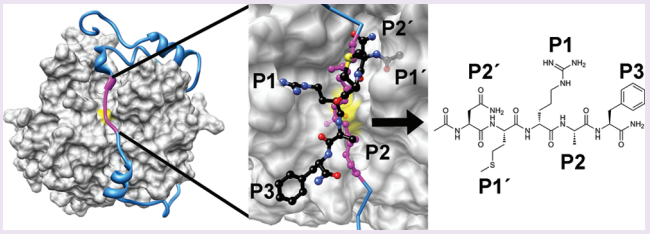
[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 16610 Prague, Czech Republic

[‡]Department of Biochemistry, Faculty of Science, Charles University in Prague, Hlavova 8, 12843 Prague, Czech Republic

[§]Sandler Center for Drug Discovery, California Institute for Quantitative Biosciences (QB3), University of California San Francisco, 1700 Fourth Street, San Francisco, California 94158, United States

Supporting Information

ABSTRACT: Blood flukes of the genus *Schistosoma* cause the disease schistosomiasis that infects over 200 million people worldwide. Treatment relies on just one drug, and new therapies are needed should drug resistance emerge. *Schistosoma mansoni* cathepsin B1 (SmCB1) is a gut-associated protease that digests host blood proteins as source of nutrients. It is under evaluation as a therapeutic target. Enzymatic activity of the SmCB1 zymogen is prevented by the pro-peptide that sterically blocks the active site until activation of the zymogen to the mature enzyme. We investigated the structure−inhibition relationships of how the SmCB1 pro-peptide interacts with the enzyme core using a SmCB1 zymogen model and pro-peptide-derived synthetic fragments. Two regions were identified within the pro-peptide that govern its inhibitory interaction with the enzyme core: an "active site region" and a unique "heparin-binding region" that requires heparin. The latter region is apparently only found in the pro-peptides of cathepsins B associated with the gut of trematode parasites. Finally, using the active site region as a template and a docking model of SmCB1, we designed a series of inhibitors mimicking the pro-peptide structure, the best of which yielded low micromolar inhibition constants. Overall, we identify a novel glycosaminoglycan-mediated mechanism of inhibition by the pro-peptide that potentially regulates zymogen activation and describe a promising design strategy to develop antischistosomal drugs.



Schistosomiasis (bilharzia) is a chronic infectious disease that is caused by a trematode blood fluke and infects over 200 million people in tropical and subtropical areas.[1] Of the five species of schistosomes affecting humans, *Schistosoma mansoni* is a major agent of disease in the Middle East, Africa, and South America and the most convenient experimental model. Morbidity associated with the disease results from immuno-pathological reactions to parasite eggs trapped in various tissues, including the liver, intestinal tract, and bladder. Symptoms include decreased physical and cognitive performance, abdominal pain, and lassitude.[2] Treatment and control of schistosomiasis now relies on mass chemotherapy with just one drug, praziquantel, a tenuous situation should drug resistance emerge and spread.[1,3] Accordingly, there is pressure to identify new schistosomal protein targets for both chemo- and immuno-therapeutic interventions.

Adult schistosomes live in the cardiovascular system, and host blood proteins are a primary source of nutrients required for growth, development, and reproduction. In the schistosome gut, a network of proteases contributes to the digestion of host proteins to absorbable peptides and amino acids.[4,5] For *S. mansoni*, the component digestive proteases thus far characterized include (i) the Clan CA papain-like cathepsins B1 and L1, L2, and L3, dipeptidyl peptidase I (cathepsin C), (ii) the Clan CD asparaginyl endopeptidase (legumain), and (iii) the Clan AA aspartic protease, cathepsin D.[5−11] Similar acid pH-based proteolytic networks are found across phylogenetically diverse organisms such as nematodes, *Plasmodium* (malaria), and ticks.[12−15] Given their central importance to schistosome nutrition, therefore, gut proteases represent potential chemotherapeutic targets. The present research focuses on *S. mansoni* cathepsin B1 (SmCB1), which has been validated in a murine model of schistosomiasis mansoni as a molecular target for therapy with a peptidomimetic, vinyl sulfone protease inhibitor.[16] SmCB1 is a central digestive protease due first to its relative abundance and second to its complex proteolytic activity comprising both endopeptidase and exopeptidase (peptidyl dipeptidase) modes of action.[4,7]

Members of the Clan CA, Family C1 (papain-like) cysteine proteases,[17] including cathepsins B, are biosynthesized as inactive

Received: October 22, 2010
Accepted: March 4, 2011
Published: March 04, 2011

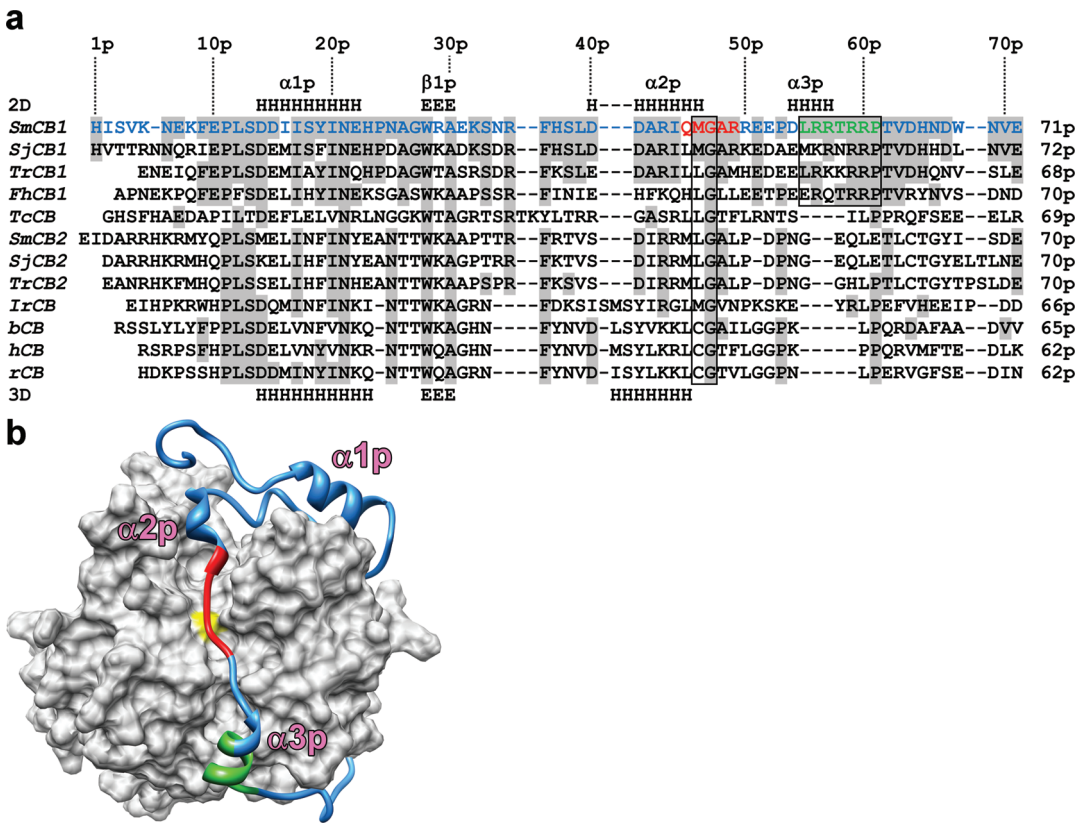


Figure 1. (a) Structure-based alignment of the pro-peptides of cathepsin B-type proteases from the human blood flukes *S. mansoni* (SmCB1 and SmCB2), *S. japonicum* (SjCB1 and SjCB2), the bird fluke *Trichobilharzia regenti* (TrCB1 and TrCB2), the liver fluke *Fasciola hepatica* (FhCB1), the protozoan *Trypanosoma cruzi* (TcCB), the tick *Ixodes ricinus* (IrCB), and bovine, human and rat cathepsins B (bCB, hCB and rCB; Uniprot accessions Q8MNY2, Q9SPM1, P43157, C7TYR4, Q4VRW9, A7L844, Q8I7B2, O61066, A4GTA7, P07688, P07858, and P00787, respectively). The secondary structure of the SmCB1 pro-peptide (2D) is based on prediction analysis, whereas the secondary structure of rCatB and hCatB pro-peptides (3D) is derived from crystallographic data (PDB entry 1MIR and 3PBH): H, α-helix; E, *β*-strand. The SmCB1 pro-peptide (blue) contains two inhibitory regions: an heparin-binding region (green) with a heparin-binding motif (boxed) and an active site region (red) containing residues in contact with the S1 and S1′ subsites of the active site (boxed). Residues identical with those of SmCB1 are shaded in gray; numbers above the alignment refer to the pro-peptide of SmCB1, the suffix "p" indicates pro-peptide numbering. A larger multiple alignment is provided in Supplementary Figure S1. (b) The tertiary structure of the SmCB1 zymogen was modeled using the X-ray structure of rat pro-cathepsin B (PDB entry 1MIR). The enzyme core is shown as a gray surface, and the active site Cys is in yellow. The pro-peptide (blue ribbon) contains two inhibitory regions color-coded as in the alignment (panel a): a heparin-binding region (residues 55p−61p in green) and an active site region (residues 45p−49p in red). The α-helices are indicated in magenta bold typeface.

zymogens (pro-cathepsins) in which the pro-peptide operates as an intramolecular inhibitor by blocking the active site. Activation to the mature, catalytically active forms occurs by proteolytic removal of the N-terminal pro-peptide (also termed the "activation peptide"); removal is either autocatalytic or involves other protease(s). After cleavage, the released pro-peptide can remain bound and inhibit its cognate enzyme.[18] The genomes of many organisms encode pro-peptide-like proteins that may act as intermolecular inhibitors of cysteine cathepsins.[19] Most studies on pro-peptide inhibition have been performed with full length pro-peptides and have demonstrated that inhibition of both parental and nonparental enzymes by pro-peptides is a function of the latter's structural complementarity.[20−23]

The application of pro-peptide fragments and crystal structures of cathepsin zymogens has shown that pro-peptides contain structural regions that govern their interactions with the mature enzyme.[24−30] As reported for the human cathepsin L pro-peptide, the structural principles defining pro-peptide inhibition can be used to construct small peptidomimetic inhibitors.[31] Also, studies with pro-peptides and their fragments have shown how they regulate the activity of cognate cysteine proteases from a diversity of protozoan and metazoan parasites including *Trypanosoma*, *Plasmodium*, and *Fasciola*[21,22,25,27] and, importantly, that they can suppress the development of *Plasmodium*.[28] Yet, there is a lack of detailed structure−activity studies for small pro-peptide mimetics of parasite peptidases and their potential as chemical starting points for therapeutic intervention of disease.

The activity of cysteine cathepsins is modulated by glycosaminoglycans (GAGs), sulfated polysaccharides with a strong negative charge that are widely distributed in tissues. The interaction of GAGs with mammalian cathepsins is complex and includes, for example, induction of collagenolytic activity of cathepsin K,[32] stabilization of cathepsin B at neutral pH,[33] and facilitation of autocatalytic activation of several cathepsins including cathepsins B, L, and S.[34−36] Mutation studies have shown that the GAG-accelerated activation involves interaction of GAGs with pro-peptides.[34,36] However, the effect of GAGs on inhibition by pro-peptides or their truncated synthetic derivatives has not been investigated.

Here, we define a structure−activity relationship (SAR) for SmCB1 and its pro-peptide. Using synthetic pro-peptide

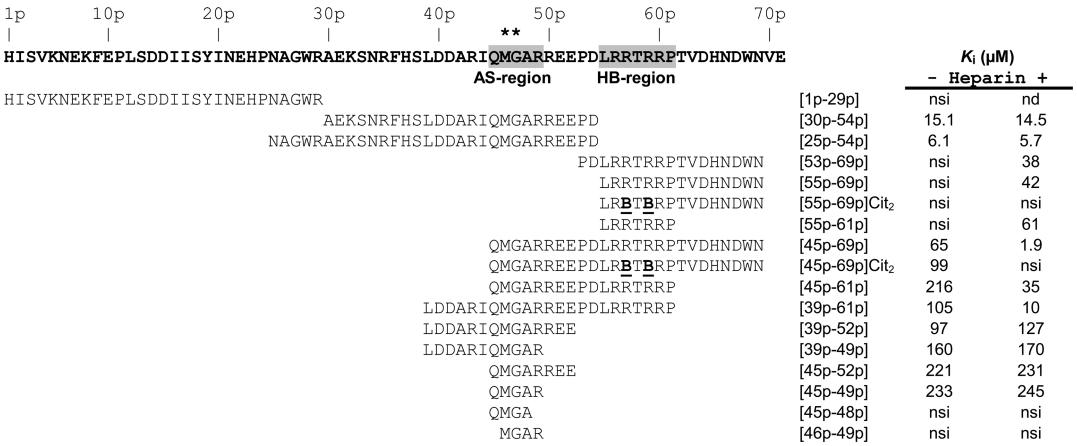


Figure 2. Design and inhibitory activity of synthetic peptide fragments derived from the pro-peptide of SmCB1. In the pro-peptide sequence, the active site (AS) and heparin-binding (HB) inhibitory regions are highlighted in gray, and those residues (Met46p and Gly47p) in contact with the S1 and S1′ subsites of the active site are marked (*). Synthetic peptide fragments are coded according to their location in the sequence; the suffix "p" indicates pro-peptide numbering. Substituted residues are in bold and underlined; B denotes citrulline. The inhibition constants (K_i) were determined in a kinetic activity assay with SmCB1 and the fluorogenic peptide substrate, Z-Phe-Arg-AMC, at pH 5.5 in the absence and presence of heparin. Mean values are given for triplicate measurements (SE values are within 10% of the mean); nsi, no significant inhibition with 400 μM peptide; nd, not determined due to low solubility.

fragments, we identify structural regions within the pro-peptide that account for its inhibition of the mature enzyme and uncover a new mechanism for their functional regulation by GAGs. In addition, one of these structural regions was employed as a scaffold to design small inhibitory peptides. These inhibitors of SmCB1 represent a chemical basis from which new antischistosomal drugs might be developed.

RESULTS AND DISCUSSION

SmCatB1 Pro-Peptide Contains a Phylogenetically Constrained α-Helix Incorporating a Heparin-Binding Motif. As the amino acid sequence of pro-SmCB1 (Uniprot accession Q8MNY2) is substantially homologous (48−57% identity) with those of mammalian pro-cathepsins B, we constructed a spatial model of the SmCB1 zymogen as a tool for structure−function analysis. We employed the X-ray structure of rat pro-cathepsin B (PDB entry 1MIR) as a template (Figure 1b). The SmCB1 pro-peptide comprises two α-helices (α1p and α2p) and a short β-strand (β1p) that is shared between zymogens of the cathepsin B subfamily. A significant difference occurs toward the C-terminal of the SmCB1 pro-peptide where there is an extra α-helical segment, denoted α3p. Figure 1a shows an alignment of pro-peptides of SmCB1 and other cathepsins B. The predicted secondary structure of α3p is associated with a sequence insertion that contains a consensus sequence for heparin binding, namely, XBBXBBX (where B is a basic amino acid and X is a hydropathic amino acid).[37] This new structural segment may be involved in intermolecular interactions between the SmCB1 pro-peptide and GAGs, as represented by heparin.

In addition, using blastp tools at NCBI, we inspected the cathepsin B-type sequences from other organisms and found that the predicted α3p helix including the heparin-binding motif is phylogenetically distributed in the pro-peptides of some, but not all, cathepsins B of parasitic trematodes, including *Schistosoma*, *Trichobilharzia*, and *Fasciola* spp. (Supplementary Figure S1). The feature is apparently absent from cathepsin B pro-peptides of nematodes, insects, and vertebrates. For those enzymes that contain the feature, the literature suggests they are gut localized and involved in digestion of host blood and/or tissues. By contrast, other trematode enzymes not associated with this function (*e.g.*, SmCB2, homologous to SmCB1 but found in the tegument (worm surface)), do not contain this structural feature.[38] Accordingly, it seems that α3p is a novel structural feature in the pro-peptide of SmCB1 and certain other cathepsins B involved in the digestion of host proteins by parasitic trematodes.

Design of Pro-Peptide Fragments for Structure−Inhibition Analysis. With the help of the 3D model (Figure 1), we designed a panel of peptide fragments derived from the sequence of the SmCB1 pro-peptide (Figure 2). The fragments were synthesized with capping groups (N-terminal acetyl and C-terminal amide) to mimic peptide linkages and to protect peptides against exopeptidase degradation by SmCB1. The main fragments were designed to reflect the secondary structure pattern, *i.e.*, the fragments [1p−29p], [30p−54p], and [53p−69p] contain α1p, α2p, and α3p α-helices, respectively. Fragment [25p−54p] combines α2p with the β-strand β1p that is involved in the formation of an adjacent hairpin structure. Met46p and Gly47p fill the S1′ and S1 pockets of the active site of SmCB1. Both Met46p and Gly47p are centrally positioned in a series of peptides that were synthesized to extend in both the N-terminal (to include α2p) and C-terminal directions (to include α3p and its heparin-binding motif, and the pro-peptide C-terminus). A number of the synthesized fragments contain sequence stretches that form α-helices according to the prediction analysis and the 3D model (Figures 1 and 2). For representative fragments, we demonstrated by circular dichroism spectroscopy the presence of a partial α-helical conformation (Supplementary Figure S2). This underscores the utility of the model and designed fragments to analyze structure−inhibition relationships in the SmCB1 pro-peptide.

Pro-Peptide Mapping of SmCB1 Identifies Two Inhibitory Regions with Differing Sensitivities to Heparin. The inhibition constants (K_i) for an array of synthetic pro-peptide fragments against mature SmCB1 was measured using a kinetic assay with the fluorogenic substrate Z-Phe-Arg-AMC (Figure 2).

The long N- and C-terminal pro-peptide fragments ([1p−29p] and [53p−69p]) did not inhibit SmCB1 activity. In contrast, the central fragment [25p−54p] had a K_i of 6 μM. The potency of the central fragment was partially supported by the β1p structure as demonstrated by its deletion in peptide [30p−54p] that resulted in a K_i of 15 μM. This β-strand contains the conserved hydrophobic residues Trp28p and Ala30p that interact with the "pro-peptide binding loop", a short loop found on the surface of the core of mammalian pro-cathepsins B.[29] The central fragment [30p−54p] that traverses the active site was gradually truncated at both termini in a set of peptides spanning the S and S′ subsites. The extensive trimming of the 25-mer peptide did not abolish inhibition but did result in increased K_i values up to 233 μM as measured for the shortest pentapeptide QMGAR [45p−49p]. This segment represents a minimal inhibitory region, the active site region, which surrounds Met46p−Gly47p residues occupying the S1′ and S1 subsites.

Next, we tested the inhibitory potency of the synthetic peptides in the presence of heparin, as the heparin-binding motif is located in the C-terminal region of the SmCB1 pro-peptide (residues 55p−61p). Interestingly, the large C-terminal fragment [53p−69p] required heparin to display inhibition, which was substantial (K_i = 38 μM; Figure 2). The same heparin-dependent inhibition was observed for the shorter peptide LRRTRRP [55p−61p] that is equivalent to the length of the heparin-binding motif (K_i = 61 μM in the presence of heparin). Accordingly, we identify a novel heparin-binding region [55p−61p] that acts as an inhibitor of SmCB1, but only in the presence of heparin.

The peptides spanning both the heparin-binding and active site regions ([45p−69p], [39p−61p], [45p−61p]) inhibited with K_i values in the range 65−216 μM in the absence of heparin. In the presence of heparin, however, these values were decreased by 1 order of magnitude (K_i ~2−35 μM). This demonstrates that the heparin-binding region significantly improves the inhibitory potency of the active site region. The properties of the most potent peptide fragment [45p−69p] were studied in detail. First, the mode of inhibition was found to be competitive (Figure 3), which is consistent with the expected interaction of this peptide in the substrate-binding site through the active site region. Second, the inhibitory selectivity was investigated. It is known that pro-peptides of cysteine proteases and their truncated synthetic forms bind not just to the mature cognate enzymes but also to homologous enzymes with varying potencies (for a review, see ref 20). The pro-SmCB1 peptide [45p−69p] inhibited bovine cathepsin B with K_i values of 512 and 257 μM in the presence and absence of heparin, respectively. This compares with the K_i values for SmCB1 of 2 and 65 μM, respectively. The weaker inhibition of bovine cathepsin B activity can be explained by its partial homology (51% identity) to SmCB1. Importantly, heparin does not improve the inhibition of the mammalian enzyme by the peptide [45p−69p], pointing to the specificity of the interaction between SmCB1 and the fragment through the heparin-binding region.

Overall, mapping the pro-peptide of SmCB1 with synthetic peptides identifies two distinct small inhibitory regions. First, the active site region is topologically similar (*i.e.*, in contact with the active site residues) to pro-peptide fragments that inhibit other cysteine proteases, including mammalian cathepsin B.[24−26] In those reports, 7- to 15-mer peptides had K_i values in the range of those described here. For SmCB1, inhibition by peptides covering the active site region is independent of the presence

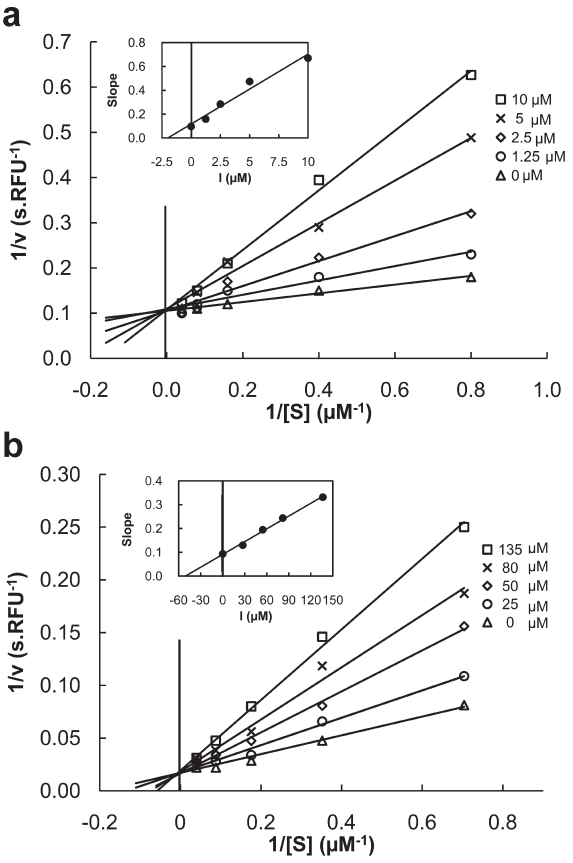


Figure 3. Competitive mode of inhibition of SmCB1 by fragments [45p−69p] (a) and [55p−61p] (b) derived from the SmCB1 pro-peptide. A Lineweaver−Burk plot is presented together with a secondary plot of the same data (inset); K_i values are 2.0 and 55 μM for [45p−69p] and [55p−61p], respectively. The activity assay with the fluorogenic peptidyl substrate, Z-Phe-Arg-AMC, was performed at pH 5.5 in the presence of 25 μg mL^{-1} heparin. Mean values are given for triplicate measurements (SE were within 10%).

or absence of heparin. Second, it is now clear that the heparin-binding region, unique to certain trematode cathepsins B, can also inhibit proteolysis by SmCB1, but that heparin (and possibly other GAGs) needs to be present.

Heparin Interacts with Both the Heparin-Binding Region of the Pro-Peptide and the Enzyme Core. The properties of the heparin-binding region's interaction with heparin were examined. For fragments [55p−69p] and [45p−69p], we synthesized derivatives in which the heparin-binding motif was altered by substituting two positively charged arginine residues with citrulline isosteres that possess uncharged side chains (peptides [55p−69p]Cit$_2$ and [45p−69p]Cit$_2$). This abolished the heparin-dependent inhibition of SmCB1 (Figure 2). The result was confirmed by heparin affinity chromatography (Figure 4): the wild-type peptide [45p−69p] bound to immobilized heparin, whereas its citruline derivative [45p−69p]Cit$_2$ did not.

In addition, mature, *i.e.*, fully activated, SmCB1 was bound by the heparin affinity column (Figure 4). This can be rationalized on the basis that SmCB1, being more basic then mammalian cathepsins B (theoretical pI values of 8.7 and 5.1, respectively), can interact with the positively charged heparin. The mature SmCB1 sequence does not contain a canonical heparin-binding

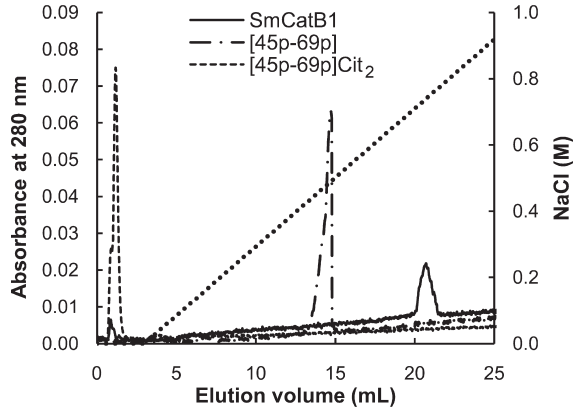


Figure 4. Heparin affinity chromatography of SmCB1, its pro-peptide-derived fragment [45p−69p] containing the heparin-binding motif, and the same fragment but in which the two positively charged arginine residues have been substituted with citrulline [45p−69p]Cit_2. The column was equilibrated in 25 mM sodium acetate, pH 5.5, and eluted with a linear gradient of 0−1 M NaCl.

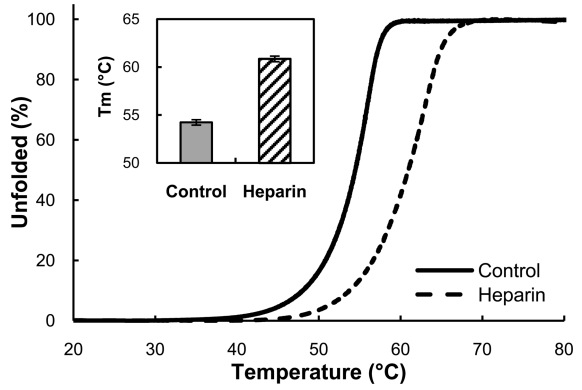


Figure 5. Thermal denaturation of SmCB1 using the thermofluor technique. The curves were recorded for SmCB1 in the presence (dashed line) and absence (solid line, control) of heparin (25 μg mL^{-1}). The fraction of unfolded protein is displayed. Inset: the melting temperature (T_m) values determined from the denaturation curves. Mean values ± SE are given for triplicate measurements.

motif; however, inspection of the 3D model reveals several positively charged patches. One of these involves the occluding loop on the "front site" surface of SmCB1 close to the active site (Supplementary Figure S3). Basic surface patches are reported to mediate heparin−protein interactions.[39] The interaction of SmCB1 with heparin was also studied using the thermofluor technique[40,41] in order to understand whether heparin alters the thermal stability of SmCB1. In the presence of heparin, the thermal stability of SmCB1 was increased (melting temperature (T_m) shifted from 54 to 61 °C (Figure 5)), thus indicating the formation of a complex between SmCB1 and heparin.

On the basis of the above results, we propose a mechanism by which heparin promotes the inhibitory functionality of the heparin-binding region. The heparin polymer interconnects the pro-peptide heparin-binding region through the heparin-binding motif and the enzyme core through a surface basic patch. This facilitates docking of the heparin-binding region in a proper orientation, which finally leads to steric hindrance of the adjacent active site by a bulky complex formed from the heparin-binding peptide and heparin. This is in line with the competitive mode of inhibition of SmCB1 by the synthetic heparin-binding region [55p−61p] in the presence of heparin (Figure 3b). The electrostatic character of the involved interactions is supported by the finding that the heparin-dependent inhibition of SmCB1 by [45p−69p] is abolished by 0.3 M NaCl (data not shown). The specific requirements for the heparin structure were tested using heparins ranging from 2 to 17 kDa in size. The inhibition by peptide [45p−69p] did not vary significantly (K_i ∼1.0−1.9 μM), demonstrating that a 2 kDa hexasaccharide fragment (K_i = 1.2 μM) is as effective as 17 kDa heparin (K_i = 1.9 μM) in mediating the interaction.

What is the function of the heparin-binding region? We speculate that it may participate in regulating the activation of the SmCB1 zymogen. For several mammalian cysteine cathepsins, including cathepsin B, it was reported that heparin and other GAGs accelerate zymogen autoactivation through destabilization of the pro-peptide interaction with the mature enzyme, thus facilitating its autocatalytic cleavage.[34,36] The present data suggest that the modulatory activities of GAGs are more complex. In the case of SmCB1, these polysaccharides might produce increased stabilization of the pro-peptide interaction *via* the heparin-binding region. GAGs are widespread in host tissues and are also found in *Schistosoma*.[42] Interestingly, and unlike mammalian cathepsins B, the SmCB1 zymogen only partially autoactivates and the assistance of processing proteases, such as an asparaginyl endopeptidase, is necessary for complete zymogen conversion.[7] The partially processed SmCB1 zymogen is catalytically inactive and still retains a pro-peptide fragment (residues 39p−71p) that includes both the active site and heparin-binding regions.[7] Our future research will test the hypothesis that GAGs regulate the activation of pro-SmCB1 using purified recombinant zymogen and various heparin-binding region mutants.

Pro-Peptide Mimetics Derived from the Active Site Region as a Design Basis for Novel Anti-Schistosomals. We found that the active site region [45p−49p] synthesized as a synthetic pentapeptide retains some inhibitory activity (K_i = 233 μM) against SmCB1 (Figure 2). Also, neither this pentapeptide nor variants extended in either the N- ([30p−54p]) or C-terminal directions ([45p−69p]) are hydrolyzed by SmCB1 (see Supplementary Figure S4). We suggest that the synthetic active site-directed pentapeptide likely binds in the same manner to the active site as the native pro-peptide, *i.e.*, in the opposite orientation than substrates, and thus prevents cleavage. Accordingly, we set out to use the active site region as a starting scaffold for the design of more potent inhibitors of SmCB1.

In our structure-based design, we employed a 3D model of mature active SmCB1 that was built by homology modeling using the X-ray structure of rat cathepsin B (PDB entry 1CPJ) as template (Figure 6). The wild-type pentapeptide [45p−49p], Ac-QMGAR, was docked into the active site cleft where it occupied the S2′ through S3 subsites. The corresponding P2′ through P3 residues in the inhibitor structure were substituted by L- and D-amino acids, and the modified structures were redocked and evaluated for their interaction potential. The best candidates selected *in silico* were synthesized. These compounds are listed in Table 1 together with their K_i values for SmCB1. Initially, we substituted the P1 residue (Gly) with D-Arg that is known to prevent interaction of peptidic ligands in the substrate-like binding mode.[31] Then, we used Phe at P3 to fill the S3 pocket of SmCB1 as bulky hydrophobic groups in this position are generally favored in SmCB1 substrates.[43] Next, the P2′ residue

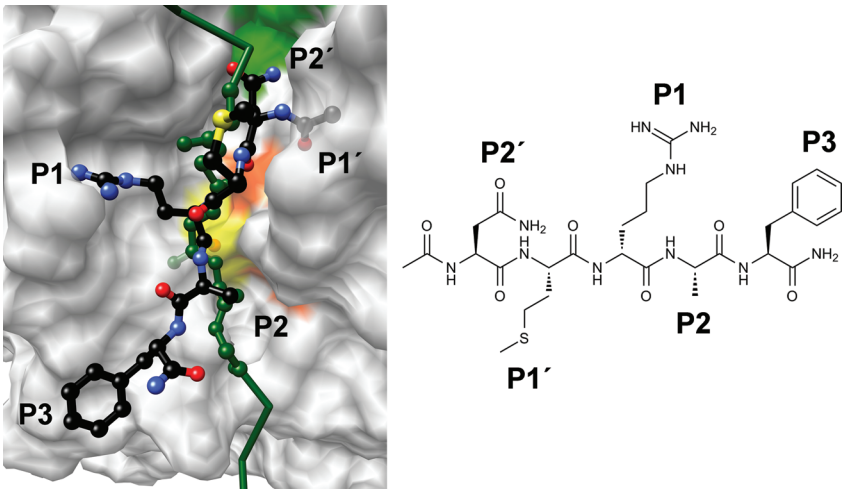


Figure 6. Chemical structure and docking model of the Ac-NMrAF inhibitor, a synthetic mimetic of the pro-peptide active site region. The inhibitor (black ball-and-stick) was docked into the active site of SmCB1 (gray surface) and residues in the P2′−P3 positions are indicated. The Ac-NMrAF binding mode is superimposed on that of the SmCB1 pro-peptide (dark green, with active site region backbone in ball-and-stick). Heteroatoms of Ac-NMrAF are colored red for oxygen, blue for nitrogen, and yellow for sulfur. In the SmCB1 active site, the catalytic residues Cys100 and His270 are highlighted in yellow and orange, respectively. The histidine residues of the occluding loop (His180 and His181) are in green. The SmCB1 molecule is oriented as in Figure 1.

that interacts with the occluding loop of SmCB1 was optimized. During this step, we also evaluated N-terminal capping, which for several P2′ substituents was required for efficient inhibition. This step defined two routes for the further iterative design of compounds containing either an acetylated P2′ Asn or an N-terminally free Glu. Finally, employing mostly hydrophobic residues at P1′ and P2 yielded Trp and Ala, respectively, as generally suitable substituents. In summary, the screening of 48 inhibitors (Table 1) yielded 5 inhibitors with K_i values lower than 50 μM and one (EYrAF) with a value of 14 μM, *i.e.*, 16-fold more potent than the native active site region [45p−49p] (K_i = 233 μM). The 5 best ranked inhibitors were also screened against human cathepsins B, L, and S under analogous conditions, and no significant inhibition was observed (Table 1).

Docking Ac-NMrAF, one of the more potent inhibitors (K_i = 32 μM), in the active site of SmCB1 (Figure 6) illustrates a network of potential interactions (contacts within 4 Å) that is formed between enzyme and inhibitor. These include P3 (Gly67, Leu68, Ile74), P2 (Gly72, Gly73), P1 (Gly27, Cys39, Cys70, Glu71, Ile122, His199), P1′ (Gly197, Gly198, Leu196), and P2′ (His110, Val176, Leu181, Leu196, Trp221). The P2′ residue (Ac-Asn) interacts with His110 of the occluding loop. A similar interaction of the epoxysuccinyl peptidyl inhibitor, CA-074, with His110 and His111 was critical for its potent inhibition of bovine cathepsin B.[44]

In conclusion, we demonstrate that the active site region is a useful starting scaffold in the design of inhibitors selective for SmCB1 over host-derived cathepsins B, L, and S. The K_i values of the most potent pentapeptide inhibitors are in the mid to low micromolar range and are thus an improvement on the inhibition produced by the original active site region [45p−49p]. Moreover, the K_i values are comparable with those measured for the large central 25-mer pro-peptide fragment [30p−54p] (Figure 2). We will continue the optimization process by introducing unnatural amino acids and building blocks. These modifications may confer low or submicromolar inhibition as was found for the pro-peptide-derived inhibitors of human cathepsin L.[31] We will also attempt to cocrystallize optimized inhibitors with SmCB1 for X-ray analysis to provide direct evidence for the modeled binding mode. This will further the rational design of inhibitors that may help identify molecules for consideration as antischistosomal drugs.

METHODS

Materials. Heparins from porcine intestinal mucosa with an average molecular weight of 17, 5, and 3 kDa were from Sigma, and heparin hexasaccharide was from Dextra Laboratories. The fluorogenic substrate Z-Phe-Arg-AMC was from Bachem. Amino acid precursors and resins for peptide synthesis were from Iris Biotech. *S. mansoni* asparaginyl endopeptidase (SmAE) was produced recombinantly in *Pichia pastoris* as described in ref 10. Bovine cathepsin B was purified from bovine spleen,[45] human cathepsin B was obtained from Merck, and human cathepsins L and S from Sigma.

Recombinant Expression and Purification of SmCB1. SmCB1 was expressed in the X33 strain of the methylotrophic yeast *P. pastoris* as described previously.[7] The induction media were concentrated and desalted on a Sephadex G-25 column equilibrated with 25 mM sodium acetate, pH 5.5. Recombinant protein (a mixture of zymogen and intermediate forms) was separated on a Mono S HR 5/5 column (GE Healthcare Bio-Sciences) equilibrated with 25 mM sodium acetate, pH 5.5 and eluted using a linear gradient of 0−1 M NaCl. The activated SmCB1 was obtained by proteolytic treatment with SmAE in 0.1 M sodium acetate, pH 5.0, containing 2.5 mM dithiothreitol.[7] The conversion to active SmCB1 was monitored using the Z-Phe-Arg-AMC assay (see below) and SDS-PAGE. SmCB1 was finally purified on a Mono S HR 5/5 column as above. The preparation was concentrated and desalted into 20 mM sodium acetate, pH 5.5 using Amicon Ultracel-10k (Millipore). The typical yield was 2 mg of SmCB1 from 1 L of induction media.

Synthesis of Peptides. Peptides were synthesized by Fmoc solid phase chemistry in an ABI 433A Peptide Synthesizer (Applied Biosystems) in the form of peptidyl amides with free or acetylated N-termini as described previously.[46] Peptides were purified by RP-HPLC over a C18 column using a trifluoroacetic acid (TFA) elution system and characterized by ESI mass spectrometry on an LCQ Classic

Table 1. Structure-Based Optimization of Synthetic Inhibitors Derived from the Active Site Region of the SmCB1 Pro-Peptide

peptide[a]	structure[b]	K_i (μM)[c] Ac	K_i (μM)[c] NH_2
	QMGARREE	221	268
	QMrARREE	83	159
	QMGAR	233	nsi
XMrAF	QMrAF	210	nsi
	NMrAF	32[d]	nsi
	EMrAF	nsi	161
	DMrAF	nsi	nsi
XWrAF	PWrAF	134	265
	HWrAF	159	nsi
	RWrAF	164	189
	QWrAF	266	291
	SWrAF	299	nsi
	TWrAF	nsi	66
NXrAF	NWrAF	22[d]	
	NYrAF	nsi	
	NFrAF	nsi	
EXrAF	EYrAF		14[d]
	EWrAF		37[d]
	ECrAF		97
	EFrAF		112
	EIrAF		177
	ELrAF		225
	EArAF		230
	EGrAF		240
	ESrAF		255
	EHrAF		nsi
	ERrAF		nsi
	ER(Tos)rAF		nsi
EMXAF	EMkAF		299
	EMVAF		nsi
	EMLAF		nsi
	EMFAF		nsi
EWXAF	EWwAF		46[d]
	EWiAF		nsi
	EWlAF		nsi
	EWfAF		nsi
EMrXF	EMrIF		231
	EMrFF		nsi
	EMrRF		nsi

[a] X positions (P2′−P3 residues) were substituted with the selected L- and D-amino acids (upper- and lower-case letters, respectively). [b] Peptides were synthesized as carboxy-amides with free (NH_2) or acetylated (Ac) N-termini; R(Tos), tosyl-arginine. [c] The inhibition constants (K_i) were determined in a kinetic activity assay with SmCB1 and the fluorogenic peptidyl substrate, Z-Phe-Arg-AMC. Mean values are given for triplicate measurements (SE values were within 10% of means); nsi, no significant inhibition with 300 μM peptide. [d] Peptides were screened against human cathepsins B, L, and S under conditions analogous to those used for SmCB1, and no significant inhibition was observed at a 300 μM peptide concentration.

Finnigan Mat device (Thermo Finnigan). The characterization of the synthesized peptides is presented in Supplementary Table S1.

SmCB1 Activity and Inhibition Assays. SmCB1 activity was measured with the fluorogenic substrate Z-Phe-Arg-AMC.[7] The reaction was performed in a 96-well microplate format in a total assay volume of 100 μL in 0.1 M sodium acetate, pH 5.5, containing 2 mM dithiothreitol and 0.1% (w/v) PEG 6000. For inhibition measurements, the enzyme (0.6 nM) was preincubated with synthetic peptide (0−400 μM) at 37 °C for 15 min followed by the addition of 25 μM Z-Phe-Arg-AMC. Where indicated, the 17 kDa heparin was added to the preincubation mixture to give a final concentration of 25 μg mL^{-1} in the assay; this value was selected based on the fact that the potential of heparin to promote the SmCB1 inhibition (by peptides [45p−69p] and [55p−61p]) did not vary significantly in the tested range of 0.5−50 μg mL^{-1} (data not shown). The kinetics of product release was continuously monitored in an Infinite M200 microplate reader (Tecan) at excitation and emission wavelengths of 360 and 465 nm, respectively. Each measurement was performed in triplicate. The inhibition constants (K_i) were determined from the residual velocities using the dose−response plot (v_i/v_0 vs [I]) to obtain IC_{50} values that were then converted to K_i values by Cheng and Prusoff relationships.[47] The inhibition mode was determined using an analogous activity assay and the initial velocities of product release were interpreted using the Lineweaver−Burk plot. The active site concentration of SmCB1 was determined by titration with E-64, and the peptide solutions were quantified by amino acid analysis. The same activity assay was employed with bovine and human cathepsins B and human cathepsins L and S.

Stability of Peptides. Resistance of pro-peptide-derived fragments to proteolysis was investigated as described previously[48] under conditions similar to those used in the inhibition assay. The reaction mixtures contained 0.6 or 6.0 nM SmCB1 and 100 μM synthetic peptide in 0.1 mM sodium acetate, pH 5.5, containing 2 mM dithiothreitol (and 25 μg mL^{-1} 17 kDa heparin where indicated). The reaction was incubated at 37 °C for 16 h and stopped by the addition of 10 μM E-64. Mixtures were analyzed by RP-HPLC, and the obtained profiles were compared to those of controls not subjected to enzyme treatment. The RP-HPLC separation was performed on a Vydac C18 column (218TP54, Vydac) equilibrated in 5% (v/v) acetonitrile solution in 0.1% (v/v) TFA and eluted with a 2% min^{-1} gradient of a 90% (v/v) acetonitrile solution in 0.1% (v/v) TFA.

Secondary Structure Prediction. Prediction of secondary structure was performed using the PORTER and PSIPRED web services (http://distill.ucd.ie/porter/ and http://bioinf.cs.ucl.ac.uk/psipred/).

Molecular Modeling. A model of SmCB1 was constructed using the X-ray structure of rat cathepsin B (PDB entry 1CPJ) as a template and a pairwise sequence alignment generated by the BLAST program using a BLOSUM62 substitution matrix. The homology module of the MOE program was used for the modeling of the SmCB1 structure.[49] The Ac-NMrAF inhibitor was built into the SmCB1 active site cleft based on the position of three known active site ligands: (i) Compound 13 inhibitor from the human cathepsin L complex (PDB entry 1MHW) defined the orientation in the P1 to P3 positions; (ii) CA-074 inhibitor from the bovine cathepsin B complex (PDB entry 1QDQ) defined the orientation in the P2′ position; and (iii) CGT (42p−44p) segment in the pro-peptide of rat pro-cathepsin B (PDB accession 1MIR) defined the orientation in the P1′−P2 positions. The conformation of Ac-NMrAF was refined by applying the LigX module of the MOE for the optimization procedure. The final binding mode of the inhibitor was selected by the best-fit model based on the London dG scoring function and the generalized Born method.[50] The same strategy was used for the docking of other compounds derived from the active site region to evaluate their interaction potential with SmCB1. The structure of pro-SmCB1 was modeled on the basis of the X-ray structure of the rat pro-cathepsin B (PDB accession 1MIR) as template as described for the SmCB1 model. The multiple sequence alignment of the pro-peptide segments that was used for the SmCB1 modeling was generated by ClustalW2 (Figure 1a). Molecular images were generated with UCSF Chimera (http://www.cgl.ucsf.edu/chimera/).

Heparin Affinity Chromatography. SmCB1 (10 μg) or synthetic peptides (50 nmol) were applied to a 1 mL HiTrap Heparin HP column (GE Healthcare Bio-Sciences) equilibrated with 25 mM sodium acetate, pH 5.5. A linear 0−1 M NaCl gradient was used to elute the bound material at a flow rate of 1 mL min^{-1}. Chromatography was monitored by measuring the absorbance at 280 nm.

Thermal Stability Assay. The thermal denaturation curves of SmCB1 were recorded in a thermofluor assay using the LightCycler 480 System

(Roche) and a thin-wall PCR plate format.[40] A temperature increment of 1 °C min^{-1} was applied. Samples (25 μL) contained 0.8 μg of SmCB1 in 0.1 M sodium acetate, pH 5.5, in the presence or absence of 25 μg mL^{-1} of 17 kDa heparin. Protein unfolding was monitored by measuring the fluorescence signal of the hydrophobic reporter dye Sypro Orange (Invitrogen) with excitation and emission wavelengths of 465 and 580 nm, respectively. Melting temperature (T_m) values were calculated using the first derivative method.[40]

■ ASSOCIATED CONTENT

Supporting Information. This material is available free of charge *via* the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*Tel: (+420) 220183356. Fax: (+420) 220183578. E-mail: mares@uochb.cas.cz.

■ ACKNOWLEDGMENT

This work was supported by grants 203/09/1585 from the Grant Agency of the Czech Republic (GACR), KJB400550516 and IAA400550705 from the Grant Agency of the Academy of Sciences of the Czech Republic (GAASCR), institutional research project Z40550506, the Academy of Science of the Czech Republic, and the Sandler Center for Drug Discovery (C.R.C. and M.H.). C.R.C. and M.H. were recipients of a NATO Collaborative Linkage Grant (NATO LST/CLG 980187). We thank Miroslava Blechová for peptide synthesis, Lucie Bednárová for CD spectroscopy, and Irena Pražáková for technical assistance.

■ REFERENCES

(1) Utzinger, J., N'goran, E. K., Caffrey, C. R., Keiser, J. (2010) From innovation to application: Social-ecological context, diagnostics, drugs and integrated control of schistosomiasis, *Acta Trop.* E-pub ahead of print; DOI: 10.1016/j.actatropica.2010.08.020.

(2) Gryseels, B., Polman, K., Clerinx, J., and Kestens, L. (2006) Human schistosomiasis. *Lancet 368*, 1106–1118.

(3) Caffrey, C. R. (2007) Chemotherapy of schistosomiasis: present and future. *Curr. Opin. Chem. Biol. 11*, 433–439.

(4) Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvořák, J., Hsieh, I., Bahgat, M., Dissous, C., and McKerrow, J. H. (2006) A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J. Biol. Chem. 281*, 39316–39329.

(5) Caffrey, C. R., McKerrow, J. H., Salter, J. P., and Sajid, M. (2004) Blood 'n' guts: an update on schistosome digestive peptidases. *Trends Parasitol. 20*, 241–248.

(6) Dvořák, J., Mashiyama, S. T., Sajid, M., Braschi, S., Delcroix, M., Schneider, E. L., McKerrow, W. H., Bahgat, M., Hansell, E., Babbitt, P. C., Craik, C. S., McKerrow, J. H., and Caffrey, C. R. (2009) SmCL3, a gastrodermal cysteine protease of the human blood fluke Schistosoma mansoni. *PLoS Neglected Trop. Dis. 3*, e449.

(7) Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M. A., Lucas, K. D., Hsieh, I., Greenbaum, D., Bogyo, M., Salter, J. P., Lim, K. C., Franklin, C., Kim, J. H., and Caffrey, C. R. (2003) Functional expression and characterization of Schistosoma mansoni cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol. Biochem. Parasitol. 131*, 65–75.

(8) Brady, C. P., Brinkworth, R. I., Dalton, J. P., Dowd, A. J., Verity, C. K., and Brindley, P. J. (2000) Molecular modeling and substrate specificity of discrete Cruzipain-like and cathepsin L-like cysteine proteinases of the human blood fluke Schistosoma mansoni. *Arch. Biochem. Biophys. 380*, 46–55.

(9) Hola-Jamriska, L., Dalton, J. P., Aaskov, J., and Brindley, P. J. (1999) Dipeptidyl peptidase I and III activities of adult schistosomes. *Parasitology 118* (Pt 3), 275–282.

(10) Caffrey, C. R., Mathieu, M. A., Gaffney, A. M., Salter, J. P., Sajid, M., Lucas, K. D., Franklin, C., Bogyo, M., and McKerrow, J. H. (2000) Identification of a cDNA encoding an active asparaginyl endopeptidase of Schistosoma mansoni and its expression in *Pichia pastoris*. *FEBS Lett. 466*, 244–248.

(11) Brindley, P. J., Kalinna, B. H., Wong, J. Y., Bogitsh, B. J., King, L. T., Smyth, D. J., Verity, C. K., Abbenante, G., Brinkworth, R. I., Fairlie, D. P., Smythe, M. L., Milburn, P. J., Bielefeldt-Ohmann, H., Zheng, Y., and McManus, D. P. (2001) Proteolysis of human hemoglobin by schistosome cathepsin D. *Mol. Biochem. Parasitol. 112*, 103–112.

(12) Williamson, A. L., Lecchi, P., Turk, B. E., Choe, Y., Hotez, P. J., McKerrow, J. H., Cantley, L. C., Sajid, M., Craik, C. S., and Loukas, A. (2004) A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. *J. Biol. Chem. 279*, 35950–35957.

(13) Goldberg, D. E. (2005) Hemoglobin degradation. *Curr. Top. Microbiol. Immunol. 295*, 275–291.

(14) Sojka, D., Franta, Z., Horn, M., Hajdůšek, O., Caffrey, C. R., Mareš, M., and Kopáček, P. (2008) Profiling of proteolytic enzymes in the gut of the tick Ixodes ricinus reveals an evolutionarily conserved network of aspartic and cysteine peptidases. *Parasit. Vectors 1*, 7.

(15) Horn, M., Nussbaumerova, M., Šanda, M., Kovářová, Z., Srba, J., Franta, Z., Sojka, D., Bogyo, M., Caffrey, C. R., Kopáček, P., and Mareš, M. (2009) Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. *Chem. Biol. 16*, 1053–1063.

(16) Abdulla, M. H., Lim, K. C., Sajid, M., McKerrow, J. H., and Caffrey, C. R. (2007) Schistosomiasis mansoni: novel chemotherapy using a cysteine protease inhibitor. *PLoS. Med. 4*, e14.

(17) Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010) MEROPS: the peptidase database. *Nucleic Acids Res. 38*, D227–D233.

(18) Mach, L., Mort, J. S., and Glossl, J. (1994) Noncovalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass forms of the enzyme. *J. Biol. Chem. 269*, 13036–13040.

(19) Yamamoto, Y., Kurata, M., Watabe, S., Murakami, R., and Takahashi, S. Y. (2002) Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. *Curr. Protein Pept. Sci. 3*, 231–238.

(20) Wiederanders, B., Kaulmann, G., and Schilling, K. (2003) Functions of propeptide parts in cysteine proteases. *Curr. Protein Pept. Sci. 4*, 309–326.

(21) Roche, L., Tort, J., and Dalton, J. P. (1999) The propeptide of Fasciola hepatica cathepsin L is a potent and selective inhibitor of the mature enzyme. *Mol. Biochem. Parasitol. 98*, 271–277.

(22) Reis, F. C., Costa, T. F., Sulea, T., Mezzetti, A., Scharfstein, J., Bromme, D., Menard, R., and Lima, A. P. (2007) The propeptide of Cruzipain—a potent selective inhibitor of the trypanosomal enzymes Cruzipain and brucipain, and of the human enzyme cathepsin F. *FEBS J. 274*, 1224–1234.

(23) Fox, T., de, M. E., Mort, J. S., and Storer, A. C. (1992) Potent slow-binding inhibition of cathepsin B by its propeptide. *Biochemistry 31*, 12571–12576.

(24) Chagas, J. R., Ferrer-Di, M. M., Gauthier, F., and Lalmanach, G. (1996) Inhibition of cathepsin B by its propeptide: use of overlapping peptides to identify a critical segment. *FEBS Lett. 392*, 233–236.

(25) Lalmanach, G., Lecaille, F., Chagas, J. R., Authie, E., Scharfstein, J., Juliano, M. A., and Gauthier, F. (1998) Inhibition of trypanosomal cysteine proteinases by their propeptides. *J. Biol. Chem. 273*, 25112–25116.

(26) Horn, M., Dolečková-Marešová, L., Rulíšek, L., Máša, M., Vasiljeva, O., Turk, B., Gan-Erdene, T., Baudyš, M., and Mareš, M. (2005) Activation processing of cathepsin H impairs recognition by its propeptide. *Biol. Chem. 386*, 941–947.

(27) Pandey, K. C., Barkan, D. T., Sali, A., and Rosenthal, P. J. (2009) Regulatory elements within the prodomain of Falcipain-2, a cysteine protease of the malaria parasite *Plasmodium falciparum*. *PLoS One 4*, e5694.

(28) Korde, R., Bhardwaj, A., Singh, R., Srivastava, A., Chauhan, V. S., Bhatnagar, R. K., and Malhotra, P. (2008) A prodomain peptide of *Plasmodium falciparum* cysteine protease (falcipain-2) inhibits malaria parasite development. *J. Med. Chem. 51*, 3116–3123.

(29) Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure 4*, 405–416.

(30) Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., and Turk, V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Angstroms resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett. 384*, 211–214.

(31) Chowdhury, S. F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Menard, R., Lefebvre, J., Konishi, Y., Cygler, M., Sulea, T., and Purisima, E. O. (2002) Design of noncovalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides. *J. Med. Chem. 45*, 5321–5329.

(32) Li, Z., Hou, W. S., Escalante-Torres, C. R., Gelb, B. D., and Bromme, D. (2002) Collagenase activity of cathepsin K depends on complex formation with chondroitin sulfate. *J. Biol. Chem. 277*, 28669–28676.

(33) Almeida, P. C., Nantes, I. L., Chagas, J. R., Rizzi, C. C., Faljoni-Alario, A., Carmona, E., Juliano, L., Nader, H. B., and Tersariol, I. L. (2001) Cathepsin B activity regulation. Heparin-like glycosaminogylcans protect human cathepsin B from alkaline pH-induced inactivation. *J. Biol. Chem. 276*, 944–951.

(34) Caglic, D., Pungercar, J. R., Pejler, G., Turk, V., and Turk, B. (2007) Glycosaminoglycans facilitate procathepsin B activation through disruption of propeptide-mature enzyme interactions. *J. Biol. Chem. 282*, 33076–33085.

(35) Vasiljeva, O., Dolinar, M., Pungercar, J. R., Turk, V., and Turk, B. (2005) Recombinant human procathepsin S is capable of autocatalytic processing at neutral pH in the presence of glycosaminoglycans. *FEBS Lett. 579*, 1285–1290.

(36) Fairhead, M., Kelly, S. M., and van der Walle, C. F. (2008) A heparin binding motif on the pro-domain of human procathepsin L mediates zymogen destabilization and activation. *Biochem. Biophys. Res. Commun. 366*, 862–867.

(37) Cardin, A. D., and Weintraub, H. J. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis 9*, 21–32.

(38) Caffrey, C. R., Salter, J. P., Lucas, K. D., Khiem, D., Hsieh, I., Lim, K. C., Ruppel, A., McKerrow, J. H., and Sajid, M. (2002) SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. *Mol. Biochem. Parasitol. 121*, 49–61.

(39) Gandhi, N. S., and Mancera, R. L. (2008) The structure of glycosaminoglycans and their interactions with proteins. *Chem. Biol. Drug Des. 72*, 455–482.

(40) Uniewicz, K. A., Ori, A., Xu, R., Ahmed, Y., Wilkinson, M. C., Fernig, D. G., and Yates, E. A. (2010) Differential scanning fluorimetry measurement of protein stability changes upon binding to glycosaminoglycans: a screening test for binding specificity. *Anal. Chem. 82*, 3796–3802.

(41) Matulis, D., Kranz, J. K., Salemme, F. R., and Todd, M. J. (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry 44*, 5258–5266.

(42) Hamed, R. R., Maharem, T. M., and El-Guindy, A. S. (1997) Proteoglycans from adult worms of *Schistosoma haematobium*. *J. Helminthol. 71*, 151–160.

(43) Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Bromme, D., Ellman, J. A., and Craik, C. S. (2006) Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *J. Biol. Chem. 281*, 12824–12832.

(44) Yamamoto, A., Tomoo, K., Hara, T., Murata, M., Kitamura, K., and Ishida, T. (2000) Substrate specificity of bovine cathepsin B and its inhibition by CA074, based on crystal structure refinement of the complex. *J. Biochem. 127*, 635–643.

(45) Pohl, J., Baudyš, M., Tomášek, V., and Kostka, V. (1982) Identification of the active site cysteine and of the disulfide bonds in the N-terminal part of the molecule of bovine spleen cathepsin B. *FEBS Lett. 142*, 23–26.

(46) Dolečková-Marešová, L., Pavlík, M., Horn, M., and Mareš, M. (2005) De novo design of alpha-amylase inhibitor: a small linear mimetic of macromolecular proteinaceous ligands. *Chem. Biol. 12*, 1349–1357.

(47) Cheng, Y., and Prusoff, W. H. (1973) Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50% inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol. 22*, 3099–3108.

(48) Máša, M., Marešová, L., Vondrášek, J., Horn, M., Ježek, J., and Mareš, M. (2006) Cathepsin D propeptide: mechanism and regulation of its interaction with the catalytic core. *Biochemistry 45*, 15474–15482.

(49) MOE: Chemical Computing Group, 1010 Sherbrooke St.West, Suite 910, Montreal, Canada H3A 2R7; http://www.chemcomp.com.

(50) Labute, P. (2008) The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem. 29*, 1693–1698.

