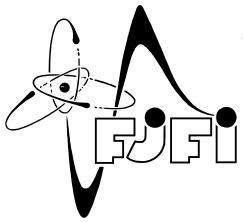
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ

FAKULTA JADERNÁ A FYZIKÁLNĚ INŽENÝRSKÁ



Pozorování interakcí bílkovin s ligandy pomocí makromolekulární rentgenové krystalografie

Protein:ligand interactions viewed by macromolecular X-ray crystallography

Autor: Ing. Petr Kolenko, PhD.

Rok: 2017

Prohlášení

Prohlašuji, že jsem tuto práci vypracoval samostatně a použil jsem pouze podklady (literaturu, projekty, SW atd.) uvedené v přiloženém seznamu.

.....	.....
Místo a datum	Ing. Petr Kolenko, PhD.

Poděkování

V prvé řadě chci poděkovat Prof. Miltonu T. Stubbsovi za příležitost pracovat v jeho laboratoři během mé post-doktorské stáže, za jeho podporu v mé samostatné vědecké činnosti a za mnoho inspirujících myšlenek. Dále chci poděkovat Ing. Janu Dohnálkovi, Ph.D. za příležitost spolupracovat na mnoha zajímavých projektech. Nemohu opomenout vedoucího mé diplomové a doktorské práce RNDr. Jindřicha Haška, DrSc., který mne jako osobnost významně oslovil a přitáhl můj zájem o tuto nádhernou vědeckou disciplínu. Největší dík patří vedoucímu Katedry inženýrství pevných látek Doc. Ing. Ladislavu Kalvodovi, CSc., bez jehož podpory a poskytnuté příležitosti by tato práce vůbec nemohla vzniknout. Na závěr chci poděkovat své rodině za celkovou podporu v mé práci, výzkumných aktivitách a osobních zájmech.

Ing. Petr Kolenko, PhD.

Title:

Protein:ligand interactions viewed by macromolecular X-ray crystallography

Author: Ing. Petr Kolenko, PhD.

Abstract: Ligands in macromolecular structures are used as tools to decipher the structure:function relationship of the target molecules. They proved to be useful in design of novel inhibitors of glutaminyl cyclases, inhibitors of bovine cationic trypsin, inhibitors of biotechnologic nucleases, and overall understanding of human interferon γ receptor 2.

Key words: protein, ligand, X-ray structure analysis

Název práce:

Pozorování interakcí bílkovin s ligandy pomocí makromolekulární rentgenové krystalografie

Autor: Ing. Petr Kolenko, PhD.

Abstrakt: Ligandy ve strukturách makromolekul pomáhají odhalovat vtah mezi strukturou a funkcí. Hrály důležitou roli při návrhu inhibitorů enzymů glutaminylcyklasa, hovězí trypsin a u dvou biotechnologických nukleas. Přispějí rovněž k celkovému pochopení role receptoru 2 lidského interferonu γ.

Klíčová slova: bílkovina, ligand, rentgenová strukturní analýza

Obsah

Kapitola 1

Předmluva

Nejen chemické složení, ale především znalost vnitřní struktury je zcela nezbytnou součástí vědomostí, které vedou k pochopení vlastnosti jakéhokoliv materiálu. Toto tvrzení je platné pro sloučeniny s malým počtem atomů i pro mnohem větší makromolekulární objekty. Jeden z nejznámějších příkladů je uhlík ve dvou rozdílných strukturních formách: grafit a diamant. Oba tyto materiály mají stejné atomární složení, ale jejich vnitřní struktura a uspořádání vazeb je zcela jiné a předurčuje oba materiály k jiné aplikaci. Podobně u makromolekulárních systémů hraje důležitou úlohu struktura. Pomocí struktury lze odhalit, proč enzym o složení $\mathrm{C_{2468}N_{651}O_{719}H_{\mathit{x}}S_8P_2}$ je aktivní, zatímco mutovaná forma lišící se pouze o jeden atom kyslíku je takřka neaktivní. Odhalení takovéto skutečnosti vyžaduje strukturní analýzu s velmi vysokým rozlišením. Jedinou metodou, která pro obor makromolekul takto vysokého rozlišení doposud dosahuje, je monokrystalová difrakční analýza.

Výzkum v oblasti strukturní biologie je typickým příkladem mezioborového výzkumu, který vzájemně propojuje moderní fyzikální techniky s poznatky z oblasti organické chemie, biochemie, biologie aj. Hlavní metodou strukturní biologie je makromolekulární krystalografie. Více než 80% všech známých struktur makromolekul bylo vyřešeno právě touto metodou. Výsledky difrakčních experimentů často přináší nové impulzy pro rozvoj nových biomateriálů, biotechnologií apod. Metoda sama významně přispívá k odhalení širokého spektra buněčných procesů na molekulární úrovni. Výsledky těchto analýz navíc

mohou najít uplatnění i v široké škále biotechnologií a průmyslu. Většina nových léčiv je založena na znalosti struktur cílových makromolekul.

Tato habilitační práce je předkládána k posouzení na Fakultě jaderné a fyzikálně inženýrské Českého vysokého učení technického v Praze. Předložená práce se stává z úvodu, krátkého komentáře k publikacím z let 2013-2017, životopisu a seznamu významné publikační aktivity doplněného o kopie publikací zmíněných v textu, na nichž se habilitant tvůrčí činností podílel.