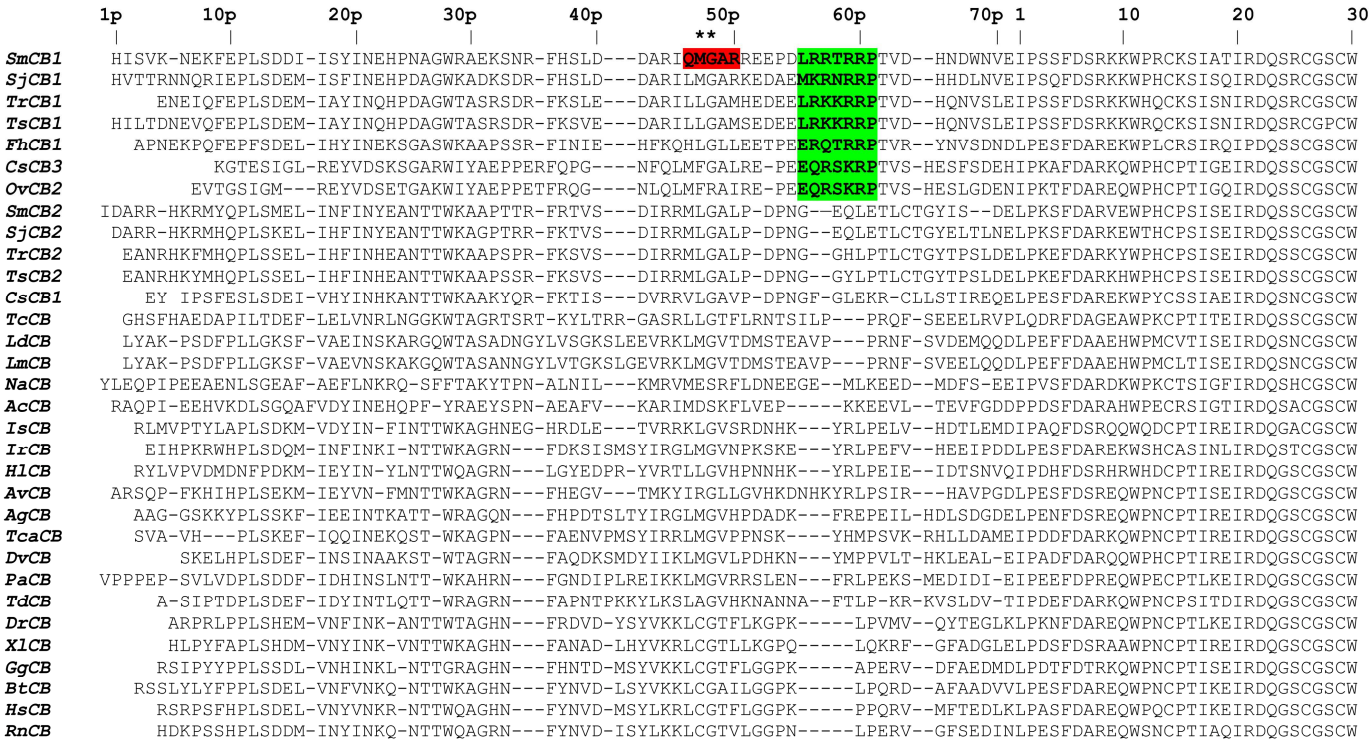
Supporting information

for

Mapping the pro-peptide of the *Schistosoma mansoni* cathepsin B1 drug target: modulation of inhibition by heparin and design of mimetic inhibitors

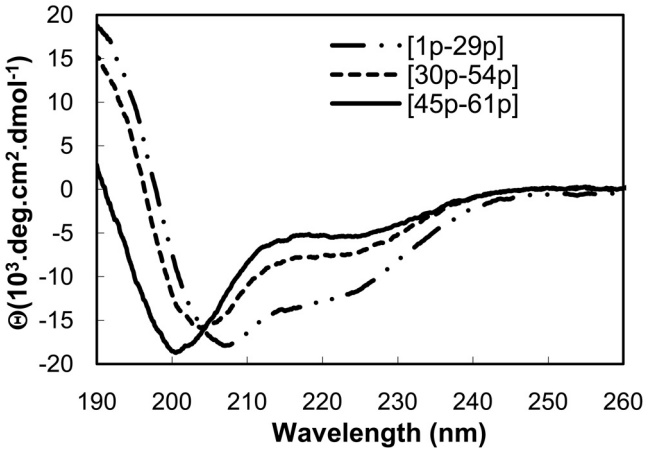
Martin Horn, Adéla Jílková, Jiří Vondrášek, Lucie Marešová,

Conor R. Caffrey, Michael Mareš



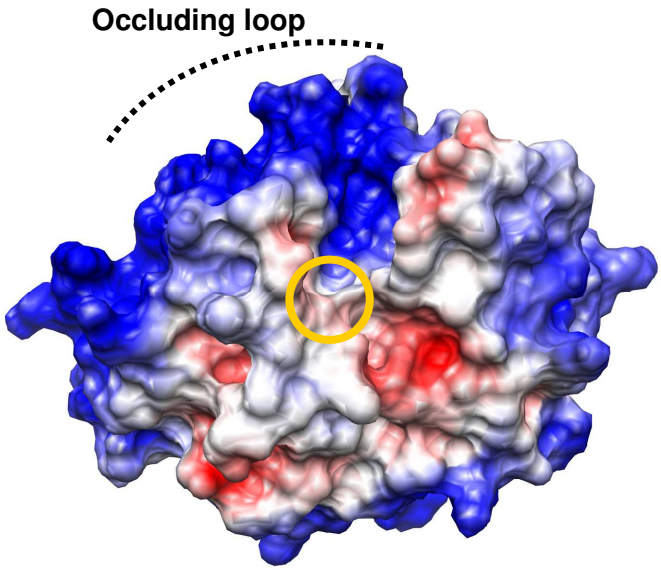
Supplementary Figure S1.

Sequence alignment of pro-SmCB1 with representative members of cathepsin B-type proteases. Only sequences for the pro-peptide and N-terminal portions of mature enzyme are presented. The active site and heparin-binding regions of SmCB1 are highlighted in red and green, respectively. Residues in contact with S1 and S1´ subsites of the active site are marked with asterisks. Every tenth amino acid residue in the sequence is numbered and the suffix 'p' refers to the pro-peptide. The sources of the sequences are as follows: trematodes: *Schistosoma mansoni* (SmCB1 and SmCB2, Uniprot accessions Q8MNY2 and Q95PM1, respectively), *Schistosoma japonicum* (SjCB1 and SjCB2, P43157 and C7TYR4), *Trichobilharzia regenti* (TrCB1 and TrCB2, Q4VRW9 and A7L844), *Trichobilharzia szidati* (TsCB1 and TsCB2, B5AXI3 and B5AXI4), *Fasciola hepatica* (FhCB1, Q8I7B2), *Chlonorchis sinensis* (CsCB3, A3R0V6) and *Opisthorchis viverrini* (OvCB2, C7EXK1); protozoa: *Trypanosoma cruzi* (TcCB, O61066), *Leishmania donovani* (LdCB, Q9GQM5) and *Leishmania major* (LdCB, P90627); nematodes: *Necator americanus* (NaCB, A1YUM5) and *Ancylostoma caninum* (AcCB, Q11007); arthropods: *Ixodes ricinus* (IrCB, A4GTA7), *Ixodes scapularis* (IsCB, B7PF28), *Araneus ventricosus* (AvCB,Q7Z0Z2), *Haemaphysalis longicornis* (HlCB, A1IHG0), *Anopheles gambiae* (AgCB, Q7Q9Y3), *Tribolium castaneum* (TcaCB, D6WGZ1), *Diabrotica virgifera* (DvCB,Q70EX1), *Periplaneta americana* (PaCB, Q5MBV5) and *Triatoma dimidiata* (TdCB, Q5GH35); vertebrates: *Danio rerio* (DrCB, Q6PH75), *Xenopus laevis* (XlCB, Q7ZXM4), *Gallus gallus* (GgCB, P43233), *Bos taurus* (BtCB, P07688), *Homo sapiens* (HsCB, P07858) and *Rattus norvegicus* (RnCB, P00787). Of the above trematode enzymes, the following have been localized in the digestive tract; SmCB1 (Sajid *et al.*, Mol. Biochem. Parasitol. 131, 65-75, 2003), SjCB1 (Caffrey and Ruppel, J. Parasitol. 83,1112-1118, 1997), TrCB1 (Dvořák *et al.*, Int. J. Parasitol. 35, 895-910, 2005), and FhCB1 (Beckham *et al.*, Int. J. Parasitol. 41, 1601-1612, 2009).



Supplementary Figure S2.

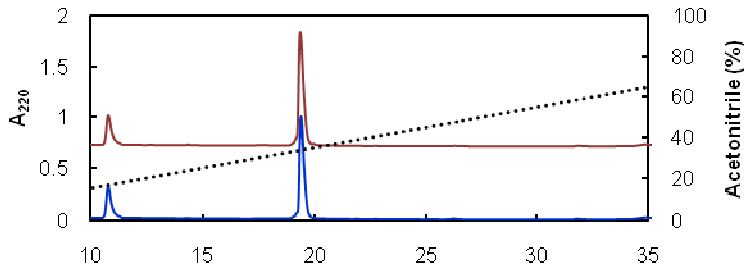
Circular dichroism spectra of α-helical fragments derived from the SmCB1 pro-peptide. The peptides [1p-29p], [30p-54p], and [45p-61p] contain the α-helices α1p, α2p, and α3p, respectively (see Figure 2). The spectra were measured with 0.3 mM peptide in 50 mM sodium phosphate, pH 5.8 in the presence of 50% (v/v) trifluoroethanol (TFE) at 26℃ using a CD6 Dichrograph (Jobin Yvon). The CD spectra were also measured in the absence of TFE and with 20% (v/v) TFE. The relative content of helical structure was determined from the spectra by Dichroweb (Whitmore and Wallace, Nucleic Acids Res 32, W668-W673, 2004); the values for 0%, 20%, and 50% TFE, respectively, are as follows: 22%, 31%, and 39% for peptide [1p-29p]; 15%, 23%, and 27% for peptide [30p-54p]; 17%, 19%, and 22% for peptide [45p-61p].



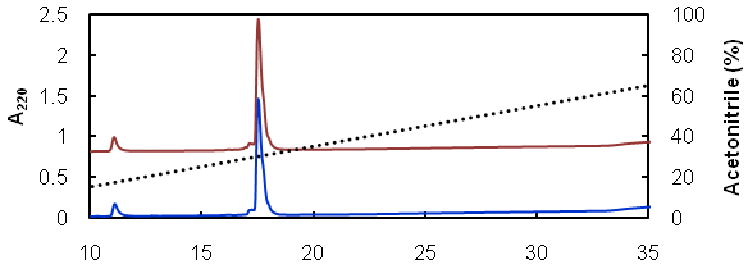
Supplementary Figure S3.

The surface of the mature SmCB1 molecule is colored by its electrostatic potential displayed at a scale from –5 kT/e (red) to +5 kT/e (blue). The protein is in the same 'front site' orientation as in Figure 1. The open yellow circle indicates the position of the active site Cys residue. The model was constructed using the X-ray structure of the rat cathepsin B (PDB entry 1CPJ) (see Methods). The electrostatic potential surface was calculated for pH 5.5 using the APBS software (Baker *et al.,* Proc. Natl. Acad. Sci. U.S.A. 98, 10037-10041, 2001) and input data were prepared with the PDB2PQR (Dolinsky *et al.*, Nucleic Acids Res 32, W665-W667, 2004) and PROPKA programs (Li *et al.*, Proteins 61, 704-721, 2005).

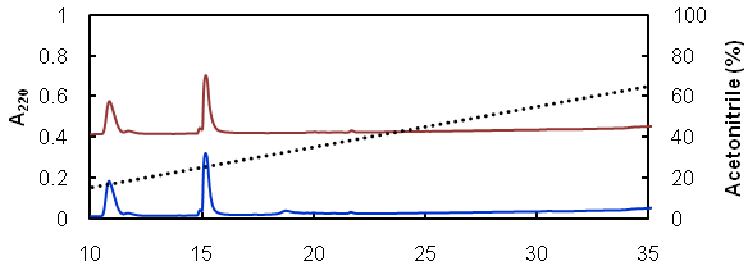
a


A220
Acetonitrile (%)

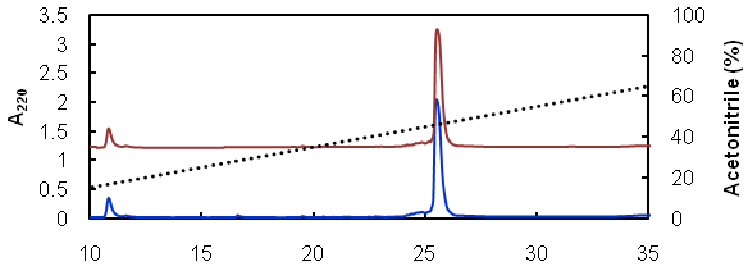
b


A220
Acetonitrile (%)

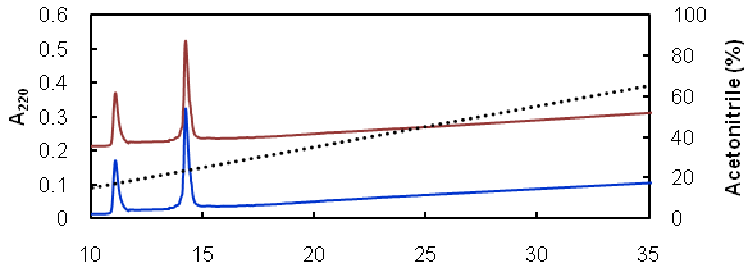
c


A220
Acetonitrile (%)

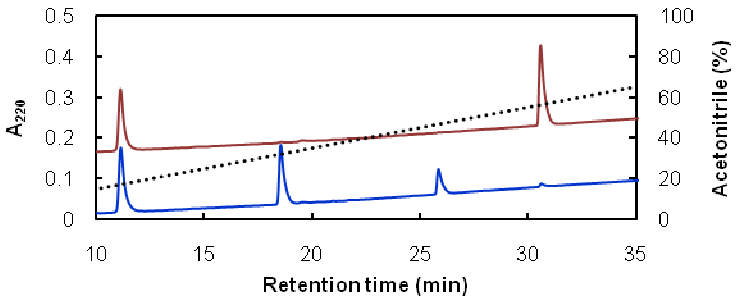
d


A220
Acetonitrile (%)

e


A220
Acetonitrile (%)

f



Supplementary Figure S4.

Stability of synthetic pro-peptide-derived inhibitors to proteolysis by SmCB1. The analysis is presented for the following peptide inhibitors: (a) [30p-54p], (b) [55p-69p], (c) [55p-61p], (d) [45p-69p], (e) [45p-49p], and for (f) a fluorogenic peptidic substrate of SmCB1 Z-Phe-Arg-AMC. The reaction mixtures (200 µM) containing 6.0 nM SmCB1 and 100 µM peptide in 0.1 mM sodium acetate, pH 5.5, containing 2 mM dithiothreitol and 25 µg mL^{-1} heparin (for [55p-61p] and [45p-65p]), were incubated at 37 °C for 16 h. The reaction was stopped by the addition of 10 µM E-64 and mixtures were analyzed by RP-HPLC (blue lines). In the control experiments (brown lines), the peptide was added to the reaction mixture after the incubation was stopped by E-64. Chromatography was performed on a Vydac C18 column equilibrated in 5% (v/v) acetonitrile solution in 0.1% (v/v) TFA and eluted with a 2% min^{-1} gradient (dotted line) of 90% (v/v) acetonitrile solution in 0.1% (v/v) TFA. The RP-HPLC elution profile was monitored by measuring absorbance at 220 nm. The peak at the retention time of around 11 min is associated with the assay system and not with the tested peptides.

Supplementary Table S1. Characterization of purified peptides prepared by Fmoc solid phase chemistry.

Peptide[a]	Theoretical Mass (Da)	Measured Mass (Da)[b] m/z	HPLC purity[c,d] %
Pro-peptide derived fragments from Figure 2			
[1p-29p]	3448.7	$[M+3H]^{3+}$ 1150.4	91.3
[30p-54p]	2968.4	$[M+2H]^{2+}$ 1485.3	95.1
[25p-54p]	3552.7	$[M+3H]^{3+}$ 1185.4	91.5
[53p-69p]	2189.37	$[M+2H]^{2+}$ 1095.4	99.8
[55p-69p]	1976.0	$[M+2H]^{2+}$ 989.5	94.3
[55p-69p]Cit_2	1978.1	$[M+2H]^{2+}$ 990.3	97.7
[55p-61p]	994.6	$[M+H]^{+}$ 995.6	92.5
[45p-69p]	3145.5	$[M+3H]^{3+}$ 1050.1	91.1
[45p-69p]Cit_2	3147.6	$[M+3H]^{3+}$ 1050.9	94.4
[45p-61p]	2164.2	$[M+3H]^{3+}$ 722.4	98.3
[39p-61p]	2847.5	$[M+3H]^{3+}$ 950.3	95.6
[39p-52p]	1699.8	$[M+2H]^{2+}$ 850.8	97.1
[39p-49p]	1285.6	$[M+H]^{+}$ 1286.4	94.5
[45p-52p]	1016.5	$[M+H]^{+}$ 1017.5	98.1
[45p-49p]	602.3	$[M+H]^{+}$ 603.3	98.9
[45p-48p]	446.2	$[M+H]^{+}$ 447.2	94.5
[46p-49p]	474.2	$[M+H]^{+}$ 475.2	98.6
Active site region derived peptides from Table 2			
QMGARREE	974.5	$[M+H]^{+}$ 975.5	95.6
Ac-QMrARREE	1115.6	$[M+H]^{+}$ 116.5	98.2
QMrARREE	1073.6	$[M+H]^{+}$ 1074.6	96.8
QMGAR	560.3	$[M+H]^{+}$ 561.3	92.7
Ac-QMrAF	692.3	$[M+H]^{+}$ 693.3	99.8
QMrAF	650.3	$[M+H]^{+}$ 651.4	98.4
Ac-NMrAF	678.3	$[M+H]^{+}$ 679.4	96.1
NMrAF	636.3	$[M+H]^{+}$ 673.3	94.9
Ac-EMrAF	693.3	$[M+H]^{+}$ 694.3	98.8
EMrAF	651.3	$[M+H]^{+}$ 652.3	98.3
Ac-DMrAF	679.3	$[M+H]^{+}$ 680.4	98.4
DMrAF	637.3	$[M+H]^{+}$ 638.4	97.6
Ac-PWrAF	716.4	$[M+H]^{+}$ 717.4	98.1
PWrAF	674.4	$[M+H]^{+}$ 675.4	97.5
Ac-HWrAF	756.4	$[M+H]^{+}$ 757.4	96.5
HWrAF	714.4	$[M+H]^{+}$ 715.4	97.4
Ac-RWrAF	775.4	$[M+H]^{+}$ 776.4	96.8
RWrAF	733.4	$[M+H]^{+}$ 734.4	99.1
Ac-QWrAF	747.4	$[M+H]^{+}$ 748.4	95.6
QWrAF	705.4	$[M+H]^{+}$ 706.4	97.4
Ac-SWrAF	706.4	$[M+H]^{+}$ 707.4	96.3
SWrAF	664.4	$[M+H]^{+}$ 665.4	97.2
Ac-TWrAF	720.4	$[M+H]^{+}$ 721.4	98.2
TWrAF	678.4	$[M+H]^{+}$ 679.4	97.4
Ac-NWrAF	733.4	$[M+H]^{+}$ 734.4	98.5
Ac-NYrAF	710.3	$[M+H]^{+}$ 711.3	97.6
Ac-NFrAF	694.4	$[M+H]^{+}$ 695.4	98.3
EYrAF	683.3	$[M+H]^{+}$ 684.4	96.1
EWrAF	706.4	$[M+H]^{+}$ 707.4	96.6
ECrAF	623.3	$[M+H]^{+}$ 624.3	95.1
EFrAF	667.3	$[M+H]^{+}$ 668.4	97.8
EIrAF	633.3	$[M+H]^{+}$ 634.3	97.7
ELrAF	633.3	$[M+H]^{+}$ 634.3	96.8
EArAF	591.3	$[M+H]^{+}$ 592.3	98.7
EGrAF	577.3	$[M+H]^{+}$ 578.3	95.9
ESrAF	607.3	$[M+H]^{+}$ 608.4	97.5
EHrAF	657.3	$[M+H]^{+}$ 658.3	96.8
ERrAF	676.4	$[M+H]^{+}$ 677.4	99.1
ER(Tos)AF	830.4	$[M+H]^{+}$ 831.4	98.1
EMkAF	623.3	$[M+H]^{+}$ 624.4	96.1

EMVAF	594.3	$[M+H]^+$ 595.3	95.9
EMLAF	608.3	$[M+H]^+$ 609.2	97.1
EMFAF	642.3	$[M+H]^+$ 643.1	97.3
EWwAF	736.3	$[M+H]^+$ 737.3	97.5
EWiAF	663.3	$[M+H]^+$ 664.2	96.8
EWlAF	663.3	$[M+H]^+$ 664.2	97.4
EWfAF	697.3	$[M+H]^+$ 698.2	95.1
EMrIF	693.4	$[M+H]^+$ 694.6	98.5
EMrFF	727.3	$[M+Na]^+$ 750.5	96.2
EMrRF	736.4	$[M+H]^+$ 737.7	97.3

[a] Pro-peptide derived fragments from Figure 2 were synthesized as carboxy-amides with acetylated N-termini. Active site region derived peptides from Table 2 were synthesized as carboxy-amides with free or acetylated (Ac-) N-termini. R(Tos), tosyl-arginine; upper- and lower-case letters, L- and D-amino acids, respectively.

[b] ESI mass spectra were measured on an LCQ Classic Finnigan Mat device (Thermo Finnigan).

[c] HPLC analysis was performed on a Vydac C18 column equilibrated in 0.1% (v/v) TFA and eluted with a 1% min^{-1} gradient of 90% (v/v) acetonitrile solution in 0.1% (v/v) TFA.

[d] The relative amino acid composition of peptides was determined by amino acid analysis on Biochrom 20 (Amersham Biosciences); the obtained experimental values were within 10% of the theoretical values.

5.2. Publikace č. 2: Structural Basis for Inhibition of Cathepsin B Drug Target from the Human Blood Fluke, *Schistosoma mansoni*

5.2.1. Souhrn

Tato publikace nabízí odlišný přístup pro hledání inhibitorů SmCB1, než byl použit v publikaci č. 1. Jedná se o plošné testování sady syntetických peptidomimetických vinylsulfonů, které se kovalentně váží do aktivního místa cysteinových proteas. Publikace se zabývá komplexní strukturně funkční charakterizací SmCB1, která poskytuje detailní popis jednotlivých vazebných podmíst aktivního místa pomocí (1) analýzy získaných krystalových struktur komplexu SmCB1 s inhibitory, (2) analýzy vazby inhibitorů do aktivního místa SmCB1 pomocí metod výpočetní chemie, (3) určení inhibiční specifity SmCB1 pomocí vinylsulfonových inhibitorů a (4) určení substrátové specifity SmCB1 pomocí fluorogenních substrátů a fragmentace přirozeného substrátu hemoglobinu.

Tato publikace přináší vůbec první krystalovou strukturu proteasy z parazita krevničky. Struktura SmCB1 vykazuje obecné znaky proteas rodiny papainu a obsahuje smyčku "occluding loop" charakteristickou pro katepsiny B. Byly určeny celkem tři krystalové struktury SmCB1 v komplexu s inhibitory K11777 (PDB kód: 3S3R), K11017 (PDB kód: 3S3Q) a CA074 (PDB kód: 3QSD), které dávají nové informace o vazbě inhibitorů v aktivním místě. K11777 je vinylsulfon a byl vybrán jako účinný inhibitor SmCB1, který v r. 2007 poskytl zásadní důkaz o SmCB1 jako cílové molekule pro terapeutický zásah při léčbě schistosomózy na myším modelu [23]. K11017 je derivát K11777, který se liší substituenty v P2 a P3 pozici. CA074 je epoxidový inhibitor specifický pro katepsin B s mechanismem inhibice odlišným od vinylsulfonů. Krystalové struktury komplexů SmCB1 s inhibitory byly dále použity pro analýzu interakce substituentů v jednotlivých vazebných podmístech enzymu pomocí kvantově chemických výpočtů.

V další části práce byla testována sada 20 vinylsulfonových inhibitorů cysteinových proteas odvozených ze struktury K11777, které obsahují různé substituenty v pozicích P3 až P1´. Inhibitory pocházejí z knihovny inhibitorů ze spolupracujícího pracoviště UCSF. Tato sada byla testována *in vitro* v aktivitním testu s rekombinantním SmCB1 a *ex vivo* na živých larvách *S. mansoni*. Bylo prokázáno, že *in vitro* i *ex vivo* výsledky korelují a účinnost inhibice podle parametru IC_{50} odpovídá míře suprese parazita.

Proteolytická aktivita SmCB1 byla studována pomocí fyziologického substrátu hemoglobinu a pomocí syntetických peptidových substrátů určených pro exopeptidasový a endopeptidasový mód štěpení. Oba přístupy prokázaly, že SmCB1 je schopen využívat jak

endo- tak exopeptidasovou aktivitu, ale exopeptidasová aktivita převládá. Detailní exopeptidasová specifita SmCB1 byla dále studována pomocí knihoven FRET substrátů, navržených pro analýzu aminokyselin preferovaných v P1´ a P2´ pozicích.

Tato publikace prokázala význam SmCB1 jako cílové molekuly pro vývoj inhibitorů jako potenciálních chemoterapeutik proti schistosomóze.

Můj podíl na práci zahrnoval: (1) rekombinantní expresi, aktivaci a purifikaci SmCB1, (2) přípravu inhibičních komplexů a jejich krystalizaci, analýzu krystalografických struktur (pomocí programů CONTACT a Pymol), jejich interpretaci a molekulární grafiku (pomocí programu Pymol), (3) testování knihoven FRET substrátů a analýzu degradace hemoglobinu, určení inhibičních konstant IC_{50}, *ex vivo* testování inhibitorů a vyhodnocování změn fenotypu parazita a (4) účast na přípravě manuskriptu.

5.2.2. Publikace č. 2

Jílková, **A.**, Řezáčová, P., Lepšík, M., Horn, M., Váchová, J., Fanfrlík, J., Brynda, J., McKerrow, J. H., Caffrey, C. R. a Mareš, M.

Structural Basis for Inhibition of Cathepsin B Drug Target from the Human Blood Fluke, *Schistosoma mansoni*

Journal of Biological Chemistry, 286 (41), 35770-35781 (2011)

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 41, pp. 35770–35781, October 14, 2011

Structural Basis for Inhibition of Cathepsin B Drug Target from the Human Blood Fluke, *Schistosoma mansoni*[*][S]

Received for publication, June 15, 2011, and in revised form, July 19, 2011 Published, JBC Papers in Press, August 10, 2011, DOI 10.1074/jbc.M111.271304

Adéla Jílková[‡§], Pavlína Řezáčová[‡¶], Martin Lepšík[‡], Martin Horn[‡], Jana Váchová[‡], Jindřich Fanfrlík[‡], Jiří Brynda[‡¶], James H. McKerrow[‖], Conor R. Caffrey[‖], and Michael Mareš[‡1]

From the [‡]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, 16610 Prague, Czech Republic, the [§]Department of Biochemistry, Faculty of Science, Charles University, 12843 Prague, Czech Republic, the [¶]Department of Structural Biology, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, 16610 Prague, Czech Republic, and the [‖]Sandler Center for Drug Discovery, California Institute for Quantitative Biosciences and Department of Pathology, University of California San Francisco, San Francisco, California 94158

Schistosomiasis caused by a parasitic blood fluke of the genus *Schistosoma* afflicts over 200 million people worldwide. *Schistosoma mansoni* cathepsin B1 (SmCB1) is a gut-associated peptidase that digests host blood proteins as a source of nutrients. It is under investigation as a drug target. To further this goal, we report three crystal structures of SmCB1 complexed with peptidomimetic inhibitors as follows: the epoxide CA074 at 1.3 Å resolution and the vinyl sulfones K11017 and K11777 at 1.8 and 2.5 Å resolutions, respectively. Interactions of the inhibitors with the subsites of the active-site cleft were evaluated by quantum chemical calculations. These data and inhibition profiling with a panel of vinyl sulfone derivatives identify key binding interactions and provide insight into the specificity of SmCB1 inhibition. Furthermore, hydrolysis profiling of SmCB1 using synthetic peptides and the natural substrate hemoglobin revealed that carboxydipeptidase activity predominates over endopeptidolysis, thereby demonstrating the contribution of the occluding loop that restricts access to the active-site cleft. Critically, the severity of phenotypes induced in the parasite by vinyl sulfone inhibitors correlated with enzyme inhibition, providing support that SmCB1 is a valuable drug target. The present structure and inhibitor interaction data provide a footing for the rational design of anti-schistosomal inhibitors.

Schistosomiasis (bilharzia) is a chronic infectious disease caused by trematode blood flukes that infect over 200 million people in tropical and subtropical areas (1). Of the five species of schistosomes infecting humans, *Schistosoma mansoni* is a major etiological agent of disease in parts of Asia, the Middle East, Africa, and South America. Morbidity associated with the disease arises from immunopathological reactions to parasite eggs that accumulate in various tissues, including the liver, intestinal tract, and bladder (2). Treatment and control of schistosomiasis now relies on just one drug, praziquantel, a perilous situation should drug resistance emerge and become established (1, 3). Accordingly, there is continued impetus to identify new schistosomal protein targets and chemotherapeutically active anti-schistosomals (4–6).

Adult schistosomes live in the cardiovascular system, and host blood proteins are a nutritive source for growth, development, and reproduction. In the schistosome gut, a network of peptidases (proteases) contributes to the digestion of host proteins, predominated by hemoglobin, to absorbable peptides and amino acids (7, 8). For *S. mansoni*, the component digestive peptidases thus far characterized include the following: (i) cysteine peptidases of the Clan CA (papain family), namely cathepsin B1, cathepsins L1–L3, dipeptidyl peptidase I (cathepsin C), and a Clan CD asparaginyl endopeptidase (legumain); (ii) the Clan AA aspartic peptidase, cathepsin D; and (iii) the Clan MF metallopeptidase, leucine aminopeptidase (7–11). This study focuses on *S. mansoni* cathepsin B1 (SmCB1),[2] which is the most abundant cysteine peptidase in the parasite gut (12, 13) and is necessary for normal parasite growth (14). SmCB1 is synthesized as an inactive zymogen and is converted *in vitro* to a mature, active 31-kDa enzyme by proteolytic removal of the pro-peptide that can be catalyzed by legumain (12). SmCB1 is a molecular target for cure of schistosomiasis mansoni in a mouse model using the vinyl sulfone cysteine peptidase inhibitor K11777 (15). Inhibition of SmCB1 therefore represents an attractive option for anti-schistosomal drug development; however, target-based rational design of lead compounds has been hampered by a lack of structural information for the enzyme.

* This work was supported, in whole or in part, by Grant 203/09/1585 from the Grant Agency of the Czech Republic, Institutional Research Projects Z40550506 and Z50520514, the Academy of Sciences of the Czech Republic, and by the Sandler Center for Drug Discovery. Additional support was by Grant P208/11/0295 (to M. L. and J. F.) from the Grant Agency of the Czech Republic and by Grant OC09007 (to M. M.) from the Ministry of Education, Youth and Sports of the Czech Republic. Use of the Advanced Photon Source was supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1–S5, Figs. S1–S5, and additional references.

The atomic coordinates and structure factors (codes 3QSD, 3S3Q, and 3S3R) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] To whom correspondence should be addressed: Flemingovo n. 2, 16610 Prague, Czech Republic. Tel.: 420-220183356; Fax: 420-220183578; E-mail: mares@uochb.cas.cz.

[2] The abbreviations used are: SmCB1, cathepsin B1 from *S. mansoni*; VSPh, phenyl vinyl sulfone; RP-HPLC, reverse phase HPLC; AMC, aminomethylcoumarin; Cbz, carboxybenzyl; Hph, homophenylalanine; NTS, newly transformed schistosomula; r.m.s.d., root mean square deviation; Abz, aminobenzoic acid; Nph, 4-nitrophenylalanine; EDDnp, ethylenediamine 2,4-dinitrophenyl; Tricine, *N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; *N*-Mpip, *N*-methylpiperazinylcarbonyl; Mu, morpholinylcarbonyl; Cbz, benzyloxycarbonyl.

Recently, we designed reversible inhibitors of SmCB1 based on the pro-peptide scaffold. These were effective *in vitro* in the low micromolar range (16). Here, we identify covalent nanomolar inhibitors of SmCB1 and analyze their binding mode by structural analysis. The inhibitors include the following: (i) epoxide inhibitor CA074, a specific inhibitor of cathepsin B-type peptidases (17) that has been previously structurally characterized in complex with mammalian cathepsins B (18), and (ii) vinyl sulfone inhibitors K11017 and K11777 that have not been crystallographically studied so far in complex with cathepsins B. Vinyl sulfones are effective against papain-like cysteine peptidases and were originally investigated in the context of inhibiting human cathepsins (19, 20). Later, they were demonstrated to inhibit cysteine peptidases from a variety of protozoan pathogens such as *Trypanosoma* and *Plasmodium*, and provide either parasitological cure or a temporary remission of parasitemia (21–23). As a chemotype, vinyl sulfones have acceptable pharmacokinetic attributes and *in vivo* safety profiles (24, 25). Currently, K11777 is in pre-clinical development as an anti-chagasic compound (26).

Here, we report the crystallographic structure of SmCB1, the first for a schistosomal proteolytic enzyme. A comprehensive analysis of structure-activity/inhibition relationships is provided to describe the active site of SmCB1. We demonstrate that SmCB1 is an efficient exopeptidase/endopeptidase against both synthetic peptide substrates and the physiologically relevant protein substrate, hemoglobin. Also, inhibition of SmCB1 by various vinyl sulfone inhibitors correlates with the severity of phenotypes induced in the parasite in culture. This study therefore provides both evidence that SmCB1 is a relevant drug target and the necessary structure-ligand data with which the design of anti-schistosomal SmCB1 inhibitors can be continued.

EXPERIMENTAL PROCEDURES

Cloning and Mutagenesis of SmCB1

The pPICZαA plasmid containing SmCB1 insert was constructed as described previously (12). A nonglycosylated mutant of SmCB1 was generated from the plasmid construct using site-directed mutagenesis performed by PCR according to QuikChange® system (Stratagene). A two-step PCR procedure was employed for disruption of consensus glycosylation sites Asn-His-Thr to Asn-His-Ala (residues 166–168) and Asn-Lys-Thr to Asn-Lys-Ala (residues 281–283) using forward 5′-AGTTCGAAGGAGAATCACGCAGGTTGTGAACCATATC-3′ and reverse 5′-GATATGGTTCACAACCTGCGTGATTCTCCTTCGAACT-3′ primers (for Thr to Ala-168 mutagenesis) and forward 5′-TGGGGTGTGGAAAACAAGGCTCCTTATTGGTTGATTG-3′ and reverse 5′-CAATCAACCAATAAGGAGCCTTGTTTTCCACACCCCA-3′ primers for Thr to Ala-283 mutagenesis. Constructs were sequenced to verify desired mutations.

Recombinant Expression and Purification of SmCB1

The nonglycosylated SmCB1 zymogen was expressed in the X33 strain of the methylotrophic yeast *Pichia pastoris*, purified and activated by *S. mansoni* legumain (27), as described previously (16). All purification steps were maintained under reducing conditions in the presence of 3.5 mM β-mercaptoethanol and 1 mM EDTA to prevent the active site cysteine from oxidation. The expressed nonglycosylated SmCB1 exhibited analogous activity properties as the wild-type SmCB1 produced in the *Pichia* expression system (12). The nonglycosylated SmCB1 was used in all experiments described here.

Preparation of Substrates and Inhibitors

Fluorogenic fluorescence resonance energy transfer (FRET) substrates Abz-Phe-Arg-Xaa-Nph-OH and Abz-Phe-Arg-Xaa-Nph-OH (the Xaa position contains one of the 19 proteinogenic amino acids except Cys) were synthesized by *N*-(9-fluorenyl)methoxycarbonyl (Fmoc) solid phase chemistry in an ABI 433A peptide synthesizer (Applied Biosystems) as described previously (28). Peptides were purified by RP-HPLC over a C18 column using a TFA/acetonitrile system and characterized by electrospray ionization mass spectrometry on an LCQ Classic Finnigan Mat device (Thermo Finnigan). FRET substrates Abz-Gln-Val-Val-Ala-Gly-Ala-EDDnp and Abz-Ala-Phe-Arg-Phe-Ser-Gln-EDDnp, fluorogenic substrate Cbz-Phe-Arg-AMC, and CA074 inhibitor were purchased from Bachem. Vinyl sulfone inhibitors (Table 1) with prefix K were synthesized as described previously (19, 29), with prefix WRR as described previously (30–33), and with prefix AR as described previously (34) and were kindly provided by James T. Palmer, William R. Roush, and Adam R. Renslo.

Preparation of SmCB1-Inhibitor Complexes

The freshly activated SmCB1 was incubated (10 h, 18 °C) with a 5-fold molar excess of the inhibitor in 0.1 M sodium acetate, pH 5.5, containing 15 mM cysteine and 1 mM EDTA. The enzyme inhibition was monitored with Cbz-Phe-Arg-AMC substrate. The complex was rechromatographed on an FPLC Mono S column (16), concentrated, and buffer-exchanged into 2.5 mM sodium acetate, pH 5.5, using Amicon Ultracel-10k centrifugal units (Millipore).

Protein Crystallization and Data Collection

Crystals were obtained by vapor diffusion in hanging drop. Drops consisting of 1 μl of the protein solution and 1 μl of the reservoir solution were equilibrated over 1-ml reservoir solution at 20 °C. The reservoirs solutions were 0.2 M ammonium acetate, 0.1 M sodium citrate, 30% PEG 1500, pH 6.2, for SmCB1-K11777 and SmCB1-K11017 complexes, and 0.1 M sodium citrate, 0.2 M ammonium acetate, 30% PEG 1500, pH 6.1, for SmCB1-CA074 complex. Protein concentrations of the stock solutions of the complexes were 2.5–5 mg/ml (in 2.5 mM sodium acetate, pH 5.5). Rectangle-shaped crystals reached their final size within 10 days and were flash-cooled by plunging into liquid nitrogen. Diffraction data at 100 K were collected at beamline 19-BM of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory, Argonne, IL All diffraction data were processed using the HKL-3000 suite of programs (35). Crystal parameters and data collection statistics are given in supplemental Table S1.

Structure Determination, Refinement, and Analysis

The structure of SmCB1 was determined by molecular replacement with the program Molrep (36) using the structure

of human cathepsin B (Protein Data Bank code 1HUC) (37) as the search model. The sequence alignment of SmCB1 with human cathepsin B is shown in supplemental Fig. S1. Model refinement was carried out using the program REFMAC 5.2 (38) from the CCP4 package (39), interspersed with manual adjustments using Coot (40). The final steps included TLS refinement (41). The quality of the final models was validated with Molprobity (42). Final refinement statistics are given in supplemental Table S1. Atomic coordinates and experimental structure factors have been deposited with the Protein Data Bank with the codes 3QSD, 3S3R and 3S3Q for SmCB1-CA074, SmCB1-K11777, and SmCB1-K11017, respectively. The following services were used to analyze the structures: PISA server (43) and CONTACTS (39). All figures showing structural representations were prepared with the program PyMOL (44).

Interaction Energy Calculations

The subsite interaction energy between inhibitors and SmCB1 was calculated using the quantum chemical approach. The strategy consisted of the following two steps: optimization of the crystallographic complexes and calculation of interaction energies in the individual subsites.

Model Setup and Geometry Optimization—Hydrogen atoms were added to the crystallographic complexes of SmCB1 and inhibitors to correspond to pH of ~6 and were optimized using the AMBER 10 software (45). Further optimization of the inhibitor in the active site (residues within 6 Å) was carried out using the corrected semi-empirical quantum chemical method PM6-DH2 (46, 47), including implicit water environment. The alternative conformation of VSPh in P1′ position of inhibitors was modeled using PyMOL (44).

Subsite Interaction Energies—The inhibitor structures were fragmented into the P3 to P2′ segments, with separated side chains and the main chains, and capped by hydrogens. The reactive centers of the inhibitors originating from the vinyl and epoxide moieties (located between P1 and P1′) as well as the catalytic Cys-100 that form the covalent linkage were not calculated. The subsite interaction energies were obtained as the difference between the energy of the fragment noncovalently bonded to the enzyme and the sum of energies of the enzyme and the inhibitor fragment calculated separately. The PM6-DH2 quantum chemical calculations in implicit water were applied.

SmCB1 Activity and Inhibition Assays

Activity measurements were performed in a microplate format (100-μl assay volume) at 37 °C. The reaction mixture contained enzyme (0.6 nM) and fluorogenic substrate Cbz-Phe-Arg-AMC (20 μM) in 0.1 M sodium acetate, pH 5.5, containing 2.5 mM dithiothreitol and 0.1% PEG 1500 (16). The kinetics of product release was continuously monitored in an Infinite M200 microplate reader (Tecan) at excitation and emission wavelengths of 360 and 465 nm, respectively. For inhibition measurements, the enzyme was preincubated with inhibitor (0–100 μM) for 5 min followed by the addition of substrate. The IC_{50} values were determined by nonlinear regression using GraFit software. The SmCB1 activity assay with FRET substrates was performed analogously, and the kinetics of product release was continuously monitored at excitation and emission wavelengths of 320 and 420 nm, respectively. The screening of libraries of FRET carboxydipeptidase substrates was performed with 40 μM substrates in the reaction mixture. The Michaelis-Menten kinetic parameters (supplemental Table S5) were determined by measuring the rate of hydrolysis in the substrate concentration range of 0–200 μM, and K_m and k_{cat} values were obtained by nonlinear regression using GraFit software. In all assay systems, the final concentration of DMSO did not exceed 1.5%. Each measurement was performed in triplicate. The concentration of SmCB1 was determined by active site titration with E-64; the peptide solutions were quantified by amino acid analysis.

Hemoglobin Degradation

Digestion of human hemoglobin (Sigma, H7379) with SmCB1 was performed as described previously (48). Briefly, hemoglobin (10 μg) was incubated with SmCB1 (0.25 μg) in 25 mM sodium acetate, pH 3–6, including 2.5 mM DTT in a total volume of 35 μl for 1–4 h at 37 °C. Aliquots of the digest were subjected to derivatization with fluorescamine to quantify the newly formed N-terminal ends (49). The fluorescence signal was measured using an Infinite M200 microplate reader (Tecan) at 370 nm excitation and 485 nm emission wavelengths. All measurements were performed in triplicate. For SDS-PAGE visualization, hemoglobin digests were separated in Tricine gels (16% T, 6% C) containing 6 M urea (50). For RP-HPLC analysis, hemoglobin (0.15 mg) was incubated with SmCB1 (1.25 μg) in 50 mM sodium acetate, pH 4.5, containing 2.5 mM DTT in a total volume of 200 μl for 0–15 h at 37 °C. The reaction mixture was treated with 10 μl of 10% TFA and separated by RP-HPLC on a C4 Vydac column (Vydac) equilibrated in 0.1% (v/v) TFA and eluted with a 1%/min gradient of a 99% (v/v) acetonitrile solution in 0.1% (v/v) TFA. The collected peak fractions were analyzed by FT-MS using an LTQ Orbitrap XL mass spectrometer (Thermo).

Parasite Assay and Phenotype Scoring

Newly transformed schistosomula (NTS) of *S. mansoni* were prepared from infective larvae (cercariae) as described previously (5) and incubated in the presence of protease inhibitors. Briefly, the assay was performed in a microplate 96-well format (200 μl assay volume) with 200–300 NTS in Basch Medium 169 (51) containing 5% FBS, 100 units/ml penicillin, and 100 μg/ml streptomycin (52). Final concentrations of 1 or 10 μM inhibitors in 0.5% DMSO were added and incubations continued for 3 days at 5% CO_2 and 37 °C. Phenotypes that arise as a function of time and concentration were graded as follows: grade I, dead NTS by 2 days of culture at 10 μM and dying/dead NTS by 3 days at 1 μM; grade II, dead NTS by 3 days at 10 μM and round/dark/dying by 3 days at 1 μM; grade III, round/dark/dying by 3 days at 1 and 10 μM concentrations.

RESULTS

Determination of Crystal Structures—Recombinant SmCB1 was produced as a nonglycosylated mutant in the methylotrophic *P. pastoris* expression system. The enzymatically active SmCB1 was obtained by activation processing of the

SmCB1 zymogen with legumain that removes the activation peptide (pro-peptide) (12). The activated SmCB1 contains 253 amino acid residues starting with N-terminal Val-70 (the SmCB1 zymogen numbering is used throughout the paper).

SmCB1 was crystallized in complex with three covalent active site inhibitors, namely CA074, K11017, and K11777. The structure of SmCB1 was determined by molecular replacement using the structure of human cathepsin B (37) sharing 59% sequence identity. The SmCB1-CA074 and SmCB1-K11017 complexes crystallized in the same orthorhombic space group $P2_12_12_1$ with one molecule in asymmetric unit and solvent content of ~23%. The structures were refined using data to resolution 1.3 and 1.8 Å and the final crystallographic model contains residues 71–323 and 70–323 for SmCB1-CA074 and SmCB1-K11017 complex, respectively. The electron density used for modeling of inhibitors was of excellent quality in both structures (Fig. 3). The SmCB1-K11777 complex crystallized in the orthorhombic space group $P2_12_12_1$ with three molecules (named A, B, and C) in the asymmetric unit and solvent content of ~47% and was refined using data to resolution 2.64 Å. One C-terminal residue in molecule C as well as residues 118–122 of molecule B could not be located in the electron density map and were thus not included in the final model. All three molecules of the SmCB1-K11777 complex present in the asymmetric unit were very similar. The root mean square deviations (r.m.s.d.) for superposition of the three protein molecule backbones onto each other range from 0.34 to 0.48 Å, a value within the range observed for different crystal structures of identical proteins (53). Minor structural changes were localized in the surface exposed loops, and the substrate-binding sites are structurally almost identical. The electron density used to model K11777 was of excellent quality in all three protein chains in the asymmetric unit (Fig. 3). Mutual comparison of SmCB1 in complex with K11777, K11017, and CA074 did not reveal any significant differences in protein structure (backbone r.m.s.d. values are 0.18–0.54 Å).

Overall Structure of SmCB1—SmCB1 is a single polypeptide chain that adopts a classic papain-like fold in which the molecule is divided into L and R domains (37). The active site cleft with catalytic residues Cys-100, His-270, and Asn-290 is located between both domains (Fig. 1). The SmCB1 structure clearly resembles cathepsin B-type peptidases of papain superfamily with the characteristic "occluding loop" (Phe-175 to Pro-197) that restricts access to the primed region of the active site (37). A comparison of SmCB1 with the structure of human cathepsin B shows a high degree of similarity (r.m.s.d. 0.87 Å for 247 Cα atoms). The major differences in backbone superposition (with r.m.s.d. >1 Å) are located at the surface loop segments, including residues 117–125, 164–167, 175–194 (occluding loop), 247–268, and 281 (supplemental Fig. S2*A*). On the SmCB1 surface, there are several large basic patches; the major positively charged cluster is located along the edge of the occluding loop and is absent in mammalian structures (supplemental Fig. S2*B*). This is reflected in the pI of SmCB1 that is more than 3 units higher than that of human (and other mammalian) cathepsin B based on theoretical pI values of 8.7 and 5.2, respectively.

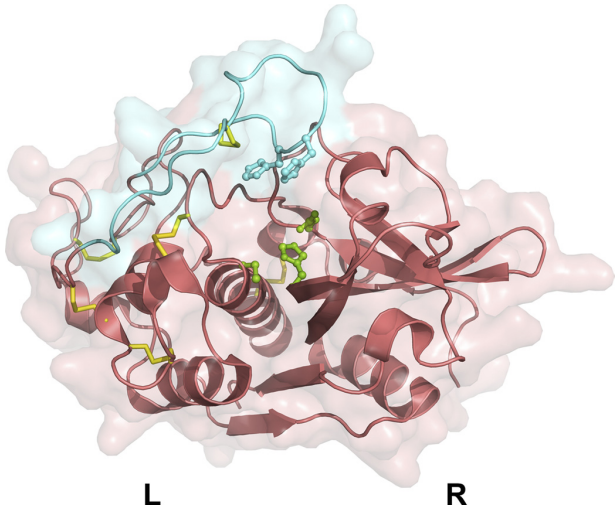


FIGURE 1. **Crystal structure of SmCB1.** The three-dimensional structure of SmCB1 is shown in a cartoon representation with a transparent surface representation overlaid. The active site is located at the interface between the L and R domains (marked *L* and *R*). The catalytic triad residues (Cys-100, His-270, and Asn-290) are depicted in the *ball-and-stick* representation in *green*. The occluding loop (residues 175–194) is colored *cyan* and contains two histidine residues involved in the substrate-enzyme interaction (His-180 and His-181; highlighted as *ball-and-sticks*). Six disulfide bridges are shown as *yellow sticks*.

In the SmCB1 occluding loop, there are two important features of structural rearrangements compared with mammalian homologs. First, a conserved segment Gly-Glu-Gly-Asp is replaced by the sequence Lys-Ile-Tyr-Lys (residues 192–195) in SmCB1. The glycine-containing segment is flexible in mammalian structures and able to move into the active site (18, 54, 55). In SmCB1, this segment is located more distally from the active site, where it is stabilized by the stacking interaction formed between Tyr-194 and Phe-175 at the side of the occluding loop (Fig. 2*A*). Second, the flexibility of the occluding loop of SmCB1 is restrained by the presence of two salt bridges that stabilize the loop in the "closed" conformation (37). The ion pair His-180/Asp-93 is conserved, whereas the pair Arg-186/Asp-295 of mammalian structures (SmCB1 numbering) is rearranged in SmCB1. The arginine is substituted by Tyr-186, which interacts with Asp-295 through Lys-185 to form a cluster stabilizing the loop conformation in SmCB1 (supplemental Fig. S3).

Mode of Binding of Inhibitors to SmCB1—Based on the crystal structures of the SmCB1-inhibitor complexes, the binding mode of the inhibitors K11777, K11017, and CA074 was described. These irreversible inhibitors form a covalent adduct with the thiol group of the catalytic residue Cys-100 and differ substantially in their positions in the active site (Fig. 3). CA074 (L-*trans*-epoxysuccinyl(propylamide)-Ile-Pro-OH) occupies the S2 to S2′ subsites and is bonded via a C6 atom after opening the epoxide ring. K11777 (*N*-Mpip-Phe-Hph-VSPh) and K11017 (Mu-Leu-Hph-VSPh) occupy the S3 to S1′ subsites making covalent bond through the Cβ atom of the vinyl sulfone moiety. In all complexes, there is a set of common interactions between the inhibitor backbone and the enzyme active site that involves predicted contacts of Gln-94, Gly-98, Gly-143, Gly-144, Gly-269 and His-270, Trp-101, Trp-292 (supplemental Table S2). Specific critical structural determi-

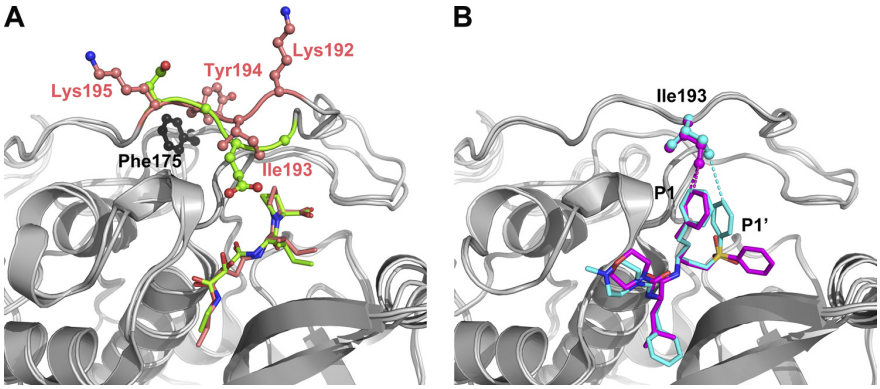


FIGURE 2. **Structural details of the occluding loop in SmCB1.** *A,* Lys-Ile-Tyr-Lys (residues 192–195) segment in SmCB1 (*salmon*) is stabilized by a stacking interaction between Tyr-194 and Phe-175; this segment replaces a flexible segment Gly-Glu-Gly-Asp in human cathepsin B (*green*). Enzyme complexes with epoxide inhibitors, SmCB1-CA074 and human cathepsin B-CA030 (Protein Data Bank code 1CSB), were used for the superposition. Loop segment residues (*ball-and-stick*) and the corresponding inhibitor (*sticks*) have the same color for carbon atoms, and other atoms are colored by the standard color coding (O, *red*; N, *blue*; S, *yellow*). *B,* superposition of the complexes SmCB1-K11017 (*magenta*) and SmCB1-K11777 (*cyan*). Both vinyl sulfone inhibitors (*sticks*) contain Hph and phenyl sulfone in P1 and P1′ positions, respectively. The Ile-193 residue (*ball-and-stick*) interacts with the P1 position of K11017 but with the P1′ position of K11777 (the contacts are highlighted). This is associated with a different orientation of the P1′ moiety that fills the S1′ subsite in K11017 but it is flipped in K11777. Heteroatoms have standard color coding.

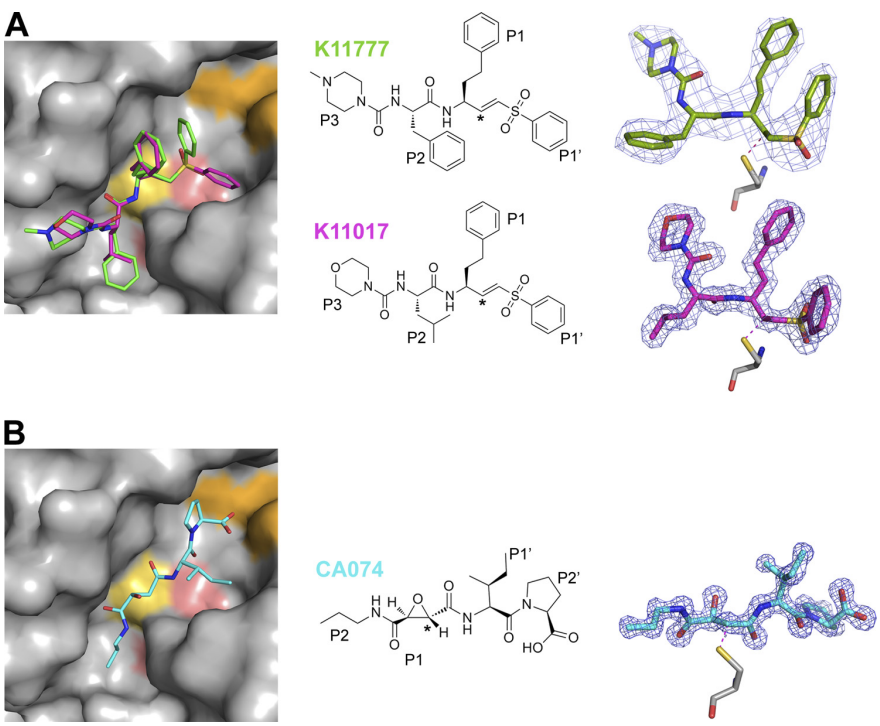


FIGURE 3. **Binding mode of inhibitors in the SmCB1 active site.** *A,* superposition of vinyl sulfone inhibitors K11777 and K11017 occupying the P3- to P1′-binding positions of SmCB1. *B,* epoxide inhibitor CA074 occupying the P2′- to P2-binding positions of SmCB1. *Left-hand panels,* surface representation of SmCB1. Highlighted are the catalytic residues Cys-100 (*yellow*), His-270 and Asn-290 (*salmon*), and the occluding loop residues His-180 and His-181 (*orange*). Inhibitors are in *stick* representation with differently colored carbon atoms (K11777, *green*; K11017, *magenta*; CA074, *cyan*). Heteroatoms have standard color coding (O, *red*; N, *blue*; S, *yellow*). *Middle panels* indicate the chemical structures of inhibitors. The C atom forming a covalent bond with the S atom of the catalytic Cys-100 is indicated with an *asterisk*. *Right-hand panels,* electron density maps for the inhibitors. The $2F_o - F_c$ electron density maps are contoured at 1σ for K11777 and K11017, and 1.5σ for CA074. The covalently bound catalytic Cys-100 is depicted.

nants of inhibitors and their interactions with SmCB1 subsites are as follows.

In the SmCB1-CA074 complex, propyl and carbamoyl groups of CA074 are in the S2 and S1 subsites, and the -Ile-Pro-OH part mimics a substrate in the S1′ and S2′ subsites (Fig. 3). The C-terminal carboxyl group of the P2′ Pro residue interacts with two His residues localized at the occluding loop of SmCB1; three charge-assisted hydrogen bonds in total were formed between carboxyl oxygens and imidazole nitrogens of His-180 and His-181 (Fig. 4). Thus, the binding mode of CA074 to SmCB1 is similar to that known in mammalian cathepsin B complexes with CA074 and related derivatives, which target the

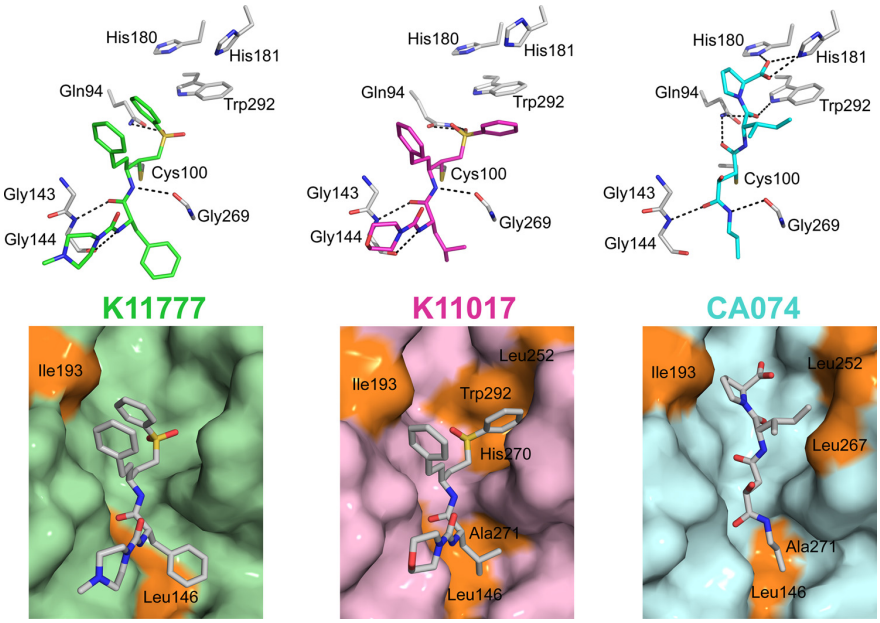


FIGURE 4. **Interactions between the SmCB1 active-site residues and inhibitors.** *Upper panels,* hydrogen bonds (*dashed lines*) formed between SmCB1 residues (*gray*) and inhibitors K11777 (*green*), K11017 (*magenta*) and CA074 (*cyan*). Heteroatoms have a standard color coding (O, *red*; N, *blue*; S, *yellow*). *Lower panels,* surface representation of the active site of SmCB1. Highlighted are the SmCB1 residues (*orange*) that form nonpolar interactions with inhibitors in *stick* representation. Carbon atoms are colored *gray*; heteroatoms have a standard color coding.

occluding loop at the S2′ subsite leading to cathepsin B specific inhibition (18, 55, 56). An additional stabilization is conferred by nonpolar interaction between Ile-193 and the P2′ proline ring. The S1′ subsite is a hydrophobic pocket (Val-247, Leu-252, Leu-267, His-270, and Trp-292) in the R domain and stably holds the P1′ Ile of CA074 through hydrogen bonding (Trp-292 and Gln-94). CA074 does not protrude deeply into the S1 subsite composed of Gln-94, Gly-98, Gly-144, and Gly-269. The inhibitor binding in this region is stabilized by the following interactions: C6 atom covalently bound to Cys-100 and two carbonyl oxygen atoms interacting with Gln-94 (in the "oxyanion hole") and Gly-144 (Fig. 4). The terminal propyl of CA074 occupies part of the S2 subsite, where it is directed toward Glu-316. This P2 group is bound through hydrophobic interactions with the backbone of Leu-146 and Ala-271.

The chemical structures of K11777 and K11017 are identical at the P1 position (Hph) and the P1′ position (VSPh) but differ at the P2 position (Phe and Leu, respectively) and P3 position (*N*-Mpip and Mu, respectively) (Fig. 3). Contrary to CA074, the vinyl sulfone inhibitors do not occupy the S2′ subsite of SmCB1. When comparing the binding mode of K11777 and K11017, a striking difference was observed for the conserved P1′ substituent; the phenyl sulfone moiety fills the S1′ subsite in K11017 but is flipped ~90° out of the active site in K11777 (for all three molecules in the asymmetric unit) (Fig. 3 and supplemental Fig. S4). Both conformations of phenyl sulfone are stabilized by polar contacts, mainly with Leu-252 and Trp-292 in K11017, and with Cys-97 and Gly-98 in K11777 (supplemental Table S2). The S1 subsite of SmCB1 binds the Hph residue of both K11777 and K11017; however, a net of polar contacts in this subsite is influenced by the situation in S1′. The residue Ile-193 located on the Lys-Ile-Tyr-Lys segment of the occluding loop has different orientation in the K11777 and K11017 complexes as it interacts with different positions in the inhibitors (Fig. 2*B*). In the SmCB1-K11017 complex, Ile-193 makes contacts with Hph in the S1 subsite and, in the SmCB1-K11777 complex, with phenyl sulfone that is flipped out from the S1′ subsite. SmCB1 contains an acidic residue (Glu-316) at the bottom of the S2 subsite similarly to mammalian cathepsins B (57). Leu in the P2 position of K11017 is able to establish a polar contact with Glu-316, whereas for the bulkier Phe in K11777, Glu-316 points out of the pocket to avoid a steric clash (supplemental Fig. S5). Similar conformational changes of the acidic residue in S2 were reported for vinyl sulfone complexes of cathepsin L-type peptidases of *Trypanosoma* and *Plasmodium* (58). The flexibility of Glu-316 in SmCB1 is further demonstrated by its dual conformation in the SmCB1-K11017 complex (supplemental Fig. S5). The S3 subsite located at the entrance of the SmCB1 active site cleft is generally hydrophobic. This wide subsite accommodated loosely the terminal groups Mu and *N*-Mpip as the P3 substituents of vinyl sulfone inhibitors.

Computational Analysis of the Inhibitor Binding Mode—The quantum chemical calculations on the crystallographic complexes were employed to determine the noncovalent interaction energy of the K11017, K11777, and CA074 inhibitors in the subsites of SmCB1. Fig. 5 shows the interaction energies of the individual side- and main-chain segments in P3 to P2′ positions.

For CA074, by far the largest favorable contribution comes from the P2′ position containing C-terminal Pro residue (−32.8 kcal/mol). The other contributions vary in range from −4.5 to 5.3 kcal/mol with favorable interactions formed by the P2/P1 and P1 segments and unfavorable interactions by the side

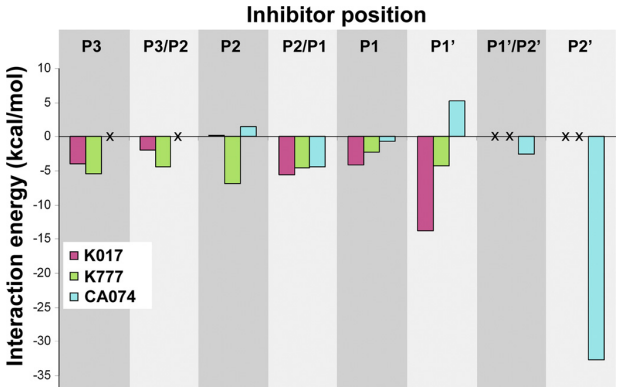


FIGURE 5. **Subsite interaction energies between inhibitors and SmCB1.** The noncovalent interaction energy was determined using quantum chemical calculations on the crystallographic complexes of SmCB1 with K11017, K11777, and CA074. The inhibitor structures were fragmented into the side-chain segments (P3 to P2′) and main-chain segments (Pi/P(i-1) connecting Pi and P(i-1)). The P1/P1′ segment forming a covalent bond with the catalytic Cys-100 was not calculated. Positions absent in inhibitor structures are marked with a ×.

chains of P2 and P1′ (Fig. 5). This is contrasted with the interaction energies of the vinyl sulfone inhibitors that represent rather smaller favorable contributions. The comparison of K11017 and K11777 revealed that there are large differences in the interaction energy between both inhibitors at the P2 and P1′ positions (Fig. 5). The side chain of the P2 residue, Leu of K11017 does not bring any favorable interaction (0.1 kcal/mol) in contrast to Phe of K11777 (−6.9 kcal/mol). The phenyl sulfone moiety at P1′ adopts a distinct conformation in each of the both vinyl sulfone-SmCB1 complexes, contributing −13.8 kcal/mol in K11017 and −4.3 kcal/mol in K11777 to binding. To evaluate the effect of the two orientations of the P1′ residue in the respective inhibitor complexes, we calculated the interaction energies for artificial complexes of K11017 and K11777, in which the side chains were built with interchanged conformations (supplemental Table S3). The calculated interaction energy of P1′ is reduced substantially to less negative values, which strongly indicates that only the crystallographic conformation is favorable for binding of the respective inhibitor to SmCB1.

In the crystallographic complexes of SmCB1 with CA074 and K11017, we found dual conformations of the side chain of Glu-316 in S2 interacting with the P2 residues. Subsite interaction energies in both alternative conformations were calculated (supplemental Table S3). The energy differences were 1.3 and 0.1 kcal/mol for K11017 and CA074, respectively. We conclude that two conformations of Glu-316 in the S2 subsite are nearly isoenergetic and might be favorable for the complex formation in terms of the conformational entropy.

Inhibitor Specificity of the SmCB1-binding Subsites—A set of 20 vinyl sulfone inhibitors was screened *in vitro* against SmCB1 to explore the structural requirements of the inhibitor-binding subsites in the SmCB1 active site cleft. These compounds are listed in Table 1 (Fig. 6) together with their IC_{50} values determined using a kinetic assay with the fluorogenic substrate Cbz-Phe-Arg-AMC. The substitutions at the positions P3 to P1′ were introduced on the inhibitor scaffold derived from the K11777 molecule.

The P1′ position of K11777 (IC_{50} = 2.09 nM) contains a phenyl group directly attached to the vinyl sulfone moiety. A modified linkage of the P1′ phenyl through a one-atom spacer (-NH-Ph, -O-Ph, -CH_2-Ph) led to inhibitors with similar or slightly weaker potency (IC_{50} in the range of 2.02–7.80 nM for WRR-282, WRR-284, and WRR-285); however, a three-atom linker (-NH-O-CH_2-Ph) provided the best inhibitor WRR-286 (IC_{50} = 0.61 nM) of the whole set. On the other hand, the inhibition dramatically decreased by 2 orders of magnitude when the P1′ phenyl was modified to give -O-Ph-O-CH_3 (WRR-347) or was replaced with -O-CH_2-CH_3 (WRR-283).

At the P1 position, Hph is the favored residue that is present in all tested inhibitors with IC_{50} of <10 nM. Its substitution with Lys (K11006) or Tyr (WRR-453) led to 1 and 4 orders of magnitude higher IC_{50}, respectively, as compared with K11002. Also, a change of configuration at P1 Hph (and adjacent P2 Phe) from *S* to *R* substantially decreased the inhibitory potency as shown for WRR-359 derived from WRR-284 (IC_{50} ~114 and 7.8 nM, respectively). Unfavorable substitutions at P1 (containing (*R*)-Ala) and P1′ resulted in low inhibition of WRR-185 and WRR-200 (compare with WRR-145).

At the P2 position, both Phe and Leu are highly effective as demonstrated with K11002 and K11017 (IC_{50} around 1.7 nM). In the K11777 scaffold, replacement of the P2 Phe by Phe-4-CH_3 (AR-198049) and Phe-4-CF_3 (AR-198048) resulted in 3- and 5-fold weaker inhibition, and His (WRR-499) and Arg (WRR-483) afforded 14- and 24-fold weaker inhibition, respectively. The N-terminal modification of inhibitors corresponding to the P3 position was by *N*-Mpip, Mu, and Cbz capping groups. The heterocycles are present in the best inhibitors and do not differ importantly in their contribution to the inhibitory effect, as shown with K11777 and K11002 (IC_{50} of 2.09 and 1.73 nM, respectively).

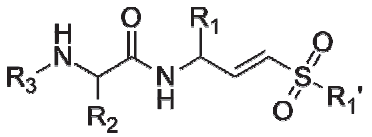
Cleavage Mode and Substrate Specificity of SmCB1—Hydrolysis by SmCB1 was analyzed with the physiological protein substrate, hemoglobin, and with a series of synthetic peptide substrates. SmCB1 degraded hemoglobin at acidic pH between 4 and 6 as measured by a fluorescamine assay that directly quantifies hemoglobin fragments (Fig. 7*A*). SDS-PAGE visualization of the hemoglobin fragmentation showed that the disappearance of the substrate band is not associated with a corresponding accumulation of large hemoglobin fragments of >3.5 kDa (Fig. 7*B*). A detailed pattern of hemoglobin-derived products was obtained by RP-HPLC separation (Fig. 7*C*). Like those after SDS-PAGE, the RP-HPLC profiles indicated that hemoglobin is gradually converted into a pool of dipeptides with little accumulation of peptides of intermediate size. The detected intermediate fragments ranging in size from 10 to 41 amino acids are mostly derived from the interior of the hemoglobin sequence (supplemental Table S4); this demonstrates the involvement of endopeptidase activity. The combined data suggest that upon endopeptidolytic cleavage of hemoglobin by SmCB1, the substrate is rapidly processed via carboxydipeptidase activity.

SmCB1 was then tested with various fluorogenic peptide substrates that allowed for the discrimination of endo- and exopeptidase activities. The supplemental Table S5 compares

TABLE 1

Inhibition of SmCB1 by vinyl sulfone inhibitors and their anti-schistosomal activity

The IC_{50} values for 20 vinyl sulfone inhibitors were determined in a kinetic activity assay with SmCB1 and the fluorogenic peptide substrate, Cbz-Phe-Arg-AMC, at pH 5.5. The epoxide inhibitor CA074 was assayed for comparison. The vinyl sulfone structures are defined by the compound core (see scheme below) and substituents R3 to R1′. Inhibitors are ranked according to their IC_{50} values. Mean values ± S.E. are given for triplicate measurements. Induction of phenotypic alterations by the inhibitors was determined with NTS of *S. mansoni*. The inhibitors were tested at 1 and 10 μm concentrations, and the resulting phenotypes, arising as a function of time and concentration, were graded I to III, with I being the most severe (see Fig. 6).



Compound name	Substituent position[a]				Inhibition	Severity of phenotype against parasite
Vinyl sulfone	**R_3**	**R_2**	**R_1**	**R_1'**	**IC_{50} (nM)**	**Grade[d]**
WRR-286	N-Mpip	Phe	Hph	NH-O-CH_2-Ph	0.61 ± 0.05	I
K11017[b]	Mu	Leu	Hph	Ph	1.71 ± 0.13	I
K11002	Mu	Phe	Hph	Ph	1.73 ± 0.21	I
WRR-282	N-Mpip	Phe	Hph	CH_2-Ph	2.02 ± 0.10	I
K11777[b]	N-Mpip	Phe	Hph	Ph	2.09 ± 0.08	II
WRR-285	N-Mpip	Phe	Hph	NH-Ph	2.13± 0.03	I
K11747	N-Mpip	Naphtyl-CH_2	Hph	Naphtyl	2.50 ± 0.12	I
AR-198048	N-Mpip	Phe-4-CH_3	Hph	Ph	6.00 ± 0.21	II
WRR-284	N-Mpip	Phe	Hph	O-Ph	7.80 ± 0.23	II
AR-198049	N-Mpip	Phe-3-CF_3	Hph	Ph	10.5 ± 0.5	II
WRR-145	Cbz	Phe	Ala	Ph	22.9 ± 0.5	III
WRR-499	N-Mpip	His	Hph	Ph	29.5 ± 0.4	III
K11006	Mu	Phe	Lys	Ph	33.2 ± 0.3	III
WRR-483	N-Mpip	Arg	Hph	Ph	48.9 ± 0.4	III
WRR-359	N-Mpip	Phe[c]	Hph[c]	O-Ph	114 ± 2	III
WRR-185	Cbz	Phe	Ala[c]	NH-CH(Bz)-COO-CH_3	126 ± 2	II
WRR-283	N-Mpip	Phe	Hph	O-CH_2-CH_3	192 ± 3	III
WRR-200	Cbz	Phe	Ala[c]	NH-CH_2-Ph-4-C-F_3	247 ± 4	II
WRR-347	N-Mpip	Phe	Hph	O-Ph-4-O-CH_3	747 ± 20	III
WRR-453	Mu	Phe	Tyr	Ph	11849 ± 101	III
Epoxide						
CA074[b]					63.9 ± 2.8	II

[a] The following abbreviations used are as follows: *N*-Mpip, *N*-methylpiperazinylcarbonyl; Mu, morpholinylcarbonyl; Cbz, benzyloxycarbonyl; Ph, phenyl; Bz, benzyl.
[b] Compounds were analyzed by x-ray crystallography in complex with SmCB1.
[c] Residues are in *R* configuration; all other residues are in *S* configuration.
[d] Grade I was the most severe; see text for details.

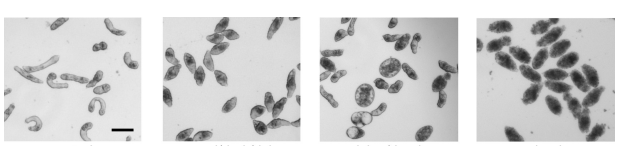


FIGURE 6. **Phenotypes induced *in vitro* in *S. mansoni* NTS by cysteine peptidase inhibitors (listed in Table 1).** Examples of three inhibitor-induced phenotypes in the parasite *versus* untreated controls. Phenotypes arise as a function of time and inhibitor concentration and were graded as follows: Grade I, dead NTS by 2 days of culture at 10 μM and dying/dead NTS by 3 days at 1 μM; Grade II, dead NTS by 3 days at 10 μM and round/dark/dying by 3 days at 1 μM; Grade III, round/dark/dying phenotype in 3 days at 1 and 10 μM concentrations. *Scale bar,* 0.2 mm.

kinetic parameters of the mode-selective substrates. The endopeptidase substrates of cathepsins B (59), Abz-Gln-Val-Val-Ala-Gly-Ala-EDDnp, and Abz-Ala-Phe-Arg-Phe-Ser-Gln-EDDnp, displayed 2 orders of magnitude lower k_{cat}/K_m values than the carboxydipeptidase substrates Abz-Phe-Arg-Val-Nph and Abz-Phe-Arg-Nph-Val (60, 61). With the minimized endopeptidase substrate, Cbz-Phe-Arg-AMC (62), k_{cat}/K_m for SmCB1 was 1 order of magnitude lower than for the carboxydipeptidase substrates. This analysis indicates that the exopeptidase catalysis is more efficient than endopeptidase catalysis.

We next designed FRET-based substrate libraries for mapping the carboxydipeptidase activity and residue preferences in

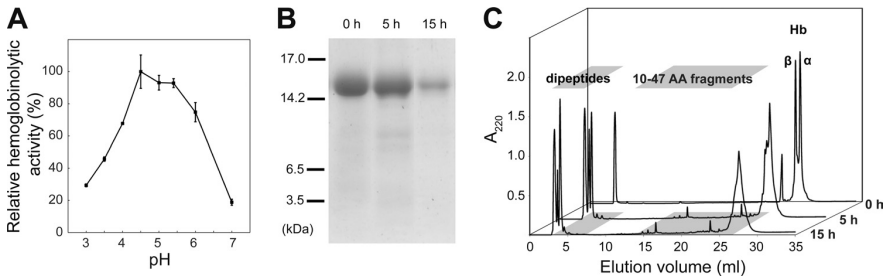


FIGURE 7. **Hydrolysis of hemoglobin by SmCB1.** *A,* human hemoglobin (Hb) was digested with SmCB1 at various pH values. The degradation rate was determined with the fluorescamine derivatization assay quantifying the liberated fragments. The mean values ± S.E. are expressed relatively to the maximum value. *B* and *C,* Hb digest at pH 4.5 was performed at two time points; the reaction mixture was electrophoretically and chromatographically separated and compared with the undigested control. *B,* Tricine-SDS-PAGE of the Hb digest visualized by protein staining. *C,* RP-HPLC of the Hb digest resolved on a C4 column using a TFA/acetonitrile system. Elution positions are indicated for the intact Hb substrate (α and β subunits) and for Hb-derived fragments, which form the pools of dipeptides and large peptides (ranging in size from 10 to 41 amino acids, see supplemental Table S4). The flow-through peak (see profile at 0 h) contains nonpeptide components of the reaction mixture.

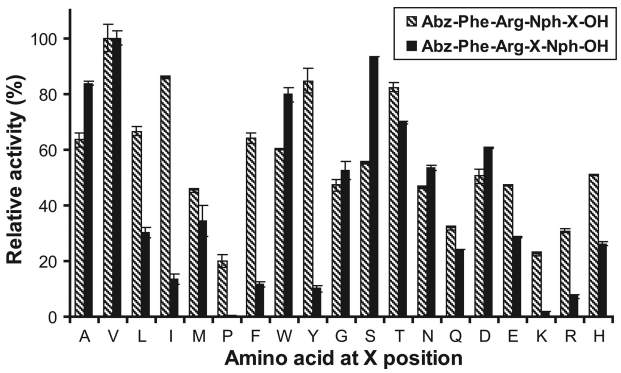


FIGURE 8. **Carboxydipeptidase specificity of SmCB1.** The libraries of FRET-based peptide substrates Abz-Phe-Arg-Xaa-Nph-OH and Abz-Phe-Arg-Nph-Xaa-OH were designed to span the P2 to P2′ positions of SmCB1 and introduce substitutions (*X*) at the P1′ and P2′ positions, respectively. The amino acids at the *X* position are grouped according to their chemical characteristics (aliphatic, aromatic, polar, acidic, and basic). The substrate hydrolysis was measured in a kinetic activity assay with SmCB1 at pH 5.5. Mean values ± S.E. for triplicate measurements are given normalized to the maximum value in each series (100%).

the primed substrate-binding subsites of SmCB1. The structure of the synthesized libraries Abz-Phe-Arg-Xaa-Nph-OH and Abz-Phe-Arg-Nph-Xaa-OH contain two fixed residues, Phe-Arg in P2-P1 that are favored by SmCB1 and other cathepsins B (60, 61, 63), and help to anchor the substrates in nonprimed subsites. The substitutions in the Xaa positions define the P1′ and P2′ residues. The screening of the libraries showed that SmCB1 has a broader specificity of the S2′ subsite than S1′ subsite (Fig. 8). All residues were accepted at the P2′ position, although Pro and basic residues led to lower substrate hydrolysis. At the P1′ position, larger hydrophobic, aromatic (except for Trp), and basic residues were unfavorable, whereas Pro was not tolerated. Generally, the primed subsites of SmCB1 differed in their ability to accommodate large hydrophobic and aromatic residues of the tested carboxydipeptidase substrates.

Severity of Phenotypes Induced in Cultured Parasites Correlates with the Potency of SmCB1 Inhibition—A panel of inhibitors of SmCB1 listed in Table 1 was screened against *S. mansoni* NTS, the post-invasive parasite stage that feeds on host blood (5). The NTS were exposed to 1 and 10 μM inhibitors, and the resultant phenotypes were graded I through III from most to least severe. About 30 and 60% of the tested compounds led to death of NTS by the 3rd day of the incubation at 1 and 10 μM, respectively. The score and images of typical disordered phenotypes are presented in Table 1 and Fig. 6. The severity of phenotypes induced by the inhibitors was statistically correlated with their potency to inhibit SmCB1 activity (Table 1; Kruskal-Wallis rank analysis of variance, $H_{(2,21)} = 12.6$, $p < 0.001$). Specifically, grade I and II phenotypes were generally caused by inhibitors with IC_{50} values less than 20 nM, and grade III phenotypes were induced by inhibitors with IC_{50} values greater than 20 nM (except for WRR-185 and WRR-200, which were grade II).

DISCUSSION

SmCB1 is one of a number of digestive peptidases in the gut of the flatworm parasite *S. mansoni* (8, 12). Both reverse genetics and chemical experiments suggest that it is a critical for parasite growth and a valuable target for the development of novel anti-schistosomal drugs (14, 15). In this study, we provide a comprehensive structure-activity analysis of SmCB1 that includes a series of crystal structure determinations. We also describe the SmCB1 interaction with inhibitors and characterize its specificity with both peptidyl and protein substrates.

Structure of SmCB1—The three-dimensional structure of SmCB1 was solved for three inhibitor complexes; the best resolution achieved was 1.3 Å. SmCB1 possesses an occluding loop that is characteristic of cathepsin B-type peptidases (37). It is known to regulate access to the active site, where it partially blocks the primed substrate-binding subsites (at S3′ and beyond), and thus confer carboxydipeptidase activity to cathepsins B (59, 61, 64). The occluding loop of SmCB1 presents local structural rearrangements compared with mammalian homologs; however, these changes retained the overall loop fold suggesting its functional competence. This was probed by determining the mode of SmCB1 action using specific peptide substrate; carboxydipeptidase activity was clearly manifested. In addition, SmCB1 displayed endopeptidase activity indicating that steric hindrance by the SmCB1 occluding loop is flexible such that the loop can move to accommodate endopeptidase substrate in the active site cleft as has been reported for human cathepsin B (59).

SmCB1, an Efficient Endo- and Exopeptidolytic Machine—The carboxydipeptidase catalytic efficiency of SmCB1 as measured with peptide substrates was greater than its endopeptidase efficiency. The screening of carboxydipeptidase substrate libraries showed that a broad range of residues is tolerated in the primed positions P1′ and especially P2′. This suggests that SmCB1 is able to trim effectively the C termini of peptides. SmCB1 has more promiscuous substrate specificity in P2′ than human cathepsin B (60). Furthermore, we identified a combined carboxydipeptidase/endopeptidase action of SmCB1 on the physiological substrate hemoglobin. Analysis of the reaction products indicates that endopeptidolytic fragments are rapidly converted into dipeptides. Thus, oligo/polypeptide fragments do not accumulate to the extent observed for hemoglobin digestion by helminth cathepsin L-type endopeptidases (65–67). With regard to hemoglobinolytic capability, SmCB1 resembles cathepsins B from the Southeast Asian liver fluke *Opisthorchis viverrini* and the hookworm *Ancylostoma caninum* (66, 68) but differs from cathepsins B of the avian fluke *Trichobilharzia regenti* and the hookworm *Necator americanus* that cannot initiate hemoglobinolysis (69, 70). We conclude that SmCB1 operates as an effective proteolytic machine to degrade the major protein in the parasite's blood meal.

Structure-based Insights for Drug Design—The interaction of SmCB1 with peptidomimetic inhibitors was investigated using the following: (i) crystal structures of three SmCB1-inhibitor complexes, (ii) computational analysis of interaction energies, and (iii) inhibition profiling with a panel of vinyl sulfones. These enabled us to evaluate the critical interactions of inhibitors in the binding subsites and provide a basic SAR for improving inhibitory potency and selectivity.

The S2′ subsite was efficiently occupied in the complex of CA074, a specific inhibitor of cathepsin B-type peptidases. The hydrogen bonding of the inhibitor's C-terminal P2′ residue with the occluding loop (especially the two His residues) was the largest favorable subsite interaction among the three crystallographic complexes (Fig. 5). The vinyl sulfone inhibitors, by contrast, do not contain a P2′ residue, and therefore, we would consider re-designing the scaffold to extend into the S2′ position, which may improve both potency and selectivity to SmCB1. For the vinyl sulfone P1′, aromatic sulfone moieties were preferred over the Ile residue in CA074 (in accordance with the low substrate specificity for Ile in P1′ (Fig. 8)). An important discovery is the strikingly different conformation between the P1′ phenyl sulfone moieties in the K11777 and K11017 complexes, which suggests a cooperativity between S1′ and other subsite(s) of SmCB1. This conformational switch should be taken into account in future docking experiments to optimize P1′ substituents, *e.g.* by aromatic groups with a longer linker (Table 1). Interestingly, two types of orientation of the P1′ phenyl sulfone were also observed in the crystal structures of several vinyl sulfone inhibitors with cathepsin L-type peptidases from protozoan parasites, as shown in supplemental Fig. S4 (58, 71, 72). In these complexes, the particular phenyl sulfone orientation was regulated by the structural environment of S1′; however, a dual (transient) conformation of this substituent was also documented (71).

In the S1 subsite of SmCB1, Hph of the vinyl sulfone inhibitors was energetically favored (Fig. 5). A basic residue at this position reduced inhibitory potency (Table 1), although basic P1 residues are preferred in SmCB1 substrates (63). This may reflect the effect of the overall scaffold of the active-site ligand that has been reported to change the P1 specificity of human cathepsin B (60, 61). Based on the structural difference, the S1 pocket can be exploited to engineer selective inhibition of SmCB1 over human cathepsin B. For this purpose, the interaction can be optimized between the P1 residue and Ile-193 located on an nonconserved sequence segment of the occluding loop of SmCB1 (Fig. 2). At P2, bulky hydrophobic residues such as Phe and Leu on the vinyl sulfone scaffold afforded highly potent inhibitors, which agrees with the known P2 substrate preferences of SmCB1 (63). The bottom of the S2 pocket of SmCB1 and other cathepsins B contains Glu, which facilitates the recognition of positively charged residues at P2 (57, 60, 63). Further focus can be placed on this P2-S2 interaction by introducing basic substituents of a suitable size to make contact without displacing the flexible side chain of Glu-316 (supplemental Fig. S5). Finally, for S3, occupation by monocyclic heterocycles of the vinyl sulfones generated favorable interaction energies. More bulky substituents can be tested at P3 to improve inhibition, as reported for human cathepsin B inhibitors (73).

SmCB1 as Priority Drug Target—Although SmCB1 is one of a number of peptidases expressed in the gut (7–9) and elsewhere in the parasite (74), the correlation between the severity of phenotypes induced by vinyl sulfone inhibitors and the potency of inhibition of SmCB1 encourage the view that SmCB1 is a valuable drug target. This is congruent with the identification of SmCB1 as a major target for inhibition by K11777 during experimental therapy in a murine model of *S. mansoni* infection (15). Given the catalytic efficiency of SmCB1 against hemoglobin described here and being the major cysteine peptidase activity in the parasite (13, 75), it might be anticipated that inhibition of this enzyme would impact the parasite's ability to thrive. Indeed, RNA interference of SmCB1 slowed the growth of the parasite both in culture and in an animal model of infection (14). To conclude, the SmCB1 crystal structures described herein provide the necessary first step in a structure-based drug development program to improve inhibitor specificity and potency, and possibly, generate new lead anti-schistosomal compounds.

Acknowledgments—We thank James T. Palmer (Biota Inc., Victoria, Australia), William R. Roush (The Scripps Research Institute, FL), and Adam R. Renslo (Department of Pharmaceutical Chemistry and the Small Molecule Discovery Center, University of California San Francisco) for providing inhibitors. We thank Mohammed Sajid (Leiden University Medical Centre, Leiden, The Netherlands) and Melaine Delcroix (School of Public Health, University of California, Berkeley) for providing the SmCB1 expression construct, and Brian Suzuki (University of California San Francisco) for tests with S. mansoni NTS. We also thank Miloslav Šanda for mass spectroscopy, Miroslava Blechová for peptide synthesis, Robert Hanus for statistical analysis, and Irena Pražáková for technical assistance (Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, Czech Republic).

REFERENCES

1. Ützinger, J., N'goran, E. K., Caffrey, C. R., and Keiser, J. (2011) *Acta Trop.,* in press
2. Gryseels, B., Polman, K., Clerinx, J., and Kestens, L. (2006) *Lancet* **368,** 1106–1118
3. Caffrey, C. R. (2007) *Curr. Opin. Chem. Biol.* **11,** 433–439
4. Caffrey, C. R., Rohwer, A., Oellien, F., Marhöfer, R. J., Braschi, S., Oliveira, G., McKerrow, J. H., and Selzer, P. M. (2009) *PLoS One* **4,** e4413
5. Abdulla, M. H., Ruelas, D. S., Wolff, B., Snedecor, J., Lim, K. C., Xu, F., Renslo, A. R., Williams, J., McKerrow, J. H., and Caffrey, C. R. (2009) *PLoS Negl. Trop. Dis.* **3,** e478
6. Rai, G., Sayed, A. A., Lea, W. A., Luecke, H. F., Chakrapani, H., Prast-Nielsen, S., Jadhav, A., Leister, W., Shen, M., Inglese, J., Austin, C. P., Keefer, L., Arnér, E. S., Simeonov, A., Maloney, D. J., Williams, D. L., and Thomas, C. J. (2009) *J. Med. Chem.* **52,** 6474–6483
7. Caffrey, C. R., McKerrow, J. H., Salter, J. P., and Sajid, M. (2004) *Trends Parasitol.* **20,** 241–248
8. Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvorák, J., Hsieh, I., Bahgat, M., Dissous, C., and McKerrow, J. H. (2006) *J. Biol. Chem.* **281,** 39316–39329
9. Brindley, P. J., Kalinna, B. H., Dalton, J. P., Day, S. R., Wong, J. Y., Smythe, M. L., and McManus, D. P. (1997) *Mol. Biochem. Parasitol.* **89,** 1–9
10. McCarthy, E., Stack, C., Donnelly, S. M., Doyle, S., Mann, V. H., Brindley, P. J., Stewart, M., Day, T. A., Maule, A. G., and Dalton, J. P. (2004) *Int. J. Parasitol.* **34,** 703–714
11. Dvorák, J., Mashiyama, S. T., Sajid, M., Braschi, S., Delcroix, M., Schneider, E. L., McKerrow, W. H., Bahgat, M., Hansell, E., Babbitt, P. C., Craik, C. S., McKerrow, J. H., and Caffrey, C. R. (2009) *PLoS Negl. Trop. Dis.* **3,** e449
12. Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M. A., Lucas, K. D., Hsieh, I., Greenbaum, D., Bogyo, M., Salter, J. P., Lim, K. C., Franklin, C., Kim, J. H., and Caffrey, C. R. (2003) *Mol. Biochem. Parasitol.* **131,** 65–75
13. Caffrey, C. R., and Ruppel, A. (1997) *Parasitol. Res.* **83,** 632–635
14. Correnti, J. M., Brindley, P. J., and Pearce, E. J. (2005) *Mol. Biochem. Parasitol.* **143,** 209–215
15. Abdulla, M. H., Lim, K. C., Sajid, M., McKerrow, J. H., and Caffrey, C. R. (2007) *PLoS Med.* **4,** e14
16. Horn, M., Jílková, A., Vondrásek, J., Maresová, L., Caffrey, C. R., and Mares, M. (2011) *ACS Chem. Biol.* **6,** 609–617
17. Murata, M., Miyashita, S., Yokoo, C., Tamai, M., Hanada, K., Hatayama, K., Towatari, T., Nikawa, T., and Katunuma, N. (1991) *FEBS Lett.* **280,** 307–310
18. Yamamoto, A., Tomoo, K., Hara, T., Murata, M., Kitamura, K., and Ishida, T. (2000) *J. Biochem.* **127,** 635–643
19. Palmer, J. T., Rasnick, D., Klaus, J. L., and Brömme, D. (1995) *J. Med. Chem.* **38,** 3193–3196
20. Brömme, D., Klaus, J. L., Okamoto, K., Rasnick, D., and Palmer, J. T. (1996) *Biochem. J.* **315,** 85–89
21. Steverding, D., Caffrey, C. R., and Sajid, M. (2006) *Mini. Rev. Med. Chem.* **6,** 1025–1032
22. Engel, J. C., Doyle, P. S., Hsieh, I., and McKerrow, J. H. (1998) *J. Exp. Med.* **188,** 725–734
23. Olson, J. E., Lee, G. K., Semenov, A., and Rosenthal, P. J. (1999) *Bioorg. Med. Chem.* **7,** 633–638
24. McKerrow, J. H., Rosenthal, P. J., Swenerton, R., and Doyle, P. (2008) *Curr. Opin. Infect. Dis.* **21,** 668–672
25. Renslo, A. R., and McKerrow, J. H. (2006) *Nat. Chem. Biol.* **2,** 701–710
26. McKerrow, J. H., Doyle, P. S., Engel, J. C., Podust, L. M., Robertson, S. A., Ferreira, R., Saxton, T., Arkin, M., Kerr, I. D., Brinen, L. S., and Craik, C. S. (2009) *Mem. Inst. Oswaldo Cruz* **104,** 263–269
27. Caffrey, C. R., Mathieu, M. A., Gaffney, A. M., Salter, J. P., Sajid, M., Lucas, K. D., Franklin, C., Bogyo, M., and McKerrow, J. H. (2000) *FEBS Lett.* **466,** 244–248
28. Mása, M., Maresová, L., Vondrásek, J., Horn, M., Jezek, J., and Mares, M. (2006) *Biochemistry* **45,** 15474–15482
29. Somoza, J. R., Zhan, H., Bowman, K. K., Yu, L., Mortara, K. D., Palmer, J. T., Clark, J. M., and McGrath, M. E. (2000) *Biochemistry* **39,** 12543–12551
30. Roush, W. R., Cheng, J., Knapp-Reed, B., Alvarez-Hernandez, A., McKerrow, J. H., Hansell, E., and Engel, J. C. (2001) *Bioorg. Med. Chem. Lett.* **11,** 2759–2762
31. Chen, Y. T., Brinen, L. S., Kerr, I. D., Hansell, E., Doyle, P. S., McKerrow, J. H., and Roush, W. R. (2010) *PLoS Negl. Trop. Dis.* **4,** e825
32. Scheidt, K. A., Roush, W. R., McKerrow, J. H., Selzer, P. M., Hansell, E., and Rosenthal, P. J. (1998) *Bioorg. Med. Chem.* **6,** 2477–2494
33. Roush, W. R., Gwaltney, S. L., Cheng, J., Scheidt, K. A., McKerrow, J. H., and Hansell, E. (1998) *J. Am. Chem. Soc.* **120,** 10994–10995
34. Jaishankar, P., Hansell, E., Zhao, D. M., Doyle, P. S., McKerrow, J. H., and Renslo, A. R. (2008) *Bioorg. Med. Chem. Lett.* **18,** 624–628
35. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62,** 859–866
36. Vagin, A., and Teplyakov, A. (2000) *Acta Crystallogr. D Biol. Crystallogr.* **56,** 1622–1624
37. Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., *et al.* (1991) *EMBO J.* **10,** 2321–2330
38. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
39. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
40. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
41. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallogr. D Biol. Crystallogr.* **57,** 122–133
42. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50,** 437–450
43. Krissinel, E., and Henrick, K. (2005) in *Detection of Protein Assemblies in Crystals* (Berthold, M. R., ed) pp. 163–174, Springer-Verlag, Berlin
44. De Lano, W. L. (2002) *The PyMOL Molecular Graphics System,* DeLano Scientific LLC, San Carlos, CA
45. Case, D. A., Darden, T. A., Cheatham, T. E., 3rd, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., and Kollman, P. A. (2008) *AMBER 10*, University of California, San Francisco
46. Rezac, J., Fanfrlik, J., Salahub, D., and Hobza, P. (2009) *J. Chem. Theory Comput.* **5,** 1749–1760
47. Korth, M., Pitonak, M., Rezac, J., and Hobza, P. (2010) *J. Chem. Theory Comput.* **6,** 344–352
48. Horn, M., Nussbaumerová, M., Sanda, M., Kovárová, Z., Srba, J., Franta, Z., Sojka, D., Bogyo, M., Caffrey, C. R., Kopácek, P., and Mares, M. (2009) *Chem. Biol.* **16,** 1053–1063
49. Sorgine, M. H., Logullo, C., Zingali, R. B., Paiva-Silva, G. O., Juliano, L., and Oliveira, P. L. (2000) *J. Biol. Chem.* **275,** 28659–28665
50. Schagger, H. (2006) *Nat. Protoc.* **1,** 16–22
51. Basch, P. F. (1981) *J. Parasitol.* **67,** 179–185
52. Stefanìc, S., Dvorák, J., Horn, M., Braschi, S., Sojka, D., Ruelas, D. S., Suzuki, B., Lim, K. C., Hopkins, S. D., McKerrow, J. H., and Caffrey, C. R. (2010) *PLoS Negl. Trop. Dis.* **4,** e850
53. Betts, M. J., and Sternberg, M. J. (1999) *Protein Eng.* **12,** 271–283
54. Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J. S., To, R., and Huber, C. P. (1995) *J. Biol. Chem.* **270,** 5527–5533
55. Watanabe, D., Yamamoto, A., Tomoo, K., Matsumoto, K., Murata, M., Kitamura, K., and Ishida, T. (2006) *J. Mol. Biol.* **362,** 979–993
56. Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R., and Turk, V. (1995) *Biochemistry* **34,** 4791–4797
57. Hasnain, S., Hirama, T., Huber, C. P., Mason, P., and Mort, J. S. (1993) *J. Biol. Chem.* **268,** 235–240
58. Kerr, I. D., Lee, J. H., Farady, C. J., Marion, R., Rickert, M., Sajid, M., Pandey, K. C., Caffrey, C. R., Legac, J., Hansell, E., McKerrow, J. H., Craik, C. S., Rosenthal, P. J., and Brinen, L. S. (2009) *J. Biol. Chem.* **284,** 25697–25703
59. Nägler, D. K., Storer, A. C., Portaro, F. C., Carmona, E., Juliano, L., and Ménard, R. (1997) *Biochemistry* **36,** 12608–12615

60. Cezari, M. H., Puzer, L., Juliano, M. A., Carmona, A. K., and Juliano, L. (2002) *Biochem. J.* **368,** 365–369
61. Krupa, J. C., Hasnain, S., Nägler, D. K., Ménard, R., and Mort, J. S. (2002) *Biochem. J.* **361,** 613–619
62. Quraishi, O., Nägler, D. K., Fox, T., Sivaraman, J., Cygler, M., Mort, J. S., and Storer, A. C. (1999) *Biochemistry* **38,** 5017–5023
63. Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Brömme, D., Ellman, J. A., and Craik, C. S. (2006) *J. Biol. Chem.* **281,** 12824–12832
64. Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J. S. (1997) *J. Biol. Chem.* **272,** 1197–1202
65. Lowther, J., Robinson, M. W., Donnelly, S. M., Xu, W., Stack, C. M., Matthews, J. M., and Dalton, J. P. (2009) *PLoS Negl. Trop. Dis.* **3,** e369
66. Sripa, J., Laha, T., To, J., Brindley, P. J., Sripa, B., Kaewkes, S., Dalton, J. P., and Robinson, M. W. (2010) *Cell. Microbiol.* **12,** 781–795
67. Brady, C. P., Dowd, A. J., Brindley, P. J., Ryan, T., Day, S. R., and Dalton, J. P. (1999) *Infect. Immun.* **67,** 368–374
68. Williamson, A. L., Lecchi, P., Turk, B. E., Choe, Y., Hotez, P. J., McKerrow, J. H., Cantley, L. C., Sajid, M., Craik, C. S., and Loukas, A. (2004) *J. Biol. Chem.* **279,** 35950–35957
69. Dvorák, J., Delcroix, M., Rossi, A., Vopálenský, V., Pospísek, M., Sedinová, M., Mikes, L., Sajid, M., Sali, A., McKerrow, J. H., Horák, P., and Caffrey, C. R. (2005) *Int. J. Parasitol.* **35,** 895–910).
70. Ranjit, N., Zhan, B., Stenzel, D. J., Mulvenna, J., Fujiwara, R., Hotez, P. J., and Loukas, A. (2008) *Mol. Biochem. Parasitol.* **160,** 90–99
71. Kerr, I. D., Wu, P., Marion-Tsukamaki, R., Mackey, Z. B., and Brinen, L. S. (2010) *PLoS Negl. Trop. Dis.* **4,** e701
72. Brinen, L. S., Hansell, E., Cheng, J., Roush, W. R., McKerrow, J. H., and Fletterick, R. J. (2000) *Structure* **8,** 831–840
73. Greenspan, P. D., Clark, K. L., Tommasi, R. A., Cowen, S. D., McQuire, L. W., Farley, D. L., van Duzer, J. H., Goldberg, R. L., Zhou, H., Du, Z., Fitt, J. J., Coppa, D. E., Fang, Z., Macchia, W., Zhu, L., Capparelli, M. P., Goldstein, R., Wigg, A. M., Doughty, J. R., Bohacek, R. S., and Knap, A. K. (2001) *J. Med. Chem.* **44,** 4524–4534
74. Caffrey, C. R., Salter, J. P., Lucas, K. D., Khiem, D., Hsieh, I., Lim, K. C., Ruppel, A., McKerrow, J. H., and Sajid, M. (2002) *Mol. Biochem. Parasitol.* **121,** 49–61
75. Caffrey, C. R., Rheinberg, C. E., Moné, H., Jourdane, J., Li, Y. L., and Ruppel, A. (1997) *Parasitol. Res.* **83,** 37–41

SUPPLEMENTAL DATA

for

STRUCTURAL BASIS FOR INHIBITION OF THE CATHEPSIN B DRUG TARGET FROM THE HUMAN BLOOD FLUKE, *SCHISTOSOMA MANSONI*

Adéla Jílková, Pavlína Řezáčová, Martin Lepšík, Martin Horn, Jana Váchová, Jindřich Fanfrlík, Jiří Brynda, James H. McKerrow, Conor R. Caffrey, Michael Mareš

SUPPLEMENTAL TABLE S1

X-ray data collection and refinement statistics. Numbers in parentheses refer to the highest-resolution shell.

Inhibitor-SmCB1 complex	SmCB1-K11777	SmCB1-K11017	SmCB1-CA074
Data collection statistics			
Wavelenght (Å)	0.9791	0.9791	0.9791
Temperature (K)	100	100	100
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a, b, c (Å)	77.63, 103.12, 100.70	33.09, 79.53, 90.27	33.17, 79.16, 90.61
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50-2.65 (2.50-2.65)	50-1.8 (1.83-1.80)	50-1.3 (1.32-1.30)
Number of unique reflections	24,218 (1,178)	22,894 (1,111)	55,759 (1,811)
Redundancy	6.0 (5.6)	6.8 (3.7)	5.5 (3.1)
Completeness (%)	99.8 (99.5)	99.8 (98.0)	93.5 (61.6)
R_{merge} [a] (%)	6.2 (66.9)	10.5 (47.9)	5.3 (28.1)
Average I/σ (I)	21.7 (1.85)	18.4 (2.30)	26.29 (2.70)
Wilson B ($Å^2$)	69.1	13.2	8.5
Refinement statistics			
Resolution range (Å)	34.20-2.64 (2.70-2.64)	29.80-1.80 (1.85-1.80)	22.33-1.30 (1.33-1.30)
No. of reflections in working set	22,916	21,548	52,880
No. of reflections in test set	1,236	1,162	2,805
R value (%)	20.46 (29.1)	16.61 (23.26)	11.4 (13.0)
R_{free} value [c] (%)	27.1 (36.2)	23.3 (35.0)	15.4 (22.7)
Number of molecules in AU	3	1	1
Number of atoms in AU[d] protein / inhibitor /solvent	5,956 / 123 / 66	2,057 / 37 / 536	2,137 / 27 / 472
Average ADP[e] for protein / inhibitor /solvent ($Å^2$)	41.1 / 53.3 / 51.56	10.72 / 11.32 / 27.1	8.39 / 10.15 / 23.93
RMSD bond length (Å)	0.011	0.012	0.018
RMSD bond angle (°)	1.287	1.268	1.788
Ramachandran plot statistics[f]			
Most favored regions (%)	84.2	87.8	86.3
Allowed regions (%)	15.5	12.2	12.7
PDB code	3S3R	3S3Q	3QSD

[a] $R_{merge}=100\sum_{hkl}\sum_i|I_i(hkl)-\langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is an individual intensity of the i^{th} observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity of reflection *hkl* with summation over all data.

[b] R-value = $||F_o| - |F_c||/|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

[c] R_{free} is equivalent to R value but is calculated for 5 % of the reflections chosen at random and omitted from the refinement proces (1).

[d] AU, asymmetric unit

[e] ADP, atomic displacement parameter, formally B-factor

[f] As determined by PROCHECK (2)

SUPPLEMENTAL TABLE S2

List of contacts formed between SmCB1 and inhibitors. Analysis of protein-inhibitor contacts between the SmCB1 active site and inhibitors (K11017, K11777, CA074) was performed using program CONTACT (3). The distance cutoffs were set to 3.3 Å and 4.0 Å for hydrogen bonds (H-bonds) and van der Waals interaction, respectively. The SmCB1 residues interacting in the individual inhibitor positions (P2 to P2') are specified. For each SmCB1 residue, total number of contacts is indicated including H-bonds (residues forming H-bonds are *in bold*).

Subsite	K11017			K11777			CA074		
	Residue	Contacts	H-bond	Residue	Contacts	H-bond	Residue	Contacts	H-bond
P2'							Gln 94	2	
							Ser 95	1	
							His 180	8	1
							His 181	6	2
							Ile 193	1	
							Trp 292	8	
P1'	**Gln 94**	4	2	**Gln 94**	5	1	**Gln 94**	4	1
				Cys 97	1				
				Gly 98	2				
	Cys 100	3		Cys 100	3		Cys 100	1	
				Ile 193	1				
	Val 247	5							
	Leu 252	1							
	Gly 269	1		Gly 269	1				
	His 270	11		His 270	8		His 270	4	
	Trp 292	8		Trp 292	2		**Trp 292**	3	1
P1	Gln 94	1		Gln 94	1		**Gln 94**	3	1
				Cys 97	1				
	Gly 98	9		Gly 98	5		Gly 98	5	
							Ser 99	1	
	Cys 100	6		Cys 100	5		**Cys 100**	15	1
							Trp 101	4	
	Cys 141	1		Cys 141	1				
	Glu 142	7		Glu 142	7				
	Gly 143	2		Gly 143	2		Gly 143	5	
							Gly 144	4	1
	Ile 193	2							
	Gly 269	1	1	**Gly 269**	1	1	**Gly 269**	6	1
							His 270	3	
P2	Gly 98	1		Gly 98	1				
	Cys 100	2		Cys 100	2				
	Trp 101	3		Trp 101	3				
	Gly 143	3		Gly 143	3				
	Gly 144	10	2	**Gly 144**	9	2	Gly 144	1	
	Leu 146	2		Leu 146	3		Leu 146	4	
				Gly 244	1				
				Phe 245	1				
	Gly 269	4		Gly 269	5		Gly 269	1	
	His 270	2		His 270	5				
							Ala 271	1	
	Glu 316	1					Glu 316	1	
P3				Gly 138	1				
	Leu 139	2		Leu 139	3				
	Glu 142	2							
	Gly 143	8		Gly 143	9				
	Gly 144	8		Gly 144	8				

SUPPLEMENTAL TABLE S3

Quantum chemical calculations of interaction energies between inhibitors and SmCB1. The non-covalent interaction energy was calculated for the following complexes: (i) crystallographic SmCB1-inhibitor complexes (K11017, K11777 and CA074), (ii) crystallographic complexes with an alternative conformation of Glu-316 in the S2 subsite (K11017(Glu-316), CA074(Glu-316)), and (iii) artificial complexes of K11017 and K11777, in which the phenyl sulfone moiety at P1' were built with interchanged conformations (K11017(P1') and K11777(P1'), respectively). The inhibitor structures were fragmented into the side-chain segments (P3 to P2') and main-chain segments (Pi/P(i-1) connecting Pi and P(i-1)). The P1/P1' segment forming a covalent bond with the catalytic Cys-100 was not calculated (n.c.).

Inhibitor segment	Inhibitor (conformation variant) / (kcal/mol)						
	K11017	K11017 (Glu-316)	K11017 (P1')	K11777	K11777 (P1')	CA074	CA074 (Glu-316)
P3	-4.0	-3.9	-3.9	-5.4	-5.3		
P3/P2	-2.0	-2.2	-2.7	-4.5	-4.4		
P2	0.1	-1.2	-0.6	-6.9	-5.5	1.4	1.6
P2/P1	-5.6	-4.2	-6.0	-4.6	-4.5	-4.5	-4.4
P1	-4.2	-3.7	-4.5	-2.2	-1.1	-0.7	-1.5
P1/P1'	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
P1'	-13.8	-11.3	-2.5	-4.3	-2.1	5.3	5.6
P1'/P2'						-2.6	-2.7
P2'						-32.8	-32.9

SUPPLEMENTAL TABLE S4

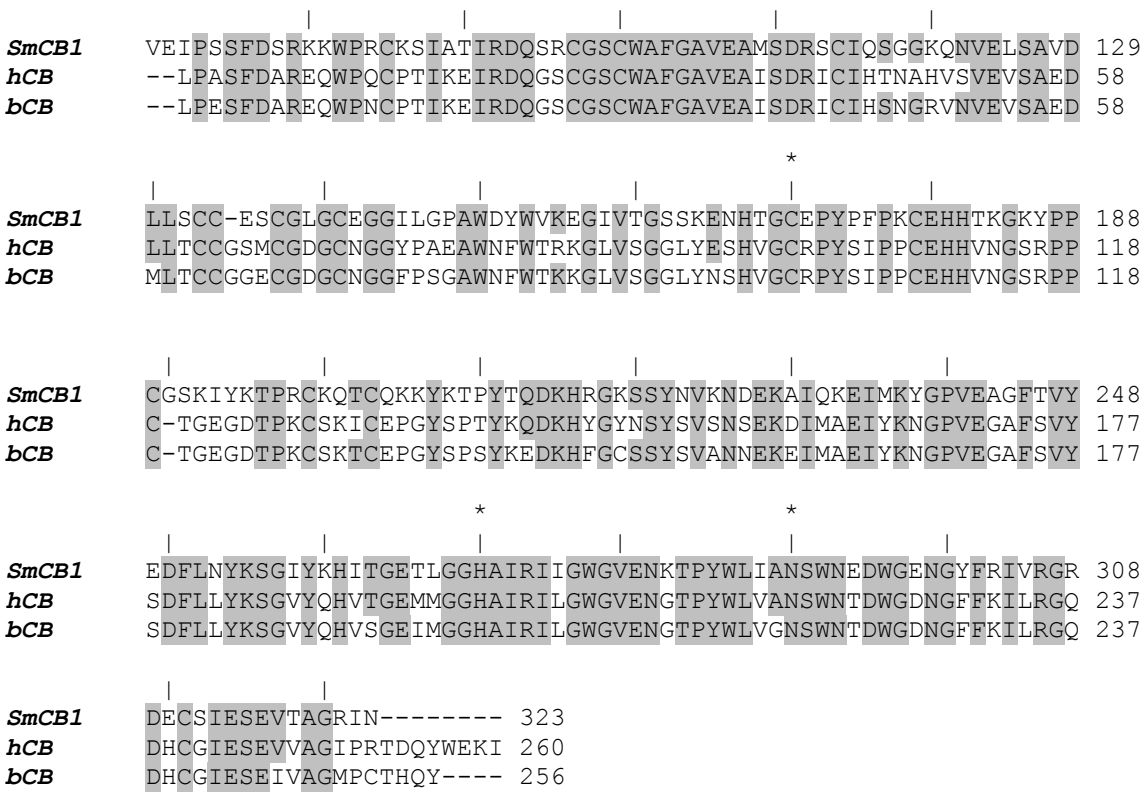
Peptide fragments of hemoglobin produced by the endopeptidase activity of SmCB1. Human hemoglobin was digested with SmCB1 and the digest was separated by RP-HPLC as presented in Fig. 6. The peak fractions of the pool of large peptides fragments (Fig. 6) were analyzed by FT-MS using an LTQ Orbitrap XL mass spectrometer (Thermo) operating in high-resolution mode (R ~10^5). Cleavage sites were searched by the MS-NonSpecific module of ProteinProspector software (University of California San Francisco) using a mass tolerance of 5 ppm. Alternative fragmentations producing isobaric peptides are indicated (/).

Peptide position (residues)	Hemoglobin chain	Measured peptide mass $[M+H]^+$ (Da)
1-32	α	3195.66
14-32	α	1971.97
37-46/38-47	α	1238.62
53-84/54-85	α	3229.66
90-136	α	5105.81
108-142	α	3741.04
111-142	α	3427.84
1-41	β	4261.30
45-68	β	2483.32
102-123	β	2542.32

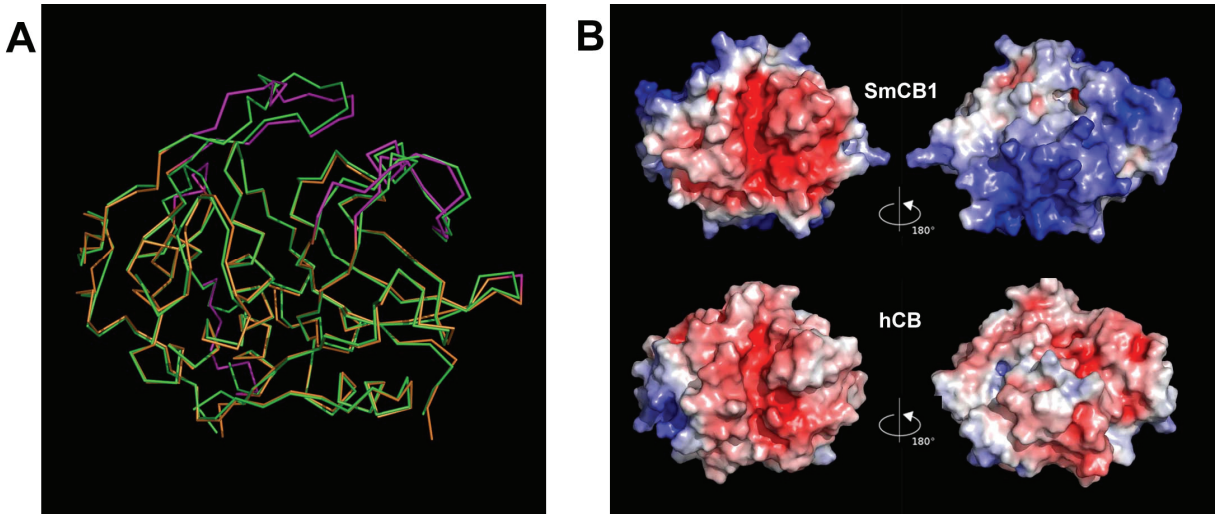
SUPPLEMENTAL TABLE S5

Hydrolysis of carboxydipeptidase and endopeptidase substrates by SmCB1. Kinetic parameters were determined for selective peptidic substrates of cathepsins B with the indicated cleavage mode. The substrate hydrolysis was measured in a kinetic activity assay with SmCB1 at pH 5.5. Mean values are given for triplicate measurements (SE values are within 10% of the mean).

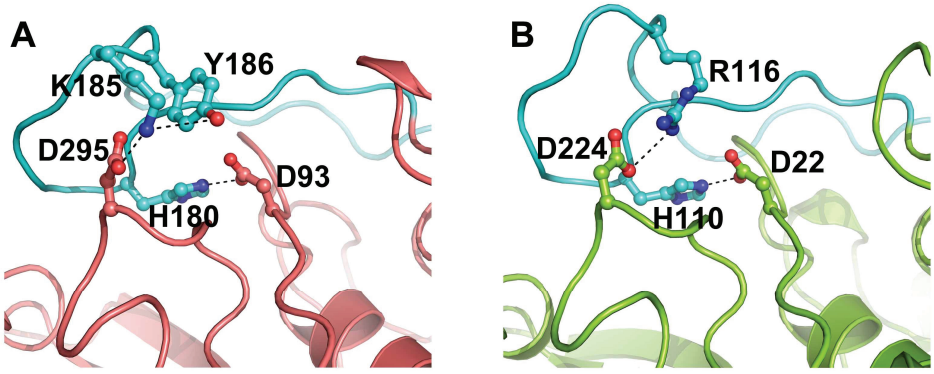
Substrate	K_m (µM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
Carboxydipeptidase			
Abz-Phe-Arg-Val-Nph-OH	1.6	0.4	241,881
Abz-Phe-Arg-Nph-Val-OH	1.8	0.8	442,772
Endopeptidase			
Abz-Gln-Val-Val-Ala-Gly-Ala-EDDnp	7.7	0.07	8,636
Abz-Ala-Phe-Arg-Phe-Ser-Gln-EDDnp	24	0.09	3,756
Cbz-Phe-Arg-AMC	25.4	1.0	37,673



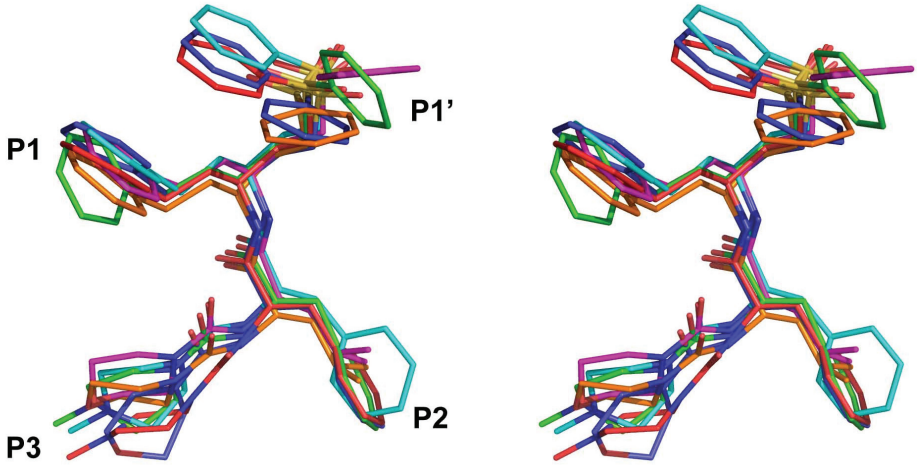
SUPPLEMENTAL FIGURE S1. **Sequence alignment of SmCB1 and mammalian cathepsins B.** The amino acid sequence of mature SmCB1 (UniProt accession Q8MNY2) was aligned with those of human cathepsin B (hCB, UniProt accession P07858) and bovine cathepsin B (bCB, UniProt accession P07688) based on the superposition of their X-ray structures (hCB, PDB code 1CSB; bCD, PDB code 1QDQ). Positions of the catalytic residues are indicated (*). Amino acids identical with those of SmCB1 are shaded. The numbering of SmCB1 is based on the zymogen sequence, whereas numberings of the mammalian cathepsins B are based on the mature proteins. The sequence identities to SmCB1 are 58% and 57% for human and bovine cathepsin B, respectively.



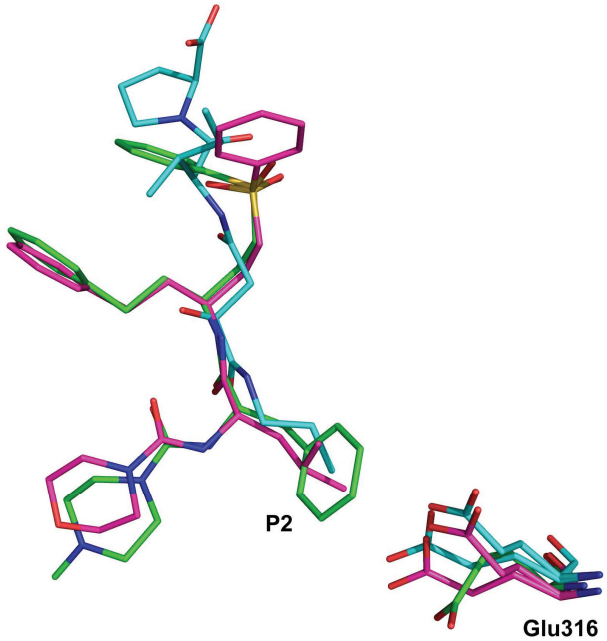
SUPPLEMENTAL FIGURE S2. **Comparison of the structures of SmCB1 and human cathepsin B.** *A*. A superposition of Cα traces of SmCB1 (*orange*) with human cathepsin B (*green*; PDB code 1CSB). The regions in SmCB1 with largest backbone deviations (RMSD >1 Å) are highlighted in *magenta*. The SmCB1 molecule is in the same orientation as in Fig. 1. *B.* Molecular surfaces of SmCB1 and human cathepsin B (hCB) are colored by electrostatic potential that is displayed at a scale from −6 kT (red) to +6 kT (blue). Right-hand views correspond to the orientation in *(A)*.



SUPPLEMENTAL FIGURE S3. **Stabilization of the SmCB1 occluding loop by salt bridges.** A comparison of SmCB (*A*) and human cathepsin B (*B*). The ion pair His-180/Asp-93 (His-110/Asp-22 in *B*) is conserved, while the pair Asp-224/Arg-115 found in human cathepsin B is replaced by a cluster Asp-295/Lys-185/Tyr-186 in SmCB1. The occluding loop is highlighted in *cyan* (*cartoon* and *ball-and-stick* representation). The enzyme structures are from the SmCB1-CA074 complex and the human cathepsin B-CA030 complex (PDB code 1CSB). Side chain heteroatoms have a standard color-coding (O, *red*; N, *blue*).



SUPPLEMENTAL FIGURE S4. **A comparison of vinyl sulfone inhibitors in the active site of SmCB1 and cathepsin L-type peptidases.** Stereo image showing a superposition of structurally related vinyl sulfone inhibitors from the following complexes: SmCB1-K11777 (carbon atoms *cyan*), SmCB1-K11017 (*magenta*), rhodesain-K11777 (*red*, PDB code 2P7U), rhodesain-K11002 (*blue*, 2P86), cruzain-K11777 (*green*, 2OZ2), falcipain 3-K11017 (*orange*, 3BWK). The inhibitor P/P' positions are indicated. Heteroatoms have a standard color-coding (O, *red*; N, *blue*; S, *yellow*).



SUPPLEMENTAL FIGURE S5. **Conformational flexibility of Glu-316 in the S2 subsite of SmCB1.** A comparison of SmCB1-inhibitor complexes: SmCB1-K11777 (*green*), SmCB1-K11017 (*magenta*) and SmCB1-CA074 (*cyan*). Visualized are inhibitors (P2 is indicated) and Glu-316 (including dual conformations) at the bottom of S2. Heteroatoms have a standard color-coding (O, *red*; N, *blue*; S, *yellow*).

SUPPLEMENTAL REFERENCES

1. Brunger, A. T. (1992) *Nature* **355,** 472-475
2. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *Journal of Applied Crystallography* **26,** 283-291
3. CCP4 (1994) *Acta Crystallogr. D. Biol. Crystallogr.* **50,** 760-763

5.3. Publikace č. 3: Quantum Mechanics-Based Scoring Rationalizes the Irreversible Inactivation of Parasitic *Schistosoma mansoni* Cysteine Peptidase by Vinyl Sulfone Inhibitors

5.3.1. Souhrn

V publikaci č. 2 byly aplikovány metody výpočetní chemie na krystalové struktury komplexu SmCB1 s třemi různými inhibitory pro analýzu interakce inhibitorů s jednotlivými vazebnými podmísty aktivního místa SmCB1. Inhibitory byly fragmentovány na dílčí segmenty, pro které byly určeny interakční energie s využitím metod pro výpočet nekovalentních interakcí. Tento přístup umožnil relativně rychlou analýzu vazebných podmíst, ale neřešil problematiku tvorby kovalentní vazby mezi enzymem a inhibitorem a mechanismus jejího vzniku. Publikace č. 3 přináší detailní studii reakčního mechanismu kovalentní inhibice SmCB1 vinylsulfonovými inhibitory s využitím enzymové kinetiky a výpočetní chemie. Hlavním cílem bylo vyvinout výpočetní metodu, tzv. skórovací funkci, pro kovalentní vinylsulfonové inhibitory, která by umožnila efektivně predikovat účinnost inhibitorů *in silico.* Tento nový přístup představuje cenný nástroj při navrhování nových inhibitorů SmCB1.

V této publikaci byla využita sada 20 vinylsulfonových inhibitorů z publikace č. 2, kde pro ně byly experimentálně určeny inhibiční parametry IC_{50}. Tato sada umožňuje studovat deriváty se stejnou reaktivní skupinou a zároveň velkou inhibiční diverzitou danou odlišnými substituenty (v rozsahu 5 řádů hodnot IC_{50}). Reaktivní vinylsulfonová skupina tvoří kovalentní vazbu s thiolovou skupinou katalytického cysteinu v aktivním místě SmCB1. Molekulární proces inhibice lze rozdělit do dvou kroků: nejprve se inhibitor váže do aktivního místa za tvorby nekovalentního intermediátu, což je popsáno parametrem *Ki*; následná tvorba kovalentní vazby je charakterizována parametrem k_{inact}. Celková rychlost reakce je určena rychlostní konstantou k_{2nd}, pro kterou platí $k_{2nd}=k_{inact}/K_i$.

Publikace obsahuje enzymologickou analýzu kinetických parametrů pro reprezentativní vinylsulfonové inhibitory a dále určení ireverzibility jejich inhibiční interakce. Skórovací funkce pro kovalentní vinylsulfonové inhibitory byla získána modifikací již známé kvantově mechanické skórovací funkce pro nekovalentní komplexy, tzv. $score_{nc}$ [188]. Nově je zahrnut popis kovalentní vazby a přidán nový $\Delta G'_{cov}$ term do skórovací funkce: $score_{cov}= score_{nc} + \Delta G'_{cov}$, kde $\Delta G'_{cov}$ odpovídá rozdílu energií mezi kovalentním a nekovalentním komplexem. Kovalentní skóre bylo vypočítáno pro studovanou sadu vinylsulfonových komplexů, přičemž experimentálně získaná krystalová struktura komplexu SmCB1 s K11017 sloužila jako templát pro výstavbu ostatních studovaných

komplexů. Po porovnání skóre pro jednotlivé inhibitory s jejich inhibičními parametry IC_{50} vyplývá, že korelace je patrná zejména pro $\Delta G'_{cov}$ složku skórovací funkce. Tato složka je zároveň jednodušeji získatelná a tím použitelná pro efektivní *in silico* testování kovalentních inhibitorů.

Můj podíl na práci zahrnoval enzymologická měření (určení kinetických parametrů a ireverzibility inhibice).

5.3.2. Publikace č. 3

Fanfrlík, J., Brahmkshatriya, P. S., Řezáč, J., **Jílková**, **A.**, Horn, M., Mareš, M., Hobza, P. a Lepšík, M.

Quantum Mechanics-Based Scoring Rationalizes the Irreversible Inactivation of Parasitic *Schistosoma mansoni* Cysteine Peptidase by Vinyl Sulfone Inhibitors

Journal of Physical Chemistry B 117(48), 14973-14982 (2013)

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Quantum Mechanics-Based Scoring Rationalizes the Irreversible Inactivation of Parasitic *Schistosoma mansoni* Cysteine Peptidase by Vinyl Sulfone Inhibitors

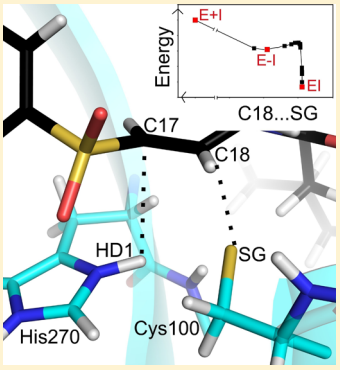
Jindřich Fanfrlík,[†] Pathik S Brahmkshatriya,[†] Jan Řezáč,[†] Adéla Jílková,[†] Martin Horn,[†] Michael Mareš,[†] Pavel Hobza,*[,†,‡] and Martin Lepšík*[,†]

[†]Institute of Organic Chemistry and Biochemistry, v.v.i., and Gilead Sciences and IOCB Research Center, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

[‡]Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc, 771 46 Olomouc, Czech Republic

Supporting Information

ABSTRACT: The quantum mechanics (QM)-based scoring function that we previously developed for the description of noncovalent binding in protein−ligand complexes has been modified and extended to treat covalent binding of inhibitory ligands. The enhancements are (i) the description of the covalent bond breakage and formation using hybrid QM/semiempirical QM (QM/SQM) restrained optimizations and (ii) the addition of the new $\Delta G_{cov}'$ term to the noncovalent score, describing the "free" energy difference between the covalent and noncovalent complexes. This enhanced QM-based scoring function is applied to a series of 20 vinyl sulfone-based inhibitory compounds inactivating the cysteine peptidase cathepsin B1 of the *Schistosoma mansoni* parasite (SmCB1). The available X-ray structure of the SmCB1 in complex with a potent vinyl sulfone inhibitor K11017 is used as a template to build the other covalently bound complexes and to model the derived noncovalent complexes. We present the correlation of the covalent score and its constituents with the experimental binding data. Four outliers are identified. They contain bulky R_1' substituents structurally divergent from the template, which might induce larger protein rearrangements than could be accurately modeled. In summary, we propose a new computational approach and an optimal protocol for the rapid evaluation and prospective design of covalent inhibitors with a conserved binding mode.



1. INTRODUCTION

The covalent inactivation of biomolecular targets is an alternative strategy in pharmaceutical research which yielded several clinical drugs, was later forsaken for potential safety issues but has recently been revived,[1−3] as exemplified by second-generation covalent proteasome inhibitors for the treatment of myeloma.[4] Computational tools for the design of covalent inhibitors are far less developed than for their noncovalent counterparts. A few examples could be algorithms for covalent docking,[5,6] analysis of noncovalent complexes,[7] covalent QSAR descriptors,[8] or enzyme mechanism-based method (EMBM).[9] However, there is still an urgent need to develop robust and efficient protocols for covalent inhibitor computational design.

The molecular processes of the covalent binding of ligands such as enzyme inhibitors can be notionally divided into two steps (eq 1). First, the inhibitor (I) binds noncovalently to the target enzyme (E); this phenomenon is characterized by the K_i value or by its constituent on- and off-rates, k_{on} and k_{off}, respectively (eq 1a). This leads to the formation of a noncovalent complex (E−I), in which the moderately reactive electrophilic group of I approaches the catalytic nucleophile of E (e.g., cysteine or serine residue). Second, a chemical reaction occurs (characterized by k_{inact}), and the resulting covalent-bond formation gives rise to a covalently inactivated complex (EI). The overall thermodynamics and kinetics can be described by the k_{2nd} parameter (eq 1b).

$$\mathrm{E} + \mathrm{I} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{E{-}I} \xrightarrow{k_{inact}} \mathrm{EI} \tag{1}$$

$$K_i = k_{off}/k_{on} \tag{1a}$$

$$k_{2nd} = k_{inact}/K_i \tag{1b}$$

The thermodynamics of the first noncovalent binding step can be described by standard computational tools based on molecular mechanics (MM), such as free-energy perturbation, linear interaction energy, MM-GBSA, or scoring functions.[10] The quantum mechanical (QM) description should also be used where the MM description fails (e.g., in the case of metals[11,12] or halogen bonds[13,14] in the E−I complexes) for an enhanced reliability of the scoring.[15] The second step, the

Received: September 26, 2013
Revised: November 3, 2013
Published: November 6, 2013

covalent-bond formation, requires the use of QM methods to explore the reaction mechanism and to search for the transition states and intermediates. This is done either in a model system using QM calculation in the gas phase or implicit solvent[16] or in the full protein−ligand system using a hybrid QM/MM setup.[17] It should be mentioned here that the use of MM in this step is generally excluded because the MM cannot describe the formation and breaking of a covalent bond unless specifically tailored to this aim.[18]

For cysteine peptidases, much computational work has been done in the group of Bernd Engels.[19−25] Importantly, the stabilization of the zwitterionic thiolate-imidazolium pair of the catalytic Cys-His dyad has been attributed to a hydrogen-bond network and 2−4 water molecules.[21] The reactivity of the thiolate is increased when the substrate binds and disrupts this hydrogen-bond network.[21] The stereo- and regiospecificities of epoxide and aziridine-based covalent inhibitors of cysteine peptidases have been elucidated using QM/MM calculations.[22,23] On the basis of the knowledge gained from computations, new inhibitors have been rationally designed and synthesized and their high potency has been demonstrated experimentally.[24,25]

In our laboratory, we have been developing semiempirical QM (SQM) methods for a reliable and accurate description of all types of noncovalent interactions[26−28] (reviewed in ref 29). Indeed, the use of the corrections for dispersion (D) and hydrogen (H) and halogen (X) bonding with the PM6 SQM method (designated as PM6-D3H4X) results in a high degree of accuracy, which is close to the much more costly DFT-D.[27] Moreover, the linear-scaling algorithm MOZYME[30,31] makes this SQM method so fast that systems of up to 10 000 atoms can be optimized. An important feature of the method is the reliable description of the solvent using the implicit COSMO model.[32]

Armed with these powerful corrected SQM methods, we have designed a QM-based scoring function and successfully applied it to two iconic examples of pharmaceutically relevant E−I complexes, namely, the HIV-1 protease and protein kinases, namely, cyclin-dependent kinase 2 (CDK2) and casein kinase (CK2).[13,33,34] Recently, we have implemented the triple-layer QM/SQM/MM methodology and applied it to another known series of CDK2 inhibitors[35] and also to a designed series of aldose reductase inhibitors.[14] In this work, we aim at extending our QM-based scoring function to be applicable in describing the process of the covalent inactivation of enzymes by inhibitory ligands.

Other laboratories have also used corrected SQM methods for drug design purposes.[29,36−38] The corrected PM6 method has been applied to evaluate the binding to a cucurbit[7]uril host,[39] vitamin D receptor,[40] NADH-dependent enoyl-acyl carrier protein reductase,[41] DNA containing zinc-finger protein,[42] avidin, factor Xa, and ferritin.[43]

Schistosomiasis, caused by a parasitic blood fluke of the genus *Schistosoma*, afflicts over 250 million people worldwide. *Schistosoma mansoni* cathepsin B1 (SmCB1) is the most abundant cysteine peptidase in the parasite gut and is necessary for normal parasite growth.[44−46] SmCB1 is a molecular target for the cure of schistosomiasis in a mouse model using the vinyl sulfone-based inhibitor K11777. Recently, we have reported three crystal structures of SmCB1 complexed with three peptidomimetic covalent inhibitors containing a vinyl sulfone or epoxide reactive moiety.[47] We have also evaluated the interactions of these inhibitors with the individual subsites of the SmCB1 active-site cleft by QM calculations.[47] These data and the inhibition profiling with a panel of 20 vinyl sulfone derivatives have identified the key binding interactions and provided initial insights into the specificity of SmCB1 inhibition.[47] Here, we investigate this panel of vinyl sulfone derivatives using a computational approach.

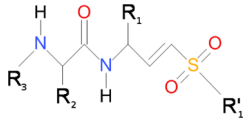
In this study, we show for the first time that the QM-based scoring function originally designed for noncovalent protein−ligand binding[33] can be readily extended to the inhibitors binding covalently to their targets. The QM character of the scoring function enables us to reliably describe both steps of the ligand binding (i.e., the formation of the noncovalent complex and the formation of the covalent bond) by the same methodology. As a proof of concept, we demonstrate this procedure here with a series of vinyl sulfone-based inhibitors of the drug-target peptidase SmCB1,[47] taking advantage of (i) the high-resolution crystal structure of SmCB1 in complex with the potent vinyl sulfone inhibitor **K11017**,[47] (ii) the structural diversity of the investigated vinyl sulfone inhibitors, and (iii) the broad range of their inhibition potency (IC_{50} from 0.61 to 11849 nM, spanning 5 orders of magnitude). Finally, we evaluate the correlation between the calculated and experimental inhibition data and discuss the potential applications and limitations of the developed QM-based procedure in the analysis of covalent protein−ligand complexes.

2. METHODS

2.1. Vinyl Sulfone Inhibitor Series. Table 1 presents 20 vinyl sulfone inhibitors studied previously biochemically[47] and computationally in this work. They are defined by the compound backbone and various substituents at the R_3, R_2, R_1, and R_1' positions (see Table 1). The inhibitors are ordered according to their IC_{50} values against SmCB1. Among the series, X-ray crystal structures have been reported for compounds **K11017** and **K11777**.[47]

2.2. Enzyme Kinetics. *An Analysis of Inhibition Kinetics.* An inhibition assay of vinyl sulfone inhibitors and SmCB1 was performed as described previously.[47] SmCB1 (40 pM) was added to a mixture of fluorogenic substrate Cbz-Phe-Arg-AMC (16 μM) and an inhibitor (2 nM to 20 μM) in the assay buffer (0.1 M Na-acetate pH 5.5, containing 3 mM DTT and 0.125% PEG 1500); the hydrolysis of the substrate was monitored for 30 min. An observed rate constant, k_{obs}, was calculated at each inhibitor concentration by fitting the progress curve to the equation $P = v_i/k_{obs}(1 - e^{(-k_{obs}t)})$ where P is product formation, v_i initial velocity, and t reaction time. The second-order rate constant, k_{second}, was determined depending on the kinetic behavior of the enzyme. In theory, two situations may occur:[48,49] (a) If k_{obs} varies linearly with the inhibitor concentration, then k_{second} can be determined by fitting to the linear equation $k_{obs} = (k_{second}[I])/(1 + [S]/K_m)$; however, this kinetic mechanism does not allow the determination of the individual k_{inact} and K_i parameters. (b) If k_{obs} varies hyperbolically with [I], then $k_{second} = k_{inact}/K_i$, and k_{inact} and K_i can be determined by nonlinear regression using the equation $k_{obs} = k_{inact}[I]/([I] + K_i(1+[S]/K_m))$. The measurement of inhibition kinetics was performed with five model inhibitors (**K11017**, **K11777**, **WRR-145**, **WRR-185**, and **WRR-347**, see Table 1), and all of them were found to follow the kinetic mechanism (a) (see Figure S1 of the Supporting Information). The linear plot of k_{obs} as a function of [I] (Figure S1 of the Supporting Information) shows that the dependence of k_{obs} on the inhibitor concentration is nonsaturating. On the basis of this,

Table 1. Structures of the SmCB1 Inhibitors Used in This Study along with Their Experimental IC_{50} Values (Adapted from Ref 47)[a]



compound	IC_{50} (nM)	substituent position R_3	R_2	R_1	R_1'
WRR-286	0.6	Mpip	Phe	Hph	NH−O−CH_2−Ph
K11017	1.7	Mu	Leu	Hph	Ph
K11002	1.7	Mu	Phe	Hph	Ph
WRR-282	2.0	Mpip	Phe	Hph	CH_2−Ph
K11777	2.1	Mpip	Phe	Hph	Ph
WRR-285	2.1	Mpip	Phe	Hph	NH−Ph
K11747	2.5	Mpip	2-Np-Ala	Hph	Np
AR-198048	6.0	Mpip	Phe-3-CH_3	Hph	Ph
WRR-284	7.8	Mpip	Phe	Hph	O−Ph
AR-198049	10.5	Mpip	Phe-3-CF_3	Hph	Ph
WRR-145	22.9	Cbz	Phe	Ala	Ph
WRR-499	29.5	Mpip	His	Hph	Ph
K11006	33.2	Mu	Phe	Lys	Ph
WRR-483	48.9	Mpip	Arg	Hph	Ph
WRR-359	114	Mpip	Phe[b]	Hph[b]	O−Ph
WRR-185	126	Cbz	Phe	Ala[b]	NH−CH(Bz)−COO−CH_3
WRR-283	192	Mpip	Phe	Hph	O−CH_2−CH_3
WRR-200	247	Cbz	Phe	Ala[b]	NH−CH_2−Ph-4-CF_3
WRR-347	747	Mpip	Phe	Hph	O−Ph-4-O−CH_3
WRR-453	11849	Mu	Phe	Tyr	Ph

[a]We use the following abbreviations for the substituents: *N*-methylpiperazinylcarbonyl (Mpip), morpholinylcarbonyl (Mu), benzyloxycarbonyl (Cbz), phenyl (Ph), benzyl (Bz), naphtyl (Np), and the standard three-letter codes for amino-acid side chains (Phe, Leu, His, Ala, Hph (homophenylalanine)). [b]These residues are in an *R* configuration; all other residues are in an *S* configuration.

we assume that the identified interaction scheme is general for vinyl sulfone inhibitors binding to the SmCB1 active site; the same scheme was previously reported for the interaction of vinyl sulfone inhibitors with several other cysteine peptidases.[48,50,51] Furthermore, the plot of v_i as a function of [I] (Figure S1 of the Supporting Information), showing decreasing v_i with [I], indicates a two-step reaction scheme.[52]

An Analysis of Irreversibility. To demonstrate the irreversibility of vinyl sulfone inhibitors, we investigated two representative compounds, **K11777** and **WRR-359** (the model inhibitors of *S* and *R* configurations, respectively), using a dilution experiment.[53] SmCB1 (400 pM) was preincubated (10 min at 37 °C) with a vinyl sulfone inhibitor and the assay buffer in a 20 μL volume; the inhibitor concentration was set to the IC_{50} or 10-times the IC_{50} level. Afterward, the mixture was diluted 10-fold with the assay buffer with an inhibitory concentration of IC_{50} (for IC_{50} preincubation concentration) or without an inhibitor (for 10× IC_{50} preincubation concentration) and was incubated for 10 min at 37 °C. The final activity was measured with a Cbz-Phe-Arg-AMC substrate as described in ref 47. This analysis showed that the activity of SmCB1 is controlled by the initial concentration of the inhibitors in the preincubation step and that the extent of inhibition is not reduced by the subsequent dilution (as known for reversible inhibitors).[53] The experiment clearly demonstrated the irreversibility of the inhibition of the inhibitors tested.

The inhibition constant values presented in Table 1 are in the form of IC_{50}, which represents the inhibitor concentration necessary to cause 50% inhibition of the enzyme under the conditions of the enzyme assay. The following relations between IC_{50} and k_{obs} hold for an irreversible inhibitor:[54]

$$k_{obs} = 0.693/t_{1/2} = 0.693/t_{assay}$$

$$k_{obs}/[I] = 0.693/[t_{assay}IC_{50}]$$

Under the assumption that IC_{50} is the inhibitor concentration necessary to reduce the enzyme activity by 50% during the time of incubation, the assay time (t_{assay}) is equal to the half-life $t_{1/2}$ for the k_{obs} rate constant.

2.3. Setup of the Calculations. Hydrogens were added to the protein using the Reduce program[55] and to the ligands using Chimera, ver. 1.5.3[56] to correspond to the experimental pH of 5.5.[47] This means that both the N- and C-termini, all the aspartates, glutamates, lysines, arginines, and all the histidine residues (except a neutral Nε-protonated His270 in the covalent EI complex, Figure 1B) were charged. Special care

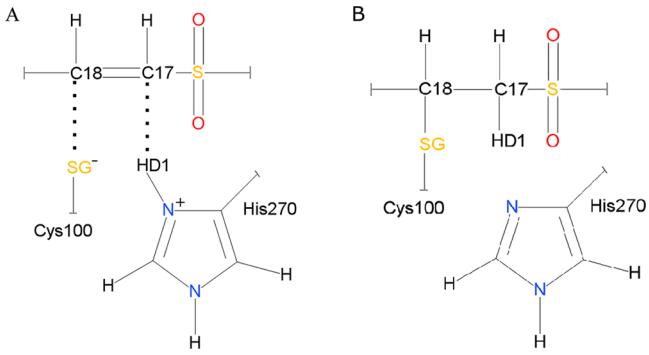


Figure 1. Scheme of noncovalent (E−I, panel A) and covalent (EI, panel B) SmCB-vinyl sulfone inhibitor complexes. The dotted lines in (A) represent the nucleophilic attack of Cys100:SG by C18 of the inhibitor and the associated proton transfer (HD1 from His270 to C17 of the inhibitor).

was taken to treat the catalytic residues Cys100 and His270. As was found earlier,[21] the protein frame provides the stabilization of the zwitterionic state, and these two residues were thus modeled as Cys100:thiolate and doubly protonated His270 in the noncovalent E−I complex, Figure 1A.

Our previous setup was a full SQM optimization of the protein−ligand complexes.[13,33,34] In a subsequent work we utilized a hybrid SQM/MM setup with a generalized Born (GB) solvent to speed up the calculations; a three-layer QM/SQM/MM-GB setup was also implemented for more quantitative results.[14,35] Herein, we have made further progress in using a hybrid QM/SQM methodology in that we coupled it with the COSMO implicit solvent model,[32] which has recently been shown to be more reliable than GB.[57]

Like in our previous works,[14,35] the coupling between the QM and SQM parts was done with an in-house program (CUBY3) using a subtractive scheme of an ONIOM type.[58,59] It calls Turbomole[60] for QM and MOPAC[30] for SQM. The QM part was treated by DFT with empirical dispersion (D for optimizations,[61] D3 for single points[62]). The B-LYP/SVP

functional/basis set combination was used for optimizations and the TPSS/TZVPP combination for single-point energies. The QM region comprised the inhibitors and residues Cys100, Gln94, and His270 (146 atoms altogether). The SQM part was treated by PM6-D3H4X. The PM6-D3H4X method (the semiempirical PM6 method[31] corrected for dispersion interactions as well as hydrogen- and halogen-bonding[26−28]) has been shown to provide accurate interaction energies for model complexes[27] and for the evaluation of protein−ligand interactions in QM-based scoring.[14] For single-point calculations, whole complexes were considered (about 4 000 atoms). For geometry optimization, only residues within 10 Å of the inhibitor were considered (about 1 600 atoms). Residues farther than 8 Å from the inhibitor were frozen during the optimization. The protein surrounding was modeled using the COSMO implicit solvent model.[32]

The QM/SQM optimizations were performed using the FIRE optimization algorithm[63] until the energy and gradient convergence criteria (ΔE = 0.005 kcal mol^{-1}, maximum gradient of 1 kcal mol^{-1} Å^{-1}, root-mean-square value of the gradient of 0.5 kcal mol^{-1} Å^{-1}) were met.

2.4. Generating Covalent Complex Structures. Covalently bound complexes were built based on the SmCB1-**K11017** crystal structure (PDB entry 3S3Q).[47] The guanidinium side chain of compound **WRR-483**, the aminobutyl side chain of **K11006**, and the *N*-methylpiperazine ring at R_3 and the imidazole ring of the **WRR-499** compound were considered in the cationic form because of their pK_a values.

The compounds **WRR-359**, **WRR-185**, and **WRR-200** differed in their stereochemistry at the carbons bearing the R_2 and R_1 substitutions (R isomers) (Table 1). We could not build these complexes from the SmCB1-**K11017** complex because of steric clashes. The modeling of these compounds would require large changes in the protein conformation, which is beyond the limits of the presented methodology.

The built inhibitor modifications were relaxed by annealing from 300 to 0 K at the MM level in AMBER. The parm03 force field was used for the protein and the GAFF force field for the ligands.[64] The atomic charges for the ligands were calculated using the RESP procedure at the HF/6-31G* level as recommended.[65] The cooling runs were 1 ps, and we used the Berendsen thermostat, 1 fs time step, GB solvent model (Bondi radii and igb = 7 sander option). Three levels of relaxation were explored: first, only the ligand modifications were treated as flexible; second, His181 and Glu313 were also treated as flexible; third, all the amino acids within 6 Å of R_1' of the **K11017** were treated as flexible. Four runs of annealing were performed for each level of relaxation. All the generated structures were optimized using the QM/SQM methodology. The most stable complex of each ligand was selected and further used.

2.5. Generating Noncovalent Complex Structures. For the generation of noncovalent complexes, it was necessary to break the covalent bond between the C18 atom of the inhibitor and the SG atom of Cys100 and transfer the HD1 atom of the inhibitor to His270, according to an assumed reaction mechanism (see also Figure 1).[21] To this aim, we essayed three types of bond breaking using harmonic restraints and relaxed scans: (i) the C18-SG bond, (ii) the C17-HD1 bond, and (iii) both C18-SG and C17-HD1 bonds. The distance ranges were 1.9−3.4 and 1.0−3.0 Å for the C18-SG and C17-HD1 bonds, respectively. The obtained noncovalent complexes (E−I) were reoptimized without any restraint. To localize the noncovalent intermediate E−I, we further extended the C18−SG distance from 2.4 to 3.4 Å using a relaxed scan. Such an approach gave us the "free" energy difference between the covalent and noncovalent complexes ($\Delta G'_{cov}$) and an estimate for the height of the reaction barrier (E_a).

2.6. QM/SQM Scoring. The noncovalent complexes were scored using our standard QM-based scoring function[33] (QM/SQM in this case). The binding free energy of the noncovalent protein−ligand complex is approximated by the noncovalent score ($score_{nc}$) expressed by eq 5 (ref 33; a new consistent notation in ref 29).

$$score_{nc} = \Delta E_{int} + \Delta\Delta G_{solv} + \Delta G'_{conf}{}^{w}(P) + \Delta G'_{conf}{}^{w}(L) - T\Delta S_{int} \quad (5)$$

where

$$\Delta\Delta G_{solv} = \Delta\Delta G_{int,solv} + (\Delta G^{low}{}_{solv}(L) - \Delta G^{high}{}_{solv}(L)) \quad (5a)$$

The individual terms describe the gas-phase interaction energy (ΔE_{int}), interaction desolvation free energy ($\Delta\Delta G_{solv}$), change of the conformational free energy of the ligand and protein ($\Delta G'_{conf}{}^{w}$(P,L)), and the entropy change upon binding ($T\Delta S_{int}$). The ΔE_{int} was calculated using the QM/SQM method on the whole complexes (optimized in water environment) by subtracting the energies of the protein and ligand alone. The $\Delta\Delta G_{int,solv}$ was determined by the COSMO solvent model[32] implemented in MOPAC. The desolvation free energy of the bare inhibitor ΔG_{solv}(L) was evaluated by a more demanding but also more reliable[57] SMD model[66] at the HF/6-31G* level of theory implemented in Gaussian09.[67] This gave rise to the correction for ligand desolvation ($\Delta G^{low}{}_{solv}$(L) − $\Delta G^{high}{}_{solv}$(L)), which is especially important for a correct description of charged inhibitors.[33]

The $T\Delta S_{int}$ term was estimated by assigning a penalty of 1 kcal/mol to each rotatable bond of the ligand that became hindered upon complex formation, which is similar to other scoring functions.[68,69] As an alternative, we applied a 1 kcal/mol penalty for the total number of rotatable bonds in the ligand and also for the rotatable bonds of the protein side chains that became hindered upon complex formation.[70]

The $\Delta G'_{conf}{}^{w}$(L) term represents the difference between the conformation free energy of the ligand in the conformation restrained by the protein surroundings and that of the free ligand in water. $\Delta G'_{conf}{}^{w}$(L) is the sum of the vacuum and solvation parts. ΔE_{def}(L) was calculated at the DFT-D3 level (see above), and $\Delta\Delta G_{conf,solv}$(L) using SMD at the HF/6-31G* level. The ligand conformation restrained by the protein surroundings was taken from the optimized E−I complex. The energy of the ligand conformation in water was obtained by using a quenching technique (DFT-D B-LYP/SVP for MD and gradient optimization followed by DFT-D3 TPSS/TZVPP calculation combined with the COSMO solvent model; a 50 ps MD simulation in 300 K, snapshots were taken every picosecond and optimized, and the obtained energies were averaged).

The $\Delta G'_{conf}{}^{w}$(P) term is the sum of the vacuum and solvation parts of the protein in the E−I complex and of the isolated optimized protein (taken from the SmCB1/**K11017** crystal structure). This simple approach can be used because all the complexes have been modeled using the same crystal structure. This term is calculated at the QM/SQM level.

In addition to these standard scoring terms for treating noncovalent protein−ligand complexes, which had been applied in several studies from our laboratory (reviewed in ref 29), the total covalent score in this work was constructed as a sum of $score_{nc}$ and $\Delta G'_{cov}$. The latter term was calculated as the "free" energy (i.e. the vacuum energy plus solvation free energy) difference between the covalent and noncovalent complexes.

The calculated total scores were compared to the experimental binding data in the form of $RTln(IC_{50})$ using the coefficient of determination (R^2) and the predictive index (PI).[71] While R^2 reflects the variability of the data by the least-squares linear fit, the PI represents a measure of the rank-order prediction: 1 stands for an always correct prediction, 0 for a random prediction, and −1 for an always incorrect prediction.[71]

3. RESULTS AND DISCUSSION

3.1. Covalent Complexes of SmCB1 and Vinyl Sulfone Inhibitors. The series of 20 vinyl sulfone inhibitors (Table 1) was modeled in the active site of SmCB1 based on the crystal structure of the SmCB1/**K11017** covalent complex. Four complexes (**WRR-359**, **WRR-185**, **WRR-200**, and **WRR-453**) could not be built because of steric clashes with the protein. For three of these inhibitors (**WRR-359**, **WRR-185**, and **WRR-200**), the simple structural reason was the opposite stereochemistry at the carbons having R_2 and R_1 substitutions (R isomers) as compared to the parental **K11017**. This finding qualitatively indicates that these compounds require large changes of the protein conformation, the modeling of which is beyond our current approach. This may also in part explain the poor activity of these compounds in the experiment (IC_{50} from 114 to 11849 nM, see Table 1).

The binding mode of the covalently bound inhibitors in the modeled complexes is shown in Figure 2. Nine of the studied inhibitors have the phenylsulfonyl moiety in the P1′ position. This chemical group can adopt two distinct conformations in the active site of SmCB1 and related cysteine peptidases.[47,48,72−74] In our approximate dynamics-based approach with the limited treatment of protein flexibility, we found the same conformation of the phenylsulfonyl moiety in P1′ in all of these complexes (see Figure 2A). It might be the case that allowing extended protein flexibility and longer simulation times would also yield different phenylsulfonyl conformations. However, we had previously found that the alternative conformation of the phenylsulfonyl moiety in P1′ had a smaller stabilization energy,[47] which is consistent with the current results.

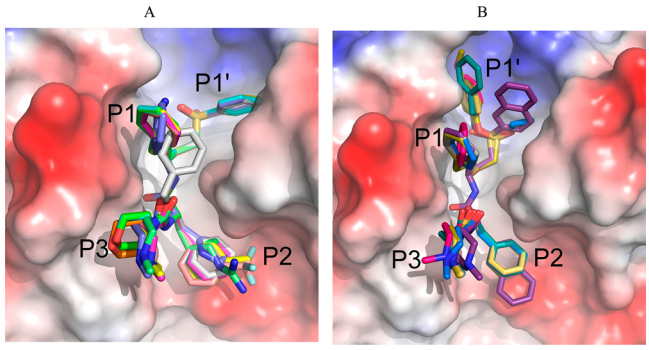


Figure 2. Binding mode of the vinyl sulfone inhibitors (the sticks of different colors) in the modeled covalent complexes with SmCB1. (A) The inhibitors with phenylsulfonyl at R_1': **K11002** (cyan), **K11017** (green), **K11777** (magenta), **AR-198048** (yellow), **AR-198049** (pink), **WRR-145** (light gray), **WRR-499** (purple-blue), **K11006** (orange), and **WRR-483** (lime-green); (B) Inhibitors with other substituents at R_1': **WRR-286**, **WRR-282** (red), **WRR-285** (yellow-orange), **K11747** (deep purple), **WRR-284** (gray), **WRR-283** (blue), **WRR-347** (sand). Hydrogens are not shown for clarity. The protein surface is colored based on the vacuum electrostatic potential. The figure was prepared with PyMol.[75]

The studied inhibitors had similar overall binding modes, and the R_3, R_2, R_1, and R_1' substituents of the inhibitor bound to the corresponding S3, S2, S1, and S1′ binding subsites of SmCB1. Of special interest was the structure of the SmCB1/**WRR-145** covalent complex. The benzyloxycarbonyl R_3 substituent folded back onto the hydrophobic R_1 substituent (Figure 2A, the inhibitor is show in the light gray color). This may be caused by the higher hydrophobicity of the R_3 substituent (benzyloxycarbonyl vs *N*-methylpiperazinylcarbonyl) and the smaller size of the R_1 substituent (Ala vs Hph) of **WRR-145**.

3.2. Noncovalent Complexes of SmCB1 and Vinyl Sulfone Inhibitors. The protocol to model the noncovalent E−I complexes was worked out on the crystal structure of the SmCB1/**K11017** covalent complex (PDB entry 3S3Q).[47] Following the presumable reaction coordinate in the reverse direction,[54,21] we used harmonic restraints and relaxed scans to break different covalent bonds: (i) the C18-SG bond, (ii) the C17-HD1 bond, and (iii) both C18-SG and C17-HD1 bonds (Figure 3A,B). The first case resulted in an unrealistic scenario, because upon the C18-SG break the presumed associated HD1 proton transfer from C17 to His270:ND1 did not occur. In contrast, the second proton-transfer approach induced a concomitant break of the C18-SG covalent bond (the interatomic distance increased from 1.9 Å to about 2.4 Å) and a change of C17 and C18 carbon hybridization from sp^3 to sp^2. This was the most appropriate description of the reaction mechanism and we have thus continued to use it. The third approach gave similar results but was more artificial because two distances had been restrained.

However, the completed C17-HD1 proton transfer to His270:ND1 did not cover the whole reaction coordinate. The C18-SG distance of 2.4 Å comprised only about 70% of the sum of the van der Waals radii and thus probably represented the presumed transition states. To localize the noncovalent intermediate E−I, we further extended the C18−SG distance from 2.4 to 3.4 Å using a relaxed scan. The noncovalent E−I complex was found at a C18-SG distance of 2.84 Å.

The energy profile of the forward reaction in the SmCB1/**K11017** complex proceeded from the separated reactants to the noncovalent intermediate E−I, which was more stable by 19.0 kcal/mol (Figure 3E). Further shortening of the C18−SG bond lead to the presumed transition state via an activation barrier E_a of 4.4 kcal/mol (see Figure 3C, D, E). The further progress to the products was highly exothermic (by about 25 kcal/mol).

This reaction profile is in qualitative agreement with kinetic measurements (see 2.2. Enzyme Kinetics), which did not allow us to determine the individual k_{inact} and K_i parameters and thus indicated a possible very fast forward-running chemical reaction. The structural and energy features of the reaction coordinates of the modeled noncovalent complexes for the other investigated vinyl sulfone inhibitors were similar, with E_a values ranging from 1.5 to 7.2 kcal/mol.

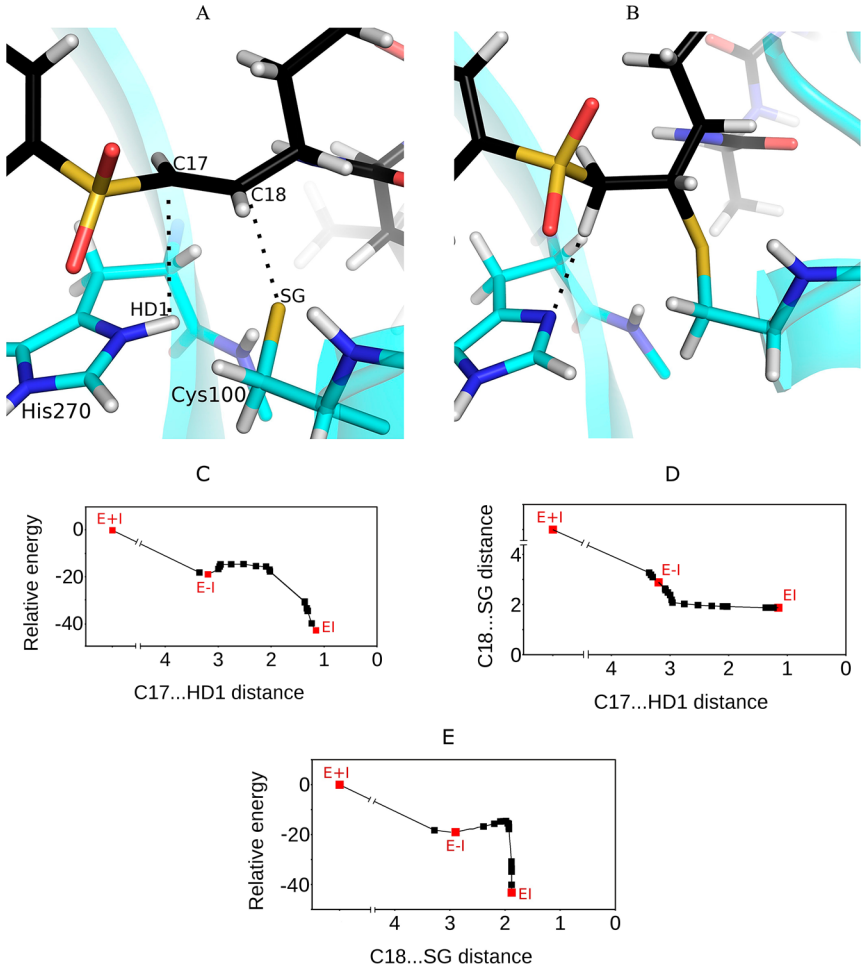


Figure 3. (A) Structural details of the modeled noncovalent SmCB1/**K11017** complex, intermediate (E−I), with the reaction coordinates, the C17−HD1 and C18−SG distances, marked by dotted lines. (B) Covalent SmCB1/**K11017** complex, the experimentally determined final product of the inhibition reaction (PDB entry 3S3Q). The color coding is as follows: red, oxygen; blue, nitrogen; white, hydrogen; yellow, sulfur; cyan, carbon atoms of SmCB1; black, carbon atoms of **K11017**. (C) Relative energy (compared to the energy of the separated reactants) plotted against the C17−HD1 distance. (D) C18−SG distance plotted against the C17−HD1 distance. (E) Relative energy plotted against the C18−SG distance. All distances are in angstroms, and energies are in kilocalories per mole.

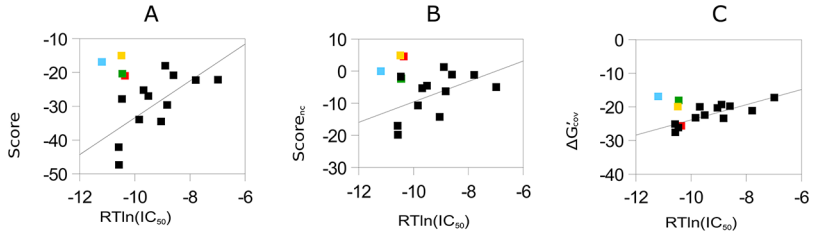


Figure 4. Correlation between the calculated and experimental binding data of the vinyl sulfone inhibitors in the modeled complexes with SmCB1. The calculated terms used for the correlation are as follows: (A) The total covalent score, where the coefficient of determination (R^2) is 0.47 and predictive index (PI) is 0.68. The outliers (**WRR-286**, cyan; **WRR-285**, green; **WRR-282**, yellow; **K11747**, red) are not considered for R^2 and PI determination. (B) The noncovalent score ($score_{nc}$), where R^2 is 0.27 and PI is 0.49. (C) The free energy difference between covalent and noncovalent complexes ($\Delta G'_{cov}$), where R^2 is 0.69 and PI is 0.81. All the energies are in kilocalories per mole.

3.3. QM-Based Scoring of SmCB1/Vinyl Sulfone Inhibitor Complexes. The calculated estimates of the binding affinities (scores) for the modeled vinyl sulfone inhibitors to SmCB1 were compared to the experimental inhibition IC_{50} values available. The IC_{50} is a suitable parameter for the enzymological evaluation of the inhibitors studied because we experimentally demonstrated that they are irreversible ligands and the kinetic measurements did not allow us to determine their individual k_{inact} and K_i parameters (see 2.2. Enzyme Kinetics and Figure S1 of the Supporting Information). The computationally estimated values of the reaction barrier E_a were low (ranging from 2 to 7 kcal/mol), which agreed with the kinetic measurements of the fast forward-running chemical reaction. We thus assumed that the conversion of noncovalent complexes to covalent is very fast for all the studied inhibitors and the k_{inact} rate should have a comparable impact on the measured IC_{50}. The estimates of E_a energies were thus neither included in the scores nor correlated with the experimental

dx.doi.org/10.1021/jp409604n | J. Phys. Chem. B 2013, 117, 14973−14982

data. The experimental binding values (from Table 1) were compared to the total covalent score values and their components which are defined in Section 2.6. QM/SQM Scoring of Methods.

The calculated scores and their terms are summarized in Table S1 of the Supporting Information and are plotted against the experimental binding data in Figure 4. The correlation for the whole series of compounds was insignificant (the predictive index, PI, ranged from −0.05 to 0.29, and R^2 from 0.02 to 0.12). However, a visual examination of the graphs showed that four outliers deteriorated the correlation. Upon exclusion of these four outliers (**WRR-286**, **WRR-282**, **WRR-285**, and **K11747**), the coefficient of determination (R^2) equaled 0.47 and PI equaled 0.68 (Figure 4A). The affinities of these inhibitors tend to be underestimated (see Figure 4). A structural analysis revealed that the common feature of these compounds was a more bulky substituent R_1' (Table 1) as compared to the template inhibitor **K11017** used for the modeling. We hypothesize that these structurally divergent substituents in the R_1' position induce larger conformational changes in the S1′ pocket of SmCB1, which cannot be accurately described with the limited treatment of the protein flexibility in the methodology applied (see Methods). However, a more adequate treatment of the protein flexibility would significantly increase the cost of the calculations, which would be impractical for high-throughput screening.

The correlation of the $\Delta G'_{cov}$ term with the experimental data was significantly higher than that of the total covalent score (compare Figures 4A and 4C). The number of outliers decreased from four to three for the correlation of the $\Delta G'_{cov}$ term. Upon exclusion of the three remaining outliers (**WRR-286**, **WRR-282**, and **WRR-285**), the R^2 equaled 0.69 and PI was 0.81. The $\Delta G'_{cov}$ term was thus the most important term of the total covalent score. When the $\Delta G'_{cov}$ term was neglected (i.e., $score_{nc}$ alone considered) there was only a weak correlation (R^2 = 0.27 and PI = 0.49, see Figure 4B).

The other terms were less important. The omission of $\Delta G'_{conf}{}^{w}(P)$ from the total score had a negligible effect (R^2 = 0.46 and PI = 0.67). The exclusion of $\Delta G'_{conf}{}^{w}(L)$ slightly decreased the correlation (R^2 = 0.40 and PI = 0.62). Similarly, when both $\Delta G'_{conf}{}^{w}(L)$ and $\Delta G'_{conf}{}^{w}(P)$ were excluded, the correlation slightly decreased (R^2 = 0.41 and PI = 0.64, see Graph S1 of the Supporting Information)

The entropy change of the ligand upon binding was estimated by the number of rotatable bonds of the ligand hindered upon complex formation. The omission of this term slightly deteriorated the correlation (R^2 of 0.40 and PI of 0.66 as compared to the original R^2 of 0.47 and PI of 0.68 for the total covalent score including the entropy term). We also tried to estimate this term by the total number of rotatable bonds in the ligand. This approach gave very similar correlation of the total covalent score (R^2 = 0.46 and PI = 0.68, see also Graph S1 of the Supporting Information). Additionally, we also applied 1 kcal/mol penalty for the rotatable bonds from the protein side chains which became hindered upon binding. The effect on the score correlation was also quite small, R^2 equaled 0.45 and PI was 0.70.

3.4. Limitations in Scoring Terms. The correlation of the $\Delta G'_{cov}$ term with the experimental binding data was considerably higher than that of the total covalent score. This may stem from the fact that the ΔE, $\Delta\Delta G_{solv}$, and $\Delta G'_{cov}$ terms are described at a significantly higher level than the other terms.[34] These approximations in our scoring procedure can deteriorate the correlation as shown previously.[13,34] For example, the lack of dynamical treatment of the ligands may lead in the case of large flexible HIV protease inhibitors to uncertainties in the computed energies of 2.5 kcal/mol.[76] Even worse, to obtain converged values for the conformational free energy changes of a protein in the binding process, an extensive sampling on the microsecond time scale is needed (see, for example, ref 77). Much simpler and faster protocols utilizing optimizations of a limited surroundings of the protein active site have thus been proposed.[78] Also, the use of the ligand-bound form of the protein instead of the apoprotein can circumvent the problem and yield reliable relative scores.[33]

In this study, the $T\Delta S_{int}$ term was estimated by assigning an energy penalty to each hindered rotatable bond in the ligand and optionally also to the hindered amino acid side chains of the protein. The changes in the vibrational entropy and zero-point vibrational energy (ZPVE) upon binding are very difficult to address accurately using computations.[79,80] In our previous studies, we used vibrational analysis at the MM level[33,34] but only with a limited success because of the inability of the empirical potential in describing the quantum phenomena, such as halogen bonds.[13] Making use of this knowledge, we did not opt for the vibrational analysis at the MM level because of the breaking of covalent bonds and proton transfer in the systems studied here. In addition, the QM/SQM methodology that is used for other terms of the covalent score is impractical for vibrational analysis and optimization to very tight convergence criteria in extended P−L complexes.

ΔE_{int} is usually negative and thus makes a favorable contribution to binding, whereas $\Delta\Delta G_{solv}$ usually disfavors binding.[29,33] The former term reflects the strength of the interaction between the ligand and protein. The latter term can be interpreted as the free energy needed to desolvate the ligand and binding site of the protein. It can also be viewed as an effect of the solvent on the first term. In the case of charged ligands and proteins, these two terms may be large and their balance can become the most important part of the score. However, when a ligand and its protein surroundings have a charge of the same sign, ΔE_{int} can become repulsive. In such a case, the $\Delta\Delta G_{solv}$ term may lower the repulsion and have a negative sign, i.e., favor the ligand binding. This is the case of **K1006** binding to SmCB1, where the solvent screens the repulsion of its lysine R_1 side chain with the positively charged His180 and His181 side chains of SmCB1 (Table S1).

3.5. Interaction Energy Decomposition. To demonstrate the power of the QM-based scoring for rational drug design, we have fragmented the vinyl sulfone SmCB1 inhibitors and calculated the interaction free energies of the R_1 substituents with the SmCB1 active site at the PM6-D3H4X/COSMO level. For the purpose of fragmentation, the Cα−Cβ bonds of the inhibitors were cut and capped by hydrogens.

The results of this analysis are summarized in Table 2. It is shown that chemically identical substituents have similar interaction energies, and the most important finding is as follows. The parental compound **K11017** as well as a majority of the other inhibitors have homophenylalanine (Hph) as R_1; **WRR-145** has Ala and **K11006** has a charged Lys in R_1. The interaction free energies of Hph in R_1 were calculated to range from −2.2 to −3.8 kcal/mol for the inhibitors studied. When Hph was changed to Ala, the interactions were lost (an interaction energy of 0.9 kcal/mol) and the charged Lys in the hydrophobic S1 pocket had a repulsive character (6.1 kcal/mol). These calculations quantitatively explain why the IC_{50} of

Table 2. Interaction Free Energies of the R_1 Substituents with the SmCB1 at the PM6-D3H4X/COSMO Level

compound	R_1	interaction free energy (kcal/mol)
WRR-286	Hph	−3.0
K11017	Hph	−3.1
K11002	Hph	−3.0
WRR-282	Hph	−2.3
K11777	Hph	−3.1
WRR-285	Hph	−2.4
K11747	Hph	−3.1
AR-198048	Hph	−2.9
WRR-284	Hph	−2.2
AR-198049	Hph	−3.1
WRR-145	Ala	0.9
WRR-499	Hph	−3.8
K11006	Lys	6.1
WRR-483	Hph	−3.3
WRR-283	Hph	−2.6
WRR-347	Hph	−3.1

both **WRR-145** and **K11006** are quite high (about 23 and 33 nM, respectively; see Table 1). However, it should be mentioned here that the fragmentation approach provides only quick hints for a rational in silico drug design process and that the $\Delta G_{cov}'$ or the whole score should be calculated for new proposed inhibitors.

4. CONCLUSIONS

We have extended the applicability of the advanced quantum mechanics (QM)-based scoring function developed previously in our laboratory for a rapid and reliable description of noncovalent protein−ligand binding. From this moment onward, it can also be used for covalent binding thanks to these enhancements: (i) QM/SQM (DFT-D/PM6-D3H4X with the COSMO implicit solvent model) restrained optimizations for the modeling of the noncovalent complexes and (ii) a new term in the score, $\Delta G'_{cov}$, describing the free energy difference between the covalent and noncovalent complexes.

This new, enhanced covalent QM-based scoring function has been applied to describe the inactivation of *S. mansoni* cathepsin B1 (SmCB1) with 20 vinyl sulfone-based inhibitors. Four outliers have appeared in the study. They all have bigger R_1' substituents which are difficult to model with our relatively fast, medium-throughput procedure. We believe that new crystallographic structures of other SmCB1/inhibitor complexes could overcome these limitations.

The analysis of the results showed that the $\Delta G_{cov}'$ term correlated with the experimental binding data better than the total covalent score (R^2 of 0.69, PI of 0.81 versus 0.47 and 0.68, respectively). In addition, the $\Delta G_{cov}'$ term is more readily accessible. Although these findings may not be general, we propose that $\Delta G_{cov}'$ may be used for faster and broader in silico screening of covalent inhibitors for activity. The total covalent score may be used to provide a more complete insight into the binding of selected inhibitors.

The developed covalent QM-scoring methodology has the advantage of being modular, i.e., it gives physically meaningful terms. Moreover, this structure-based approach offers the possibility of decomposing the interaction energies to individual inhibitor fragments. The insight obtained makes this method a powerful tool for rational in silico drug design of covalent enzyme inhibitors.

■ ASSOCIATED CONTENT

Supporting Information

Detailed enzyme inhibition kinetics for SmCB1 and the inhibitor compound K11777 and the calculated scores and scoring terms. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*M.L.: phone, (+420) 220 410 319; fax, (+420) 220 410 320; e-mail: lepsik@uochb.cas.cz.

*P.H.: phone, (+420) 220 410 311; fax, (+420) 220 410 320; e-mail, hobza@uochb.cas.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was a part of research project RVO: 61388963 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and was supported by the Gilead Sciences and IOCB Research Center, Prague. P.H., M.L., P.S.B., J.Ř., and J.F. were supported by the Czech Science Foundation [P208/12/G016]. This work was also supported by the Operational Program Research and Development for Innovations − European Science Fund (CZ.1.05/2.1.00/03.0058). A.J., M.H., and M.M. were supported by the Czech Science Foundation [203/09/1585] and grant LH12023 from the Ministry of Education, Youth and Sports of the Czech Republic. We thank James T. Palmer (Biota Inc., Victoria, Australia), William R. Roush (The Scripps Research Institute, FL) and Adam R. Renslo (Department of Pharmaceutical Chemistry and the Small Molecule Discovery Center, University of California San Francisco, San Francisco, CA) for providing inhibitors.

■ REFERENCES

(1) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discovery* **2011**, *10*, 307−317.

(2) Smith, A. J.; Zhang, X.; Leach, A. G.; Houk, K. N. Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *J. Med. Chem.* **2009**, *52*, 225−233.

(3) Potashman, M. H.; Duggan, M. E. Covalent Modifiers: An Orthogonal Approach to Drug Design. *J. Med. Chem.* **2009**, *52*, 1231−1246.

(4) Kisselev, A. F.; Van Der Linden, W. A.; Overkleeft, H. S. Proteasome Inhibitors: An Expanding Army Attacking a Unique Target. *Chem. Biol.* **2012**, *19*, 99−115.

(5) De Cesco, S.; Deslandes, S.; Therrien, E.; Levan, D.; Cueto, M.; Schmidt, R.; Cantin, L. D.; Mittermaier, A.; Juillerat-Jeanneret, L.; Moitessier, N. Virtual Screening and Computational Optimization for the Discovery of Covalent Prolyl Oligopeptidase Inhibitors with Activity in Human Cells. *J. Med. Chem.* **2012**, *55*, 6306−6315.

(6) Ouyang, X.; Zhou, S.; Su, Ch. T. T.; Ge, Z.; Li, R.; Kwoh, C. K. CovalentDock: Automated Covalent Docking with Parameterized Covalent Linkage Energy Estimation and Molecular Geometry Constraints. *J. Comput. Chem.* **2013**, *34*, 326−336.

(7) Kawamura, S.; Unno, Y.; Tanaka, M.; Sasaki, T.; Yamano, A.; Hirokawa, T.; Kameda, T.; Asai, A.; Arisawa, M.; Shuto, S. Investigation of the Noncovalent Binding Mode of Covalent Proteasome Inhibitors around the Transition State by Combined

Use of Cyclopropylic Strain-Based Conformational Restriction and Computational Modeling. *J. Med. Chem.* **2013**, *56*, 5829−5842.
(8) Shokhen, M.; Traube, T.; Vijayakumar, S.; Hirsch, M.; Uritsky, N.; Albeck, A. Differentiating Serine and Cysteine Proteases Mechanism by New Covalent QSAR Descriptors. *ChemBioChem* **2011**, *12*, 1023−1026.
(9) Traube, T.; Vijayakumar, S.; Hirsch, M.; Uritsky, N.; Shokhen, M.; Albeck, A. EMBM - A New Enzyme Mechanism-Based Method for Rational Design of Chemical Sites of Covalent Inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 2256−2265.
(10) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−43.
(11) Raha, K.; Merz, K. M. A Quantum Mechanics-Based Scoring Function: Study of Zinc Ion-Mediated Ligand Binding. *J. Am. Chem. Soc.* **2004**, *126*, 1020−1021.
(12) Ciancetta, A.; Genheden, S.; Ryde, U. A QM/MM Study of the Binding of RAPTA Ligands to Cathepsin B. *J. Comput. Aided Mol. Des.* **2011**, *25*, 729−742.
(13) Dobeš, P.; Řezáč, J.; Fanfrlík, J.; Otyepka, M.; Hobza, P. Semiempirical Quantum Mechanical Method PM6-DH2X Describes the Geometry and Energetics of CK2-Inhibitor Complexes Involving Halogen Bonds Well, While the Empirical Potential Fails. *J. Phys. Chem. B* **2011**, *115*, 8581−8589.
(14) Fanfrlik, J.; Kolar, M.; Kamlar, M.; Hurny, D.; Ruiz, F. X.; Cousido-Siah, A.; Mitschler, A.; Řezáč, J.; Munusamy, E.; Lepšík, M.; et al. The Modulation of Aldose Reductase Inhibition by Halogen Bond Tuning. *ACS Chem. Biol.*, in press. DOI: 10.1021/Cb400526n.
(15) Söderhjelm, P.; Kongsted, J.; Ryde, U. Ligand Affinities Estimated by Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 1726−1737.
(16) Siegbahn, P. E. M.; Himo, F. Recent Developments of the Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. *J. Biol. Inorg. Chem.* **2009**, *14*, 643−651.
(17) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(18) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.
(19) Helten, H.; Schirmeister, T.; Engels, B. Model Calculations about the Influence of Protic Environments on the Alkylation Step of Epoxide, Aziridine and Thiirane Based Cysteine Protease Inhibitors. *J. Phys. Chem. A* **2004**, *108*, 7691−7701.
(20) Paasche, A.; Schirmeister, T.; Engels, B. Benchmark Study for the Cysteine−Histidine Proton Transfer Reaction in a Protein Environment: Gas Phase, COSMO, QM/MM Approaches. *J. Chem. Theory Comput.* **2013**, *9*, 1311−1319.
(21) Mladenovic, M.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. On the Origin of Stabilization of the Zwitterionic Resting State of Cysteine Proteases: A Theoretical Study. *J. Am. Chem. Soc.* **2008**, *130*, 8696−8705.
(22) Mladenovic, M.; Ansorg, K.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. Atomistic Insights into the Inhibition of Cysteine Proteases: First QM/MM Calculations Clarifying the Stereoselectivity of Epoxide-Based Inhibitors. *J. Phys. Chem. B* **2008**, *112*, 11798−11808.
(23) Mladenovic, M.; Junold, K.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. Atomistic Insights into the Inhibition of Cysteine Proteases: First QM/MM Calculations Clarifying the Regiospecificity and the Inhibition Potency of Epoxide- and Aziridine-Based Inhibitors. *J. Phys. Chem. B* **2008**, *112*, 5458−5469.
(24) Vicik, V.; Helten, H.; Schirmeister, T.; Engels, B. Rational Design of Aziridine-Containing Cysteine Protease Inhibitors with Improved Potency: Studies on Inhibition Mechanism. *ChemMedChem* **2006**, *1*, 1021−1028.
(25) Mladenovic, M.; Arnone, M.; Fink, R. F.; Engels, B. Environmental Effects on Charge Densities of Biologically Active Molecules: Do Molecule Crystal Environments Indeed Approximate Protein Surroundings? *J. Phys. Chem. B* **2009**, *113*, 5072−5082.
(26) Řezáč, J.; Fanfrlik, J.; Salahub, D.; Hobza, P. Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* **2009**, *5*, 1749−1760.
(27) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141−151.
(28) Řezáč, J.; Hobza, P. A Halogen-Bonding Correction for the Semiempirical PM6 Method. *Chem. Phys. Lett.* **2011**, *506*, 286−289.
(29) Lepšík, M.; Řezáč, J.; Kolář, M.; Pecina, A.; Hobza, P.; Fanfrlík, J. The Semiempirical Quantum Mechanical Scoring Function for In-Silico Drug Design. *ChemPlusChem* **2013**, *78*, 921−931.
(30) Stewart, J. J. P. *MOPAC2009*; Stewart Computational Chemistry: Colorado Springs, CO; http://OpenMOPAC.net.
(31) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(32) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799−805.
(33) Fanfrlík, J.; Bronowska, A. K.; Řezáč, J.; Přenosil, O.; Konvalinka, J.; Hobza, P. A Reliable Docking/Scoring Scheme Based on the Semiempirical Quantum Mechanical PM6-DH2 Method Accurately Covering Dispersion and H-Bonding: HIV-1 Protease with 22 Ligands. *J. Phys. Chem. B* **2010**, *114*, 12666−12678.
(34) Dobeš, P.; Fanfrlík, J.; Řezáč, J.; Otyepka, M.; Hobza, P. Transferable Scoring Function Based on Semiempirical Quantum Mechanical PM6-DH2 Method: CDK2 with 15 Structurally Diverse Inhibitors. *J. Comput.-Aid. Mol. Des.* **2011**, *25*, 223−235.
(35) Brahmkshatriya, P.; Dobeš, P.; Fanfrlík, J.; Rezáč, J.; Paruch, K.; Bronowska, A. K.; Lepšík, M.; Hobza, P. Quantum Mechanical Scoring: Structural and Energetic Insights into Cyclin-Dependent Kinase 2 Inhibition by Pyrazolo(1,5-*a*)Pyrimidines. *Curr. Comput. Aided Drug Des.* **2012**, *9*, 118−129.
(36) Peters, M. B.; Raha, K.; Merz, K. M. Quantum Mechanics in Structure-Based Drug Design. *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 370−379.
(37) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M. The Role of Quantum Mechanics in Structure-Based Drug Design. *Drug Discovery Today* **2007**, *12*, 725−731.
(38) Mucs, D.; Bryce, R. A. The Application of Quantum Mechanics in Structure-Based Drug Design. *Expert Opin. Drug Discovery* **2013**, *8*, 263−276.
(39) Muddana, H. S.; Gilson, M. K. Calculation of Host-Guest Binding Affinities using a Quantum-Mechanical Energy Model. *J. Chem Theory. Comput.* **2012**, *8*, 2023−2033.
(40) Kamel, K.; Kolinski, A. Assessment of the Free Binding Energy of 1,25-Dihydroxyvitamin D3 and its Analogs with the Human VDR Receptor Model. *Acta Biochim. Pol.* **2012**, *59*, 653−660.
(41) Stigliani, J. L.; Bernardes-Genisson, V.; Bernadou, J.; Pratviel, G. Cross-Docking Study on InhA Inhibitors: A Combination of Autodock Vina and PM6-DH2 Simulations to Retrieve Bioactive Conformations. *Org. Biomol. Chem.* **2012**, *10*, 6341−6349.
(42) Nagy, G.; Gyurcsik, B.; Hoffmann, E. A.; Koertvelyesi, T. Theoretical Design of a Specific DNA-Zinc-Finger Protein Interaction with Semi-Empirical Quantum Chemical Methods. *J. Mol. Graph. Model.* **2011**, *29*, 928−934.
(43) Mikulskis, P.; Genheden, S.; Wichmann, K.; Ryde, U. A Semiempirical Approach to Ligand-Binding Affinities: Dependence on the Hamiltonian and Corrections. *J. Comput. Chem.* **2012**, *33*, 1179−1189.
(44) Correnti, J. M.; Brindley, P. J.; Pearce, E. J. Long-Term Suppression of Cathepsin B Levels by RNA Interference Retards Schistosome Growth. *Mol. Biochem. Parasitol.* **2005**, *143*, 209−215.

(45) Sajid, M.; McKerrow, J. H.; Hansell, E.; Mathieu, M. A.; Lucas, K. D.; Hsieh, I.; Greenbaum, D.; Bogyo, M.; Salter, J. P.; Lim, K. C. Functional Expression and Characterization of Schistosoma Mansoni Cathepsin B and its Trans-Activation by an Endogenous Asparaginyl Endopeptidase. *Mol. Biochem. Parasitol.* **2003**, *131*, 65−75.
(46) Horn, M.; Jílková, A.; Vondrášek, J.; Marešová, L.; Caffrey, C. R.; Mareš, M. Mapping the Pro-Peptide of the Schistosoma Mansoni Cathepsin B1 Drug Target: Modulation of Inhibition by Heparin and Design of Mimetic Inhibitors. *ACS Chem. Biol.* **2011**, *6*, 609−617.
(47) Jílková, A.; Řezáčová, P.; Lepšík, M.; Horn, M.; Váchová, J.; Fanfrlík, J.; Brynda, J.; McKerrow, J. H.; Caffrey, C. R.; Mareš, M. Structural Basis for Inhibition of Cathepsin B Drug Target from the Human Blood Fluke, *Schistosoma mansoni*. *J. Biol. Chem.* **2011**, *286*, 35770−35781.
(48) Kerr, I. D.; Lee, J. H.; Farady, C. J.; Marion, R.; Rickert, M.; Sajid, M.; Pandey, K. C.; Caffrey, C. R.; Legac, J.; Hansell, E. Vinyl Sulfones As Antiparasitic Agents and a Structural Basis for Drug Design. *J. Biol. Chem.* **2009**, *284*, 25697−25703.
(49) Bieth, J.G. Theoretical and Practical Aspects of Proteinase Inhibition Kinetics. *Methods Enzymol.* **1995**, *248*, 59−84.
(50) Caffrey, C. R.; Hansell, E.; Lucas, K. D.; Brinen, L. S.; Alvarez Hernandez, A.; Cheng, J.; Gwaltney, S. L.; Roush, W. R.; Stierhof, Y.-D.; Bogyo, M. Active Site Mapping, Biochemical Properties and Subcellular Localization of Rhodesain, the Major Cysteine Protease of Trypanosoma Brucei Rhodesiense. *Mol. Biochem. Parasitol.* **2001**, *118*, 61−73.
(51) Roush, W. R.; Gwaltney, S. L.; Cheng, J. M.; Scheidt, K. A.; McKerrow, J. H.; Hansell, E. Vinyl Sulfonate Esters and Vinyl Sulfonamides: Potent, Irreversible Inhibitors of Cysteine Proteases. *J. Am. Chem. Soc.* **1998**, *120*, 10994−10995.
(52) Baici, A.; Schenker, P.; Wachter, M.; Ruedi, P. 3-Fluoro-2,4-dioxa-3-phosphadecalins as Inhibitors of Acetylcholinesterase. A Reappraisal of Kinetic Mechanisms and Diagnostic Methods. *Chem. Biodiversity* **2009**, *6*, 261−282.
(53) Horn, M.; Pavlík, M.; Dolečková, L.; Baudyš, M.; Mareš, M. Arginine-Based Structures are Specific Inhibitors of Cathepsin C. Application of Peptide Combinatorial Libraries. *Eur. J. Biochem.* **2000**, *267*, 3330−3336.
(54) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chem. Rev.* **2002**, *102*, 4639−4750.
(55) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; et al. *AMBER 10*; University of California: San Francisco, CA, 2008.
(56) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera − A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−12.
(57) Kolář, M.; Fanfrlík, J.; Lepšík, M.; Forti, F.; Luque, F. J.; Hobza, P. Assessing the Accuracy and Performance of Implicit Solvent Models for Drug Molecules: Conformational Ensemble Approaches. *J. Phys. Chem. B* **2013**, *117*, 5950−5962.
(58) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357−19363.
(59) Dapprich, S.; Komaromi, I.; Suzie Byun, K.; Morokuma, K.; Frisch, M. J. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives. *J. Mol. Struct. Theochem.* **1999**, *461*−*462*, 1−21.
(60) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
(61) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. Density Functional Theory Augmented with an Empirical Dispersion Term. Interaction Energies and Geometries of 80 Noncovalent Complexes Compared with Ab Initio Quantum Mechanics Calculations. *J. Comput. Chem.* **2007**, *28*, 555−569.
(62) Ehrlich, S.; Moellmann, J.; Reckien, W.; Bredow, T.; Grimme, S. System-Dependent Dispersion Coefficients for the DFT-D3 Treatment of Adsorption Processes on Ionic Surfaces. *ChemPhysChem* **2011**, *12*, 3414−3420.
(63) Bitzek, E.; Koskinen, P.; Gahler, F.; Moseler, M.; Gumbsch, P. Structural Relaxation Made Simple. *Phys. Rev. Lett.* **2006**, *97*, 170201−170205.
(64) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General AMBER Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(65) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(66) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(67) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(68) Hayik, S. A.; Dunbrack, R., Jr.; Merz, K. M. A., Jr. Mixed QM/MM Scoring Function to Predict Protein-Ligand Binding Affinity. *J Chem. Theory Comput.* **2010**, *6*, 3079−3091.
(69) Ishchenko, A. V.; Shakhnovich, E. I. Small Molecule Growth 2001 (SMoG2001): An Improved Knowledge-Based Scoring Function for Protein-Ligand Interactions. *J. Med. Chem.* **2002**, *45*, 2770−2780.
(70) Raha, K.; Merz, K. M. Large-Scale Validation of a Quantum Mechanics Based Scoring Function: Predicting the Binding Affinity and the Binding Mode of a Diverse Set of Protein−Ligand Complexes. *J. Med. Chem.* **2005**, *48*, 4558−4575.
(71) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44*, 3417−3423.
(72) Chen, Y. T.; Brinen, L. S.; Kerr, I. D.; Hansell, E.; Doyle, P. S.; McKerrow, J. H.; Roush, W. R. In Vitro and In Vivo Studies of the Trypanocidal Properties of WRR-483 against *Trypanosoma cruzi*. *PLOS Neglected Trop. Dis.* **2010**, *4*, 825e.
(73) Brinen, L. S.; Hansell, E.; Cheng, J.; Roush, W. R.; McKerrow, J. H.; Fletterick, R. J. A Target within the Target: Probing Cruzain's P1′ Site to Define Structural Determinants for the Chagas' Disease Protease. *Structure* **2000**, *8*, 831−840.
(74) Kerr, I. D.; Wu, P.; Marion-Tsukamaki, R.; Mackey, Z. B.; Brinen, L. S. Crystal Structures of TbCatB and Rhodesain, Potential Chemotherapeutic Targets and Major Cysteine Proteases of Trypanosoma Brucei. *PLOS. Neglected Trop. Dis.* **2010**, *4*, e701.
(75) *PyMOL Molecular Graphics System*, version 1.5.0.4. Schrödinger, LLC.
(76) Kolář, M.; Fanfrlík, J.; Hobza, P. Ligand Conformational and Solvation/Desolvation Free Energy in Protein−Ligand Complex Formation. *J. Phys. Chem. B* **2011**, *115*, 4718−4724.
(77) Sadiq, S. K.; De Fabritiis, G. Explicit Solvent Dynamics and Energetics of HIV-1 Protease Flap Opening and Closing. *Proteins* **2010**, *78*, 2873−85.
(78) Pecina, A.; Přenosil, O.; Fanfrlík, J.; Řezáč, J.; Granatier, J.; Hobza, P.; Lepšík, M. On the Reliability of the Corrected Semiempirical Quantum Chemical Method (PM6-DH2) for Assigning the Protonation States in HIV-1 Protease/Inhibitor Complexes. *Collect. Czech. Chem. Commun.* **2011**, *76*, 457−479.
(79) Zhou, H.-X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **2009**, *109*, 4092−4107.
(80) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem.—Eur. J.* **2012**, *18*, 9955−9964.

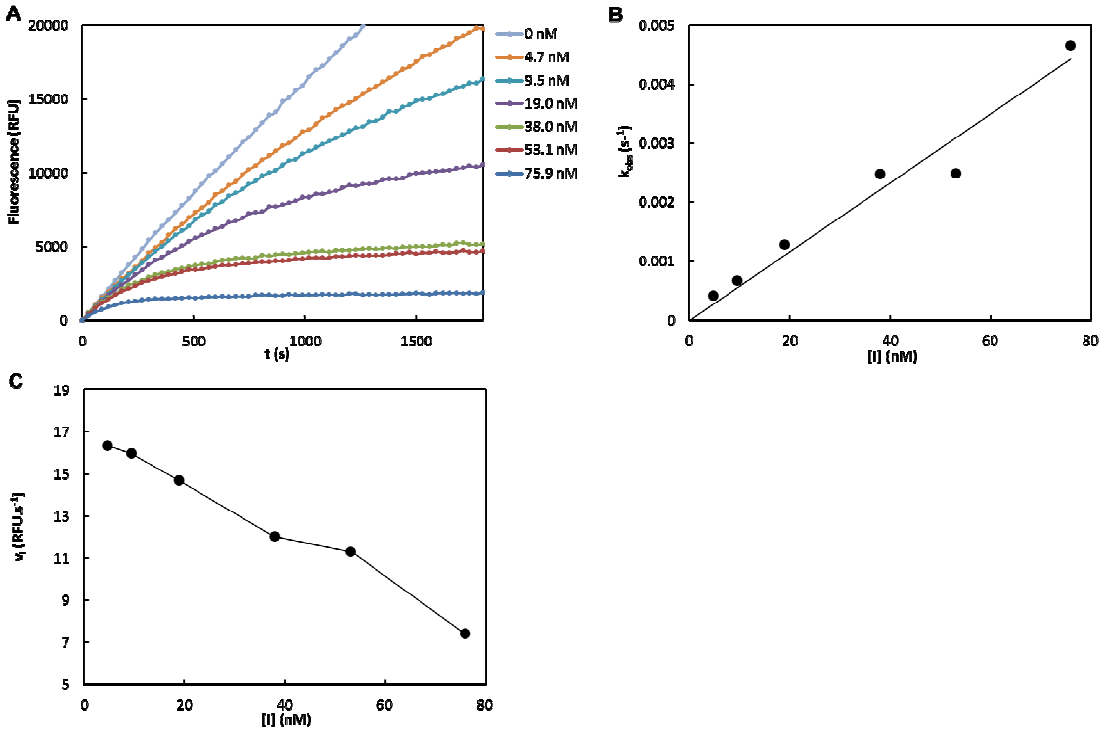
Supporting Information: Quantum Mechanics-Based Scoring Rationalizes Irreversible Inactivation of Parasitic *Schistosoma mansoni* Cysteine Peptidase by Vinyl Sulfone Inhibitors

Jindřich Fanfrlík,[1] Pathik S Brahmkshatriya,[1] Jan Řezáč,[1] Adéla Jílková[1], Martin Horn[1], Michael Mareš,[1] Pavel Hobza,[1,2*] and Martin Lepšík[1*]

[1]Institute of Organic Chemistry and Biochemistry, v.v.i. and Gilead Sciences and IOCB Research Center, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic, tel.: (+420) 220 410 319, fax: (+420) 220 410 320, E-mail: *lepsik@uochb.cas.cz* and *hobza@uochb.cas.cz*

[2]Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc, 771 46 Olomouc, Czech Republic

Supplementary Figure S1. Enzyme inhibition kinetics for SmCB1 and compound **K11777** used as a representative vinyl sulfone inhibitor. A) Time course of the hydrolysis of fluorogenic substrate Cbz-Phe-Arg-AMC in the presence and absence of **K11777**. The reaction mixture contained 40 pM SmCB1, 16 µM Cbz-Phe-Arg-AMC (K_m = 25.4 µM), 0-75 nM **K11777**, and the assay buffer (0.1 M Na-acetate pH 5.5, containing 3 mM DTT and 0.125 % PEG 1500). B) Linear relation of the used inhibitor concentrations and the observed first-order inactivation constants (k_{obs}) determined from the A plot. C) Dependence of v_i on the inhibitor concentration. Measurements were performed in duplicates. Standard errors were within 5% of the means.



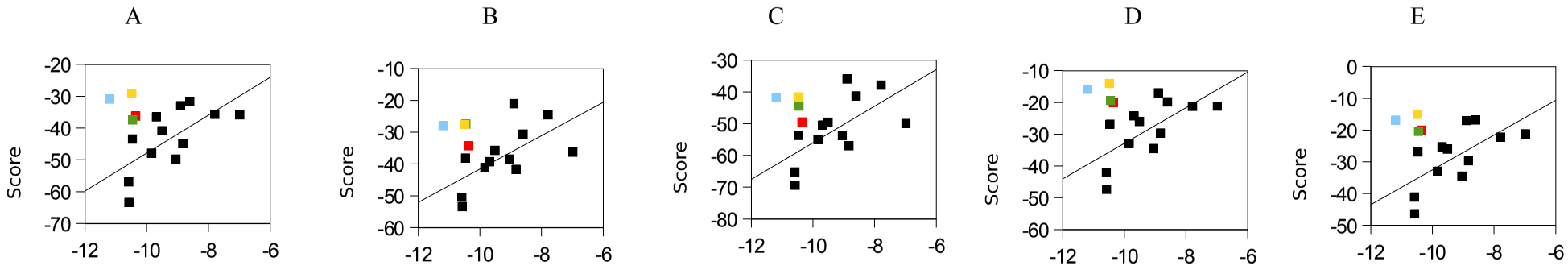
Supplementary Table S1. Summary of calculated scores and their terms for complexes of vinyl sulfone inhibitory compounds with SmCB1. The 'free' energy difference between the covalent and non-covalent complexes ($\Delta G'_{cov}$), the gas-phase interaction energy (ΔE_{int}), the interaction desolvation free energy ($\Delta\Delta G_{solv}$), the change of the conformational 'free' energy of the ligand and protein ($\Delta G'^{w}_{conf}(P,L)$) and the entropy change upon binding ($T\Delta S_{int}$) in kcal/mol.

Compound	ΔG_{cov}	ΔE_{int}	$\Delta\Delta G_{solv}$	$\Delta G^{w}_{conf}(L)$	$\Delta G^{w}_{conf}(P)$	$-T\Delta S_{int}$	$Score_{nc}$	Score
WRP-286	-16.9	-53.4	25.4	11.0	14.0	3.0	0.0	-16.9
K11017	-25.1	-128.4	86.3	8.3	14.8	2.0	-17.0	-42.1
K11002	-27.6	-126.2	84.3	6.0	16.0	0.0	-19.8	-47.3
WRR-282	-19.9	-55.6	32.9	12.5	14.1	1.0	4.9	-15.0
K11777	-26.2	-46.9	19.4	10.3	15.6	0.0	-1.7	-27.8
WRR-285	-18.0	-51.5	24.1	7.1	17.0	1.0	-2.3	-20.4
K11747	-25.6	-51.1	27.3	13.3	15.2	0.0	4.6	-21.0
AR-198048	-23.2	-55.8	23.0	7.2	13.9	1.0	-10.7	-33.9
WRR-284	-20.0	-53.8	22.3	14.0	11.2	1.0	-5.3	-25.2
AR-198049	-22.4	-46.2	18.1	8.7	13.9	1.0	-4.5	-27.0
WRR-145	-20.3	-112.2	78.8	3.9	15.3	0.0	-14.2	-34.5
WRR-499	-19.3	-40.0	23.4	3.0	14.9	0.0	1.3	-18.0
K11006	-23.4	-25.1	23.4	12.1	15.2	2.0	27.5	4.2
WRR-483	-19.7	-51.3	-10.4	9.7	10.7	3.0	-38.2	-58.0
WRR-283	-21.1	-37.1	26.7	2.3	13.4	2.0	7.3	-13.8
WRR-347	-17.2	-61.6	18.3	14.1	13.7	3.0	-12.6	-29.8

Supplementary Table S2. The count of rotatable bonds used for the estimation of the entropy change upon binding ($T\Delta S_{int}$). Each rotatable bond is assigned a penalty of 1 kcal/mol. A) the relative number of rotatable bonds of the ligand that become hindered upon complex formation, B) the total relative number of rotatable bonds in the ligands and C) the relative number of the rotatable bonds from the protein side chains that become hindered upon complex formation.

Compound	$-T\Delta S_{int}$		
	A	B	C
WRR-286	3	4	0
K11017	2	2	1
K11002	0	0	1
WRR-282	1	2	0
K11777	0	1	1
WRR-285	1	2	0
K11747	0	1	1
AR-198048	1	2	1
WRR-284	1	2	0
AR-198049	1	2	1
WRR-145	0	0	0
WRR-499	0	1	1
K11006	2	2	0
WRR-483	3	4	4
WRR-283	2	3	0
WRR-347	3	4	1

Supplementary Graph S1 Several variants of the total covalent score plotted against the experimental binding data expressed as RTln(IC_{50}). A) $\Delta G'_{conf}{}^{w}(P)$ not considered (R^2 = 0.46 and PI = 0.67), B) $\Delta G'_{conf}{}^{w}(L)$ not considered (R^2 = 0.40 and PI = 0.62), C) both $\Delta G'_{conf}{}^{w}(P)$ and $\Delta G'_{conf}{}^{w}(L)$ not considered, (R^2 = 0.41 and PI = 0.64), D) $T\Delta S_{int}$ determined by the total number of rotatable bonds in the ligand (R^2 = 0.46 and PI = 0.68), E) $T\Delta S_{int}$ determined as the rotatable bonds from the protein side chains and the ligand that become hindered upon complex formation (R^2 = 0.45 and PI = 0.70). All energies are in kcal/mol.



5.4. Publikace č. 4: Activation Route of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Structural Map with a Glycosaminoglycan Switch

5.4.1. Souhrn

Tato práce je připravena ve formě rukopisu pro revizní řízení v časopise Structure.

Cílem práce byla biochemická a strukturní analýza procesu aktivace zymogenu SmCB1. SmCB1 je syntetizován jako neaktivní zymogen (proenzym), jehož aktivita je blokována N-terminálním propeptidem fungujícím jako přirozený intramolekulární inhibitor aktivního místa. Při aktivaci je propeptid proteolyticky odštěpen za vzniku aktivního enzymu. Zymogen SmCB1 lze aktivovat působením asparaginylendopeptidasy ze *S. mansoni* (SmAE) [14]. V rámci této publikace byl identifikován a popsán nový mechanismus autokatalytické aktivace (autoaktivace) indukované sulfatovanými polysacharidy (SP).

Autoaktivace SmCB1 bez přítomnosti SP probíhá v pH 4-5 a vede k tvorbě neaktivního intermediátu, ve kterém je odštěpena jen část propeptidu. Krystalová struktura tohoto intermediátu ukazuje, že aktivní místo zůstává stále blokováno zbytkovým propeptidem. V přítomnosti SP je v pH 4-5 propeptid kompletně proteolyticky odstraněn a vzniká aktivní enzym. V průběhu jeho tvorby vzniká několik aktivních intermediátů, ve kterých je již aktivní místo přístupné substrátu. Několika experimentálními přístupy byla analyzována molekularita autoaktivační reakce, která kombinuje monomolekulární a bimolekulární děje. Studie strukturních preferencí SP prokázala, že schopnost efektivně indukovat autoaktivaci závisí jak na typu SP, tak na jeho velikosti. Proteolytická aktivace SmCB1 pomocí SmAE probíhá v rozmezí pH 5.0-6.5. Významné bylo zjištění, že aktivace katalyzovaná SmAE je působením SP potlačena. Lze tak předpokládat, že ve fyziologickém prostředí je aktivace SmCB1 komplexně regulovaná pomocí pH a interakce s SP.

Součástí studie je určení krystalových struktur 3 molekulárních forem SmCB1 v aktivační dráze, zymogenu (PDB kód: 4I04), neaktivního intermediátu (PDB kód: 4I05) a zralého enzymu (PDB kód: 4I07). Struktura zymogenu umožnila analýzu interakce propeptidu s katalytickou doménou a potvrdila přítomnost unikátního α-helixu, který byl predikován v propeptidu pomocí homologního modelování v publikaci č. 1. Tento α-helix obsahuje sekvenční motiv vázající heparin (“HB motiv”), který umožňuje interakci s SP. To bylo potvrzeno cílenými mutacemi aminokyselin v motivu, které zabránily kompletnímu odštěpení propeptidu.

Můj podíl na práci zahrnoval (1) rekombinantní expresi, aktivaci a purifikaci SmCB1 a jeho mutantních forem, (2) krystalizaci tří forem SmCB1 z aktivační dráhy, analýzu těchto krystalografických struktur (pomocí programů CONTACT a Pymol), jejich interpretaci a molekulární grafiku (pomocí programu Pymol), (3) biochemické aktivační experimenty analyzované pomocí SDS-PAGE a (4) účast na přípravě manuskriptu.

5.4.2. Publikace č. 4

Jílková, A., Horn, M., Řezáčová, P., Marešová, L., Fajtová, P., Brynda, J., Vondrášek, J., Caffrey, C. R. a Mareš, M.

Activation Route of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Structural Map with a Glycosaminoglycan Switch

Rukopis připraven k podání do časopisu *Structure*

Activation route of the *Schistosoma mansoni* cathepsin B1 drug target: structural map with a glycosaminoglycan switch

Adéla Jílková[1,2,5], Martin Horn[1,5], Pavlína Řezáčová [1,3], Lucie Marešová[1], Pavla Fajtová[1], Jiří Brynda[1,3], Jiří Vondrášek[1], James H. McKerrow[4], Conor R. Caffrey[4], and Michael Mareš[1*]

[1]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, 16610 Prague, Czech Republic,

[2]Department of Biochemistry, Faculty of Science, Charles University in Prague, 12843 Prague, Czech Republic,

[3]Department of Structural Biology, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

[4]Center for Discovery and Innovation in Parasitic Diseases, Department of Pathology, University of California San Francisco, San Francisco, CA 94158, USA

[5]These authors contributed equally to this work.

Running title: Activation of *Schistosoma mansoni* cathepsin B1

[*]Correspondence: Michael Mareš, Flemingovo n. 2, 16610 Prague, Czech Republic, (+420) 220183358, Fax: (+420) 220183578, E-mail: mares@uochb.cas.cz

SUMMARY

Cathepsin B1 (SmCB1) is a digestive protease found in the gut of the human blood fluke *Schistosoma mansoni*. It is a potential drug target for the treatment of schistosomiasis, a parasitic disease that afflicts over 200 million people worldwide. SmCB1 is synthesized as an inactive zymogen in which the pro-peptide operates as an intra-molecular inhibitor by blocking the active site. We investigated the activation and processing of the zymogen by which the pro-peptide is proteolytically removed and the regulatory role of sulfated polysaccharides (SPs) in this process. Crystal structures of three molecular forms of SmCB1 along the activation pathway were determined, namely the zymogen, an activation intermediate with a partially cleaved pro-peptide and the mature enzyme. We demonstrate that SPs are essential to autocatalytic activate SmCB1 as they interact with a specific heparin-binding domain in the pro-peptide. An alternative activation route for the SmCB1 zymogen is mediated by an *S. mansoni* asparaginyl endopeptidase which is down-regulated by SPs, indicating that SPs act as a molecular switch between both activation mechanisms.

INTRODUCTION

Schistosomiasis (bilharzia) is a chronic infectious disease caused by trematode blood flukes of the genus *Schistosoma*. The disease infects over 200 million people and is second only to malaria as a global parasitic health problem (Colley et al., 2014). *Schistosoma mansoni*, present in more than 50 countries, is the most widespread of the human-infecting schistosomes. Treatment and control of schistosomiasis now relies on just one drug, and, due to the growing concern over resistance, there is pressure to identify new schistosomal protein targets and chemotherapeutic strategies (Caffrey, 2007; Thetiot-Laurent et al., 2013).

Adult schistosomes live in the cardiovascular system and host blood proteins are a primary source of nutrients required for parasite growth, development and reproduction. In the schistosome gut, a network of proteases performs the digestion of host proteins (Delcroix et al., 2006; Brindley et al., 1997). These enzymes represent potential intervention targets. The current research focuses on *S. mansoni* cathepsin B1 (SmCB1), which is a central digestive protease due to its high abundance and complex proteolytic activity that comprises both endopeptidase and exopeptidase (peptidyl dipeptidase) modes of action (Sajid et al., 2003; Jilkova et al., 2011). SmCB1 has been validated in a murine model of schistosome infection as a molecular target for therapy with a vinyl sulfone inhibitor (Abdulla et al., 2007). Recently, we determined the crystal structure of SmCB1 in complex with vinyl sulfone inhibitors to describe the binding mode and correlated their inhibition of SmCB1 with anti-schistosomal potency (Jilkova et al., 2011).

Members of the clan CA, family C1 (papain-like) cysteine peptidases, including cathepsins B, are synthesized as inactive zymogens (pro-cathepsins), in which the pro-peptide acts as an intramolecular inhibitor by blocking the active site. Activation to the mature, catalytically active

form occurs upon proteolytic removal of the N-terminal pro-peptide (also termed the “activation peptide”) and this can be autocatalytic or involve another *trans*-activating protease (for review, see e.g. Turk et al., 2012). For SmCB1 expressed in the yeast, *Pichia pastoris*, the zymogen does not autoprocess but was *trans*-activated by *S. mansoni* asparaginyl endopeptidase/legumain (SmAE; Sajid et al., 2003), a clan CD cysteine peptidase (Caffrey et al., 2000). However, a later *in vivo* study demonstrated that SmAE was not absolutely required for SmCB1 activation (Krautz-Peterson et al., 2008), indicating that a different mechanism may also contribute to the maturation of SmCB1. We propose that this alternative mechanism is an auto-activation event mediated by sulfated polysaccharides, such as glycosaminoglycans.

Sulfated polysaccharides (SPs) have a strong negative charge and are widely distributed in tissues participating in the regulation of a number of protein functions and interactions (for example reviews, see Raman et al., 2005;Gandhi et al., 2008). SPs are also reported to modulate several mammalian cysteine cathepsins. For example, they induce the collagenolytic activity of cathepsin K (Li et al., 2002;Cherney et al., 2011), stabilize cathepsin B at neutral pH (Almeida et al., 2001) and facilitate the auto-activation of cathepsins B, L and S (Caglic et al., 2007;Vasiljeva et al., 2005;Fairhead et al., 2008).

Here, we show that SPs do more than just modulate the auto-activation of cysteine cathepsins, but are critically required for the auto-activation of SmCB1. Moreover, they act as a molecular switch in the SmCB1 activation pathway by alternating between SP-mediated auto-activation and SP-down-regulated *trans*-activation. The bi-modal system described here for SmCB1 might also be relevant to other digestive cathepsins B of parasitic metazoa, including the trematodes *Schistosoma*, *Trichobilharzia*, and *Fasciola* spp. (Horn et al., 2011;Beckham et al., 2006). We provide new insights into the mechanism of activation of cysteine cathepsins by employing a complete set of structures along the activation pathway, including the zymogen and mature forms, and a processing intermediate not previously investigated.

RESULTS

Determination of crystal structures

A recombinant zymogen of SmCB1 was produced in the *Pichia pastoris* expression system (Sajid et al., 2003; Horn et al., 2011). The inactive intermediate (iCB^{inact}) was obtained from the zymogen by autocatalytic processing, which removed 38 residues from the pro-peptide generating N-terminal Leu39 (the SmCB1 zymogen numbering is used throughout the paper). The enzymatically active mature form of SmCB1 was prepared from the zymogen by *trans*-activation (Sajid et al., 2003; Horn et al., 2011), which removed the full-length, 69 residue-long pro-peptide to reveal an N-terminal Val70. For crystallography of the zymogen, an inactivated mutant was constructed with mutated active-site cysteine (Cys100Ser) to prevent its autoprocessing.

The structures of the Cys100Ser zymogen, the intermediate iCB^{inact} and the mature enzyme were determined by molecular replacement using the structure of the mature SmCB1 complexed with the inhibitor CA-074 (Jilkova et al., 2011). The zymogen crystallized in the triclinic space group P1 with four molecules per asymmetric unit. All four molecules are very similar; the RMSD for the superposition of the four protein molecule backbones onto each other range from 0.2 to 0.3 Å. The structure was refined using data to a resolution of 1.95 Å. N-terminal residues 1–8, 1–7, 1–5 and 1–5 and residues 179–189, 181–186, 179–190 and 180–189 within the occluding loop could not be located in the electron density map in molecules A, B, C and D, respectively, and were thus not included in the final model (these two regions are disordered also in several zymogen structures of mammalian cathepsins B). For presentation purposes, the missing residues were computationally modeled for the most complete molecule B (Figure 1).

The intermediate iCB^{inact} crystallized in the monoclinic space group C2 with one molecule per asymmetric unit. The structure was refined using data to a resolution of 1.9 Å. The first 8 N-terminal residues in the molecule could not be located in the electron density map and were thus not included in the final model, which starts with Gly47.

The mCB crystallized in the orthorhombic space group $P2_12_12_1$ with one molecule per asymmetric unit. The structure was refined using data to a high resolution of 1.3 Å. No electron density was found for the N-terminal residue so the final crystallographic model starts with Glu71.

Structural characterization of the SmCB1 zymogen and its processing sites

In the SmCB1 zymogen, the enzyme core adopts the classic fold of papain-family proteases in which the molecule is divided into L and R domains with the active-site cleft located in between (Figure 1A). The active site contains the catalytic residues Cys100, His270 and Asn290 and is blocked by the pro-peptide, which binds along the cleft in the opposite direction compared to substrates. This inhibitory mechanism and the overall architecture of the pro-peptide resemble cathepsin B-type pro-peptides; however, the secondary structure pattern of the SmCB1 pro-peptide incorporates an additional α-helix, α3p (Figure S1). The interface between the pro-peptide and the enzyme core contains a net of interactions (listed in Table S1), and the pro-peptide is further stabilized by a network of internal contacts between residues within the pro-peptide (Table S2).

The pro-peptide chain (residues 8–69) wraps around the enzyme core, forming three topological regions located on the top, front and bottom of the molecule (Figure 2E): (i) a hairpin structure formed by the α1p helix (residues 13–22) and an antiparallel β1p strand (residues 28–30), (ii) the α2p helix (residues 38–44) followed by the active site-binding segment, and (iii) from the non-conserved α3p helix (residues 52–60) to the pro-peptide/mature N-terminus junction.

The α1p-β1p hairpin segment of the pro-peptide interacts with a large “propeptide-binding loop” (PBL, residues 247–265) of the enzyme core (Figure 2A). The majority of the contacts are formed by residues 27–31 from β1p, including hydrogen bonds with a short antiparallel part of the PBL (Tyr254–Tyr259); Trp28 is also a part of an aromatic cluster along with two central residues, Tyr254 and Tyr259, of the PBL.

Pro-peptide residues 33–44 follow a crevice located between the “occluding loop” (residues 178–192) and the PBL of the enzyme core (Figure 2B). Residues 33–37 are organized into a kink and residues 38–44 into the α2p helix. The interface interactions are provided mostly by hydrogen bonds and salt bridges of Arg35 from the kink and Asp41 and Ile44 from α2p. No direct hydrogen bonding was found between the pro-peptide and the occluding loop, but Asp41 in α2p is interconnected via the enzyme-core residue Arg96 with Lys192 and Tyr194 of the occluding loop (Figure 1B). The occluding loop is in “open” conformation, like in other cathepsin B-type zymogens, with a highly flexible tip (residues 181–186 have a poorly defined electron density). In this region, there are significant local changes when comparing SmCB1 with other cathepsin B-type zymogens, in which the pro-peptides are positioned closer to the occluding loop and their interactions are maintained by a set of hydrogen bonds. The cleavage site (Ser38-Leu39) that is processed to yield the intermediate form iCB^{inact} is located on the rim of the α2p helix (Figure 1C). The surrounding region (His37-Asp40) forms no hydrogen bonds with the enzyme core and is exposed to the solvent; the accessibility of this surface region arises from the flexibility of the adjacent occluding loop.

The pro-peptide occupies the substrate-binding cleft via residues 45–49, which enter the S2’-S3 subsites in the SmCB1 zymogen (Figure 2C). The S2’ subsite binds Gln45, forming hydrogen bonds with Gln94 and Trp292, and is in contact with the catalytic His270. The S1’ subsite is occupied by the Met46 hydrogen, bonded to Gln94 and Ser100, the latter of which was mutated from the catalytic Cys residue. Gly47 enters the S1 subsite and interacts via hydrogen bonds with Gly144. Ala48 and Arg49 are in the S2 and S3 subsites, respectively, where they form a total of three hydrogen bonds.

The non-conserved α3p helix contains a heparin-binding motif (based on the arginine pattern in the LRRTRRP sequence, residues 55–61), which is absent in other structurally investigated cathepsin B zymogens (Horn et al., 2011). The α-helix makes only a few contacts with the enzyme core through hydrogen bonds formed by Arg56 and Arg60, however, it is internally stabilized by hydrogen bonds between the solvent-exposed Arg59 and Arg57 of the motif and the adjacent acidic residues of the pro-peptide (Arg57-Asp54, Arg59-Glu51-Arg60) (Figure 1C and 2D). This arrangement of contacts and the position of α3p protruding above the enzyme-core surface provide a target for glycosaminoglycans; such an interaction will most likely result in conformational changes in this pro-peptide region.

Downstream of α3p is the cleavage site Asn69-Val70, which is processed to yield the mature enzyme. The processing region (residues 64–72) is weak on the hydrogen bond forming at the interface between the pro-peptide and the enzyme core (Figure 1C); the interaction is maintained by several hydrophobic residues oriented against the core while an array of four Asp and Asn residues is exposed to the solvent, thus making this stretch easily accessible to proteolysis.

Structural characterization of the SmCB1 intermediate and mature forms

The mature form of SmCB1 is derived from the zymogen by proteolytic removal of the full-length pro-peptide (Figure 1A). The RMSD for the superposition of the main-chain atoms of the zymogen and the mature enzyme is 1.1 Å, and the major difference in both protein backbones is within the occluding loop. Its "open" conformation in the zymogen imposed by the pro-peptide is changed to the "closed" conformation after pro-peptide removal (Figures 1A and 1B). In contrast to the zymogen, the tip of the occluding loop in the mature enzyme is well defined in electron density (residues 178–189) due to stabilization, which is provided by 13 interdomain hydrogen bonds with the core (Figure 1B). The largest movement (~10 Å) in the occluding loop during open-to-closed transition was found for His180. In the mature enzyme, two conserved residues His180 and His181 are oriented into the active site, where they restrict access to the primed region and can interact with the C-terminal carboxyl group of the substrate P2' residue (Jilkova et al., 2011); the histidines are hydrogen-bonded with Asp93 and Leu252. The occluding loop is further rearranged between residues 188 and 194 (~6 Å movement for Ser191/Gly190), and this conformation is stabilized by a net of hydrogen bonds from Tyr194, Pro188 and Cys189 to the core residue Arg96. The active site of the non-complexed mature enzyme resembles that determined in the complexes with peptidomimetic inhibitors (Jilkova et al., 2011), but the absence of a ligand results in higher flexibility of His270 as indicated by a double conformation, of which only one appears in the complexes. The catalytic Cys100 was found to be oxidized.

The intermediate iCB^{inact} is a structural chimera with combined features of the SmCB1 zymogen and the mature enzyme (Figure 1A). It is derived from the zymogen by proteolytic removal of the first half of the pro-peptide (residues 1–38); the newly formed N-terminal segment 39–46 is flexible and the electron density only becomes defined downstream of Gly47. The residual pro-peptide interacts with the substrate-binding cleft by three residues, Gly47, Ala48 and Arg49, located in S1, S2 and S3 subsites, respectively. The majority of iCB^{inact} pro-peptide interactions in the active site as well as of downstream pro-peptide regions, including the α3p helix and the pro-peptide/mature N-terminus junction, are similar to the zymogen. Likewise, the enzyme core of iCB^{inact} is highly similar to that in the mature enzyme with the RMSD for the superposition of the main-chain atoms of 0.4 Å. Specifically, the occluding loop has an analogous shape and position (a few discrepancies are within 1.5

Å), and interdomain interactions except for the missing hydrogen bonds between His181 and Leu252, and between Thr182 and Lys255 (Figure 1B).

Sulfated polysaccharides are essential for the auto-activation of SmCB1

The autocatalytic activation of the SmCB1 zymogen was investigated using two different assays: (1) SDS-PAGE visualization of the processing forms generated by autocatalytic proteolysis and (2) kinetic measurement of the proteolytic activity of SmCB1 with the fluorogenic peptide substrate Z-Phe-Arg-AMC. The effect of sulfated polysaccharides on the activation process was probed with dextran sulfate (DS) as a model glycosaminoglycan-like polysaccharide. In the absence of DS, the autoprocessing of the zymogen (37 kDa) was effective at acidic pH, at pH 5.0 and below, and resulted in an intermediate form of 32 kDa, which did not yield any proteolytic activity (Figures 3A and 3B). In the presence of DS, the zymogen was converted in the same pH range into mature SmCB1 (28 kDa) via an intermediate species (~30–31 kDa) (Figure 3A). This was associated with the generation of proteolytic activity that was stable at pH 5.0 (and declined at lower pH due to auto-degradation demonstrated on SDS-PAGE). At pH 5.5 with DS, the inactive 32 kDa intermediate was produced (Figures 3A and 3B).

The fluorescent activity-based probe Green-DCG-04 (Greenbaum et al., 2002) was applied to label the active site of the variously processed forms of SmCB1 which could then be visualized by SDS-PAGE (Figure 3C). The mature enzyme was intensively labeled, whereas the zymogen did not show a significant signal. Regarding the intermediates, the 30–31 kDa form but not the 32 kDa form interacted with the probe, indicating that only the former contains an accessible active site.

The cleavage sites in the SmCB1 pro-peptide during the autocatalytic processing were identified using N-terminal protein sequencing; their localization is presented in the 3D model in Figure 4. This cleavage map helps explain the difference in the activity status of both intermediates. The inactive 32 kDa-intermediate corresponds to the crystallographically characterized form iCB^{inact} (Figure 1) and is derived by the cleavage of the Ser38-Leu39 bond; this form keeps the pro-peptide region intact with Met46 and Gly47 blocking the S1' and S1 pockets of the active site. In contrast, the active 30–31 kDa-intermediate (iCB^{act}) has a heterogeneous N-terminus produced by cleavages of the Met46-Gly47 bond and in its close vicinity, thus opening the catalytic site for substrate entry.

To analyze the cleavage pattern in the pro-peptide in detail, synthetic peptides derived from the pro-peptide structure were used as a substrate for the hydrolysis catalyzed by mature SmCB1 (Figures 4 and S2). Interestingly, we found a striking difference in the distribution of the cleavage sites identified in the peptides digested in the presence and absence of DS. Whereas both cleavage patterns in the N-terminal half of the pro-peptide (upstream of Met46-Gly47) are rather similar, the C-terminal half (downstream of Met46-Gly47) was fragmented

only in the presence of DS. Residues 55–61 close to the C-terminus of the pro-peptide form the heparin-binding motif (Figure 1), and a synthetic 22-peptide bearing this motif was fragmented in the presence of DS at multiple sites, including the Asn69-Val70 bond separating the pro-peptide and mature sequence. However, the substitution of positively charged arginine residues of the motif with citrulline isosteres that possess uncharged side chains completely abolished the DS-dependent fragmentation (peptides [53–74] and [53–74]Cit_4, Figure S2). Similarly, a deletion of the heparin-binding motif in the peptide [62–74] prevented the DS-dependent fragmentation (Figure S2). It should be mentioned that under these experimental conditions DS did not increase the enzymatic activity of mature SmCB1 towards the SmCB1 substrate Z-Phe-Arg-AMC (data not shown). This suggests that DS interacts with the heparin-binding motif and increases accessibility and susceptibility to proteolysis in the same parts of the pro-peptide that are cleaved during auto-activation.

In conclusion, a sulfated polysaccharide was found to be essential for the auto-activation processing of the inactive SmCB1 zymogen to the mature active enzyme. In its absence, the auto-activation is not productive and leads to incompletely processed pro-peptide still obstructing the catalytic site.

Autoprocessing of SmCB1 involves unimolecular and bimolecular steps

The molecularity of the autocatalytic processing of SmCB1 was examined by several approaches. First, the effect of the zymogen concentration on the autoprocessing was analyzed (Figure 5A). The rate of the processing reaction was found to be concentration-independent in the absence of DS, whereas it was concentration-dependent in the presence of DS, being accelerated by higher zymogen concentration. This suggested that the critical process in the formation of the inactive intermediate without DS has a unimolecular (intramolecular) character, whereas the generation of active forms mediated by DS, i.e., the active intermediate and mature enzyme, includes bimolecular (intermolecular) reactions.

Second, the active-site inhibitor E-64 was applied in the processing regimes with/without DS (Figure 5C). In the absence of DS, the inhibitor only had a marginal effect on the production of the intermediate; in the presence of DS, however, the production of mature enzyme was completely abolished, resulting in the accumulation of the intermediate species. E-64 did not effectively inhibit the formation of the intermediates; this implies that it had limited access to the active site of the SmCB1-processed forms involved in this unimolecular proteolytic event. In contrast, the bimolecular step sensitive to E-64 was driven by the SmCB1 processed forms with an open active site.

Third, the ability of the mature enzyme to participate in the intermolecular mechanism of autoprocessing was investigated (Figure 5B). In this experiment, the active-site mutant of the SmCB1 zymogen (C100S zymogen, Figure 1), which is incapable of autoprocessing, was tested as a substrate of the mature enzyme added in a catalytic amount. The C100S zymogen

was processed only in the presence of DS to the forms of intermediate and mature size (Figures 5B and S3), indicating that the active mature enzyme is competent to function as the catalytic species in the bimolecular auto-activation of the SmCB1 zymogen.

Fourth, additional evidence for the bimolecularity of the auto-activation was provided by the experiment in the presence of glycerol, which slows down molecular diffusion by increasing the solution viscosity. In the presence of 10–20% glycerol, a suppression of the DS-mediated conversion of the zymogen to the mature enzyme (Figure S4A) was observed together with a deceleration of the sigmoidal time course of the activity (Figure S4B), both indicating the bimolecularity of the reaction.

In conclusion, the initiation of the zymogen autoprocessing involves a critical rate-limiting intramolecular step that was demonstrated in the absence of DS and resulted in a partial removal of the pro-peptide forming the intermediate form. The full auto-activation mediated by DS via the intermediate to mature enzyme is an intermolecular process, which is accelerated through catalysis by the generated mature enzyme.

The heparin-binding motif in the pro-peptide is necessary for the correct maturation of SmCB1 mediated by sulfated polysaccharides

The SmCB1 pro-peptide contains a consensus sequence for the binding of heparin (and related sulfated polysaccharides); this XBBXBBX motif (where B is a basic amino acid and X is a hydropathic amino acid) is located at residues 55–61 (LRRTRRP) (see Figures 1 and 4). To study its role in the auto-activation of the SmCB1 zymogen, we constructed SmCB1 mutants in which the function of the motif was abolished by substituting two Arg residues (R57/R59) with Ala or Asn. Both heparin-binding motif mutants were produced as correctly folded zymogens that could be *trans*-activated by legumain into mature enzymes with activities comparable to the mature enzyme derived from the zymogen bearing an intact heparin-binding motif (see Experimental procedures).

Figure 6 shows the acidic autoprocessing of the motif mutant zymogens. In the absence of DS, the mutants were converted into the intermediate form cleaved at the same position as the control zymogen with the intact heparin-binding motif (as demonstrated by N-terminal sequencing, Figure S3), although with a higher efficiency. In the presence of DS, there is a striking difference between the processing patterns of the mutant vs. control zymogens. The motif mutants were processed into a set of intermediate forms (at least 4 species) under the conditions in which the control zymogen was converted into the mature enzyme via the active intermediate (Figures 6 and S3). The N-terminal sequencing of the intermediates derived from the motif mutants revealed that some of them are processed at and downstream from the region in the pro-peptide that blocks the catalytic site (Figure S3). This is in line with the enzymatic activity generated in the processing mixtures (data not shown).

In conclusion, an intact heparin-binding motif in the SmCB1 pro-peptide and its interaction with a sulfated polysaccharide are necessary for the effective auto-processing of the zymogen to the correctly matured enzyme.

The action of sulfated polysaccharides depends on their type, size and concentration

Throughout this study, we used DS as a model sulfated polysaccharide mimicking natural glycosaminoglycans (GAGs), such as heparins and chondroitin and dermatan sulfates. Now, we compare the effect of DS, natural GAGs, and their derivatives on the SmCB1 autoprocessing.

Figure 7A shows that both DS and a similarly sized heparin mediated the regular auto-activation of the zymogen to the mature enzyme, although the heparin-mediated conversion was slower. In contrast, three chondroitin-related GAGs, chondroitin-4-sulfate, chondroitin-6-sulfate and dermatan sulfate, induced the production and accumulation of the intermediate, which was, however, not further processed to the mature enzyme. These GAGs decrease the activity of mature SmCB1 under the experimental conditions (data not shown), which might contribute to the incomplete zymogen autoprocessing.

The size of sulfated polysaccharides required for mediating the auto-activation of the SmCB1 zymogen was investigated using DS and heparin derivatives of differing masses. Heparins with a greater mass than 3 kDa (corresponding to a decasaccharide) were effective, whereas smaller octa- to disaccharides did not induce the generation of SmCB1 activity (Figure 7C). Similarly, DSs were effective over a broad mass range above 5 kDa (a 20-mer oligosaccharide); a 0.5 kDa disaccharide had no effect (Figure 7B). Furthermore, the concentration dependence of the action of sulfated polysaccharides was demonstrated with DS of 9–20 kDa (used throughout this study, unless specified otherwise), which was highly effective above 10 µg/ml and ineffective at concentrations below 0.1 µg/ml (Figure S5).

In summary, sulfated polysaccharides were found to accelerate the autoprocessing of the SmCB1 zymogen, but only specific types, such as DSs and heparins, were capable of inducing the formation of the active mature enzyme; a decasaccharide was the minimal oligosaccharide length required for their action.

Trans-activation of SmCB1 by legumain: the alternative processing pathway is down-regulated by sulfated polysaccharides

We investigated the processing and activation of the SmCB1 zymogen catalyzed by proteolytic action of asparaginyl endopeptidase (legumain) from *S. mansoni* (SmAE) (Caffrey et al., 2000). The recombinant SmAE efficiently converted the SmCB1 zymogen to the mature form, which was associated with the generation of SmCB1 activity (Figures 8A and 8C). This mature SmCB1 species was formed in the absence of DS by the cleavage of the pro-peptide at Asn69-Val70 (consistent with (Sajid et al., 2003)); importantly, the same site we found in the

mature enzyme obtained by autocatalytic processing in the presence of DS (Figure S3). The highest *trans*-activation rate of the SmCB1 zymogen by SmAE was in the pH range of 5.0 to 6.0 (Figure 8B), in which SmAE is also active towards peptide substrates (Dalton et al., 1995; Caffrey et al., 2000; Sajid et al., 2003).

In the next step, we tested the effect of DS on the *trans*-activation reaction catalyzed by SmAE. We demonstrated that the conversion of the SmCB1 zymogen was substantially decreased in the presence of DS, as shown by the processing pattern on SDS-PAGE and by the yield of SmCB1 activity (the activation rate was about one order of magnitude lower as compared to that in the absence of DS; Figures 8A and 8C). The depressive effect of DS was not due to a direct impact on the proteolytic activity of SmAE (measured with a peptide substrate; data not shown). Note that the *trans*-activation experiments were not interfered with by a DS-mediated auto-activation since they were performed at pH 6, at which the auto-activation does not proceed (Figure 3). Moreover, the active-site mutant of the SmCB1 zymogen was used to analyze the processing pattern (Figure 8A).

In conclusion, the SmCB1 zymogen was effectively matured through proteolytic processing that is mediated by SmAE at acidic pH up to pH 6.5, and this process was down-regulated by sulfated polysaccharides.

DISCUSSION

SmCB1 is a critical protease in the hemoglobinolytic pathway of the blood fluke *S. mansoni* and a promising target for anti-schistosomal drugs. In this work, we describe the activation mechanism of the SmCB1 zymogen at the molecular and the structural levels, and provide evidence for a complex control of the SmCB1 activation pathway by sulfated polysaccharides (SPs).

In contrast to the zymogens of vertebrate cathepsins B, the SmCB1 zymogen is unable to auto-catalytically process into a fully mature and active enzyme, even upon prolonged incubation at various pH values. This is in agreement with the previous findings (Sajid et al., 2003). Auto-processing truncates at the generation of an inactive intermediate iCB^{inact} whereby the first half of the pro-peptide is removed (residues 1–38). We crystallographically characterized iCB^{inact} and explain the lack of activity by showing that the remaining intact part of the pro-peptide (residues 39–69) blocks the catalytic site like in the same pose as in the inactive SmCB1 zymogen. Importantly, this nonproductive autoprocessing becomes productive in the presence of SPs (at pH 5.0 and below). Under these conditions, the zymogen is converted into a fully active mature enzyme lacking the entire pro-peptide (residues 1–69) via an enzymatically active intermediate form iCB^{act}. The pro-peptide region that binds into the active site is cleaved in the iCB^{act}, thus opening its catalytic site for substrate entry as also demonstrated by the interaction with the peptidomimetic active-site probe, DCG-04.

SPs are reported to modulate papain-family cysteine peptidases, including their ability to accelerate auto-activation as found for several cathepsin L-type peptidases and human cathepsin B (e.g. Ishidoh et al., 1995;Serveau et al., 2003;Vasiljeva et al., 2005;Caglic et al., 2007). In the case of SmCB1, however, there is a strict requirement for SPs, which are the key factor to initiate and drive productive auto-activation. As compared to human cathepsin B, it also exhibits different selectivity towards SPs: the human enzyme was rather non-selective in its interaction with various glycosaminoglycans (Caglic et al., 2007), whereas the SmCB1 auto-activation was specifically triggered by dextran sulfates and heparins but not by chondroitin sulfates or dermatan sulfate. For SmCB1, we hypothesized that a consensus heparin-binding motif in its pro-peptide is the critical binding site for SPs. The sequence motif (residues 55–61) is located on the non-conserved α3p helix (Figure S1) and is absent in other structurally investigated cathepsin B zymogens. Our recent sequence analysis revealed that this feature is specific for cathepsin B-type peptidases of parasitic trematodes (flukes) that digest host proteins (Horn et al., 2011). Our hypothesis was confirmed after we constructed two zymogen mutants lacking the consensus motif: both were defective in the SP-mediated autoprocessing and unable to produce a mature enzyme.

In order to gain structural insight into the SmCB1 activation mechanism, we solved crystal structures for three processed forms along the activation pathway, namely, the zymogen, inactive intermediate iCB^{inact} and mature enzyme. To date, similar auto-activation intermediates have not been studied in detail for cathepsins B (or other cysteine cathepsins) biochemically or at the structural level. For human cathepsin B, temporary autoprocessing intermediates were detected, but attempts to interrupt the auto-activation process and stabilize intermediates by mutations at the pro-peptide/enzyme core junction were unsuccessful (Quraishi et al., 1999;Pungercar et al., 2009). SmCB1, unable to complete the auto-activation in the absence of SPs, offered an opportunity to analyze the initial phase of the auto-activation pathway involving intermediate formation. The iCB^{inact} is a structural chimera with combined features of the mature enzyme and the zymogen (Figure 1A). The N-terminal half of the pro-peptide is removed, which leads to the formation of the mature active site architecture with the occluding loop in the “closed” conformation and positioned at the primed subsites, whereas the C-terminal half of the pro-peptide is preserved as in the zymogen, blocking the non-primed subsites.

The iCB^{inact} was generated from the SmCB1 zymogen in the absence of SPs by processing at the cleavage site Ser38-Leu39, which is located on the α2p helix, a solvent-exposed region far from the catalytic site to be accessible to direct hydrolysis (Figure 1C). It may be expected that the initial event in the autoprocessing is the destabilization of the pro-peptide binding by a conformational change induced by acidic pH, as has been proposed for other cysteine cathepsin zymogens (Mach et al., 1994;Rozman et al., 1999;McQueney et al., 1997;Menard et al., 1998). This would enable the displacement of the pro-peptide from the active site; this

rearrangement is importantly supported by the occluding loop shifting from “open” to “closed” conformation as shown in the iCB^{inact} structure (Figure 1B). The pro-peptide follows the active-site cleft in a reverse direction to that of a substrate. Therefore, the Ser38-Leu39 hydrolysis could be accomplished by an intramolecular cleavage after a major reorientation of the pro-peptide; nevertheless, an interaction of two zymogen molecules in the opposite orientation would be more sterically favorable to insert productively the pro-peptide in the neighboring active site. We found that the generation of iCB^{inact} is concentration-independent. This might reflect the initial intramolecular (monomolecular) conformational rearrangement of the pro-peptide as a slow rate-limiting step, which is in line with the slow-binding kinetics of the inhibition of cysteine cathepsins with full-length pro-peptides (Billington et al., 2000;Quraishi et al., 1999;Fox et al., 1992).

The mature SmCB1 was generated by SP-mediated autoprocessing via cleavage of Asn69-Val70 at the pro-peptide/enzyme core junction. This solvent-exposed region is located in the vicinity of the heparin-binding motif on the α3p helix of the pro-peptide. The α3p protruding above the enzyme core is stabilized by a network of contacts (both internal and to the enzyme core) formed between acidic and basic residues (Figure 1C). This suggests that the interaction between the heparin-binding motif and SPs will result in the destabilization of α3p and subsequently of the adjacent pro-peptide/enzyme core junction region. Furthermore, upstream to α3p, there is an additional arginine cluster (Arg49 and Arg50, the former located in the S3 subsite), which might interact with SPs to induce displacement of the pro-peptide from the active-site cleft. These conformational changes in the C-terminal half of the pro-peptide will expose two regions of the pro-peptide as a target for proteolysis: (i) the active site-binding region, which is cleaved to release the active intermediate iCB^{act} and (ii) the pro-peptide/enzyme core junction, which is cleaved to produce the mature enzyme (Figures 1C and 4). We found that the generation of the mature enzyme via iCB^{act} is a concentration-dependent process, indicating its bimolecular character. In the first step, the mechanism of the formation of iCB^{act} from the zymogen might resemble that of iCB^{inact} **(**both having low sensitivity to inhibition by E-64), in which, however, SPs allowed the removal of a larger part of the pro-peptide blocking the catalytic site. In the second step (sensitive to E-64), iCB^{act} cleaves another molecule of iCB^{act} (and possibly other unprocessed forms) in complex with SP to give rise to the mature enzyme. The newly formed mature enzyme enters the processing scheme as demonstrated by its ability to effectively process the active-site mutated zymogen (into intermediate and mature forms) in the presence of SP.

Our data imply that SPs convert the SmCB1 zymogen into a better substrate for autocatalytic processing. This was demonstrated by pro-peptide cleavage in the context of the zymogen molecule; moreover, we showed that the synthetic pro-peptide scaffold became a better substrate for the mature enzyme, which was largely attributed to the effect of the heparin-binding motif (Figure 4). The complexing of SP to the pro-peptide thus directly

increased its susceptibility to proteolysis. The minimal length of SPs required to induce the autoactivation processing of the SmCB1 zymogen was a decasaccharide. The critical size determined is similar to that observed for many interactions between proteins and SPs (typically 6–12 saccharide units) (for review, see Gandhi et al., 2008) but is larger than a tetrasaccharide, which has been found to accelerate the auto-activation of human cathepsin B zymogen (Caglic et al., 2007). The decasaccharide length of ~45 Å is sufficient to establish an interaction with the C-terminal half of the SmCB1 propeptide (the zymogen size is about 60x55x45 Å). The fact that larger SPs (in the tested range of 3–500 kDa) exhibited a higher induction effect on the SmCB1 autoactivation might be explained by a multivalent interaction with both the pro-peptide and (a) positively charged region(s) on the enzyme core. This is in line with our previous finding of heparin interaction with the mature SmCB1 and the identification of large basic patches on its surface, e.g. on the occluding loop (Horn et al., 2011).

The trans-activation of the SmCB1 zymogen by asparaginyl endopeptidase (legumain) leads to the mature enzyme through the cleavage of the same Asn69-Val70 site at the pro-peptide/mature core junction that is processed during the SP-mediated autoactivation. Using the *S. mansoni* legumain, SmAE, we demonstrated that the trans-activation proceeds at acidic pH up to pH 6.5, consistent with the finding by Sajid et al. (Sajid et al., 2003). Surprisingly, this reaction was strongly suppressed by SPs. We showed that SPs did not negatively influence the SmAE activity towards peptide substrates and did not interfere with pro-peptide cleavage by interacting with the heparin-binding motif of the zymogen. This suggested that SPs might sterically hinder the SmAE-protein substrate interaction by binding to the SmCB1 enzyme core and/or SmAE outside the active site.

In Figure 9, we summarize the results of this study and propose the following pathway for SmCB1 activation. There are two alternative routes for the proteolytic activation of the SmCB1 zymogen, which are under a complex control of SPs. The autocatalytic activation requires SPs for productive processing to a mature enzyme, whereas the *trans*-activation mediated by endogenous SmAE is negatively modulated by SPs. SPs thus function as a molecular switch between the two routes. Additional regulation might occur through pH – the optimum pH range for the auto-activation and *trans*-activation is about 4–5 and 5–6, respectively. Both routes lead to the same product, the mature enzyme starting by Val70; two N-terminal residues can be further trimmed by exopeptidase action as demonstrated previously (Sajid et al., 2003).

We believe that the discovered auto-activation route of SmCB1 mediated by SPs is biologically relevant. Indirect evidence can be drawn from a recent study in which SmAE was knocked down by RNA interference in adult *S. mansoni* without any deleterious effect on SmCB1 processing and activation (Krautz-Peterson et al., 2008). The required SPs, represented here by glycosaminoglycans (GAGs), are widespread in all animal tissues, including those in *Schistosoma* and the human host (e.g. heparin, effective in SmCB1 auto-

activation, is present in mast cells/basophils ingested by *Schistosoma* with the blood meal; Hamed et al., 1997). The exact location(s) of SmCB1 processing in the *Schistosoma* gut is unknown and may be both intracellular and extracellular (possibly differing in SmAE and GAGs content), and further research will be needed to evaluate the functional relationship between the auto-activation and *trans*-activation routes.

In conclusion, we discover a dual activation pathway for SmCB1, which is under complex control by SPs, and provide a structural basis to explain the proteolytic mechanisms involved. Our results give new insights into the regulation of the SmCB1 drug target and increase our understanding of GAG-protein interactions that modulate the function of pathology-associated cysteine cathepsins.

EXPERIMENTAL PROCEDURES

Materials

Dextran sulfates, heparins (above 3 kDa), dermatan sulfate and chondroitin sulfates were purchased from Sigma Aldrich, heparin-derived oligosaccharides (from 0.7 to 2.4 kDa) from Dextra Laboratories; Z-Phe-Arg-AMC and Z-Ala-Ala-Asn-AMC (Z, Benzyloxycarbonyl; AMC, 7-amino-4-methylcoumarin) from Bachem; CA-074 and E-64 from Sigma. Green-DCG-04 (Greenbaum et al., 2002) was a generous gift from Dr. M. Bogyo (Stanford University), the aza-peptide Michael acceptor CBz-Ala-Ala-(aza-Asn)-CH=CH-COOEt (Aza-N-11a) (Ekici et al., 2004) was kindly donated by Dr. J. C. Powers (Georgia Institute of Technology). *S. mansoni* asparaginyl endopeptidase (SmAE) was produced recombinantly in *P. pastoris* as described in (Caffrey et al., 2000). Specific polyclonal antibodies against SmCB1 zymogen were produced in rabbit by Moravian Biotechnology (Czech Republic) using the standard protocol; the immunoglobulin fraction purified from the serum on a HiTrap Protein A column (GE Healthcare Life Sciences) according to the manufacturer's protocol was used in all experiments.

Cloning and Site-directed Mutagenesis of SmCB1

Altogether five pPICZαA constructs for expression in *P. pastoris* were prepared: the pPICZαA plasmid containing a wild-type SmCB1 proenzyme (pCB^{WT}) insert was constructed as described previously (Horn et al., 2011); a nonglycosylated mutant (Thr168Ala, Thr283Ala) of SmCB1 zymogen (pCB) was generated as described in (Jilkova et al., 2011). The sequence of pCB was used to prepare (1) mutants with mutations (Arg57Ala, Arg59Ala or Arg57Asn, Arg59Asn) in the heparin-binding motif (pCB^{hA} and pCB^{hN}, respectively) and (2) an inactive mutant (pCB^{as}) with a mutated catalytic cysteine residue (Cys100Ser). Site-directed mutagenesis was performed as described previously (Jilkova et al., 2011) using the PCR method with the following mutagenic oligonucleotide primers 5'-AGAAGAACCAGATTTGAGGGCGACGGCACGTCCCACAGTTGACCAC-3' for Arg57Ala,

Arg59Ala; 5'-GAAGAGAAGAACCAGATTTGAGGAATACGAATCGTCCCACAGTTGACCACAAT-3' for Arg57Asn, Arg59Asn; and 5'-TCACGCTGTGGATCAAGTTGGGCTTTCGGTG-3' for Cys100Ser. After mutagenesis, the constructs were sequenced to verify the desired mutations.

Recombinant Expression and Purification of SmCB1

SmCB1 zymogen and its mutants were expressed in *P. pastoris* and purified as described previously (Horn et al., 2011). The mature forms were obtained by proteolytic treatment of the respective zymogens with SmAE and purified as described previously (Horn et al., 2011). The mature forms of SmCB1 mutants derived from pCB, pCB^{hA}, pCB^{hN} exhibited analogous activity properties to the mature wild-type SmCB1 derived from pCB^{WT} (K_m in the range of 25–30 μM and k_{cat} in the range of 0.8–1.1 s^{-1}). The zymogens of the wild-type and nonglycosylated SmCB1 (pCB^{WT} and pCB) had analogous properties with regard to the activation processing (Figure S6). Unless otherwise stated, the nonglycosylated version of SmCB1 was used in all experiments described in the paper.

Preparation of Proteins for Crystallization

The mature SmCB1 (mCB) was derived from the nonglycosylated zymogen pCB by trans-activation using SmAE as described previously (Horn et al., 2011). The SmCB1 intermediate (iCB^{inact}) was obtained by the autoprocessing of pCB in 0.1 M sodium acetate, pH 4.0, containing 2.5 mM dithiothreitol, 0.1 M PEG 6000, and 1 mM EDTA for 18 h at 37°C. The conversion to the intermediate and mature forms was monitored using SDS-PAGE and in the case of mature enzyme also using a kinetic assay with the fluorogenic substrate Z-Phe-Arg-AMC. Both processing forms as well as the active-site mutated zymogen (pCB^{as}) were finally purified using a Mono S HR 5/5 column under the conditions described in (Horn et al., 2011). The purified proteins were concentrated and buffer-exchanged into 10 mM sodium acetate, pH 5.5 using Amicon Ultracel-10k (Millipore).

Protein Crystallization and Data Collection

Crystallization was accomplished by vapor diffusion in hanging drop. The ratio of protein to reservoir solution in drops was 2:1, 1:0.75 and 1:1 for the zymogen (pCB^{as}), intermediate (iCB^{inact}) and mature (mCB) forms of SmCB1, respectively. Drops were equilibrated over a 1 ml reservoir solution at 20°C for 1–2 weeks. The reservoir solutions were 0.1 M ammonium acetate, 0.1 M Bis-Tris, 17% PEG 10000, pH 6.0 for pCB and 0.2 M ammonium acetate, 0.1 M sodium citrate, 30% PEG 4000, pH 5.6 for iCB^{inact} and mCB. The protein concentrations of the stock solutions of the complexes were 1.5–5.25 mg/ml (in 10 mM sodium acetate, pH 5.5). The crystals were flash-cooled by plunging into liquid nitrogen and then stored in liquid nitrogen until used for X-ray diffraction experiments. pCB^{as} crystals were cryoprotected in

mother liquor containing 20% ethylene glycol before flash cooling. Diffraction data from the crystals of iCB^{inact} and mCB were collected at 100 K at beamline 19-BM of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, U.S.A; diffraction data from the pCB^{as} crystal at 100 K at beamline MX14.2 operated by the Helmholtz-Zentrum Berlin (HZB) at the Bessy II electron storage ring, Berlin-Adlershof, Germany (Mueller et al., 2012). All diffraction data were processed using the HKL-3000 suite of programs (Minor et al., 2006). The crystal parameters and data collection statistics are given in the Supplemental Table S3.

Structure Determination, Refinement and Analysis

The structures were determined by molecular replacement with the program Molrep (Vagin et al., 2000) using the structure of mature SmCB1 in complex with CA-074 (PDB code 3QSD) as the search model. Model refinement was carried out using the program REFMAC 5.2 from the CCP4 package (CCP4, 1994), interspersed with manual adjustments using Coot (Emsley et al., 2004). The structure of mCB at a resolution of 1.3 Å was refined using anisotropic atomic displacement parameter (ADP, formally B-factor) refinement for all protein atoms. The final steps of the isotropic ADP refinement of iCB^{inact} and pCB^{as} included TLS refinement (Winn et al., 2001). The quality of the final models was validated with Molprobity (Lovell et al., 2003). The final refinement statistics are given in the Supplemental Table S3. All figures showing structural representations were prepared with the program PyMOL (De Lano et al., 2002).

Molecular Modeling

The part of the occluding loop of the pCB^{as} crystal structure without adequate electron density (residues 181–186: HTKGKY) was modeled in the MOE program using the following strategy: The loop was modeled utilizing a conformation from libraries of protein fragments and according to the procedure suggested by (Claessens et al., 1989;Summers et al., 1990;Levitt, 1992). The basis of the approach is a selection of fragments from the PDB that overlap with the framework at both ends. The existing ArchDB library was utilized. It is a compilation of the structural classifications of the loops extracted from known protein structures (http://sbi.imim.es/cgi-bin/archdb//loops.pl). The closest of the loops found was 1qmv_A_24, Subclass 6.9.1 in the ArchDB95 dataset. First, the loop was attached into the incomplete pCB^{as} crystal structure, then two amino acids from the loop were replaced by the original pCB^{as} residues and solvent was added to the whole structure. The solvent around the protein was optimized, followed by further optimization of the loop and the solvent with all other protein atoms fixed at crystallography positions in the ff99SB parameterization of the AMBER force field.

Autocatalytic Processing of SmCB1

SmCB1 zymogen (various mutants, 75 nM final concentration) was autoprocessed by incubation at 37°C in 1500 µl of appropriate activation buffer (see below) in the absence and presence of DS of 9–20 kDa (used throughout this study, unless specified otherwise) at a final concentration of 10 µg/ml. The following buffers were used: 0.1 M sodium acetate (pH between 4.0 and 5.5) and 0.1 M sodium phosphate (pH between 6.0 and 8.0). All activation buffers contained 2.5 mM dithiothreitol, 1 mM EDTA, and 0.1% PEG 6000. Aliquots of 5 µl were taken from the reaction mixture at different time intervals and activity of SmCB1 was measured in kinetic assay with the fluorogenic substrate Z-Phe-Arg-AMC. Simultaneously, 300 µL aliquots of the reaction mixture were withdrawn, processing was stopped by adding E-64 to 10 µM concentration and samples were analyzed by SDS-PAGE.

The effect of DS concentration on pCB autoprocessing was studied as described above in 0.1 M sodium acetate, pH 5.0, 2.5 mM dithiothreitol, 1 mM EDTA, 0.1% PEG 6000, in the presence of DS at the final concentration in the range of 0.1 to 50 µg/ml. The autoprocessing in the presence of different types of SPs was studied in the same buffer with SP at a final concentration of 10 µg/ml; for autoprocessing in the presence of E-64 or glycerol, the reaction mixtures contained 5 µM E-64 or 10-20% glycerol, respectively. Aliquots of the reaction mixtures were withdrawn at different time intervals and analyzed by SDS-PAGE or SmCB1 activity assay with Z-Phe-Arg-AMC substrate.

The molecularity of autoprocessing was studied by incubating pCB at 75 nM and 750 nM concentrations in 0.1 M sodium acetate, pH 4.0, 2.5 mM dithiothreitol, 1 mM EDTA, 0.1% PEG 6000, in the absence of DS, or in 0.1 M sodium acetate, pH 5.0, 2.5 mM dithiothreitol, 1 mM EDTA, 0.1% PEG 6000 in the presence of DS at a final concentration of 10 µg/ml DS (for 75 nM pCB) or 100 µg/ml (for 750 nM pCB). Aliquots of the reaction mixture (300 µl for 75 nM and 30 µl for 750 nM pCB) were withdrawn at different time intervals and analyzed by SDS-PAGE.

The pCB^{as} zymogen (75 nM) was processed by mature SmCB1 (2.5 nM) at 37°C for 30 h in 300 µl of 0.1 M sodium acetate, pH 5.0, 2.5 mM dithiothreitol, 1 mM EDTA, 0.1% PEG in the presence or absence of DS (10 µg/ml). The reaction mixtures were analyzed by SDS-PAGE.

Trans-activation of SmCB1

The pCB zymogen (75 nM) was *trans*-activated by incubation with SmAE (5 nM) at 37°C in the presence or absence of DS (10 µg/ml) in 100 µl of the following buffers: 0.1 M sodium acetate (pH 4.0-5.5) and 0.1 M sodium phosphate (pH 6.0-8.0). All buffers contained 2.5 mM dithiothreitol, 1 mM EDTA, and 0.1% PEG 6000. Aliquots of 5 µl were taken from the reaction mixture at different time intervals and SmCB1 activity was measured using Z-Phe-Arg-AMC substrate.

The pCB^{as} zymogen (75 nM) was processed by SmAE (5 nM) at 37°C for 6 h in 300 µl of 0.1 M sodium phosphate, pH 6.0, 2.5 mM dithiothreitol, 1 mM EDTA, 0.1% PEG in the presence or absence of DS (10 µg/ml). The reaction mixtures were resolved by SDS-PAGE and analyzed by immunoblotting using SmCB1 specific antibodies.

Proteolytic activity measurement

Enzymatic activity of SmCB1 was measured in a kinetic continuous assay using fluorogenic substrate Z-Phe-Arg-AMC. Assay was performed at 37°C in 96-well microplate. Aliquots of reaction mixtures (5 µl) were added to 200 µl of the solution containing 25 µM Z-Phe-Arg-AMC in 0.1 M sodium phosphate pH 6.0, 0.1% PEG 6000 and 1 mM EDTA. Hydrolysis of substrates was measured continuously using an Infinite M1000 microplate reader (Tecan) at excitation and emission wavelengths of 360 and 465 nm, respectively. All experiments were performed in duplicates. Activity of SmAE was measured analogously in 100 µl solution containing 50 µM Z-Ala-Ala-Asn-AMC in 0.1 M sodium phosphate, pH 6.5, 2.5 mM DTT, 0.1% PEG 6000 and 1 mM EDTA in the presence or absence of DS (10 µg/ml). The Michaelis-Menten kinetic parameters for SmCB1 (and various mutants) were determined using assay system described above by measuring the rate of Z-Phe-Arg-AMC hydrolysis in the concentration range of 0–200 µM, and K_m and k_{cat} values were obtained by nonlinear regression using the GraFit software.

The molar concentration of active SmCB1 and SmAE was determined by the stoichiometric titration method as described previously (Horn et al., 2009;Franta et al., 2010). The following inhibitors were applied as active-site titrants: CA-074 for SmCB1 and Aza-N-11a (Ekici et al., 2004) for SmAE. An aliquot of protease solution was incubated with various amounts of the titrant for 30 min at 37°C in 0.1 M sodium phosphate (pH 6.0 for SmCB1, pH 6.5 for SmAE), 0.1% PEG 6000 and 1 mM EDTA. Residual peptidase activities were measured using the assay systems as described in the previous paragraph, and plotted into the titration curves.

Analytical methods

SDS-PAGE was performed under reducing conditions on 15% Laemmli gels. Aliquots of reaction mixtures were protein precipitated with acetone, resolved by SDS-PAGE, and the gel was visualized by protein staining with Coomassie Brilliant Blue or silver nitrate. Processing sites were identified by N-terminal protein sequencing after electroblotting of SDS-PAGE gels. Immunoblots were stained using rabbit anti-SmCB1 IgG (1:1000), anti-rabbit IgG conjugated with horseradish peroxidase (1:10 000) and SuperSignal West Pico chemiluminescent substrate (Pierce), and visualized with an ImageQuant LAS-4000 Mini luminescent image analyzer (GE Healthcare Life Sciences).

Active-Site Labeling of SmCB1

The zymogen, intermediate forms and mature SmCB1 were incubated with 0.5 μM Green-DCG-04 (Greenbaum et al., 2002) in 0.1 M sodium phosphate, pH 6.0, containing 2.5 mM dithiothreitol, 1 mM EDTA for 1 h. The labeled proteins were resolved by SDS-PAGE and visualized in a Typhoon 9410 Imager (GE Healthcare) as described in (Horn et al., 2009).

Processing of Pro-peptide Derived Fragments

Pro-peptide fragments listed in Figure S2 were synthesized by Fmoc solid-phase chemistry in an ABI 433A Peptide Synthesizer (Applied Biosystems) in the form of peptidyl amides with acetylated N-termini (except the N-terminal pro-peptide fragment with free N-terminus) as described previously (Horn et al., 2011;Doleckova-Maresova et al., 2005). Synthetic peptides (100 μM) were incubated (1) with mature SmCB1 (600 nM) in 0.1 M sodium acetate, pH 4.0, 1 mM EDTA, 2.5 mM DTT in the absence of DS, or in 0.1 mM sodium acetate, pH 5.0, 2.5 mM dithiothreitol in the presence of DS at a final concentration of 10 μg/ml, and (2) with SmAE (5 nM) in 0.1 M MES pH 6.0, 1 mM EDTA, 2.5 mM DTT in the absence or presence of DS at a final concentration of 10 μg/ml. The reaction mixtures were incubated at 37°C for 1 to 16 h. The resulting fragments were purified by RP-HPLC over C18 column (Vydac, 25 x 0.46 cm) using the TFA/acetonitrile system and characterized by LC-MS/MS on the UltiMate 3000 RSLCnano system (Dionex) coupled to a TripleTOF 5600 mass spectrometer with a NanoSpray III source (AB Sciex).

ACCESSION NUMBERS

The atomic coordinates and experimental structure factors for the zymogen, intermediate and mature forms of SmCB1 have been deposited in the Protein Data Bank with the codes 4I04, 4I05 and 4I07, respectively.

ACKNOWLEDGEMENTS

This work was supported, in whole or in part, by grant 203/09/1585 from the Czech Science Foundation, grant LH12023 and NPU I project LO 1302 from the Ministry of Education, Youth and Sports of the Czech Republic, and institutional project RVO 61388963. Synchrotron data collection was supported by the Helmholtz-Zentrum Berlin (HZB) at the Bessy II electron storage ring, Berlin-Adlershof, Germany, and by the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory, Argonne, U.S.A. We thank Matthew Bogyo (Stanford University School of Medicine, Stanford, U.S.A.) for providing activity-based probe, James C. Powers (Georgia Institute of Technology, Atlanta, U.S.A.) for Aza-N-11a, Jana Horakova for mass spectroscopy and Zdenek Voburka for protein sequencing (Institute of Organic Chemistry and Biochemistry AS CR, Prague, Czech Republic).

REFERENCES

Abdulla,M.H., Lim,K.C., Sajid,M., McKerrow,J.H., and Caffrey,C.R. (2007). Schistosomiasis mansoni: novel chemotherapy using a cysteine protease inhibitor. PLoS. Med. *4*, e14.

Almeida,P.C., Nantes,I.L., Chagas,J.R., Rizzi,C.C., Faljoni-Alario,A., Carmona,E., Juliano,L., Nader,H.B., and Tersariol,I.L. (2001). Cathepsin B activity regulation. Heparin-like glycosaminoglycans protect human cathepsin B from alkaline pH-induced inactivation. J. Biol. Chem. *276*, 944-951.

Beckham,S.A., Law,R.H., Smooker,P.M., Quinsey,N.S., Caffrey,C.R., McKerrow,J.H., Pike,R.N., and Spithill,T.W. (2006). Production and processing of a recombinant *Fasciola hepatica* cathepsin B-like enzyme (FhcatB1) reveals potential processing mechanisms in the parasite. Biol. Chem. *387*, 1053-1061.

Billington,C.J., Mason,P., Magny,M.C., and Mort,J.S. (2000). The slow-binding inhibition of cathepsin K by its propeptide. Biochem. Biophys. Res. Commun. *276*, 924-929.

Brindley,P.J., Kalinna,B.H., Dalton,J.P., Day,S.R., Wong,J.Y., Smythe,M.L., and McManus,D.P. (1997). Proteolytic degradation of host hemoglobin by schistosomes. Mol. Biochem. Parasitol. *89*, 1-9.

Caffrey,C.R. (2007). Chemotherapy of schistosomiasis: present and future. Curr. Opin. Chem. Biol. *11*, 433-439.

Caffrey,C.R., Mathieu,M.A., Gaffney,A.M., Salter,J.P., Sajid,M., Lucas,K.D., Franklin,C., Bogyo,M., and McKerrow,J.H. (2000). Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. FEBS Lett. *466*, 244-248.

Caglic,D., Pungercar,J.R., Pejler,G., Turk,V., and Turk,B. (2007). Glycosaminoglycans facilitate procathepsin B activation through disruption of propeptide-mature enzyme interactions. J. Biol. Chem. *282*, 33076-33085.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D. Biol. Crystallogr. *50*, 760-763.

Cherney,M.M., Lecaille,F., Kienitz,M., Nallaseth,F.S., Li,Z., James,M.N., and Bromme,D. (2011). Structure-activity analysis of cathepsin K/chondroitin 4-sulfate interactions. J. Biol. Chem. *286*, 8988-8998.

Claessens,M., Van,C.E., Lasters,I., and Wodak,S. (1989). Modelling the polypeptide backbone with 'spare parts' from known protein structures. Protein Eng *2*, 335-345.

Colley,D.G., Bustinduy,A.L., Secor,W.E., and King,C.H. (2014). Human schistosomiasis. Lancet. *383*, 2253-2264

Dalton,J.P., Hola-Jamriska,L., and Brindley,P.J. (1995). Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. Parasitology *111*, 575-580.

De Lano, W. L. 2002 The PyMOL Molecular Graphics System. De Lano Scientific LLC, San Carlos, CA, USA.

Delcroix,M., Sajid,M., Caffrey,C.R., Lim,K.C., Dvorak,J., Hsieh,I., Bahgat,M., Dissous,C., and McKerrow,J.H. (2006). A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. J. Biol. Chem. *281*, 39316-39329.

Doleckova-Maresova,L., Pavlik,M., Horn,M., and Mares,M. (2005). De novo design of alpha-amylase inhibitor: a small linear mimetic of macromolecular proteinaceous ligands. Chem. Biol. *12*, 1349-1357.

Ekici,O.D., Gotz,M.G., James,K.E., Li,Z.Z., Rukamp,B.J., Asgian,J.L., Caffrey,C.R., Hansell,E., Dvorak,J., McKerrow,J.H., Potempa,J., Travis,J., Mikolajczyk,J., Salvesen,G.S., and Powers,J.C. (2004). Aza-peptide Michael acceptors: a new class of inhibitors specific for caspases and other clan CD cysteine proteases. J. Med. Chem. *47*, 1889-1892.

Emsley,P., and Cowtan,K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D. Biol. Crystallogr. *60*, 2126-2132.

Fairhead,M., Kelly,S.M., and van der Walle,C.F. (2008). A heparin binding motif on the pro-domain of human procathepsin L mediates zymogen destabilization and activation. Biochem. Biophys. Res. Commun. *366*, 862-867.

Fox,T., de,M.E., Mort,J.S., and Storer,A.C. (1992). Potent slow-binding inhibition of cathepsin B by its propeptide. Biochemistry *31*, 12571-12576.

Franta,Z., Frantova,H., Konvickova,J., Horn,M., Sojka,D., Mares,M., and Kopacek,P. (2010). Dynamics of digestive proteolytic system during blood feeding of the hard tick *Ixodes ricinus*. Parasit. Vectors *3*, 119.

Gandhi,N.S., and Mancera,R.L. (2008). The structure of glycosaminoglycans and their interactions with proteins. Chem. Biol. Drug Des *72*, 455-482.

Greenbaum,D., Baruch,A., Hayrapetian,L., Darula,Z., Burlingame,A., Medzihradszky,K.F., and Bogyo,M. (2002). Chemical approaches for functionally probing the proteome. Mol. Cell Proteomics *1*, 60-68.

Hamed,R.R., Maharem,T.M., and El-Guindy,A.S. (1997). Proteoglycans from adult worms of *Schistosoma haematobium*. J. Helminthol. *71*, 151-160.

Horn,M., Jilkova,A., Vondrasek,J., Maresova,L., Caffrey,C., and Mares,M. (2011). Mapping the pro-peptide of the *Schistosoma mansoni* cathepsin B1 drug target: modulation of inhibition by heparin and design of mimetic inhibitors. ACS Chem. Biol. *6*, 609–617.

Horn,M., Nussbaumerova,M., Sanda,M., Kovarova,Z., Srba,J., Franta,Z., Sojka,D., Bogyo,M., Caffrey,C.R., Kopacek,P., and Mares,M. (2009). Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. Chem. Biol. *16*, 1053-1063.

Ishidoh,K., and Kominami,E. (1995). Procathepsin L degrades extracellular matrix proteins in the presence of glycosaminoglycans in vitro. Biochem. Biophys. Res. Commun. *217*, 624-631.

Jilkova,A., Rezacova,P., Lepsik,M., Horn,M., Vachova,J., Fanfrlik,J., Brynda,J., McKerrow,J.H., Caffrey,C.R., and Mares,M. (2011). Structural basis for inhibition of cathepsin B drug target from the human blood fluke, *Schistosoma mansoni*. J. Biol. Chem. *286*, 35770-35781.

Krautz-Peterson,G., and Skelly,P.J. (2008). Schistosome asparaginyl endopeptidase (legumain) is not essential for cathepsin B1 activation in vivo. Mol. Biochem. Parasitol. *159*, 54-58.

Levitt,M. (1992). Accurate modeling of protein conformation by automatic segment matching. J. Mol. Biol. *226*, 507-533.

Li,Z., Hou,W.S., Escalante-Torres,C.R., Gelb,B.D., and Bromme,D. (2002). Collagenase activity of cathepsin K depends on complex formation with chondroitin sulfate. J. Biol. Chem. *277*, 28669-28676.

Lovell,S.C., Davis,I.W., Arendall,W.B., III, de Bakker,P.I., Word,J.M., Prisant,M.G., Richardson,J.S., and Richardson,D.C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins *50*, 437-450.

Mach,L., Mort,J.S., and Glossl,J. (1994). Maturation of human procathepsin B. Proenzyme activation and proteolytic processing of the precursor to the mature proteinase, in vitro, are primarily unimolecular processes. J. Biol. Chem. *269*, 13030-13035.

McQueney,M.S., Amegadzie,B.Y., D'Alessio,K., Hanning,C.R., McLaughlin,M.M., McNulty,D., Carr,S.A., Ijames,C., Kurdyla,J., and Jones,C.S. (1997). Autocatalytic activation of human cathepsin K. J. Biol. Chem. *272*, 13955-13960.

Menard,R., Carmona,E., Takebe,S., Dufour,E., Plouffe,C., Mason,P., and Mort,J.S. (1998). Autocatalytic processing of recombinant human procathepsin L. Contribution of both intermolecular and unimolecular events in the processing of procathepsin L in vitro. J. Biol. Chem. *273*, 4478-4484.

Minor,W., Cymborowski,M., Otwinowski,Z., and Chruszcz,M. (2006). HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes. Acta Crystallogr. D. Biol. Crystallogr. *62*, 859-866.

Mueller,U., Darowski,N., Fuchs,M.R., Forster,R., Hellmig,M., Paithankar,K.S., Puhringer,S., Steffien,M., Zocher,G., and Weiss,M.S. (2012). Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. J. Synchrotron. Radiat. *19*, 442-449.

Pungercar,J.R., Caglic,D., Sajid,M., Dolinar,M., Vasiljeva,O., Pozgan,U., Turk,D., Bogyo,M., Turk,V., and Turk,B. (2009). Autocatalytic processing of procathepsin B is triggered by proenzyme activity. FEBS J. *276*, 660-668.

Quraishi,O., Nagler,D.K., Fox,T., Sivaraman,J., Cygler,M., Mort,J.S., and Storer,A.C. (1999). The occluding loop in cathepsin B defines the pH dependence of inhibition by its propeptide. Biochemistry *38*, 5017-5023.

Raman,R., Sasisekharan,V., and Sasisekharan,R. (2005). Structural insights into biological roles of protein-glycosaminoglycan interactions. Chem. Biol. *12*, 267-277.

Rozman,J., Stojan,J., Kuhelj,R., Turk,V., and Turk,B. (1999). Autocatalytic processing of recombinant human procathepsin B is a bimolecular process. FEBS Lett. *459*, 358-362.

Sajid,M., McKerrow,J.H., Hansell,E., Mathieu,M.A., Lucas,K.D., Hsieh,I., Greenbaum,D., Bogyo,M., Salter,J.P., Lim,K.C., Franklin,C., Kim,J.H., and Caffrey,C.R. (2003). Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. Mol. Biochem. Parasitol. *131*, 65-75.

Serveau,C., Boulange,A., Lecaille,F., Gauthier,F., Authie,E., and Lalmanach,G. (2003). Procongopain from *Trypanosoma congolense* is processed at basic pH: an unusual feature among cathepsin L-like cysteine proteases. Biol. Chem. *384*, 921-927.

Summers,N.L., and Karplus,M. (1990). Modeling of globular proteins. A distance-based data search procedure for the construction of insertion/deletion regions and Pronon-Pro mutations. J. Mol. Biol. *216*, 991-1016.

Thetiot-Laurent,S.A., Boissier,J., Robert,A., and Meunier,B. (2013). Schistosomiasis chemotherapy. Angew. Chem. Int. Ed Engl. *52*, 7936-7956.

Turk,V., Stoka,V., Vasiljeva,O., Renko,M., Sun,T., Turk,B., and Turk,D. (2012). Cysteine cathepsins: from structure, function and regulation to new frontiers. Biochim. Biophys. Acta *1824*, 68-88.

Vagin,A., and Teplyakov,A. (2000). An approach to multi-copy search in molecular replacement. Acta Crystallogr. D. Biol. Crystallogr. *56*, 1622-1624.

Vasiljeva,O., Dolinar,M., Pungercar,J.R., Turk,V., and Turk,B. (2005). Recombinant human procathepsin S is capable of autocatalytic processing at neutral pH in the presence of glycosaminoglycans. FEBS Lett. *579*, 1285-1290.

Winn,M.D., Isupov,M.N., and Murshudov,G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. D. Biol. Crystallogr. *57*, 122-133.

FIGURES AND FIGURE LEGENDS

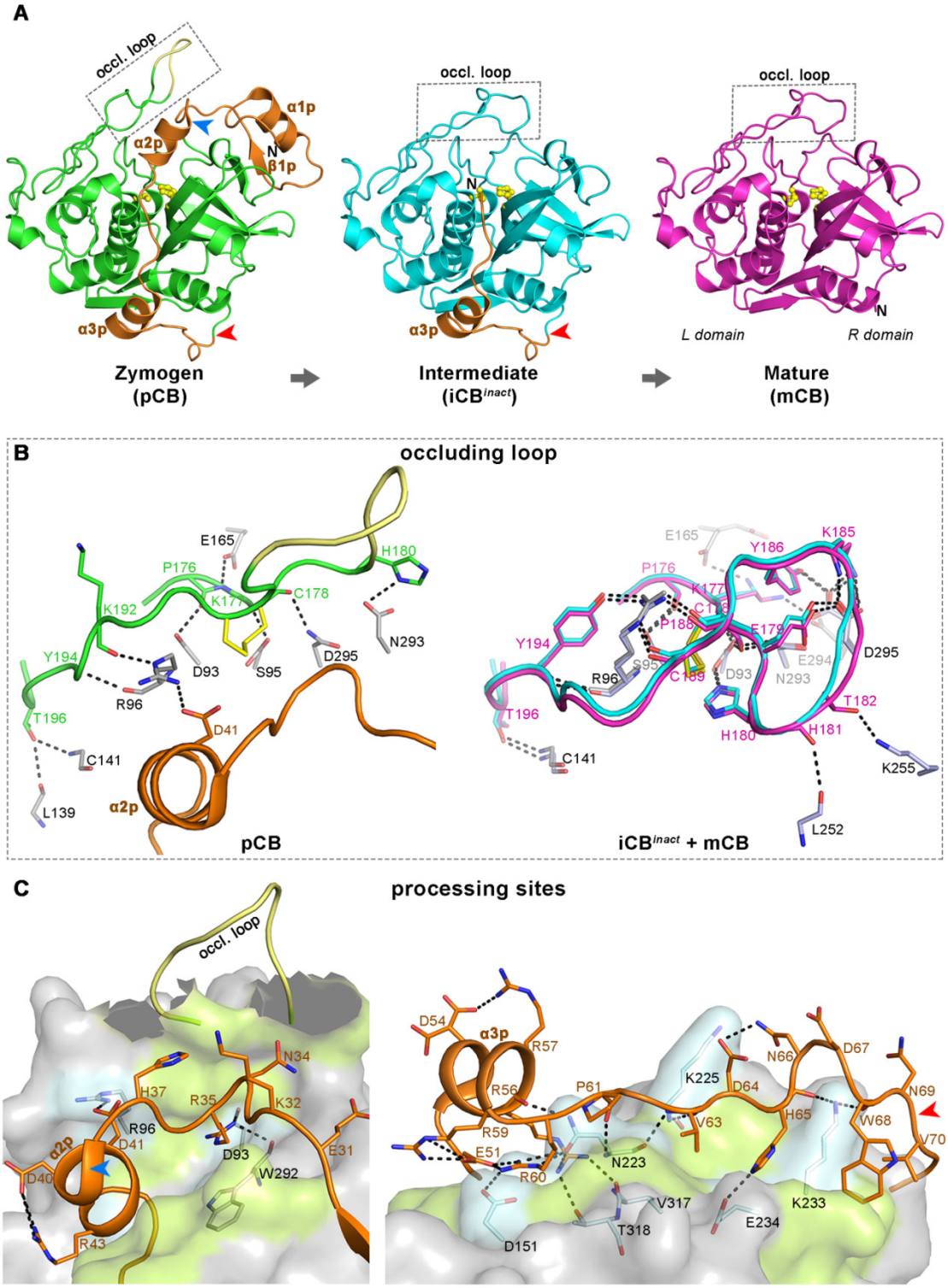


Figure 1. Structural comparison of the processing forms of SmCB1 along the activation pathway

(A) Overall crystal structures of three processing forms of SmCB1, including the zymogen (pCB) with full-length pro-peptide (residues 1–69), the inactive intermediate (iCB^{inact}) with partially processed pro-peptide (residues 39–69), and the mature enzyme (with N-terminal residue 70). The

3D structures are shown in cartoon representation with the pro-peptide in orange and the enzyme core in green (pCB), cyan (iCB^{inact}) and magenta (mCB). The catalytic residues Cys100 and His270 are depicted in ball-and-stick representation in yellow. The arrowheads mark the positions of the cleavage sites, which are processed giving rise to iCB^{inact} and mCB (blue and red arrowheads, respectively). The secondary structure elements in the pro-peptide are labeled; the occluding loop is boxed; N-terminus is marked by N.

(B) Rearrangement of the occluding loop associated with the pro-peptide processing. The residues forming hydrogen bonds between the occluding loop and the enzyme core/pro-peptide are shown as sticks (the main chains or side chains are presented; heteroatoms are red and blue for O and N). *Left panel:* The occluding loop of pCB (green; the flexible modeled part is yellow) is in the "open" conformation dictated by the pro-peptide (orange); the enzyme core residues are in gray. *Right panel:* The occluding loop of iCB^{inact} (cyan) and mCB (magenta) is in the "closed" conformation; the enzyme core residues are in gray and light blue (iCB^{inact} and mCB, respectively) and occluding loop residues are labeled in magenta. The yellow sticks represent disulfide bridges.

(C) Structural details on proteolytic processing sites within the pro-peptide (for orientation, see panel A). The pro-peptide in pCB is shown in the orange cartoon representation with side chains as sticks (the heteroatoms are colored like in (B)). Depicted are the hydrogen bonds between the pro-peptide and the enzyme core, and the internal hydrogen bonds within the pro-peptide (black dashed lines). The enzyme core is shown as a transparent gray surface with hydrogen-bonded residues represented as gray sticks; the surface of the residues forming hydrogen bonds with the pro-peptide is colored in pale cyan, the other contact residues (<4.0 Å) are in light green. *Left panel:* The cleavage site where the pro-peptide is processed to yield iCB^{inact} is surface-exposed at the α2p helix; the hydrolyzed Ser38-Leu39 bond is indicated by the blue arrowhead. The flexible tip of the occluding loop is shown in the yellow cartoon representation. *Right panel:* The cleavage site where the pro-peptide is processed to yield mCB is surface-exposed at the pro-peptide/enzyme core junction; the hydrolyzed Asn69-Val70 bond is indicated by the red arrowhead. The α3p helix with the heparin-binding motif (containing Arg56, Arg57, Arg59, and Arg60) is located in the vicinity.

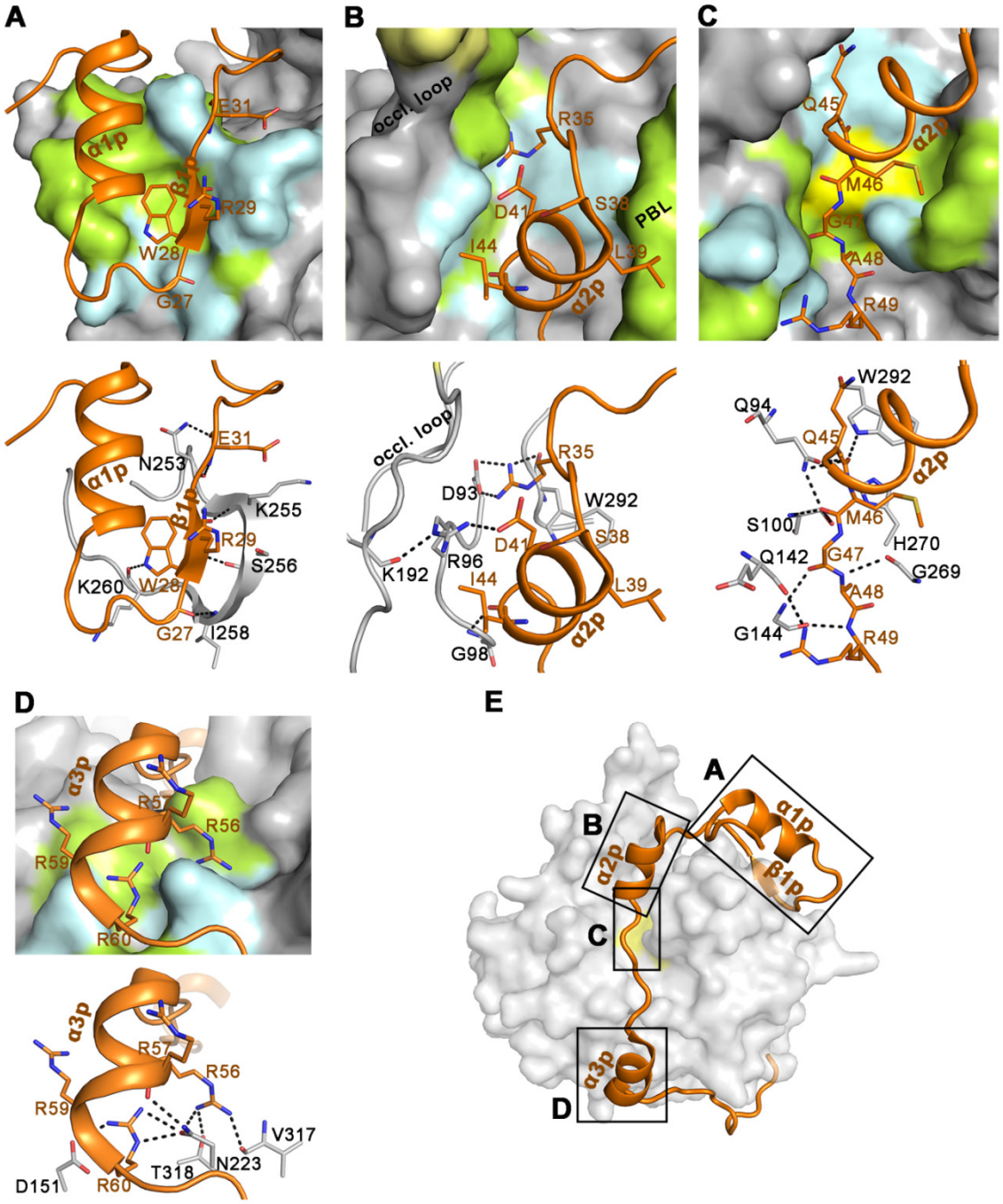


Figure 2. The interface between the pro-peptide and the enzyme core

(A–D) In the SmCB1 zymogen, the pro-peptide interacts with several regions of the enzyme core presented in the following panels: (A) the “pro-peptide binding loop” (PBL), (B) a crevice between PBL and the occluding loop, (C) the active site cleft, and (D) the region binding the α3p helix. The pro-peptide is shown in cartoon representation in orange; the residues forming hydrogen bonds (or having a specific structural role) are depicted as sticks. In the upper panels, the enzyme core is shown as a gray surface with the residues forming the interface hydrogen bonds in pale cyan and the contact residues (<4.0 Å) in light green. In the lower panels, the residues of the enzyme core

forming the interface hydrogen bonds are shown in stick representation in gray. The hydrogen bonds between the pro-peptide and the enzyme core are indicated (black dashed lines).
(E) The regions presented in panels A–D are boxed in the SmCB1 zymogen model. The pro-peptide is shown as an orange cartoon and the enzyme core as a gray surface; the position of the active site is highlighted by catalytic residues Cys100 and His270 in yellow.

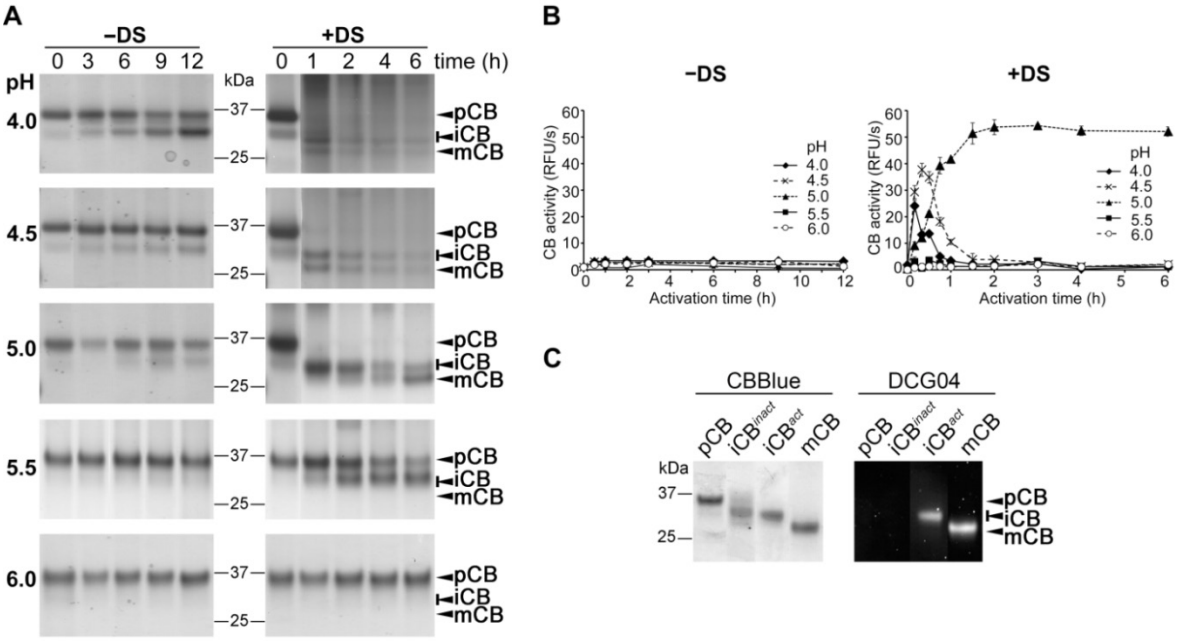


Figure 3. The SmCB1 zymogen is auto-activated in the presence of a sulfated polysaccharide

The SmCB1 zymogen (75 nM) was incubated in the presence or absence of dextran sulfate (DS) (10 µg/ml) at various pH values, and the reaction mixture was analyzed at different time points.

(A) The autoprocessing forms were resolved by SDS-PAGE and visualized by protein staining. The position is indicated for the zymogen (pCB), mature enzyme (mCB), and intermediates (iCB, including an inactive 32 kDa form iCB^{inact} and an active 30–31 kDa form iCB^{act}).

(B) Proteolytic activities generated during the autoprocessing were determined in a kinetic assay with the fluorogenic peptide substrate Z-Phe-Arg-AMC at pH 6.0. Mean values ±SE are expressed in relative fluorescence units (RFU/s).

(C) Imaging of the individual processing forms of SmCB1 was performed with the active-site probe Green-DCG-04. The proteins were treated with the probe, resolved by SDS-PAGE, and visualized with a fluorescence imager (DCG04) and by protein staining (CBBlue). The processing forms were as follows: purified zymogen (pCB) and mature enzyme (mCB); intermediates obtained by the autoprocessing of pCB at pH 4.0 in the presence (iCB^{act}) and at pH 5.0 in the absence (iCB^{inact}) of DS.

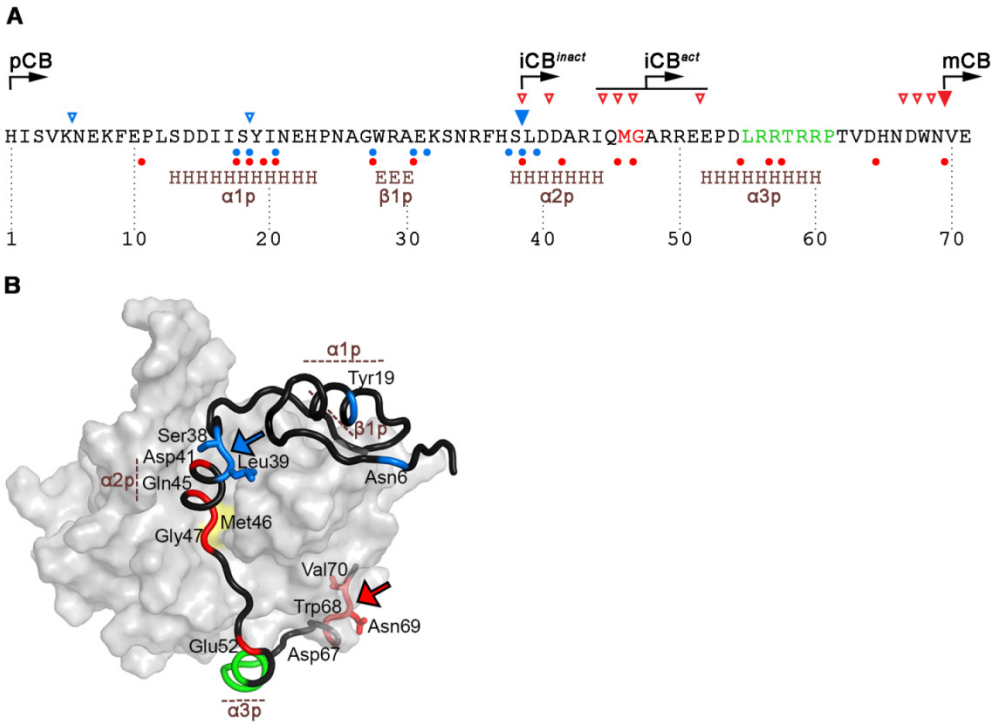


Figure 4. Autoprocessing cleavage sites in the pro-peptide of SmCB1

(A) The amino acid sequence of the SmCB1 pro-peptide is shown with secondary structure elements derived from the crystallographic model (H, α-helix; E, β-strand). The heparin-binding motif is highlighted in green and the residues that interact with the S1 and S1´ subsites of the active site are in red. The triangles above the sequence indicate cleavage sites determined by the N-terminal protein sequencing of molecular forms generated during the acidic autoprocessing of the SmCB1 zymogen. Processing products were obtained in the presence (red triangles) or absence (blue triangles) of DS at various pH and time points. The solid triangles mark the major final cleavage sites, whereas the small open triangles denote the intermediate cleavage sites. The position of N-termini is indicated for the zymogen (pCB), inactive (iCB^{inact}) and active (iCB^{act}) intermediates, and mature enzyme (mCB). The dots below the sequence indicate the cleavage sites, which were identified in synthetic pro-peptide segments digested by mature SmCB1 in the presence (red dots) or absence (blue dots) of DS (Figure S2).

(B) The 3D structure of the SmCB1 zymogen shows the location of the autoprocessing sites represented by the triangles in (A). The pro-peptide is depicted as a black cartoon with the heparin-binding motif in green, the mature enzyme core is shown as a gray surface with the catalytic residues Cys100 and His270 in yellow. In the pro-peptide, the DS-dependent (red) and DS-independent (blue) cleavage sites are highlighted and marked (as C-terminal residues of cleaved bonds). The arrows indicate the processing sites generating the inactive intermediate iCB^{inact} (blue) and mature enzyme (red); the side chains of the residues flanking the cleavage site are shown.

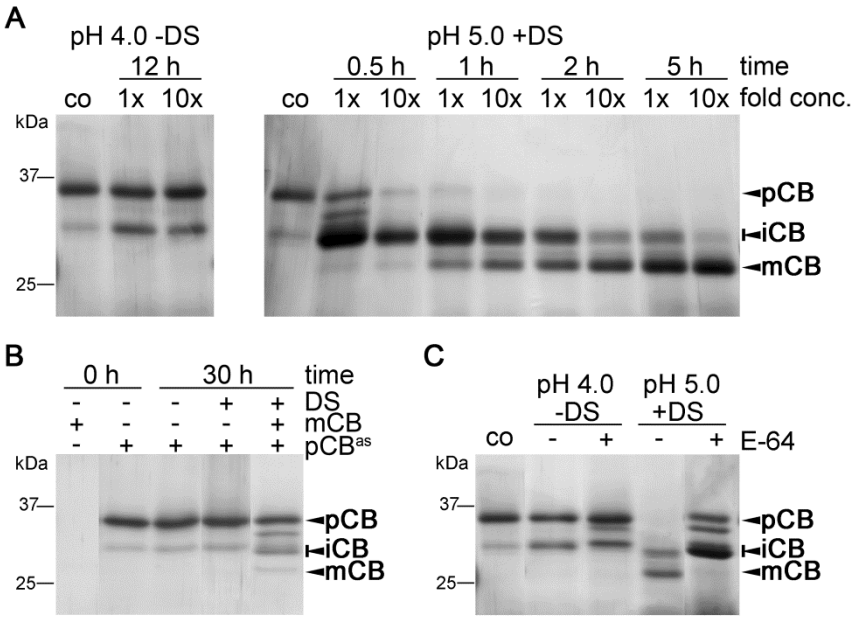


Figure 5. Molecularity of the autocatalytic processing of the SmCB1 zymogen

The reaction mixtures with autoprocessing forms were resolved by SDS-PAGE and visualized by protein staining. The position is marked for the zymogen (pCB), intermediates (iCB), and mature enzyme (mCB); co, unprocessed pCB control.

(A) The effect of the zymogen concentration on the autoprocessing. The pCB was incubated at 75 nM and 750 nM concentrations (1x and 10x, respectively) and analyzed at the indicated times. The reaction was performed in the presence and absence of DS (10 and 100 µg/ml for 1x and 10x reactions, respectively) at pH 4.0 and 5.0, optimal for the DS-independent and -dependent processing, respectively.

(B) The intermolecular processing of the active-site mutated zymogen (C100S zymogen, pCB^{as}) by the endogenous mature enzyme (mCB). The pCB^{as} and endogenous mCB (30 : 1 molar ratio) were incubated at pH 5.0 for the indicated time in the presence and absence of DS (10 µg/ml). Note that the mCB band is derived from pCB^{as} since endogenous mCB was added in a catalytic amount.

(C) The effect of the active-site inhibitor E-64 on the zymogen autoprocessing. The pCB (75 nM) was incubated in the presence and absence of E-64 (5 µM) with DS (10 µg/ml) for 5 h or without DS for 12 h. The reaction was performed at pH 4.0 and 5.0, optimal for the DS-independent and -dependent processing, respectively.

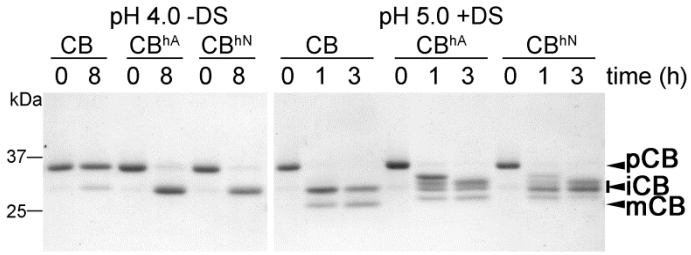


Figure 6. Autoprocessing of the SmCB1 zymogen containing a mutated heparin-binding motif in the pro-peptide

The zymogens of SmCB1 (CB) and its heparin-binding motif mutants, A57/A59 and N57/N59 (CB^{hA} and CB^{hN}, respectively) (75 nM) were incubated in the presence and absence of DS (10 µg/ml) and analyzed at the indicated times. The reaction was performed at pH 4.0 and 5.0, optimal for DS-independent and -dependent processing, respectively. The reaction mixtures were resolved by SDS-PAGE and visualized by protein staining. The position is marked for the zymogen (pCB), intermediates (iCB), and mature enzyme (mCB). The mCB form was not processed from pCB^{hA} and pCB^{hN} even at longer incubation times.

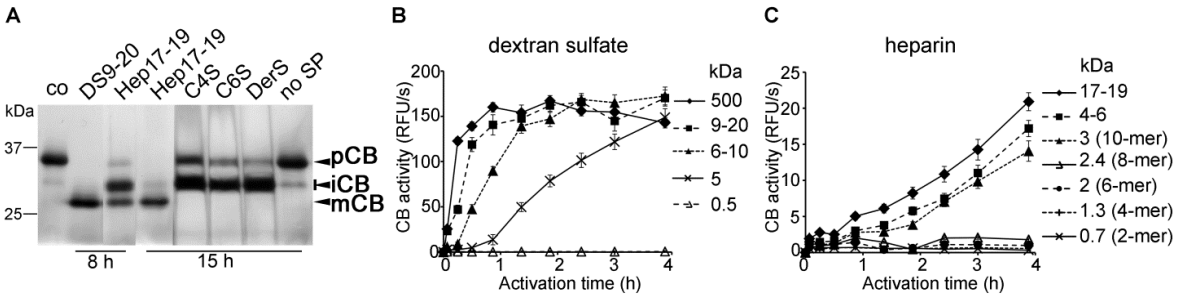


Figure 7. The effect of different sulfated polysaccharides on the auto-activation processing of the SmCB1 zymogen

The SmCB1 zymogen (75 nM) was incubated in the presence of different sulfated polysaccharides (10 µg/ml) at pH 5.0, and the reaction mixtures were analyzed.

(A) The reaction mixtures incubated for the indicated times were resolved by SDS-PAGE and visualized by protein staining. The position is marked for the zymogen (pCB), intermediates (iCB), and mature enzyme (mCB); co, unprocessed pCB control. The applied sulfated polysaccharides: dextran sulfate of 9–20 kDa (DS9–20), heparin of 17–19 kDa (Hep17-19), chondroitin-4-sulfate (C4S), chondroitin-6-sulfate (C6S), and dermatan-sulfate (DerS). (C4S, C6S and DerS were in the range of 20–60 kDa.)

(B, C) At the indicated time points, proteolytic activities generated in the reaction mixtures were measured in a kinetic assay with the fluorogenic peptide substrate Z-Phe-Arg-AMC at pH 6.0. Mean values ±SE are expressed in relative fluorescence units (RFU/s); the scale in (C) is zoomed to display slower heparin-mediated activation. The molecular mass of the applied dextran sulfates (B) and heparins (C) is indicated; the number of saccharide units in oligosaccharides is shown in parentheses.

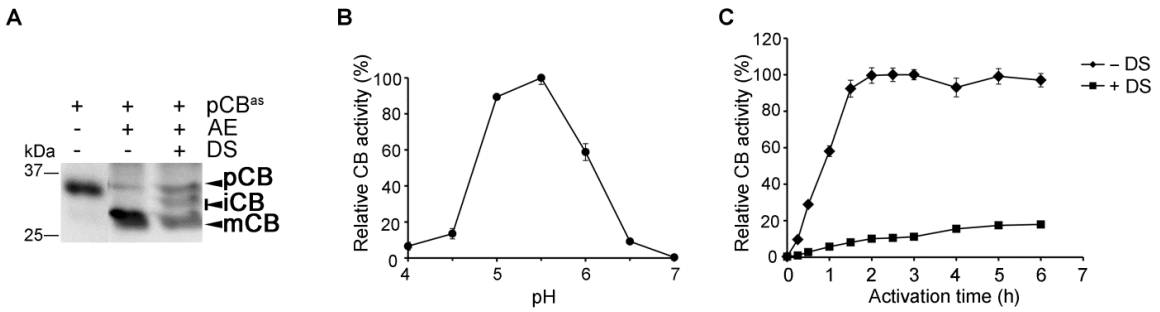


Figure 8. Trans-activation of the SmCB1 zymogen by asparaginyl endopeptidase (legumain)

The SmCB1 zymogen (75 nM) was incubated with SmAE, an asparaginyl endopeptidase (legumain) from *S. mansoni* (5 nM), and the reaction mixture was analyzed.

(A) The reaction was performed in the presence or absence of DS (10 µg/ml) at pH 6.0 for 6 h. The composition of the reaction mixtures is as indicated: the active-site mutant of the SmCB1 zymogen (pCB^{as}), SmAE (AE), and DS. The SmCB1 processing forms were resolved by SDS-PAGE and visualized by immunostaining. The position is marked for the zymogen (pCB), intermediates (iCB), and mature enzyme (mCB).

(B) The SmCB1 zymogen and SmAE were incubated at various pH values for 2 h. The generated proteolytic activity of SmCB1 was measured in a kinetic assay with the fluorogenic peptide substrate Z-Phe-Arg-AMC at pH 6.0. Mean values ±SE are expressed in relative fluorescence units (RFU/s).

(C) The SmCB1 zymogen and SmAE were incubated at pH 6.0 in the presence or absence of DS (10 µg/ml), and the SmCB1 activities were measured as in (B) and analyzed in a time course. The activity of the generated mature SmCB1 is not significantly affected by DS under the assay conditions.

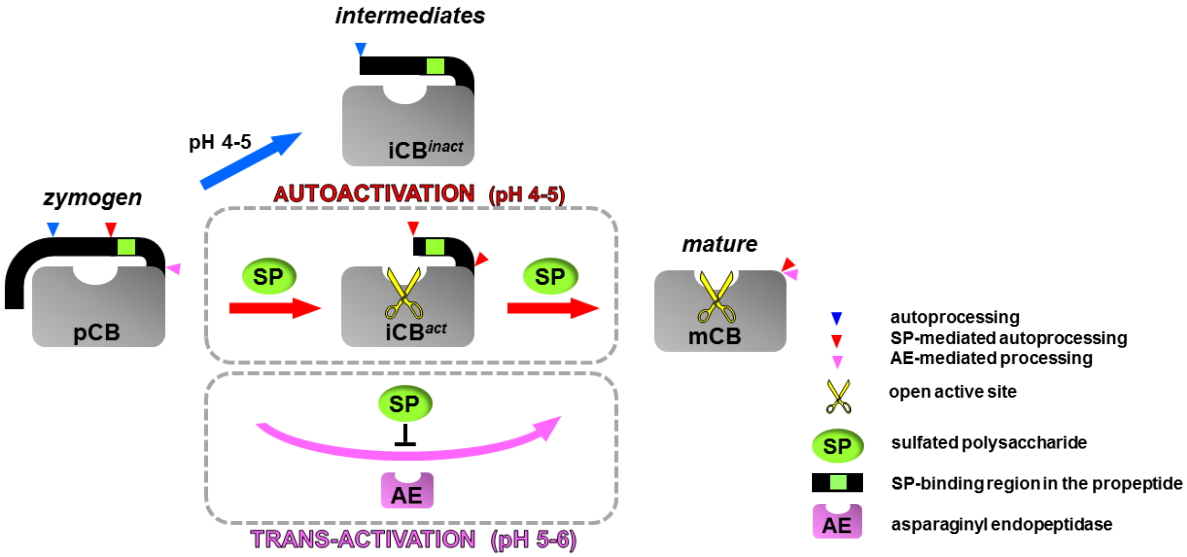


Figure 9. The activation pathway of SmCB1

A mechanistic model for the proteolytic pathway of the activation of the SmCB1 zymogen (pCB) with two alternative routes. The autocatalytic processing at acidic pH is nonproductive, leading to the inactive intermediate iCB^{inact} with partially removed pro-peptide blocking the active site. The interaction of sulfated polysaccharides (SP) with the heparin-binding motif in the pro-peptide triggers productive autoactivation, which generates the active intermediate iCB^{act} (having the accessible active site unhindered by the residual pro-peptide), and finally the mature enzyme (mCB) with fully removed pro-peptide. The trans-activation mediated by the proteolytic action of *S. mansoni* asparaginyl endopeptidase/legumain (AE) leads to the same mature enzyme, but this processing step is suppressed by SPs. Therefore, SPs function as a molecular switch between both activation routes.

Supplemental Information

Activation route of the *Schistosoma mansoni* cathepsin B1 drug target: structural map with a glycosaminoglycan switch

Adéla Jílková, Martin Horn, Pavlína Řezáčová, Lucie Marešová, Pavla Fajtová, Jiří Brynda, Jiří Vondrášek, James H. McKerrow, Conor R. Caffrey, and Michael Mareš

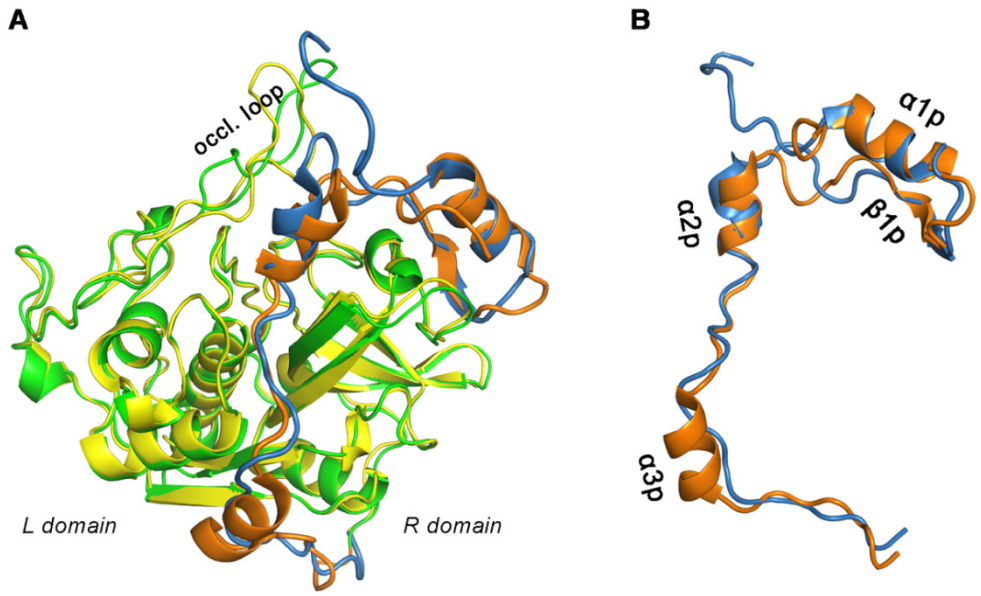


Figure S1, Related to Figure 1. Comparison of the zymogen structures of SmCB1 and human cathepsin B

(A) Superposition of the zymogens of SmCB1 and human cathepsin B (PDB code: 2PBH) in cartoon representation. The pro-peptide is shown in orange and blue for SmCB1 and human cathepsin B, respectively. The enzyme core is shown in green and yellow for SmCB1 and human cathepsin B, respectively; the occluding loop and core domains are indicated.

(B) Side view of the superimposed pro-peptides of SmCB1 (orange) and human cathepsin B (blue) shown without the enzyme cores. The secondary structure elements in the pro-peptides are labeled; note that the α3p helix (with the heparin-binding motif) of SmCB1 is absent in human cathepsin B.

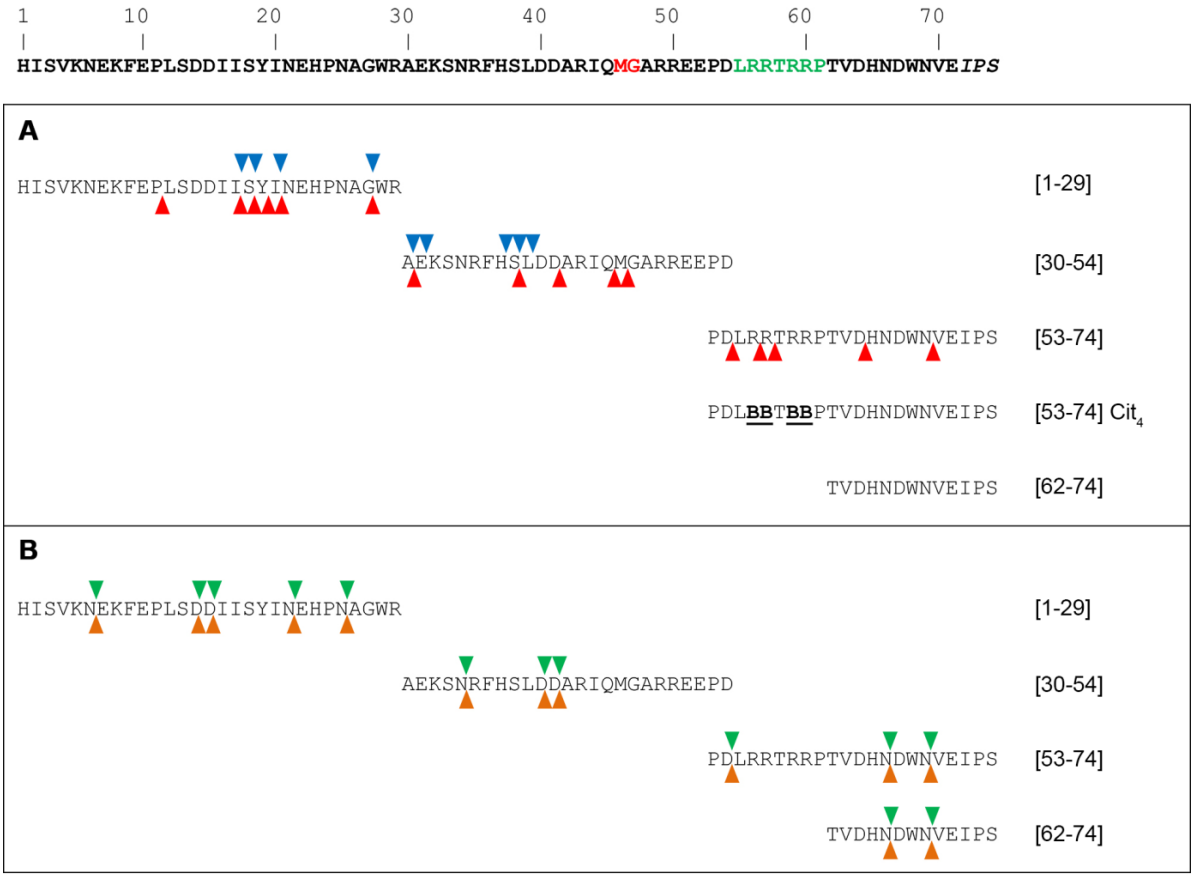


Figure S2, Related to Figure 4. The cleavage map of synthetic peptides derived from the SmCB1 pro-peptide

In the SmCB1 pro-peptide sequence (above the panels), the heparin-binding motif (green) and the residues that interact with the S1 and S1´ subsites of the SmCB1 active site (red) are highlighted; the residues corresponding to the N-terminus of the mature enzyme are in italics. The peptides in the panels were synthesized with capping groups, N-terminal acetyl (except for [1-29]) and C-terminal amide, to mimic peptide linkages and to protect peptides against exopeptidase trimming. Peptides are coded according to their location in the sequence; the substituted residues are in bold and underlined (B denotes citrulline). The cleavage sites (triangles) in the digested peptides were identified by mass spectrometry.

(A) The peptides were digested by mature SmCB1 at pH 5.0 in the presence of DS (red triangles) or at pH 4.0 in the absence of DS (blue triangles). The cleavage sites resulting from the endopeptidase activity of SmCB1 are presented; the secondary C-terminal trimming of endopeptidase fragments (by the peptidyl-dipeptidase activity of SmCB1) is omitted for simplicity.

(B) The peptides were digested by SmAE at pH 6.0 in the presence (orange triangles) or absence of DS (green triangles).

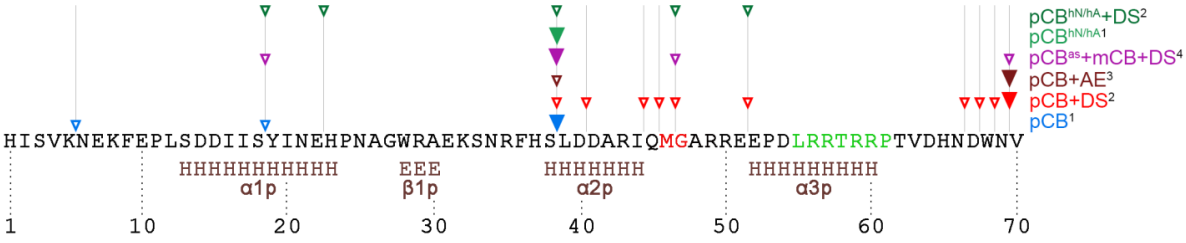


Figure S3, Related to Figure 4, 6, 8. The processing sites in the pro-peptide of the SmCB1 zymogen determined by N-terminal protein sequencing

The amino acid sequence of the SmCB1 pro-peptide is shown with secondary-structure elements derived from the crystallographic model (H, α-helix; E, β-strand). The heparin-binding motif is highlighted in green and the residues that interact with the S1 and S1´ subsites of the active site are in red. The triangles above the sequence indicate the cleavage sites determined by the N-terminal protein sequencing of the molecular forms generated during the proteolytic processing of the SmCB1 zymogen. The solid triangles mark the major final cleavage sites, whereas the open triangles mark the minor cleavage sites. The processing reactions were performed for the SmCB1 zymogen (pCB) and its active-site mutant (pCB^{as}) and heparin-binding site mutants (pCB^{hA}, pCB^{hN}). The trans-activation experiments were assisted by mature SmCB1 (mCB) and SmAE (AE). The reaction conditions analyzed at various time points were as follows: [1] pH 4.0 in the absence of DS, [2] pH 5.0 in the presence of DS, [3] pH 6.0 in the presence of AE, and [4] pH 5.0 in the presence of DS and mCB.

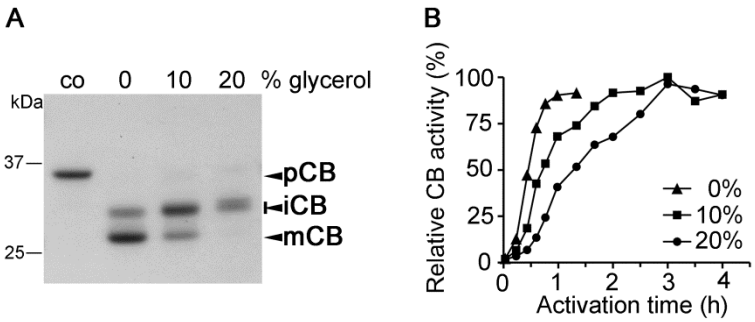


Figure S4, Related to Figure 5. The effect of glycerol on the autoactivation processing of the SmCB1 zymogen

The SmCB1 zymogen (75 nM) was incubated with DS (10 µg/ml) at pH 5.0 in the absence or presence of glycerol (10 or 20%), and the reaction mixtures were analyzed.

(A) The reaction mixtures incubated for 5 h were resolved by SDS-PAGE and visualized by protein staining. The position is marked for the zymogen (pCB), intermediates (iCB), and mature enzyme (mCB); co, unprocessed pCB control.

(B) At the indicated time points, proteolytic activities were measured in a kinetic assay with the fluorogenic peptide substrate Z-Phe-Arg-AMC at pH 6.0. Mean values are expressed in relative fluorescence units (RFU/s); SE values are within 10% of the mean. Note that the applied glycerol concentrations do not significantly influence the activity of the purified mature enzyme under these experimental conditions.

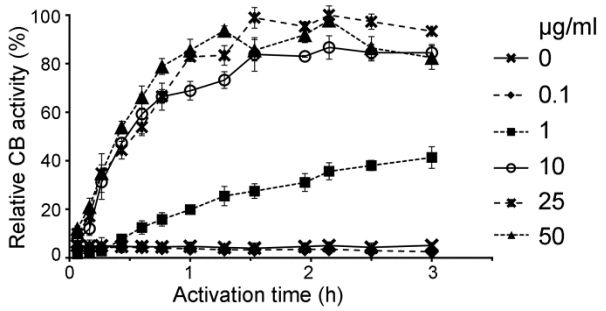


Figure S5, Related to Figure 7. The Effect of dextran-sulfate concentration on the autoactivation of the SmCB1 zymogen

The SmCB1 zymogen (75 nM) was incubated with various concentrations (0–50 µg/ml) of dextran sulfate at pH 5.0. At the indicated time points, proteolytic activities generated in the reaction mixture were measured in a kinetic assay with the fluorogenic peptide substrate Z-Phe-Arg-AMC at pH 6.0. Mean values ±SE are expressed in relative fluorescence units (RFU/s).

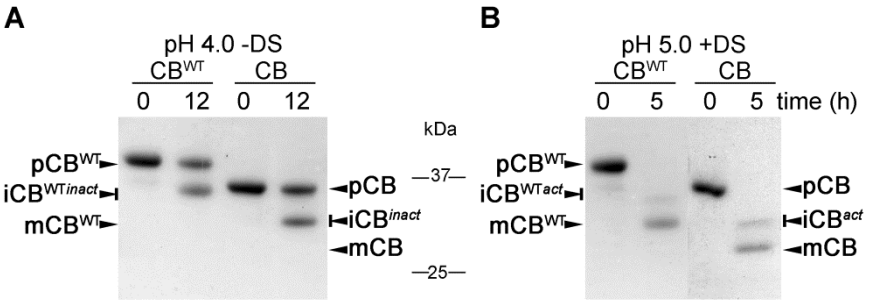


Figure S6, related to Figure 3. N-glycosylation of the SmCB1 zymogen has no effect on its autocatalytic processing

Two forms of SmCB1 zymogens were used for analysis: the wild-type form (pCB^{WT}) of 42 kDa and its non-glycosylated mutant (pCB) of 37 kDa with two mutated N-glycosylation signals. The zymogens (75 nM) were incubated (A) in the absence of dextran sulfate (-DS) for 12 h at pH 4.0, and (B) in the presence of dextran sulfate (+DS, 10 µg/ml) for 5 h at pH 5.0; these optimized conditions were selected according to Figure 3. The reaction mixtures were resolved by SDS-PAGE and visualized by protein staining. The positions are marked for the zymogens (pCB^{WT}, pCB), inactive ($iCB^{WTinact}$, iCB^{inact}) and active (iCB^{WTact}, iCB^{act}) intermediates, and mature enzymes (mCB^{WT}, mCB). The processing forms derived from pCB^{WT} are cleaved at the same sites as those derived from pCB (Figure 3); their identity was confirmed by N-terminal sequencing.

Table S1, Related to Figure 1. List of contacts formed between SmCB1 pro-peptide and mature enzyme in SmCB1 zymogen and intermediate form. Analysis of contacts between the SmCB1 pro-peptide and the enzyme core was performed using program CONTACT (CCP4, 1994). The distance cutoffs were set between 0 Å and 4.0 Å. For each SmCB1 residue, total number of contacts is indicated including H-bonds (residues forming H-bonds are *in bold*).

Zymogen SmCB1				Intermediate SmCB1			
Residue		Contacts	H-bond	Residue		Contacts	H-bond
Phe	9	1					
Pro	11	5					
Leu	12	5					
Ile	16	3					
Ile	17	3					
Tyr	19	4					
His	23	2					
Asn	25	2					
Ala	26	6					
Gly	**27**	8	1				
Trp	**28**	39	1				
Arg	**29**	12	2				
Ala	30	3					
Glu	**31**	14	2				
Lys	32	1					
Ser	33	3					
Asn	34	4					
Arg	**35**	33	3				
Phe	36	12					
Leu	39	5					
Asp	**41**	4	1				
Ala	42	2					
Arg	43	4					
Ile	**44**	11	1				
Gln	**45**	15	3				
Met	**46**	29	3				
Gly	**47**	20	1	Gly	47	8	
Ala	**48**	16	1	**Ala**	**48**	18	1
Arg	**49**	16	2	**Arg**	**49**	18	2
Arg	50	6		**Arg**	**50**	18	2
Glu	**51**	15	1	**Glu**	**51**	10	1
Arg	**56**	30	4	**Arg**	**56**	30	4
Arg	59	1		Arg	59	4	
Arg	**60**	26	3	**Arg**	**60**	40	3
Pro	**61**	11	1	**Pro**	**61**	9	1
Thr	62	7		Thr	62	8	
Val	**63**	16	2	**Val**	**63**	13	2
Asp	64	5		Asp	64	5	
His	**65**	17	2	**His**	**65**	17	2
Trp	**68**	12	1	**Trp**	**68**	14	1
Val	70	4		Val	70	2	
Glu	**71**	18	1	**Glu**	**71**	18	1

Table S2, Related to Figure 1. List of contacts stabilizing propart in SmCB1 zymogen and intermediate form. Analysis of contacts between residues within the pro-peptide was performed using program CONTACT (CCP4, 1994). The distance cutoffs were set between 0 Å and 4.0 Å. For each SmCB1 residue, total number of contacts is indicated including H-bonds (residues forming H-bonds are *in bold*).

Zymogen SmCB1				Intermediate SmCB1			
Residue		Contacts	H-bond	Residue		Contacts	H-bond
Lys	8	4					
Phe	9	19					
Glu	**10**	24	1				
Pro	**11**	22	1				
Leu	12	18					
Ser	**13**	36	6				
Asp	**14**	17	2				
Asp	**15**	22	2				
Ile	**16**	34	4				
Ile	**17**	37	3				
Ser	**18**	29	3				
Tyr	**19**	34	3				
Ile	**20**	34	4				
Asn	**21**	38	3				
Glu	**22**	19	2				
His	**23**	40	6				
Pro	**24**	26	1				
Asn	**25**	18	2				
Ala	**26**	15	2				
Gly	27	5					
Trp	**28**	24	1				
Arg	29	19					
Ala	**30**	18	2				
Glu	31	11					
Lys	32	11					
Ser	**33**	24	2				
Asn	34	9					
Arg	**35**	22	3				
Phe	**36**	41	1				
His	**37**	21	1				
Ser	**38**	31	1				
Leu	**39**	17	2				
Asp	**40**	28	4				
Asp	**41**	37	3				
Ala	**42**	30	4				
Arg	**43**	27	4				
Ile	**44**	19	2				
Gln	**45**	35	5				
Met	46	18					
Gly	47	3		Gly	47	4	
Ala	48	7		Ala	48	4	
Arg	49	9		Arg	49	11	
Arg	50	8		Arg	50	7	
Glu	**51**	35	3	**Glu**	**51**	39	3
Glu	**52**	34	3	**Glu**	**52**	29	2
Pro	**53**	31	2	**Pro**	**53**	33	1
Asp	**54**	25	4	**Asp**	**54**	29	2
Leu	**55**	38	3	**Leu**	**55**	36	4
Arg	**56**	31	1	**Arg**	**56**	31	1
Arg	**57**	21	2	**Arg**	**57**	32	1
Thr	**58**	21	2	**Thr**	**58**	25	3

Arg	**59**	31	2	**Arg**	**59**	28	3
Arg	**60**	32	1	**Arg**	**60**	29	2
Pro	61	14		Pro	61	14	
Thr	62	11		Thr	62	10	
Val	63	12		Val	63	11	
Asp	**64**	29	4	**Asp**	**64**	25	4
His	**65**	21	3	**His**	**65**	21	3
Asn	**66**	23	3	**Asn**	**66**	18	3
Asp	**67**	9	1	Asp	67	9	
Trp	**68**	26	1	**Trp**	**68**	27	1
Asn	69	12		Asn	69	10	
Val	70	20		Val	70	19	
Glu	71	5		Glu	71	5	

Table S3, Related to Experimental procedures. X-ray data collection and refinement statistics. Numbers in parentheses refer to the highest-resolution shell.

	Zymogen SmCB1	Intermediate SmCB1	Mature SmCB1
Data collection statistics			
Wavelenght (Å)	0.914	0.979	0.97857
Temperature (K)	100	100	100
Space group	P1	*C*2	$P2_12_12_1$
a, b, c (Å)	54.3, 83.2, 84.6	93.86, 50.99, 62.55	33.14, 79.17, 90.48
α, β, γ (°)	76.9, 86.6, 71.1	90.0, 91.31, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50-1.95 (1.98-1.95)	50-1.9 (1.93-1.90)	50-1.30 (1.32-1.30)
Number of unique reflections	78650	23559	59658
Redundancy	2.3 (1.9)	3.0 (2.7)	7.3 (5.7)
Completeness (%)	78.4 (42.1)	99.4 (98.7)	89.9 (49.3)
R_{merge} [a]	3.7 (29.6)	8.2 (35.0)	7.7 (43.8)
Average I/σ (I)	25.9 (3.1)	18.2 (3.13)	33.7 (4.0)
Wilson B ($Å^2$)	25.2	20.2	13.5
Refinement statistics			
Resolution range (Å)	38.4-1.95 (1.99-1.95)	26.6-1.9 (1.95-1.90)	19.79-1.3 (1.33-1.30)
No. of reflections in working set	71614 (1585)	22109 (1573)	50812 (2086)
No. of reflections in test set	3756 (84)	1201 (89)	2711 (116)
R value (%)	19.93 (24.7)	16.31 (23.2)	13.74 (17.1)
R_{free} value [c] (%)	25.24 (28.5)	21.64 (27.1)	17.67 (21.5)
Number of atoms in AU [d] all / protein / solvent	10371 / 9828 / 519	2619/2278/340/1Na ion	2531/2099/437
Average ADP [e] for all/main-chain/side-chain and solvent atoms ($Å^2$)	25.29 / 24.44 / 26.43	23.2/21.9/32.1	11.5/7.82/14.09
RMSD bond length (Å)	0.012	0.013	0.013
RMSD bond angle (°)	1.303	1.487	1.437
Ramachandran plot statistics [f]			
Residues in favored regions (%)	96.1	96.6	96.3
Residues in allowed regions (%)	99.9	100	100
PDB code	4I04	4I05	4I07

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where the $I_i(hkl)$ is an individual intensity of the i^{th} observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity of reflection *hkl* with summation over all data.
[b] *R*-value = $||F_o| - |F_c||/|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[c] R_{free} is equivalent to *R* value but is calculated for 5 % of the reflections chosen at random and omitted from the refinement process (Brunger, 1992).
[d] AU, asymmetric unit
[e] ADP, atomic displacement parameter, formally B-factor; determined by BAVERAGE (CCP4, 1994).
[f] as determined by Molprobity (Lovell et al., 2003).

SUPPLEMENTAL REFERENCES

Brunger,A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472-475.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D. Biol. Crystallogr. *50*, 760-763.

Lovell,S.C., Davis,I.W., Arendall,W.B., III, de Bakker,P.I., Word,J.M., Prisant,M.G., Richardson,J.S., and Richardson,D.C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins *50*, 437-450.

5.5 Publikace č. 5: Katepsinové proteasy v patologii

5.5.1 Souhrn

Cílem této publikace je podat přehled současných trendů týkajících se problematiky katepsinových proteas, které se účastní patologických procesů se zaměřením na témata řešená na ÚOCHB AV ČR. V práci jsou přehledně zpracovány trávicí katepsiny hematofágních parazitů s krevničkou jako modelovým druhem a jsou zde prezentovány původní výsledky z publikací č. 1 až 4.

Můj podíl na práci zahrnoval analýzu krystalových struktur, molekulární grafiku a účast na přípravě manuskriptu.

5.5.2 Publikace č. 5

Horn, M., **Jílková**, **A.**, Mareš, M.

Katepsinové proteasy v patologii

Chemické listy, 4, 358-363 (2014)

Chem. Listy *108*, 358–363 (2014) Referát

KATEPSINOVÉ PROTEASY V PATOLOGII

MARTIN HORN[a], ADÉLA JÍLKOVÁ[a,b] a MICHAEL MAREŠ[a]

[a] *Ústav organické chemie a biochemie AV ČR, v.v.i., Flemingovo nám. 2, 166 10 Praha,* [b] *Katedra biochemie, Přírodovědecká fakulta, Univerzita Karlova v Praze, Hlavova 2030/8, 128 40 Praha 2*
mares@uochb.cas.cz

Došlo 6.12.13, přijato 6.1.14.

Klíčová slova: katepsiny, proteasy, proteolýza, proteasové inhibitory, chemoterapeutika, hematofágní parazité

Obsah

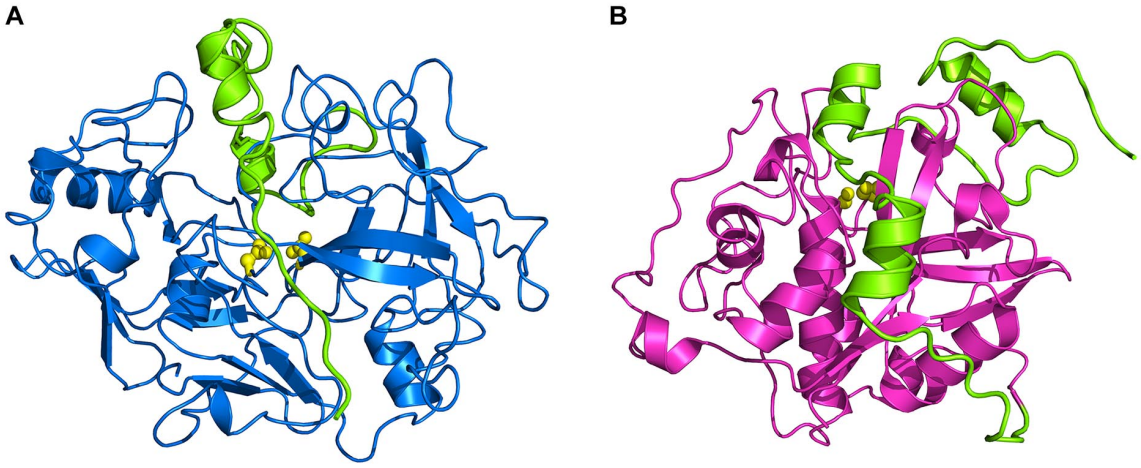
1. Úvod

Jméno katepsiny je odvozeno z řeckého „kathepsein“ – trávit a bylo původně použito jako označení proteas aktivních v kyselém prostředí lysosomů savčích buněk. Tato organela obsahuje kolem 50 kyselých hydrolytických enzymů, které se podílejí na intracelulárním katabolismu. Katepsiny představují soubor hlavních proteolytických enzymů (proteas nebo obecněji peptidas) lysosomů, jenž je např. v lidském genomu tvořen 15 katepsiny. Tato skupina je přitom heterogenní z pohledu molekulární evoluce a jednotlivé katepsiny spadají do tří nepříbuzných tříd proteas klasifikovaných jako cysteinové, aspartátové a serinové proteasy. Savčí lysosomální katepsiny byly původně považovány pouze za nástroj pro vnitrobuněčnou nespecifickou degradaci proteinů v kyselém prostředí endodosomálního-lysosomálního kompartmentu, kde odbourávají intracelulární nebo endocytované extracelulární proteiny. Tento pohled na katepsiny se nicméně změnil v posledních dvou desetiletích v několika směrech. Jednak byly aktivní katepsiny lokalizovány také v dalších buněčných kompartmentech, jako je jádro, cytoplasma nebo plasmatická membrána a jednak byly nalezeny sekretované do extracelulárního prostředí. V různých kompartmentech pak katepsiny plní odlišné a specifické fyziologické úlohy jako je např. prezentace antigenů, přestavba extracelulární matrix nebo procesování hormonů. Jejich aktivita je přitom přesně regulována a chyby v tomto kontrolním procesu jsou spojené se vznikem řady patologií, jako jsou např. nádorová a neurodegenerativní onemocnění, artritida a osteoporóza. Další nový směr v pochopení funkcí katepsinů vychází z objevu příbuzných proteas katepsinového typu u řady organismů. Studovány jsou zejména trávicí katepsiny u parazitů a patogenů, které se klíčovým způsobem účastní interakcí s hostitelem, a u herbivorních hmyzích škůdců. Katepsinové proteasy byly identifikovány také u rostlin, kde se podílejí na obranyschopnosti proti patogenům. Uvedené rozšiřování skupiny původních savčích katepsinů a velká rozmanitost jejich funkcí mění původní historické vymezení pojmu katepsin a vede k jeho volnějšímu používání při klasifikaci identifikovaných eukaryotických proteas. Tato práce se soustředí na katepsiny člověka a parazitů, které představují perspektivní cílové molekuly v biomedicíně, na něž se zaměřuje vývoj nových chemoterapeutik.

2. Struktura a aktivita katepsinů

2.1. Aspartátové katepsiny

Lidské katepsiny D a E patří do třídy aspartátových proteas, konkrétně rodiny pepsinu (rodina A1 z klanu AA podle proteasové databáze MEROPS, http://merops.sanger.ac.uk). Oba aspartátové katepsiny působí jako endopeptidasy (tj. štěpí substrát uvnitř polypeptidového řetězce) s maximální aktivitou v kyselé oblasti pH 3,5 až 5,0. Ve struktuře substrátu jsou preferovány hydrofobní aminokyseliny kolem štěpené peptidové vazby v pozicích P1 a P1′. Katepsiny D a E jsou biosyntetizovány ve formě neaktivního proenzymu (zymogenu), který je z endoplasmatického retikula transportován do lysosomů a v jejich kyselém prostředí aktivován. Proces aktivace zahrnuje proteolytické odštěpení aktivačního peptidu (tzv. propeptidu) z N-konce molekuly zymogenu. Prostorová struktura aktivního enzymu je tvořena dvěma doménami, které jsou odděleny štěrbinou aktivního místa, kde se nacházejí dva katalytické aspartátové zbytky (obr. 1A). Propeptid se ve struktuře zymogenu váže do aktivního místa a zamezuje tak přístupu substrátu. Lidský katepsin D je monomerní glykoprotein s molekulovou hmotností 48 kDa, jehož aktivace probíhá částečně autokatalyticky a dále působením jiných lysosomálních proteas. Lidský katepsin E je aktivován autokatalyticky a poté vytváří dimer (84 kDa), ve kterém jsou dvě monomerní jednotky spojeny disulfidovým můstkem.



Obr. 1. **3D struktura neaktivního prekurzoru (zymogenu) aspartátových a cysteinových katepsinů: (A) lidský katepsin D, (B) katepsin B1 z krevničky střevní (*Schistosoma mansoni*).** Aktivační peptid (propeptid, vyznačen zeleně) blokuje aktivní místo a interaguje s katalytickými aminokyselinovými zbytky (žlutě); během aktivace enzymu je tento peptidový segment proteolyticky odštěpen

2.2. Cysteinové katepsiny

Genom člověka kóduje celkem 11 katepsinů (tab. I), které patří do třídy cysteinových proteas z rodiny papainu (rodina C1 klanu CA). Jedná se o monomerní glykoproteiny s molekulovou hmotností v rozmezí 25–35 kDa; výjimkou je katepsin C, který tvoří tetramerní oligomery (180 až 200 kDa). Cysteinové katepsiny mají optimální aktivitu v mírně kyselém pH, pouze katepsin S je funkční i v neutrální oblasti pH. Jednotlivé cysteinové katepsiny jsou endopeptidasy nebo exopeptidasy, které štěpí substrát uvnitř polypeptidového řetězce resp. na jeho koncích. Typickými endopeptidasami jsou např. katepsiny L, K a S. Mezi exopeptidasy patří (1) aminopeptidasy katepsin H a katepsin C (dipeptidylaminopeptidasa I), které odštěpují po jedné resp. dvou aminokyselinách z N-konce substrátu a (2) karboxypeptidasy katepsin X a katepsin B, které odštěpují po jedné resp. dvou aminokyselinách z C-konce substrátu (katepsin B navíc tuto karboxydipeptidasovou aktivitu kombinuje s aktivitou endopeptidasovou).

Cysteinové katepsiny jsou stejně jako ostatní skupiny katepsinů biosyntetizovány ve formě neaktivního zymogenu, který je aktivován proteolytickým odstraněním N-koncového propeptidu, jenž blokuje aktivní místo (obr. 1B). Prostorová struktura aktivovaného katepsinu je tvořena dvěma doménami; mezi nimi se nachází štěrbina aktivního místa se dvěma katalytickými zbytky cysteinu a histidinu. Polypeptidový řetězec substrátu se váže do aktivního místa aminokyselinovými zbytky P3-P2-P1-P1′-P2′ (štěpena je vazba mezi P1 a P1′), kterým odpovídají vazebná podmísla S3 až S2′. U katepsinů s endopeptidasovou aktivitou umožňuje aktivní místo vazbu do všech podmíst, zatímco u katepsinů s exopeptidasovou aktivitou je počet podmíst omezen. Dochází k tomu pomocí strukturních prvků, které blokují vazebná podmísla S3/S2 u aminopeptidas nebo S3′/S2′ u karboxypeptidas, což dovoluje vazbu pouze N- resp. C-konce řetězce substrátu (obr. 2). Konkrétní design vazebných podmíst určuje preferované aminokyselinové zbytky substrátu a tak řídí specifitu jednotlivých cysteinových katepsinů.

2.3. Serinové katepsiny

Katepsin A a G patří ve třídě serinových proteas do odlišných rodin S10 (rodina karboxypeptidasy Y), resp. S1 (rodina chymotrypsinu). Aktivní místo těchto serinových katepsinů obsahuje katalytickou triádu aminokyselin serinu, histidinu a kyseliny asparagové. Lidský katepsin A (karboxypeptidasa A, lysosomální protektivní protein) je multifunkčním lysosomálním enzymem, který má v kyselém pH karboxypeptidasovou aktivitu a v neutrálním pH deamidasovou a esterasovou aktivitu. Katepsin A se vyskytuje ve formě dimeru (98 kDa) a tvoří makrokomplexy s dalšími třemi lysosomálními hydrolasami, které stabilizuje. Lidský katepsin G je monomerním glykoproteinem (28 kDa) s pH optimem cca 8. Není typickou lysosomální proteasou, ale vyskytuje se pouze v azurofilních granulích neutrofilních leukocytů příbuzných lysosomům. Více informací k této kapitole lze nalézt v pramenech[1–6].

3. Lidské katepsiny ve fyziologii a patologii

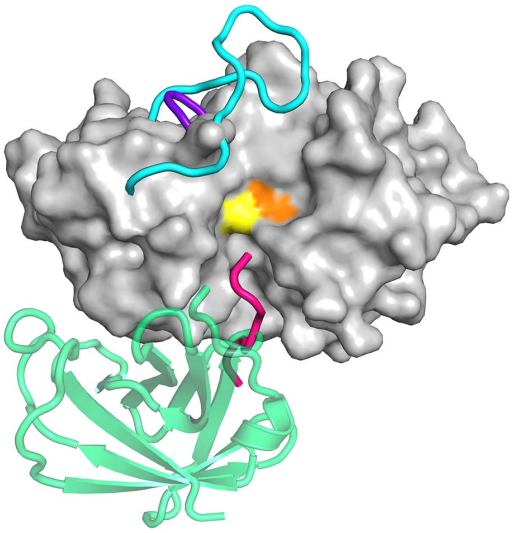
Lidské katepsiny jsou nejlépe prostudovanou skupinou katepsinů pro jejich významnou roli v mnoha fyziologických procesech a přímý vztah k řadě vážných patologií (tab. I). Jsou multifunkčními proteasami, které mohou působit jednak jako nespecifické degradační enzymy a jednak jako vysoce specializované proteasy při modifikaci specifických proteinů. Mezi základní funkce katepsinů patří především (*1*) endosomální/lysosomální odbourávání proteinových substrátů (včetně fragmentace antigenů), (*2*) degradace extracelulární matrix a (*3*) proteolytické děje a kaskády s limitovanou proteolýzou, jako je aktivace zymogenů enzymů, zpracování neuropeptidů, hormonů

Tabulka I
Přehled katepsinových proteas kódovaných v genomu člověka a jejich základní biochemické a biologické charakteristiky

Katepsin	Proteolytická aktivita	Výskyt	Patologie
Aspartátové katepsiny (rodina A1)			
Katepsin D	endopeptidasa	nespecifický	rakovina, revmatoidní artritida, neurodegenerativní onemocnění
Katepsin E	endopeptidasa	buňky imunitního systému, osteoklast	rakovina
Cysteinové katepsiny (rodina C1)			
Katepsin B	endopeptidasa/ karboxydipeptidasa	nespecifický	rakovina, revmatoidní artritida. pankreatitida
Katepsin C (dipeptidyl aminopeptidasa I)	aminodipeptidasa	nespecifický	Papillon-Lefevrův syndrom, Haim-Munkův syndrom
Katepsin F	endopeptidasa	nespecifický	rakovina, neuronální lipofuscinóza
Katepsin H	aminopeptidasa/ endopeptidasa	nespecifický	rakovina, ateroskleróza
Katepsin K	endopeptidasa	osteoklast	osteoporóza, revmatoidní artritida, osteoartróza, pyknodysostóza[a]
Katepsin L	endopeptidasa	nespecifický	rakovina, ateroskleróza, obezita, epidermální hyperplazie, kardiomyopatie, zánětlivá onemocnění
Katepsin O	endopeptidasa	nespecifický	rakovina
Katepsin S	endopeptidasa	antigen prezentující buňky	bronchiální astma, zánět kloubů (artritida), ateroskleróza, lupénka, rakovina, neurodegenerativní onemocnění
Katepsin V	endopeptidasa	brzlík, varlata, rohovka	autoimunitní onemocnění, myasthenia gravis
Katepsin W (lymfopain)	endopeptidasa	T-lymfocyty, NK buňky	není známo
Katepsin X	karboxypeptidasa	nespecifický	rakovina, zánětlivá onemocnění
Serinové katepsiny (rodina S10 a S1)			
Katepsin A (karboxypeptidasa A)	karboxypeptidasa	nespecifický	galaktosialidóza[a]
Katepsin G	endopeptidasa	neutrofilní granulocyty (azurofilní granule)	zánětlivá onemocnění, Papillon-Lefevrův syndrom[a]

[a] Genetické autosomálně recesivní onemocnění

a růstových faktorů. Katepsiny se účastní řady důležitých fyziologických procesů jako např. zpracování a prezentace antigenů na MHC-II molekulách při adaptivní imunitě, aktivace neutrofilů a cytotoxických imunitních buněk, buněčná proliferace a apoptóza, remodelace tkání a resorpce kostí, angiogeneze a homeostáza kůže. Chybná regulace katepsinů je spojená s patologickými stavy, jako jsou např. neurodegenerativní onemocnění, obezita, ateroskleróza, plicní emfyzém, revmatická artritida, osteoporóza a rakovina. Zvláště dvěma posledním uvedeným je věnována v současné době značná pozornost z pohledu vývoje nových strategií léčby. V případě katepsinu K, který se klíčovým způsobem podílí na vzniku osteoporózy, jsou uváděny na trh syntetické inhibitory této proteasy jako antiosteoporotická léčiva. V nádorové patologii se většina katepsinů účastní vzniku a progrese alespoň jednoho typu nádoru, např. žaludku (katepsin X a E), tlustého střeva (katepsin B) nebo prsu a prostaty (katepsin D), a proto jsou katepsiny intenzivně studovány jako diagnostické markery a potenciální cílové molekuly pro vývoj protinádorové terapie. Působení katepsinů v patofyziologii nádorů je značně komplexní a např. katepsin D se podílí nejen svou proteolytickou aktivitou, ale také jako proteinový mitogen. U katepsinů je dále známo několik genetických autosomálně rece-



Obr. 2. **Celková 3D architektura cysteinových katepsinů určuje jejich endopeptidasovou a exopeptidásovou aktivitu.** Šedě je zobrazen povrch endopeptidasy katepsinu L; katalytické aminokyselinové zbytky (cystein žlutě, histidin oranžově) vyznačují oblast štěpení substrátu v aktivním místě, které je umístěno vertikálně po celé čelní straně molekuly. U exopeptidas je aktivní místo po stranách omezeno pomocí specifických strukturních segmentů, což umožňuje vstup pouze *N*- nebo C-konce substrátu a jeho štěpení v aminopeptidasovém resp. karboxypeptidasovém módu. Klíčové segmenty s touto funkcí jsou zobrazeny stužkovým modelem podle 3D struktur aminopeptidas katepsinu H („minichain“ purpurově) a katepsinu C („exclussion domain“ zeleně) a karboxypeptidas katepsinu B („occluding loop“ azurově) a katepsinu X („miniloop“ fialově)

sivních onemocnění způsobených mutací katepsinových genů (tab. I).

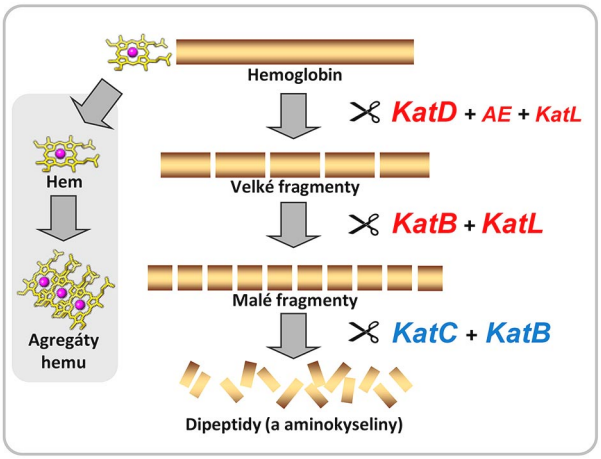
Regulace katepsinů je nezbytná pro jejich kontrolované fyziologické působení a porušení této rovnováhy je spojené s patologiemi. Kontrola funkce probíhá na několika úrovních, zejména regulací genové exprese, řízeným transportem katepsinů do cílových kompartmentů, regulací procesu aktivace zymogenů a modulací aktivity katepsinů pomocí pH, specifických modulátorů a inhibitorů. Mezi molekuly s komplexním modulačním působením patří glykosaminoglykany, které ovlivňují aktivaci, aktivitu, stabilitu a substrátovou specifitu řady katepsinů. Zásadní je role selektivních inhibitorů katepsinů, které patří do několika rodin proteinů: cysteinové katepsiny jsou inhibovány cystatiny, tyropiny a některými serpiny, serinové katepsiny inhibují především serpiny. Pro lidské aspartátové katepsiny není znám žádný endogenní proteinový inhibitor, ale jejich aktivita je blokována pomocí specifických sfingolipidů. Více informací k této kapitole lze získat v citovaných článcích[5–11].

4. Katepsinové proteasy parazitů

Proteasy parazitů patří mezi hlavní faktory při patogenezi parazitických onemocnění a účastní se řady klíčových procesů včetně invaze hostitele, migrace v jeho tkáních, procesu trávení krevních proteinů, překonání imunitního systému a modulace zánětu. Katepsinové proteasy jsou nezbytnou součástí proteolytické výbavy parazitů, zejména byly identifikovány cysteinové proteasy (typu katepsinu B, L, F a C) a aspartátové proteasy (typu katepsinu D). Byly studovány u řady endoparazitů patřících mezi prvoky a helminty a dále u ektoparazitických členovců, kteří jsou přenašeči patogenů (bakterií, prvoků a virů); jejich typičtí představitelé jsou uvedeni v tab. II.

Parazitická onemocnění představují globální zdravotní problém; chorobami jako malárie, schistosomóza, leishmanióza a trypanosomózy jsou infikovány stovky milionů lidí po celém světě, způsobují každoročně miliony úmrtí a obrovskou sociální a ekonomickou zátěž (tab. II). Katepsinové proteasy vzhledem ke svému významu pro životaschopnost parazitů jsou perspektivními cílovými molekulami pro nové strategie léčby parazitárních chorob.

Jako konkrétní příklad uveďme dva hematofágní parazity, krevničku střevní (*Schistosoma mansoni*) a klíště obecné (*Ixodes ricinus*). Krevnička je parazitický helmint způsobující schistosomózu, kterou je v tropických a subtropických zemích nakaženo přes 240 milionů osob. Dospělci krevničky žijí v cévním systému a živí se krví člověka, především hlavním krevním proteinem hemoglobinem. Trávicí systém krevničky zodpovědný za trávení hemoglobinu je tvořen sadou aspartátových a cysteinových katepsinů uvedených v obr. (3), jejichž proteolytické aktivity jsou komplementární a vytvářejí dokonalý hemoglobinolytický aparát. Schistosomální katepsin B1 působí jako endopeptidasa i exopeptidasa a byl identifikován jako cílová molekula pro racionální vývoj potenciálních léčiv proti schisto-

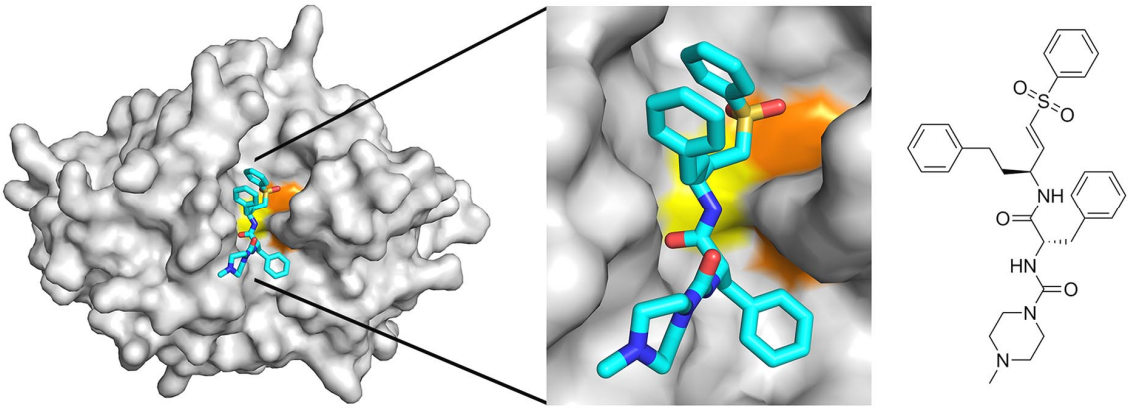


Obr. 3. **Proteolytické trávení hemoglobinu z krve hostitele je ve střevě parazitických krevniček a klíšťat řízeno systémem aspartátových a cysteinových katepsinů.** Hemoglobin je postupně degradován proteasami s endopeptidasovou aktivitou (červeně) a exopeptidasovou aktivitou (modře). Oblast působení katepsinů D, L, B a C je označena KatD, KatL, KatB a KatC (AE je asparaginylendopeptidasa). Uvolněný toxický hem je odstraněn tvorbou agregátů

Tabulka II
Příklady významných hematofágních parazitů, kteří využívají proteasy katepsinového typu pro trávení proteinů hostitele

Parazit	Klasifikace	Trávicí proteasy		Onemocnění (milióny nakažených osob)
		cysteinové proteasy rodiny papainu	aspartátové proteasy rodiny pepsinu	
Trypanosoma americká a spavičná (*Trypanosoma cruzi*, *T. brucei gambiense/rhodesiense)*	protisté (prvoci)	cruzain, rhodesain, TbCB		Chagasova choroba, spavá nemoc[a] (10)
Ničivka útrobní (*Leishmania donovani*)	protisté (prvoci)	CPA, CPB		leishmanióza (1,5)
Zimnička tropická (*Plasmodium falciparum*)	protisté (prvoci)	falcipain 2, 3, 2′, DPAP1	plasmepsin I, II, IV	malárie (350)
Krevnička střevní (*Schistosoma mansoni*)	helminti (ploštěnci)	SmCB1, SmCC, SmCL	SmCD	schistosomóza (bilharzióza) (240)
Motolice jaterní (*Fasciola hepatica*)	helminti (ploštěnci)	FhCB, FhCL		fasciolóza[a] (5)
Měchovec americký a lidský (*Necator americanus, Ancylostoma duodenale*)	helminti (hlístice)	necpain, NaCP3, AdCP2	NaAPR-1, 2, AdAPR-1	ankylostomóza (700)
Klíště obecné (*Ixodes ricinus*)	členovci (pavoukovci)	IrCB1, IrCL1, IrCC	IrCD1	vektor lymeské borreliózy a klíšťové encefalitidy

[a] Tato onemocnění jsou také ekonomicky vysoce významné parazitózy hospodářských zvířat (odhadovaná množství nakažených zvířat jsou více než 50 miliónů pro spavou nemoc a 600 miliónů pro fasciolózu)



Obr. 4. **3D struktura katepsinu B1 z krevničky střevní (*S. mansoni*) v komplexu s inhibitorem K11777**. Povrch katepsinu je šedý a zvýrazněny jsou katalytické aminokyselinové zbytky (cystein žlutě, histidin oranžově), tyčinkový model inhibitoru je tyrkysový (barevné kódování atomů: uhlík tyrkysově, dusík modře, kyslík červeně, síra žlutě). Zobrazen je celkový komplex a v detailu pohled do aktivního místa, vpravo je chemická struktura inhibitoru

somóze na bázi syntetických inhibitorů tohoto enzymu (obr. 4).

Klíště obecné je hlavním přenašečem lymeské borreliózy a klíšťové encefalitidy v Evropě. Trávicí systém klíšťat je překvapivě podobný tomu, který využívá krevnička, a obsahuje analogické katepsiny pro degradaci hemoglobinu. Interakce klíštěte s hostitelem vyžaduje ještě další systém související s katepsiny – sliny klíšťat obsahují řadu biologicky aktivních proteinů, které mají antiproliferační, antikoagulační, imunomodulační a protizánětlivé působení a umožňují úspěšné sání krve na hostiteli. Mezi nimi hrají významnou roli cystatiny a serpiny, které fungují jako inhibitory cysteinových a serinových katepsinů hostitele, a regulují imunitní a zánětlivou reakci ve prospěch parazi-

ta. Uvedené trávicí katepsiny a slinné katepsinové inhibitory představují nové cílové molekuly pro vývoj tzv. protiklíštěcích vakcín, které zabraňují sání klíšťat a přenosu patogenů. Více informací k této kapitole podávají cit.[12–18].

5. Výzkum katepsinů na ÚOCHB AV ČR

Výzkum katepsinů má za sebou více než šedesátiletou historii, silně akceleroval v posledních dvou desetiletích a v současné době je jednou z nejrychleji se rozvíjejících oblastí proteolýzy. Na Ústavu organické chemie a biochemie AV ČR jsou katepsiny studovány v kontinuitě od konce 60. let a tato práce významně přispěla zejména k určování primárních struktur a pochopení funkce savčích katepsinů a přirozených inhibitorů katepsinových proteas. V současné době se náš výzkum soustředí na katepsiny spojené s patologií a na medicinálně významné proteolytické systémy založené na katepsinech. Příkladem jsou katepsiny účastnící se nádorových a alergických onemocnění a dále proteolytické systémy parazitů způsobujících helmintózy a klíšťat přenášejících patogeny. Cílem řešených projektů je analýza vztahu struktury a funkce katepsinů a určení molekulárních mechanismů regulace jejich aktivity jako základ pro vývoj nových terapeutických přístupů a potenciálních léčiv. Více informací o problematice katepsinů a tématech výzkumu na ÚOCHB AV ČR je na webových stránkách http://www.uochb.cz/web/structure/186.html.

Referát vznikl v rámci řešení grantových projektů P302/11/1481, a výzkumného záměru RVO 61388963.

LITERATURA

1. Zaidi N., Kalbacher H., v knize: *Handbook of Proteolytic Enzymes* (Rawlings N. D., Salvesen G., ed.), sv. I, kap. 6, 3. vydání. Elsevier, Amsterdam 2013.
2. Fusek M., Mareš M., Větvička V,. v knize: *Handbook of Proteolytic Enzymes* (Rawlings N. D., Salvesen G., ed.), sv. I, kap. 8, 3. vydání. Elsevier, Amsterdam 2013.
3. Pshezhetsky A., v knize: *Handbook of Proteolytic Enzymes* (Rawlings N. D., Salvesen G., ed.), sv. III, kap. 754, 3. vydání. Elsevier, Amsterdam 2013.
4. Korkmaz B., Horwitz M. S., Jenne D. E., Gauthier F.: Pharmacol. Rev. *62*, 726 (2010).
5. Turk V., Stoka V., Vasiljeva O., Renko M., Sun T., Turk B., Turk D.: Biochim. Biophys. Acta *1824*, 68 (2012).
6. Novinec M., Lenarčič B.: Biomolecular Concepts *4*, 287 (2013).
7. Žebrakovská I., Máša M., Srp J., Horn M., Vávrová K., Mareš M.: Biochim. Biophys. Acta *1811*, 1097 (2011).
8. Masson O., Bach A. S., Derocq D., Prebois C., Laurent-Matha V., Pattingre S., Liaudet-Coopman E.: Biochimie *92*, 1635 (2010).
9. Mohamed M., Sloane B. F.: Nat. Rev. Cancer *6*, 764 (2006).
10. Mason S. D., Joyce J. A.: Trends Cell Biol. *21*, 228 (2011).
11. Colbert J. D., Matthews S. P., Miller G., Watts C.: Eur. J. Immunol. *39*, 2955 (2009).
12. McKerrow J. H., Caffrey C., Kelly B., Loke P., Sajid M.: Annu. Rev. Pathol. *1*, 497 (2006).
13. Sojka D., Franta Z., Horn M., Caffrey C. R., Mareš M., Kopáček P.: Trends Parasitol. *29*, 276 (2013).
14. Kašný M., Mikeš L., Hampl V., Dvořák J., Caffrey C. R., Dalton J. P., Horák P.: Adv. Parasitol. *69*, 205 (2009).
15. Jílková A., Řezáčová P., Lepšík M., Horn M., Váchová J., Fanfrlík J., Brynda J., McKerrow J. H., Caffrey C. R., Mareš M.: J. Biol. Chem. *286*, 35770 (2011).
16. Horn M., Nussbaumerová M., Šanda M., Kovářová Z., Srba J., Franta Z., Sojka D., Bogyo M., Caffrey C. R., Kopáček P., Mareš M.: Chem. Biol. *16*, 1053 (2009).
17. Chmelař J., Oliveira C. J., Řezáčová P., Francischetti I. M., Kovářová Z., Pejler G., Kopáček P., Ribeiro J. M., Mareš M., Kopecký J., Kotsyfakis M.: Blood *117*, 736 (2011).
18. Salát J., Paesen G. C., Řezáčová P., Kotsyfakis M., Kovářová Z., Šanda M., Majtán J., Grunclová L., Horká H., Andersen J. F., Brynda J., Horn M., Nunn M. A., Kopáček P., Kopecký J., Mareš M.: Biochem. J. *429*, 103 (2010).

M. Horn[a], A. Jílková[a,b], and M. Mareš[a] *([a] Institute of Organic Chemistry and Biochemistry Academy of Sciences of the Czech Republic, Prague, [B] Department of Biochemistry, Faculty of Science, Charles University, Prague):* **Cathepsin Proteases in Pathology**

Cathepsin research is one of the most progressive areas in proteolysis. Cathepsins are a diverse group of eukaryotic peptidases belonging to the aspartic, cysteine and serine classes. They act as both non-specific proteases and as specific processing enzymes in numerous physiological and pathological processes. In humans, their dysregulation is associated with severe pathologies, such as cancer, cardiovascular and neurodegenerative diseases, arthritis, and osteoporosis. Proteolytic systems controlled by cathepsins are a critical part of host-pathogen and host-parasite interactions as demonstrated for cathepsin-like peptidases from, e.g., hematophagous parasites, herbivorous insects and plants. Because of the broad involvement of cathepsins in pathological processes, they are among today's top-priority drug targets. This review provides an update on the structure and function of pathology-associated cathepsins in humans and human parasites.

6. Diskuze

Katepsin B1 z krevničky *S. mansoni* (SmCB1) je jednou z perspektivních cílových molekul pro nová léčiva proti schistosomóze, kterou je nakaženo více než 200 milionů lidí. Doposud však nebyly k dispozici relevantní strukturní a biochemické informace o této protease nezbytné pro pochopení mechanismu její regulace a racionální vývoj účinných inhibitorů. Cílem dizertační práce bylo detailně charakterizovat SmCB1 z hlediska vztahu struktury a aktivity a popsat principy inhibice. Práce přináší první krystalovou strukturu a strukturní analýzy proteolytického enzymu z parazitických krevniček. Je zde uvedeno celkem 6 krystalových struktur SmCB1 deponovaných v současné době v databázi PDB umožňujících jednak mapovat proces aktivace a jednak analyzovat mechanismy inhibice.

Objev nového autoaktivačního mechanismu SmCB1

SmCB1 je biosyntetizován ve formě neaktivního zymogenu, ve kterém N-koncový propeptid blokuje aktivní místo. Mechanismus aktivace SmCB1 nebyl do této doby detailně studován a nepodařilo se získat plně aktivní SmCB1 autoaktivací v kyselém prostředí jako je to možné u savčích katepsinů B [40;65;105]. Dosud byla popsána pouze *in vitro* aktivace SmCB1 pomocí proteolytického enzymu asparaginylendopeptidasy (SmAE) [14], který je přítomný ve střevě krevničky a jeho podíl na aktivaci SmCB1 je fyziologicky relevantní. V krevničkách, v kterých byla exprese SmAE potlačena pomocí RNAi experimentů, dochází však i nadále k produkci aktivního SmCB1, což dokazuje existenci alternativního mechanismu aktivace SmCB1 [189]. V rámci dizertační práce byl identifikován tento nový způsob autoaktivace SmCB1 na aktivní enzym indukovaný působením sulfatovaných polysacharidů jako jsou glykosaminoglykany (GAG), které jsou široce rozšířené ve všech živočišných tkáních. Na základě provedené biochemické a strukturní analýzy bylo navrženo schéma aktivace SmCB1 se dvěma alternativními cestami (autoaktivací a trans-aktivací pomocí asparaginylendopeptidasy), které jsou přepínány pomocí GAG a pH (publikace č. 4). Schéma aktivace bylo na strukturní úrovni zmapováno pomocí unikátní sady krystalových struktur tří aktivačních forem, od neaktivního zymogenu po aktivní SmCB1. Struktura intermediátní formy SmCB1 s nekompletně odštěpeným propeptidem, která vzniká v průběhu aktivace, je zvláště významná, protože se jí dosud nepodařilo získat pro žádnou cysteinovou proteasu.

Vztah struktury a aktivity SmCB1

Aktivita SmCB1 byla detailně studována v publikaci č. 2. Analýza aktivity SmCB1 pomocí syntetických substrátů ukázala, že SmCB1 je komplexní enzym, který může zastávat funkci endopeptidasy i exopeptidasy. Oba aktivitní módy byly prokázány i při štěpení přirozeného substrátu hemoglobinu, který byl pomocí SmCB1 rychle degradován na dipeptidy, aniž by docházelo k výrazné akumulaci větších fragmentů. Lze usuzovat, že SmCB1 je účinný nástroj pro degradaci hemoglobinu (a dalších proteinů z krve hostitele) v trávicím ústrojí krevničky. Duální endopeptidasová i exopeptidasová aktivita je u SmCB1 umožněna flexibilitou smyčky "occluding loop", jež moduluje tvar aktivního místa. Tato smyčka je charakteristická pro katepsiny B [47;109] a u SmCB1 byla prokázána krystalovou strukturou (publikace č. 2). Když je smyčka v uzavřené konformaci, blokuje přístup substrátu do aktivního místa za S2´ podmístem a tím způsobuje exopeptidasovou (peptidyldipeptidasovou) aktivitu. Vyklonění smyčky z aktivního místa umožní endopetidasovou aktivitu a štěpení větších substrátů. Tento pohyb smyčky je patrný i během procesu aktivace, kdy v zymogenu SmCB1 je smyčka vykloněná, protože je v aktivním místě navázán propeptid, zatímco u aktivního enzymu je smyčka v uzavřené konformaci a blokuje část aktivního místa (publikace č. 4). Je zajímavé, že ve struktuře neaktivní intermediátní formy SmCB1 již smyčka zaujímá finální pozici jako v aktivním enzymu v exopeptidasovém módu, což umožnilo odštěpení části propeptidu (publikace č. 4).

Dva přístupy pro vývoj inhibitorů SmCB1

Pro studium mechanismu inhibiční regulace SmCB1 byly v dizertační práci využity dva odlišné přístupy. Prvním byl racionální design pentapeptidových inhibitorů odvozených z přirozeného minimálního inhibičního motivu, který byl identifikován v propeptidu SmCB1 (publikace č. 1). Druhým bylo mapování specifity aktivního místa pomocí peptidomimetických reaktivních vinylsulfonových inhibitorů obsahujících sérii strukturních modifikací (publikace č. 2). První metodou byly navrženy nejúčinnější inhibitory s mikromolární inhibiční konstantou, což je o tři řády horší hodnota v porovnání s vinylsulfony, kde nejlepší inhibitory vykazovaly nanomolární inhibiční parametry. Obě skupiny inhibitorů se lišily i mechanismem účinku a selektivitou.

Inhibitory odvozené od propeptidu jsou reverzibilní a váží se do aktivního místa v orientaci opačné než substrát, analogicky jako je vázán propeptid v molekule zymogenu. Tyto inhibitory prokázaly vysokou selektivitu pro SmCB1 v porovnání s lidskými

katepsiny B, L a S (publikace č. 1). Toto zjištění je v souladu s částečně selektivním rozpoznáváním vlastních propeptidů, které bylo pozorováno u lidských katepsinů [38]. V současné době probíhá krystalizace SmCB1 s těmito inhibitory pro určení přesné interakce s aktivním místem, která umožní další cílené modifikace struktury inhibitorů ke zlepšení jejich inhibičních vlastností.

Vinylsulfonová peptidomimetika působí jako ireverzibilní inhibitory tvorbou kovalentní vazby s katalytickým cysteinem v aktivním místě SmCB1, kam se váží ve stejné orientaci jako substrát (publikace č. 2). Jejich selektivita vůči SmCB1 oproti lidským homologům zatím nebyla testována, tato *in vitro* studie se v současné době připravuje na ÚOCHB AV ČR. Testované vinylsulfony byly vybrány z knihovny derivátů UCSF a zahrnovaly i inhibitor cysteinových proteas K11777. Tento modelový vinylsulfon je ve třetí fázi klinických studií pro léčbu Chagasovy choroby způsobované parazitickým prvokem *Trypanosoma cruzi* [34] a byl použit i při validaci SmCB1 jako cílové molekuly na myším modelu schistosomózy [23]. Publikace č. 2 ukázala, že *in vitro* inhibice SmCB1 pomocí vinylsulfonů koreluje s jejich *ex vivo* účinností při supresi larev *S. mansoni* v kultivačních mediích. Tento výsledek dává cennou informaci o SmCB1 jako vhodném cílovém proteinu pro chemoterapeutický zásah a o vinylsulfonech jako antischistosomálních molekulách. Publikace č. 2 dále přinesla krystalové struktury dvou komplexů SmCB1 s vinylsulfony, což umožnilo identifikovat kritické inhibiční interakce ve vazebných podmístech SmCB1. Tyto strukturní údaje a nově vyvinutá *in silico* výpočetní metoda skórování účinnosti inhibitorů (publikace č. 3) vytvářejí klíčový nástroj pro racionální design nové generace inhibitorů SmCB1 s vyšším inhibičním potenciálem a selektivitou.

Interakce SmCB1 se sulfatovanými polysacharidy

V propeptidu SmCB1 byl pomocí krystalové struktury zymogenu objeven unikátní α-helix, který není přítomen u savčích katepsinů B [103;190;191] a obsahuje sekvenční motiv vázající heparin ("heparin-binding motif" – "HB motiv") a jiné sulfatované polysacharidy. Bylo zjištěno, že tento sekvenční segment je přítomen pouze u katepsinů B parazitických motolic, které se účastní trávení krve hostitele (publikace č. 1). V publikaci č. 4 byla prokázána zásadní role "HB motivu" v procesu autoaktivace SmCB1 pomocí sulfatovaných polysacharidů, včetně přirozených GAG. Na základě analýzy struktury zymogenu SmCB1 a mutací v "HB motivu" lze předpokládat, že uvedený α-helix při interakci s GAG mění konformaci propeptidu a umožňuje autokatalytické štěpení vazby na rozhraní propeptidu a katalytické domény. Přitom lze předpokládat, že GAG interaguje

nejen s “HB motivem”, ale současně i s další bazickou oblastí na povrchu katalytické domény, jak uvádí publikace č. 1. Tento mechanismus vzájemného propojení propeptidu a katalytické domény prostřednictvím GAG by bylo možné využít při navrhování inhibičních ligandů SmCB1 využívajících vazebná místa pro GAG.

7. Závěry

Dizertační práce se zabývá katepsinem B1 (SmCB1) z krevničky střevní (*Schistosoma mansoni*), která způsobuje závažné parazitární onemocnění schistosomózu. Výsledky jsou shrnuty ve čtyřech publikacích a v jednom rukopise připraveném k podání do impaktovaného časopisu. Práce přináší nové informace o struktuře, aktivitě a inhibiční regulaci SmCB1, které umožňují lépe pochopit funkci tohoto medicinálně významného enzymu a navrhovat jeho inhibitory jako potenciální léčiva proti schistosomóze.

V rámci disertační práce byly splněny zadané cíle s následujícími závěry:

1. Bylo připraveno pět přirozených a mutantních forem SmCB1 rekombinantní expresí v kvasinkách *Pichia pastoris* a byl vypracován postup pro jejich chromatografickou purifikaci.

2. Bylo vyřešeno šest krystalových struktur SmCB1, konkrétně (i) molekulární formy aktivační dráhy: neaktivní zymogen, intermediátní aktivační forma, aktivní enzym a (ii) tři komplexy enzymu s peptidomimetickými inhibitory. Koordináty struktur byly uloženy v databázi PDB (kódová označení 4I04, 4I05, 4I07, 3S3R, 3S3Q, 3QSD).

3. Analýza proteolytické aktivity SmCB1 pomocí fyziologického substrátu hemoglobinu a knihoven syntetických fluorogenních peptidových substrátů umožnila (i) určit substrátovou specifitu a preference pro štěpené vazby a (ii) identifikovat duální aktivitu SmCB1 jako proteasy působící v exopeptidasovém (peptidyldipeptidasovém) a endopeptidasovém módu. Tyto aktivity dělají z SmCB1 účinný nástroj pro komplexní degradaci hemoglobinu ve střevě krevničky.

4. Analýza inhibice SmCB1 pomocí vinylsulfonových (VS) inhibitorů:
   a) Inhibiční specifita SmCB1 byla analyzována pomocí sady VS inhibitorů s odlišnými substituenty v pozicích P3-P1´. Nejúčinnější inhibitory dosahovaly subnanomolárních hodnot IC_{50}.

   b) Kritické interakce mezi SmCB1 a VS inhibitory byly kvantifikovány *in silico* pomocí metod výpočetní chemie s využitím krystalových struktur inhibičních komplexů.

   c) Na základě enzymologické analýzy inhibice SmCB1 pomocí VS inhibitorů byla přístupy výpočetní chemie získána skórovací funkce pro predikci inhibiční

účinnosti VS inhibitorů, která umožní efektivní navrhování nových inhibitorů SmCB1 *in silico*.

d) Biologická aktivita VS inhibitorů byla určena *ex vivo* v experimentech na larvách *S. mansoni* v kultivačním mediu. Nalezená účinnost při supresi parazita koreluje s *in vitro* inhibicí SmCB1, což podává důkaz o SmCB1 jako vhodné cílové molekule a VS jako antischistosomálních molekulách.

5. Funkční analýza propeptidu SmCB1 byla provedena pomocí syntetických fragmentů odvozených z propeptidu a umožnila identifikaci dvou strukturních oblastí v propeptidu, které působí jako inhibitory aktivity SmCB1. Oblast interagující s aktivním místem byla využita pro navrhování pentapeptidových inhibitorů s účinností *in vitro* v mikromolární oblasti hodnot K_i.

6. Pomocí biochemické a strukturní analýzy byl popsán unikátní mechanismus aktivace SmCB1 z neaktivního zymogenu na aktivní enzym, který probíhá dvěma alternativními dráhami: (i) autoaktivací indukovanou působením sulfatovaných polysacharidů (např. glykosaminoglykanů) nebo (ii) trans-aktivací katalyzovanou působením asparaginylendopeptidasy. Sulfatované polysacharidy fungují jako přepínač mezi oběma dráhami. Strukturní aspekty aktivace SmCB1 a interakce se sulfatovanými polysacharidy byly vysvětleny pomocí krystalových struktur tří forem SmCB1 v aktivační dráze.

Prohlášení spoluautorů

Mgr. Adéla Jílková se významně podílela na vzniku všech publikací uvedených v disertační práci, konkrétně:

Publikace č. 1 : Mapping the Pro-Peptide of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Modulation of Inhibition by Heparin and Design of Mimetic Inhibitors

(1) Produkce SmCB1 k enzymologickým pokusům, což zahrnovalo vývoj metody pro aktivaci SmCB1 na zralý enzym a jeho chromatografickou purifikaci. (2) Analýza interakcí heparinu s SmCB1.

Publikace č. 2: Structural Basis for Inhibition of Cathepsin B Drug Target from the Human Blood Fluke, *Schistosoma mansoni*

(1) Rekombinantní exprese, aktivace a purifikace SmCB1. (2) Příprava inhibičních komplexů a jejich krystalizace, analýza krystalografických struktur (pomocí programů CONTACT a Pymol), jejich interpretace a molekulární grafika (pomocí programu Pymol). (3) Testování knihoven FRET substrátů a analýza degradace hemoglobinu, určení inhibičních konstant IC_{50}, *ex vivo* testování inhibitorů a vyhodnocování změn fenotypu parazita. (4) Účast na přípravě manuskriptu.

Publikace č. 3: Quantum Mechanics-Based Scoring Rationalizes the Irreversible Inactivation of Parasitic *Schistosoma mansoni* Cysteine Peptidase by Vinyl Sulfone Inhibitors

Enzymologická měření (určení kinetických parametrů a ireverzibility inhibice).

Publikace č. 4: Activation route of the *Schistosoma mansoni* Cathepsin B1 Drug Target: Structural Map with a Glycosaminoglycan Switch

(1) Rekombinantní exprese, aktivace a purifikace SmCB1 a jeho mutantních forem. (2) Krystalizace tří forem SmCB1 z aktivační dráhy, analýza těchto krystalografických struktur (pomocí programů CONTACT a Pymol), jejich interpretace a molekulární grafika (pomocí programu Pymol). (3) Biochemické aktivační experimenty analyzované pomocí SDS-PAGE a proteinového sekvenování. (4) Účast na přípravě manuskriptu.

Publikace č. 5: Katepsinové proteasy v patologii

Analýza krystalových struktur, molekulární grafika a účast na přípravě manuskriptu.

……..

RNDr. Michael Mareš, CSc.

Školitel a spoluautor všech uvedených publikací

8. Seznam použité literatury

1. WHO (2012): Weekly epidemiological record No.4, 87, World Health Organisation, Geneva, 37-44.
2. WHO (2013): Weekly epidemiological record No.8, 88, World Health Organisation, Geneva, 81-88.
3. Chitsulo,L., Engels,D., Montresor,A., and Savioli,L. (2000) The global status of schistosomiasis and its control. *Acta Trop.*, **77**, 41-51.
4. Steinmann,P., Keiser,J., Bos,R., Tanner,M., and Utzinger,J. (2006) Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect.Dis.*, **6**, 411-425.
5. Gryseels,B., Polman,K., Clerinx,J., and Kestens,L. (2006) Human schistosomiasis. *Lancet*, **368**, 1106-1118.
6. Mostafa,M.H., Sheweita,S.A., and O'Connor,P.J. (1999) Relationship between schistosomiasis and bladder cancer. *Clin.Microbiol.Rev.*, **12**, 97-111.
7. Thetiot-Laurent,S.A., Boissier,J., Robert,A., and Meunier,B. (2013) Schistosomiasis chemotherapy. *Angew.Chem.Int.Ed Engl.*, **52**, 7936-7956.
8. Caffrey,C.R. (2007) Chemotherapy of schistosomiasis: present and future. *Curr.Opin.Chem.Biol.*, **11**, 433-439.
9. Ismail,M., Botros,S., Metwally,A., William,S., Farghally,A., Tao,L.F., Day,T.A., and Bennett,J.L. (1999) Resistance to praziquantel: direct evidence from *Schistosoma mansoni* isolated from Egyptian villagers. *Am.J.Trop.Med.Hyg.*, **60**, 932-935.
10. Melman,S.D., Steinauer,M.L., Cunningham,C., Kubatko,L.S., Mwangi,I.N., Wynn,N.B., Mutuku,M.W., Karanja,D.M., Colley,D.G., Black,C.L., Secor,W.E., Mkoji,G.M., and Loker,E.S. (2009) Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS.Negl.Trop.Dis.*, **3**, e504.
11. Cioli,D., Botros,S.S., Wheatcroft-Francklow,K., Mbaye,A., Southgate,V., Tchuente,L.A., Pica-Mattoccia,L., Troiani,A.R., El-Din,S.H., Sabra,A.N., Albin,J., Engels,D., and Doenhoff,M.J. (2004) Determination of ED50 values for praziquantel in praziquantel-resistant and -susceptible *Schistosoma mansoni* isolates. *Int.J.Parasitol.*, **34**, 979-987.
12. Fallon,P.G. and Doenhoff,M.J. (1994) Drug-resistant schistosomiasis: resistance to praziquantel and oxamniquine induced in *Schistosoma mansoni* in mice is drug specific. *Am.J.Trop.Med.Hyg.*, **51**, 83-88.
13. Sajid,M. and McKerrow,J.H. (2002) Cysteine proteases of parasitic organisms. *Molecular and Biochemical Parasitology*, **120**, 1-21.
14. Sajid,M., McKerrow,J.H., Hansell,E., Mathieu,M.A., Lucas,K.D., Hsieh,I., Greenbaum,D., Bogyo,M., Salter,J.P., Lim,K.C., Franklin,C., Kim,J.H., and Caffrey,C.R. (2003) Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol.Biochem.Parasitol.*, **131**, 65-75.
15. Hola-Jamriska,L., Dalton,J.P., Aaskov,J., and Brindley,P.J. (1999) Dipeptidyl peptidase I and III activities of adult schistosomes. *Parasitology*, **118**, 275-282.
16. Dvorak,J., Mashiyama,S.T., Sajid,M., Braschi,S., Delcroix,M., Schneider,E.L., McKerrow,W.H., Bahgat,M., Hansell,E., Babbitt,P.C., Craik,C.S., McKerrow,J.H., and Caffrey,C.R. (2009) SmCL3, a gastrodermal cysteine protease of the human blood fluke *Schistosoma mansoni*. *PLoS.Negl.Trop.Dis.*, **3**, e449.

17. Brindley,P.J., Kalinna,B.H., Wong,J.Y., Bogitsh,B.J., King,L.T., Smyth,D.J., Verity,C.K., Abbenante,G., Brinkworth,R.I., Fairlie,D.P., Smythe,M.L., Milburn,P.J., Bielefeldt-Ohmann,H., Zheng,Y., and McManus,D.P. (2001) Proteolysis of human hemoglobin by schistosome cathepsin D. *Mol.Biochem.Parasitol.*, **112**, 103-112.

18. Caffrey,C.R., Mathieu,M.A., Gaffney,A.M., Salter,J.P., Sajid,M., Lucas,K.D., Franklin,C., Bogyo,M., and McKerrow,J.H. (2000) Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. *FEBS Lett.*, **466**, 244-248.

19. McCarthy,E., Stack,C., Donnelly,S.M., Doyle,S., Mann,V.H., Brindley,P.J., Stewart,M., Day,T.A., Maule,A.G., and Dalton,J.P. (2004) Leucine aminopeptidase of the human blood flukes, *Schistosoma mansoni* and *Schistosoma japonicum*. *Int.J.Parasitol.*, **34**, 703-714.

20. Caffrey,C.R., McKerrow,J.H., Salter,J.P., and Sajid,M. (2004) Blood 'n' guts: an update on schistosome digestive peptidases. *Trends Parasitol.*, **20**, 241-248.

21. Delcroix,M., Sajid,M., Caffrey,C.R., Lim,K.C., Dvorak,J., Hsieh,I., Bahgat,M., Dissous,C., and McKerrow,J.H. (2006) A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J.Biol.Chem.*, **281**, 39316-39329.

22. Caffrey,C.R. and Ruppel,A. (1997) Cathepsin B-like activity predominates over cathepsin L-like activity in adult *Schistosoma mansoni* and *S. japonicum*. *Parasitol.Res.*, **83**, 632-635.

23. Abdulla,M.H., Lim,K.C., Sajid,M., McKerrow,J.H., and Caffrey,C.R. (2007) Schistosomiasis mansoni: novel chemotherapy using a cysteine protease inhibitor. *PLoS.Med.*, **4**, e14.

24. Twining,S.S. (1994) Regulation of proteolytic activity in tissues. *Crit Rev.Biochem.Mol.Biol.*, **29**, 315-383.

25. Rawlings,N.D., Salvesen,G.S. (2013): *Handbook of Proteolytic enzymes*. 3rd ed., Elsevier, Amsterdam, 3932 stran.

26. Goulet,B., Baruch,A., Moon,N.S., Poirier,M., Sansregret,L.L., Erickson,A., Bogyo,M., and Nepveu,A. (2004) A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor. *Mol.Cell*, **14**, 207-219.

27. Chapman,H.A., Riese,R.J., and Shi,G.P. (1997) Emerging roles for cysteine proteases in human biology. *Annu.Rev.Physiol*, **59**, 63-88.

28. Brix,K., Dunkhorst,A., Mayer,K., and Jordans,S. (2008) Cysteine cathepsins: cellular roadmap to different functions. *Biochimie*, **90**, 194-207.

29. http://merops.sanger.ac.uk/ (7.7.2014).

30. Novinec,M., Lenarcic,B. (2013) Papain-like peptidases: structure, function, and evolution. *Biomol.Concepts*, **4**, 287-308.

31. Vasiljeva,O., Reinheckel,T., Peters,C., Turk,D., Turk,V., and Turk,B. (2007) Emerging roles of cysteine cathepsins in disease and their potential as drug targets. *Curr.Pharm.Des*, **13**, 387-403.

32. Fonovic,M. and Turk,B. (2014) Cysteine cathepsins and their potential in clinical therapy and biomarker discovery. *Proteomics Clin.Appl.*, **8**, 416-426

33. Turk,B. (2006) Targeting proteases: successes, failures and future prospects. *Nat.Rev.Drug Discov.*, **5**, 785-799.

34. McKerrow,J.H., Doyle,P.S., Engel,J.C., Podust,L.M., Robertson,S.A., Ferreira,R., Saxton,T., Arkin,M., Kerr,I.D., Brinen,L.S., and Craik,C.S. (2009) Two approaches to discovering and developing new drugs for Chagas disease. *Mem.Inst.Oswaldo Cruz*, **104 Suppl 1**, 263-269.

35. Marco,M. and Coteron,J.M. (2012) Falcipain inhibition as a promising antimalarial target. *Curr.Top.Med.Chem.*, **12**, 408-444.

36. http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/MACiE/entry/getPage.pl?id=M0174.stg03 (7.7.2014).

37. Turk,V., Stoka,V., Vasiljeva,O., Renko,M., Sun,T., Turk,B., and Turk,D. (2012) Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim.Biophys.Acta*, **1824**, 68-88.

38. Wiederanders,B., Kaulmann,G., and Schilling,K. (2003) Functions of propeptide parts in cysteine proteases. *Curr.Protein Pept.Sci.*, **4**, 309-326.

39. Nishimura,Y., Kawabata,T., Furuno,K., and Kato,K. (1989) Evidence That Aspartic Proteinase Is Involved in the Proteolytic Processing Event of Procathepsin-l in Lysosomes. *Archives of Biochemistry and Biophysics*, **271**, 400-406.

40. Rowan,A.D., Mason,P., Mach,L., and Mort,J.S. (1992) Rat procathepsin B. Proteolytic processing to the mature form *in vitro*. *J.Biol.Chem.*, **267**, 15993-15999.

41. Menard,R., Carmona,E., Takebe,S., Dufour,E., Plouffe,C., Mason,P., and Mort,J.S. (1998) Autocatalytic processing or recombinant human procathepsin L - Contribution of both intermolecular and unimolecular events in the processing of procathepsin L *in vitro*. *Journal of Biological Chemistry*, **273**, 4478-4484.

42. Karrer,K.M., Peiffer,S.L., and Ditomas,M.E. (1993) 2 Distinct Gene Subfamilies Within the Family of Cysteine Protease Genes. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 3063-3067.

43. Wex,T., Levy,B., Wex,H., and Bromme,D. (1999) Human cathepsins F and W: A new subgroup of cathepsins. *Biochemical and Biophysical Research Communications*, **259**, 401-407.

44. Drenth,J., Jansonius,J.N., Koekoek,R., Swen,H.M., and Wolthers,B.G. (1968) Structure of papain. *Nature*, **218**, 929-932.

45. Turk,D., Janjic,V., Stern,I., Podobnik,M., Lamba,D., Dahl,S.W., Lauritzen,C., Pedersen,J., Turk,V., and Turk,B. (2001) Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. *EMBO J.*, **20**, 6570-6582.

46. Turk,D., Guncar,G., Podobnik,M., and Turk,B. (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biological Chemistry*, **379**, 137-147.

47. Musil,D., Zucic,D., Turk,D., Engh,R.A., Mayr,I., Huber,R., Popovic,T., Turk,V., Towatari,T., Katunuma,N., and . (1991) The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.*, **10**, 2321-2330.

48. Guncar,G., Podobnik,M., Pungercar,J., Strukelj,B., Turk,V., and Turk,D. (1998) Crystal structure of porcine cathepsin H determined at 2.1 A resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure.*, **6**, 51-61.

49. Baici,A., Muntener,K., Willimann,A., and Zwicky,R. (2006) Regulation of human cathepsin B by alternative mRNA splicing: homeostasis, fatal errors and cell death. *Biol.Chem.*, **387**, 1017-1021.

50. Caserman,S., Kenig,S., Sloane,B.F., and Lah,T.T. (2006) Cathepsin L splice variants in human breast cell lines. *Biol.Chem.*, **387**, 629-634.

51. Maubach,G., Schilling,K., Rommerskirch,W., Wenz,I., Schultz,J.E., Weber,E., and Wiederanders,B. (1997) The inhibition of cathepsin S by its propeptide - Specificity and mechanism of action. *European Journal of Biochemistry*, **250**, 745-750.

52. McQueney,M.S., Amegadzie,B.Y., D'Alessio,K., Hanning,C.R., McLaughlin,M.M., McNulty,D., Carr,S.A., Ijames,C., Kurdyla,J., and Jones,C.S. (1997) Autocatalytic activation of human cathepsin K. *J.Biol.Chem.*, **272**, 13955-13960.

53. Chowdhury,S.F., Sivaraman,J., Wang,J., Devanathan,G., Lachance,P., Qi,H.T., Menard,R., Lefebvre,J., Konishi,Y., Cygler,M., Sulea,T., and Purisima,E.O. (2002) Design of noncovalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides. *Journal of Medicinal Chemistry*, **45**, 5321-5329.

54. Mach,L., Mort,J.S., and Glossl,J. (1994) Noncovalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass forms of the enzyme. *J.Biol.Chem.*, **269**, 13036-13040.

55. Fox,T., de,M.E., Mort,J.S., and Storer,A.C. (1992) Potent slow-binding inhibition of cathepsin B by its propeptide. *Biochemistry*, **31**, 12571-12576.

56. Guay,J., Falgueyret,J.P., Ducret,A., Percival,M.D., and Mancini,J.A. (2000) Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. *Eur.J.Biochem.*, **267**, 6311-6318.

57. Kirschke,H., Wiederanders,B., Bromme,D., and Rinne,A. (1989) Cathepsin-S from Bovine Spleen - Purification, Distribution, Intracellular-Localization and Action on Proteins. *Biochemical Journal*, **264**, 467-473.

58. Barrett,A.J. and Kirschke,H. (1981) Cathepsin B, Cathepsin H, and cathepsin L. *Methods Enzymol.*, **80**, 535-561.

59. Turk,B., Bieth,J.G., Bjork,I., Dolenc,I., Turk,D., Cimerman,N., Kos,J., Colic,A., Stoka,V., and Turk,V. (1995) Regulation of the activity of lysosomal cysteine proteinases by pH-induced inactivation and/or endogenous protein inhibitors, cystatins. *Biol.Chem.Hoppe Seyler*, **376**, 225-230.

60. Caffrey,C.R., Hansell,E., Lucas,K.D., Brinen,L.S., Alvarez,H.A., Cheng,J., Gwaltney,S.L., Roush,W.R., Stierhof,Y.D., Bogyo,M., Steverding,D., and McKerrow,J.H. (2001) Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of *Trypanosoma brucei rhodesiense*. *Mol.Biochem.Parasitol.*, **118**, 61-73.

61. Almeida,P.C., Nantes,I.L., Chagas,J.R., Rizzi,C.C., Faljoni-Alario,A., Carmona,E., Juliano,L., Nader,H.B., and Tersariol,I.L. (2001) Cathepsin B activity regulation. Heparin-like glycosaminogylcans protect human cathepsin B from alkaline pH-induced inactivation. *J.Biol.Chem.*, **276**, 944-951.

62. Li,Z., Hou,W.S., Escalante-Torres,C.R., Gelb,B.D., and Bromme,D. (2002) Collagenase activity of cathepsin K depends on complex formation with chondroitin sulfate. *J.Biol.Chem.*, **277**, 28669-28676.

63. Costa,T.F., dos Reis,F.C., and Lima,A.P. (2012) Substrate inhibition and allosteric regulation by heparan sulfate of *Trypanosoma brucei* cathepsin L. *Biochim.Biophys.Acta*, **1824**, 493-501.

64. Caglic,D., Pungercar,J.R., Pejler,G., Turk,V., and Turk,B. (2007) Glycosaminoglycans facilitate procathepsin B activation through disruption of propeptide-mature enzyme interactions. *J.Biol.Chem.*, **282**, 33076-33085.

65. Rozman,J., Stojan,J., Kuhelj,R., Turk,V., and Turk,B. (1999) Autocatalytic processing of recombinant human procathepsin B is a bimolecular process. *FEBS Lett.*, **459**, 358-362.

66. Fairhead,M., Kelly,S.M., and van der Walle,C.F. (2008) A heparin binding motif on the pro-domain of human procathepsin L mediates zymogen destabilization and activation. *Biochem.Biophys.Res.Commun.*, **366**, 862-867.

67. Vasiljeva,O., Dolinar,M., Pungercar,J.R., Turk,V., and Turk,B. (2005) Recombinant human procathepsin S is capable of autocatalytic processing at neutral pH in the presence of glycosaminoglycans. *FEBS Lett.*, **579**, 1285-1290.

68. Dubin,G. (2005) Proteinaceous cysteine protease inhibitors. *Cell Mol.Life Sci.*, **62**, 653-669.

69. Brzin,J., Rogelj,B., Popovic,T., Strukelj,B., and Ritonja,A. (2000) Clitocypin, a new type of cysteine proteinase inhibitor from fruit bodies of mushroom *Clitocybe nebularis*. *J.Biol.Chem.*, **275**, 20104-20109.

70. Sabotic,J., Popovic,T., Puizdar,V., and Brzin,J. (2009) Macrocypins, a family of cysteine protease inhibitors from the basidiomycete *Macrolepiota procera*. *FEBS J.*, **276**, 4334-4345.

71. de,O.C., Santana,L.A., Carmona,A.K., Cezari,M.H., Sampaio,M.U., Sampaio,C.A., and Oliva,M.L. (2001) Structure of cruzipain/cruzain inhibitors isolated from *Bauhinia bauhinioides* seeds. *Biol.Chem.*, **382**, 847-852.

72. Murata,M., Miyashita,S., Yokoo,C., Tamai,M., Hanada,K., Hatayama,K., Towatari,T., Nikawa,T., and Katunuma,N. (1991) Novel epoxysuccinyl peptides. Selective inhibitors of cathepsin B, in vitro. *FEBS Lett.*, **280**, 307-310.

73. Powers,J.C., Asgian,J.L., Ekici,O.D., and James,K.E. (2002) Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem.Rev.*, **102**, 4639-4750.

74. Fonovic,M. and Bogyo,M. (2008) Activity-based probes as a tool for functional proteomic analysis of proteases. *Expert.Rev.Proteomics*, **5**, 721-730.

75. Xing,R., Addington,A.K., and Mason,R.W. (1998) Quantification of cathepsins B and L in cells. *Biochem.J.*, **332**, 499-505.

76. Okamura-Oho,Y., Zhang,S., Callahan,J.W., Murata,M., Oshima,A., and Suzuki,Y. (1997) Maturation and degradation of beta-galactosidase in the post-Golgi compartment are regulated by cathepsin B and a non-cysteine protease. *FEBS Lett.*, **419**, 231-234.

77. Matsunaga,Y., Saibara,T., Kido,H., and Katunuma,N. (1993) Participation of cathepsin B in processing of antigen presentation to MHC class II. *FEBS Lett.*, **324**, 325-330.

78. Friedrichs,B., Tepel,C., Reinheckel,T., Deussing,J., von,F.K., Herzog,V., Peters,C., Saftig,P., and Brix,K. (2003) Thyroid functions of mouse cathepsins B, K, and L. *J.Clin.Invest*, **111**, 1733-1745.

79. Sohar,N., Hammer,H., and Sohar,I. (2002) Lysosomal peptidases and glycosidases in rheumatoid arthritis. *Biol.Chem.*, **383**, 865-869.

80. Baici,A., Lang,A., Horler,D., Kissling,R., and Merlin,C. (1995) Cathepsin B in osteoarthritis: cytochemical and histochemical analysis of human femoral head cartilage. *Ann.Rheum.Dis.*, **54**, 289-297.

81. Mohamed,M.M. and Sloane,B.F. (2006) Cysteine cathepsins: multifunctional enzymes in cancer. *Nat.Rev.Cancer*, **6**, 764-775.

82. Bervar,A., Zajc,I., Sever,N., Katunuma,N., Sloane,B.F., and Lah,T.T. (2003) Invasiveness of transformed human breast epithelial cell lines is related to cathepsin B and inhibited by cysteine proteinase inhibitors. *Biol.Chem.*, **384**, 447-455.

83. Tu,C., Ortega-Cava,C.F., Chen,G.S., Fernandes,N.D., Cavallo-Medved,D., Sloane,B.F., Band,V., and Band,H. (2008) Lysosomal Cathepsin B Participates in the Podosome-Mediated Extracellular Matrix Degradation and Invasion via Secreted Lysosomes in v-Src Fibroblasts. *Cancer Research*, **68**, 9147-9156.

84. Campo,E., Munoz,J., Miquel,R., Palacin,A., Cardesa,A., Sloane,B.F., and Emmert-Buck,M.R. (1994) Cathepsin B expression in colorectal carcinomas correlates with tumor progression and shortened patient survival. *Am.J.Pathol.*, **145**, 301-309.

85. Lah,T.T., Cercek,M., Blejec,A., Kos,J., Gorodetsky,E., Somers,R., and Daskal,I. (2000) Cathepsin B, a prognostic indicator in lymph node-negative breast carcinoma patients: comparison with cathepsin D, cathepsin L, and other clinical indicators. *Clin.Cancer Res.*, **6**, 578-584.

86. Gabrijelcic,D., Svetic,B., Spaic,D., Skrk,J., Budihna,M., Dolenc,I., Popovic,T., Cotic,V., and Turk,V. (1992) Cathepsins B, H and L in human breast carcinoma. *Eur.J.Clin.Chem.Clin.Biochem.*, **30**, 69-74.

87. Talieri,M., Papadopoulou,S., Scorilas,A., Xynopoulos,D., Arnogianaki,N., Plataniotis,G., Yotis,J., and Agnanti,N. (2004) Cathepsin B and cathepsin D expression in the progression of colorectal adenoma to carcinoma. *Cancer Lett.*, **205**, 97-106.

88. Cavallo-Medved,D., Rudy,D., Blum,G., Bogyo,M., Caglic,D., and Sloane,B.F. (2009) Live-cell imaging demonstrates extracellular matrix degradation in association with active cathepsin B in caveolae of endothelial cells during tube formation. *Experimental Cell Research*, **315**, 1234-1246.

89. Stoka,V., Turk,V., and Turk,B. (2007) Lysosomal cysteine cathepsins: signaling pathways in apoptosis. *Biol.Chem.*, **388**, 555-560.

90. Stoka,V., Turk,B., Schendel,S.L., Kim,T.H., Cirman,T., Snipas,S.J., Ellerby,L.M., Bredesen,D., Freeze,H., Abrahamson,M., Bromme,D., Krajewski,S., Reed,J.C., Yin,X.M., Turk,V., and Salvesen,G.S. (2001) Lysosomal protease pathways to apoptosis. Cleavage of bid, not pro-caspases, is the most likely route. *J.Biol.Chem.*, **276**, 3149-3157.

91. Gilroy,E.M., Hein,I., van der Hoorn,R., Boevink,P.C., Venter,E., McLellan,H., Kaffarnik,F., Hrubikova,K., Shaw,J., Holeva,M., Lopez,E.C., Borras-Hidalgo,O., Pritchard,L., Loake,G.J., Lacomme,C., and Birch,P.R. (2007) Involvement of cathepsin B in the plant disease resistance hypersensitive response. *Plant J.*, **52**, 1-13.

92. McLellan,H., Gilroy,E.M., Yun,B.W., Birch,P.R., and Loake,G.J. (2009) Functional redundancy in the Arabidopsis Cathepsin B gene family contributes to basal defence, the hypersensitive response and senescence. *New Phytol.*, **183**, 408-418.

93. McKerrow,J.H., Caffrey,C., Kelly,B., Loke,P., and Sajid,M. (2006) Proteases in parasitic diseases. *Annu.Rev.Pathol.*, **1**, 497-536.

94. Nobrega,O.T., Santos Silva,M.A., Teixeira,A.R., and Santana,J.M. (1998) Cloning and sequencing of tccb, a gene encoding a *Trypanosoma cruzi* cathepsin B-like protease. *Mol.Biochem.Parasitol.*, **97**, 235-240.

95. Somanna,A., Mundodi,V., and Gedamu,L. (2002) Functional analysis of cathepsin B-like cysteine proteases from *Leishmania donovani* complex. Evidence for the activation of latent transforming growth factor beta. *J.Biol.Chem.*, **277**, 25305-25312.

96. Que,X., Ngo,H., Lawton,J., Gray,M., Liu,Q., Engel,J., Brinen,L., Ghosh,P., Joiner,K.A., and Reed,S.L. (2002) The cathepsin B of *Toxoplasma gondii*, toxopain-1, is critical for parasite invasion and rhoptry protein processing. *J.Biol.Chem.*, **277**, 25791-25797.

97. Dvorak,J., Delcroix,M., Rossi,A., Vopalensky,V., Pospisek,M., Sedinova,M., Mikes,L., Sajid,M., Sali,A., McKerrow,J.H., Horak,P., and Caffrey,C.R. (2005) Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition. *Int.J.Parasitol.*, **35**, 895-910.

98. Smooker,P.M., Jayaraj,R., Pike,R.N., and Spithill,T.W. (2010) Cathepsin B proteases of flukes: the key to facilitating parasite control? *Trends Parasitol.*, **26**, 506-514.

99. Cancela,M., Acosta,D., Rinaldi,G., Silva,E., Duran,R., Roche,L., Zaha,A., Carmona,C., and Tort,J.F. (2008) A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie*, **90**, 1461-1475.

100. Ranjit,N., Zhan,B., Stenzel,D.J., Mulvenna,J., Fujiwara,R., Hotez,P.J., and Loukas,A. (2008) A family of cathepsin B cysteine proteases expressed in the gut of the human hookworm, *Necator americanus*. *Mol.Biochem.Parasitol.*, **160**, 90-99.

101. Harrop,S.A., Sawangjaroen,N., Prociv,P., and Brindley,P.J. (1995) Characterization and localization of cathepsin B proteinases expressed by adult *Ancylostoma caninum* hookworms. *Mol.Biochem.Parasitol.*, **71**, 163-171.

102. Sojka,D., Franta,Z., Horn,M., Hajdusek,O., Caffrey,C.R., Mares,M., and Kopacek,P. (2008) Profiling of proteolytic enzymes in the gut of the tick *Ixodes ricinus* reveals an evolutionarily conserved network of aspartic and cysteine peptidases. *Parasit.Vectors*, **1**, 7.

103. Turk,D., Podobnik,M., Kuhelj,R., Dolinar,M., and Turk,V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Angstroms resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett.*, **384**, 211-214.

104. vanderStappen,J.W.J., Williams,A.C., Maciewicz,R.A., and Paraskeva,C. (1996) Activation of cathepsin B, secreted by a colorectal cancer cell line requires low pH and is mediated by cathepsin D. *International Journal of Cancer*, **67**, 547-554.

105. Mach,L., Mort,J.S., and Glossl,J. (1994) Maturation of human procathepsin B. Proenzyme activation and proteolytic processing of the precursor to the mature proteinase, *in vitro*, are primarily unimolecular processes. *J.Biol.Chem.*, **269**, 13030-13035.

106. Pungercar,J.R., Caglic,D., Sajid,M., Dolinar,M., Vasiljeva,O., Pozgan,U., Turk,D., Bogyo,M., Turk,V., and Turk,B. (2009) Autocatalytic processing of procathepsin B is triggered by proenzyme activity. *FEBS J.*, **276**, 660-668.

107. Quraishi,O., Nagler,D.K., Fox,T., Sivaraman,J., Cygler,M., Mort,J.S., and Storer,A.C. (1999) The occluding loop in cathepsin B defines the pH dependence of inhibition by its propeptide. *Biochemistry*, **38**, 5017-5023.

108. Shirahama-Noda,K., Yamamoto,A., Sugihara,K., Hashimoto,N., Asano,M., Nishimura,M., and Hara-Nishimura,I. (2003) Biosynthetic processing of cathepsins and lysosomal degradation are abolished in asparaginyl endopeptidase-deficient mice. *J.Biol.Chem.*, **278**, 33194-33199.

109. Nagler,D.K., Storer,A.C., Portaro,F.C., Carmona,E., Juliano,L., and Menard,R. (1997) Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry*, **36**, 12608-12615.

110. Kirschke,H., Barrett,A.J., and Rawlings,N.D. (1995) Proteinases 1: lysosomal cysteine proteinases. *Protein Profile.*, **2**, 1581-1643.

111. Khouri,H.E., Plouffe,C., Hasnain,S., Hirama,T., Storer,A.C., and Menard,R. (1991) A model to explain the pH-dependent specificity of cathepsin B-catalysed hydrolyses. *Biochem.J.*, **275**, 751-757.

112. Turk,B., Dolenc,I., Zerovnik,E., Turk,D., Gubensek,F., and Turk,V. (1994) Human cathepsin B is a metastable enzyme stabilized by specific ionic interactions associated with the active site. *Biochemistry*, **33**, 14800-14806.

113. Mort,J.S. (2013): Cathepsin B. In: *Handbook of Proteolytic enzymes*. 3rd ed., Elsevier, Amsterdam, 1784-1791.

114. Cezari,M.H., Puzer,L., Juliano,M.A., Carmona,A.K., and Juliano,L. (2002) Cathepsin B carboxydipeptidase specificity analysis using internally quenched fluorescent peptides. *Biochem.J.*, **368**, 365-369.

115. Hasnain,S., Hirama,T., Huber,C.P., Mason,P., and Mort,J.S. (1993) Characterization of cathepsin B specificity by site-directed mutagenesis. Importance of Glu245 in the S2-P2 specificity for arginine and its role in transition state stabilization. *J.Biol.Chem.*, **268**, 235-240.

116. Shaw,E. (1990) Cysteinyl proteinases and their selective inactivation. *Adv.Enzymol.Relat Areas Mol.Biol.*, **63**, 271-347.

117. Krantz,A. (1994) Peptidyl (acyloxy)methanes as quiescent affinity labels for cysteine proteinases. *Methods Enzymol.*, **244**, 656-671.

118. Pliura,D.H., Bonaventura,B.J., Smith,R.A., Coles,P.J., and Krantz,A. (1992) Comparative behaviour of calpain and cathepsin B toward peptidyl acyloxymethyl ketones, sulphonium methyl ketones and other potential inhibitors of cysteine proteinases. *Biochem.J.*, **288**, 759-762.

119. Taralp,A., Kaplan,H., Sytwu,I.I., Vlattas,I., Bohacek,R., Knap,A.K., Hirama,T., Huber,C.P., and Hasnain,S. (1995) Characterization of the S3 subsite specificity of cathepsin B. *J.Biol.Chem.*, **270**, 18036-18043.

120. Menard,R., Carmona,E., Plouffe,C., Bromme,D., Konishi,Y., Lefebvre,J., and Storer,A.C. (1993) The specificity of the S1' subsite of cysteine proteases. *FEBS Lett.*, **328**, 107-110.

121. Krupa,J.C., Hasnain,S., Nagler,D.K., Menard,R., and Mort,J.S. (2002) S2' substrate specificity and the role of His110 and His111 in the exopeptidase activity of human cathepsin B. *Biochem.J.*, **361**, 613-619.

122. Lah,T.T., Buck,M.R., Honn,K.V., Crissman,J.D., Rao,N.C., Liotta,L.A., and Sloane,B.F. (1989) Degradation of laminin by human tumor cathepsin B. *Clin.Exp.Metastasis*, **7**, 461-468.

123. Buck,M.R., Karustis,D.G., Day,N.A., Honn,K.V., and Sloane,B.F. (1992) Degradation of extracellular-matrix proteins by human cathepsin B from normal and tumour tissues. *Biochem.J.*, **282**, 273-278.

124. Mai,J., Sameni,M., Mikkelsen,T., and Sloane,B.F. (2002) Degradation of extracellular matrix protein tenascin-C by cathepsin B: an interaction involved in the progression of gliomas. *Biol.Chem.*, **383**, 1407-1413.

125. Turk,D., Podobnik,M., Popovic,T., Katunuma,N., Bode,W., Huber,R., and Turk,V. (1995) Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry*, **34**, 4791-4797.

126. Yamamoto,A., Hara,T., Tomoo,K., Ishida,T., Fujii,T., Hata,Y., Murata,M., and Kitamura,K. (1997) Binding mode of CA074, a specific irreversible inhibitor, to bovine cathepsin B as determined by X-ray crystal analysis of the complex. *J.Biochem.*, **121**, 974-977.

127. Redzynia,I., Ljunggren,A., Abrahamson,M., Mort,J.S., Krupa,J.C., Jaskolski,M., and Bujacz,G. (2008) Displacement of the occluding loop by the parasite protein, chagasin, results in efficient inhibition of human cathepsin B. *J.Biol.Chem.*, **283**, 22815-22825.

128. Renko,M., Pozgan,U., Majera,D., and Turk,D. (2010) Stefin A displaces the occluding loop of cathepsin B only by as much as required to bind to the active site cleft. *FEBS J.*, **277**, 4338-4345.

129. Roberts,L.S., Janovy,J. (2006) *Foundations of Parasitology*. 7th ed., McGraw-Hill, New York, U.S.A., 702 stran.

130. Jira,J. (1999) Klasifikace zooparazitů a parazitóz člověka. *Klinická mikrobiologie*, **3 Suppl 1**, S1-S32.

131. http://www.who.int/mediacentre/factsheets/fs115/en/ (7.7.2014).

132. Mair,G.R., Maule,A.G., Fried,B., Day,T.A., and Halton,D.W. (2003) Organization of the musculature of schistosome cercariae. *Journal of Parasitology*, **89**, 623-625.

133. Haas,W., Grabe,K., Geis,C., Pach,T., Stoll,K., Fuchs,M., Haberl,B., and Loy,C. (2002) Recognition and invasion of human skin by *Schistosoma mansoni* cercariae: the key-role of L-arginine. *Parasitology*, **124**, 153-167.

134. Knudsen,G.M., Medzihradszky,K.F., Lim,K.C., Hansell,E., and McKerrow,J.H. (2005) Proteomic analysis of *Schistosoma mansoni* cercarial secretions. *Molecular & Cellular Proteomics*, **4**, 1862-1875.

135. http://www.merckmanuals.com/professional/infectious_diseases/trematodes_flukes/schistosomiasis.html (7.7.2014).

136. http://www.nhm.ac.uk/nature-online/species-of-the-day/scientific-advances/disease/schistosoma-mansoni/lifecycle/index.html (7.7.2014).

137. http://www.path.cam.ac.uk/~schisto/schistosoma/schisto_lifecycle_miracidium.html (7.7.2017).

138. http://www.path.cam.ac.uk/~schisto/schistosoma/schisto_lifecycle_schistosomula.html (7.7.2014).

139. Schadewaldt,H. (1962) Theodor Bilharz (1825-1862), one of the founders of German tropical medicine research. *Munch.Med.Wochenschr.*, **104**, 1730-1734.

140. Utzinger,J., N'goran,E.K., Caffrey,C.R., and Keiser,J. (2011) From innovation to application: social-ecological context, diagnostics, drugs and integrated control of schistosomiasis. *Acta Trop.*, **120 Suppl 1**, S121-S137.

141. Zhou,X.N., Yang,G.J., Yang,K., Wang,X.H., Hong,Q.B., Sun,L.P., Malone,J.B., Kristensen,T.K., Bergquist,N.R., and Utzinger,J. (2008) Potential impact of climate change on schistosomiasis transmission in China. *Am.J.Trop.Med.Hyg.*, **78**, 188-194.

142. http://www.who.int/neglected_diseases/about/en/ (7.7.2014).

143. Hotez,P.J., Molyneux,D.H., Fenwick,A., Kumaresan,J., Sachs,S.E., Sachs,J.D., and Savioli,L. (2007) Control of neglected tropical diseases. *N.Engl.J.Med.*, **357**, 1018-1027.

144. Hotez,P.J. and Kamath,A. (2009) Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden. *PLoS.Negl.Trop.Dis.*, **3**, e412.

145. Hotez,P.J. and Fenwick,A. (2009) Schistosomiasis in Africa: an emerging tragedy in our new global health decade. *PLoS.Negl.Trop.Dis.*, **3**, e485.

146. http://www.globalnetwork.org/schistosomiasis (7.7.2014).

147. Harnett,W. and Kusel,J.R. (1986) Increased exposure of parasite antigens at the surface of adult male *Schistosoma mansoni* exposed to praziquantel *in vitro*. *Parasitology*, **93 (Pt 2)**, 401-405.

148. Jeziorski,M.C. and Greenberg,R.M. (2006) Voltage-gated calcium channel subunits from platyhelminths: potential role in praziquantel action. *Int.J.Parasitol.*, **36**, 625-632.

149. Utzinger,J., Chollet,J., You,J., Mei,J., Tanner,M., and Xiao,S. (2001) Effect of combined treatment with praziquantel and artemether on *Schistosoma japonicum* and *Schistosoma mansoni* in experimentally infected animals. *Acta Trop.*, **80**, 9-18.

150. De,C.D., Vercruysse,J., Verle,P., Kongs,A., and Diop,M. (2000) What is the effect of combining artesunate and praziquantel in the treatment of *Schistosoma mansoni* infections? *Trop.Med.Int.Health*, **5**, 744-746.

151. Riveau,G., Deplanque,D., Remoue,F., Schacht,A.M., Vodougnon,H., Capron,M., Thiry,M., Martial,J., Libersa,C., and Capron,A. (2012) Safety and immunogenicity of rSh28GST

antigen in humans: phase 1 randomized clinical study of a vaccine candidate against urinary schistosomiasis. *PLoS.Negl.Trop.Dis.*, **6**, e1704.

152. McManus,D.P. and Loukas,A. (2008) Current status of vaccines for schistosomiasis. *Clin.Microbiol.Rev.*, **21**, 225-242.

153. Gryseels,B. (2000) Schistosomiasis vaccines: a devils' advocate view. *Parasitol.Today*, **16**, 46-48.

154. Boros,D.L. (1989) Immunopathology of *Schistosoma mansoni* infection. *Clin.Microbiol.Rev.*, **2**, 250-269.

155. Abdulla,M.H., Ruelas,D.S., Wolff,B., Snedecor,J., Lim,K.C., Xu,F., Renslo,A.R., Williams,J., McKerrow,J.H., and Caffrey,C.R. (2009) Drug discovery for schistosomiasis: hit and lead compounds identified in a library of known drugs by medium-throughput phenotypic screening. *PLoS.Negl.Trop.Dis.*, **3**, e478.

156. Berriman,M., Haas,B.J., LoVerde,P.T., Wilson,R.A., Dillon,G.P., Cerqueira,G.C., Mashiyama,S.T., Al-Lazikani,B., Andrade,L.F., Ashton,P.D., Aslett,M.A., Bartholomeu,D.C., Blandin,G., Caffrey,C.R., Coghlan,A., Coulson,R., Day,T.A., Delcher,A., DeMarco,R., Djikeng,A., Eyre,T., Gamble,J.A., Ghedin,E., Gu,Y., Hertz-Fowler,C., Hirai,H., Hirai,Y., Houston,R., Ivens,A., Johnston,D.A., Lacerda,D., Macedo,C.D., McVeigh,P., Ning,Z.M., Oliveira,G., Overington,J.P., Parkhill,J., Pertea,M., Pierce,R.J., Protasio,A.V., Quail,M.A., Rajandream,M.A., Rogers,J., Sajid,M., Salzberg,S.L., Stanke,M., Tivey,A.R., White,O., Williams,D.L., Wortman,J., Wu,W.J., Zamanian,M., Zerlotini,A., Fraser-Liggett,C.M., Barrell,B.G., and El-Sayed,N.M. (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature*, **460**, 352-U65.

157. Protasio,A.V., Tsai,I.J., Babbage,A., Nichol,S., Hunt,M., Aslett,M.A., De,S.N., Velarde,G.S., Anderson,T.J., Clark,R.C., Davidson,C., Dillon,G.P., Holroyd,N.E., LoVerde,P.T., Lloyd,C., McQuillan,J., Oliveira,G., Otto,T.D., Parker-Manuel,S.J., Quail,M.A., Wilson,R.A., Zerlotini,A., Dunne,D.W., and Berriman,M. (2012) A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl.Trop.Dis.*, **6**, e1455.

158. Le,T.H., Humair,P.F., Blair,D., Agatsuma,T., Littlewood,D.T., and McManus,D.P. (2001) Mitochondrial gene content, arrangement and composition compared in African and Asian schistosomes. *Mol.Biochem.Parasitol.*, **117**, 61-71.

159. Nahum,L.A., Mourao,M.M., and Oliveira,G. (2012) New frontiers in schistosoma genomics and transcriptomics. *J.Parasitol.Res.*, **2012**, 849132.

160. Lawton,S.P., Hirai,H., Ironside,J.E., Johnston,D.A., and Rollinson,D. (2011) Genomes and geography: genomic insights into the evolution and phylogeography of the genus *Schistosoma*. *Parasit.Vectors.*, **4**, 131.

161. Horn,M., Fajtova,P., Rojo,A.L., Ulrychova,L., Bartosova-Sojkova,P., Franta,Z., Protasio,A.V., Opavsky,D., Vondrasek,J., McKerrow,J.H., Mares,M., Caffrey,C.R., and Dvorak,J. (2014) Trypsin- and Chymotrypsin-like serine proteases in *Schistosoma mansoni*- 'the undiscovered country'. *PLoS.Negl.Trop.Dis.*, **8**, e2766.

162. Bos,D.H., Mayfield,C., and Minchella,D.J. (2009) Analysis of regulatory protease sequences identified through bioinformatic data mining of the *Schistosoma mansoni* genome. *BMC.Genomics*, **10**, 488.

163. Lawrence,J.D. (1973) The ingestion of red blood cells by *Schistosoma mansoni*. *J.Parasitol.*, **59**, 60-63.

164. Kloetzel,K. and Lewert,R.M. (1966) Pigment formation in *Schistosoma mansoni* infections in the white mouse. *Am.J.Trop.Med.Hyg.*, **15**, 28-31.

165. Horn,M., Nussbaumerova,M., Sanda,M., Kovarova,Z., Srba,J., Franta,Z., Sojka,D., Bogyo,M., Caffrey,C.R., Kopacek,P., and Mares,M. (2009) Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. *Chem.Biol.*, **16**, 1053-1063.

166. Williamson,A.L., Lecchi,P., Turk,B.E., Choe,Y., Hotez,P.J., McKerrow,J.H., Cantley,L.C., Sajid,M., Craik,C.S., and Loukas,A. (2004) A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. *J.Biol.Chem.*, **279**, 35950-35957.

167. Ruppel,A., Shi,Y.E., Wei,D.X., and Diesfeld,H.J. (1987) Sera of *Schistosoma japonicum*-infected patients cross-react with diagnostic 31/32 kD proteins of *S. mansoni*. *Clin.Exp.Immunol.*, **69**, 291-298.

168. Chappell,C.L. and Dresden,M.H. (1986) *Schistosoma mansoni*: proteinase activity of "hemoglobinase" from the digestive tract of adult worms. *Exp.Parasitol.*, **61**, 160-167.

169. Jolly,E.R., Chin,C.S., Miller,S., Bahgat,M.M., Lim,K.C., DeRisi,J., and McKerrow,J.H. (2007) Gene expression patterns during adaptation of a helminth parasite to different environmental niches. *Genome Biol.*, **8**, R65.

170. Planchart,S., Incani,R.N., and Cesari,I.M. (2007) Preliminary characterization of an adult worm "vomit" preparation of *Schistosoma mansoni* and its potential use as antigen for diagnosis. *Parasitol.Res.*, **101**, 301-309.

171. Losada,S., Chacon,N., Colmenares,C., Bermudez,H., Lorenzo,A., Pointier,J.P., Theron,A., Alarcon de,N.B., and Noya,O. (2005) *Schistosoma*: cross-reactivity and antigenic community among different species. *Exp.Parasitol.*, **111**, 182-190.

172. Ruppel,A., Diesfeld,H.J., and Rother,U. (1985) Immunoblot analysis of *Schistosoma mansoni* antigens with sera of schistosomiasis patients: diagnostic potential of an adult schistosome polypeptide. *Clin.Exp.Immunol.*, **62**, 499-506.

173. Klinkert,M.Q., Bommert,K., Moser,D., Felleisen,R., Link,G., Doumbo,O., and Beck,E. (1991) Immunological analysis of cloned *Schistosoma mansoni* antigens Sm31 and Sm32 with sera of schistosomiasis patients. *Trop.Med.Parasitol.*, **42**, 319-324.

174. Klinkert,M.Q., Ruppel,A., and Beck,E. (1987) Cloning of diagnostic 31/32 kilodalton antigens of *Schistosoma mansoni*. *Mol.Biochem.Parasitol.*, **25**, 247-255.

175. Caffrey,C.R., Salter,J.P., Lucas,K.D., Khiem,D., Hsieh,I., Lim,K.C., Ruppel,A., McKerrow,J.H., and Sajid,M. (2002) SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. *Mol.Biochem.Parasitol.*, **121**, 49-61.

176. Klinkert,M.Q., Felleisen,R., Link,G., Ruppel,A., and Beck,E. (1989) Primary structures of Sm31/32 diagnostic proteins of *Schistosoma mansoni* and their identification as proteases. *Mol.Biochem.Parasitol.*, **33**, 113-122.

177. Baig,S., Damian,R.T., and Peterson,D.S. (2002) A novel cathepsin B active site motif is shared by helminth bloodfeeders. *Exp.Parasitol.*, **101**, 83-89.

178. Felleisen,R. and Klinkert,M.Q. (1990) *In vitro* translation and processing of cathepsin B of *Schistosoma mansoni*. *EMBO J.*, **9**, 371-377.

179. Klinkert,M.Q., Cioli,D., Shaw,E., Turk,V., Bode,W., and Butler,R. (1994) Sequence and structure similarities of cathepsin B from the parasite *Schistosoma mansoni* and human liver. *FEBS Lett.*, **351**, 397-400.

180. Gotz,B., Felleisen,R., Shaw,E., and Klinkert,M.Q. (1992) Expression of an active cathepsin B-like protein Sm31 from *Schistosoma mansoni* in insect cells. *Trop.Med.Parasitol.*, **43**, 282-284.

181. Lipps,G., Fullkrug,R., and Beck,E. (1996) Cathepsin B of *Schistosoma mansoni*. Purification and activation of the recombinant proenzyme secreted by Saccharomyces cerevisiae. *J.Biol.Chem.*, **271**, 1717-1725.

182. Choe,Y., Leonetti,F., Greenbaum,D.C., Lecaille,F., Bogyo,M., Bromme,D., Ellman,J.A., and Craik,C.S. (2006) Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *J.Biol.Chem.*, **281**, 12824-12832.

183. Stack,C.M., Dalton,J.P., Cunneen,M., and Donnelly,S. (2005) De-glycosylation of *Pichia pastoris*-produced *Schistosoma mansoni* cathepsin B eliminates non-specific reactivity with IgG in normal human serum. *J.Immunol.Methods*, **304**, 151-157.

184. Ghoneim,H. and Klinkert,M.Q. (1995) Biochemical properties of purified cathepsin B from *Schistosoma mansoni*. *Int.J.Parasitol.*, **25**, 1515-1519.

185. Correnti,J.M., Brindley,P.J., and Pearce,E.J. (2005) Long-term suppression of cathepsin B levels by RNA interference retards schistosome growth. *Mol.Biochem.Parasitol.*, **143**, 209-215.

186. McKerrow,J.H. (2005) Designing drugs for parasitic diseases of the developing world. *PLoS.Med.*, **2**, e210.

187. Cardin,A.D. and Weintraub,H.J. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis*, **9**, 21-32.

188. Fanfrlik,J., Bronowska,A.K., Rezac,J., Prenosil,O., Konvalinka,J., and Hobza,P. (2010) A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 protease with 22 ligands. *J.Phys.Chem.B*, **114**, 12666-12678.

189. Krautz-Peterson,G. and Skelly,P.J. (2008) Schistosome asparaginyl endopeptidase (legumain) is not essential for cathepsin B1 activation in vivo. *Mol.Biochem.Parasitol.*, **159**, 54-58.

190. Podobnik,M., Kuhelj,R., Turk,V., and Turk,D. (1997) Crystal structure of the wild-type human procathepsin B at 2.5 A resolution reveals the native active site of a papain-like cysteine protease zymogen. *J.Mol.Biol.*, **271**, 774-788.

191. Cygler,M., Sivaraman,J., Grochulski,P., Coulombe,R., Storer,A.C., and Mort,J.S. (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure.*, **4**, 405-416.