Kapitola 2

Úvod

Monokrystalová strukturní analýza je dominantní metodou strukturní biologie, která umožňuje zobrazovat struktury makromolekul na atomární úrovni. Znalost struktury je zásadní pro plné pochopení vlastností makromolekul a jejich komplexů získaných na základě jejich biochemické a biofyzikální analýzy.

Ligandy jsou molekuly, které interagují s cílovou molekulou. Mohou mít velikost od několika atomů až po makromolekulární charakter. Biofyzikální a strukturní charakterizace těchto interakcí umožňuje pozorovat cílové molekuly v určitém stádiu jejich aktivity. Ke strukturním změnám může docházet na cílové makromolekule i na samotném ligandu. V případě enzymů ligandy (např. inhibitory) pomáhají odhalovat mechanismus katalýzy. Strukturní analýza je obzvlášť důležitá při návrhu a optimalizaci nových léčiv nebo při cíleném návrhu mutovaných forem enzymů pro aplikaci ve specifických laboratorních i průmyslových procesech.

Mnoho biologických procesů je řízeno interakcemi mezi makromolekulami a tvorbou komplexů mezi dvěma a více makromolekulami. Struktury takových komplexů pomohly odhalit molekulární mechanismy mnoha buněčných procesů od transkripce či translace, přes dýchání až po procesy imunitního systému. Krystalizace makromolekulárních komplexů pro účely strukturní analýzy je často obtížná, v některých případech zatím nemožná. Proto byly vyvinuty doplňkové experimentální metody, např. kryoelektronová mikroskopie, maloúhlový rozptyl rentgenového záření i neutronů a hmotnostní spektrometrie, které

v některých případech poskytují nové informace o strukturách krystalograficky zatím neřešitelných. Tyto metody však poskytují ve srovnání s makromolekulární krystalografií struktury s výrazně nižším rozlišením. Schopnost vyřešit strukturu ohromných komplexů těmito metodami také často závisí na předchozí krystalografické znalosti struktur podjednotek cílového komplexu. I proto zůstává makromolekulární krystalografie hlavním nástrojem strukturní biologie poskytujícím detailní informaci o struktuře makromolekul obecně, případně podjednotek velkých komplexů pro výše zmíněné doplňkové metody.

Tato práce je převážně zaměřena na strukturní studie čtyř různých projektů: glutaminyl cyklasa z *Drosophila melanogaster*, inhibice mutovaných variant hovězího trypsinu, biotechnologické nukleasy a receptor 2 lidského interferonu gamma. První dva projekty jsou zaměřeny na vazebné a strukturní studie nízkomolekulárních ligandů enzymů důležitých pro lidské zdraví, třetí je zaměřen na enzymy s potenciálním využitím v průmyslu a čtvrtý projekt popisuje strukturní studii důležité molekuly imunitního systému.

1. Hlavní část výzkumu týkající se návrhu inhibitorů enzymu glutaminylcyklasa (QC) z organismu *Drosophila melanogaster* byla publikována před posuzovaným habilitačním řízením (tj. před lety 2013 až 2017) [1]. Poslední z publikací několik let probíhajícího výzkumu popisuje strukturu neinhibovaného enzymu s novou prostorovou grupou $I4$ [2].

2. Čtyři mutované formy hovězího trypsinu byly navrženy k analýze vlivu dvou poloh z druhé interakční slupky oblasti aktivního místa. Mutované varianty obsahovaly vůči původnímu trypsinu další mutace, které napodobovaly vazebná místa faktoru Xa, neboť tento enzym hraje důležitou roli v koagulační kaskádě. Práce se zaměřovala především na strukturní změny aktivního místa, schopnost inhibitorů inhibovat trypsin a dále na vazebné módy sady devíti různých inhibitorů [3]. Inhibitory faktoru Xa se používají jako léky proti trombóze.

3. Nukleasy jsou enzymy, které štěpí fosfodiesterovou vazbu mezi monomery nuk-

leových kyselin. Reprezentují širokou škálu enzymů s velkým potenciálem pro biotechnologické využití v průmyslu. Předložená práce se zabývá dvěma enzymy s odlišným biotechnologickým využitím: nukleasa TBN1 z organismu *Solanum lycopersicum* a nukleasa S1 z organismu *Aspergillus oryzae*. TBN1 nukleasa má potenciální využití v nádorových terapiích, S1 nukleasa je již používána v laboratorních protokolech a je průmyslově distribuována k analýze sekundární struktury nukleových kyselin. U obou nukleas byl studován mechanismus katalýzy a byly provedeny studie vazby ligandů do jejich aktivního místa.

4. Interferon γ je významná molekula lidského imunitního systému. Účastní se v různých procesech vrozené i adaptivní části imunitního systému. Aktivita molekuly je vázána na tvorbu ternárního komplexu interferonu γ s receptorem 1 a 2. Byla vyřešena struktura receptoru 2 s vysokým rozlišením [4]. Bylo provedeno mnoho neúspěšných pokusů zkrystalizovat ternární komplex. Současná dostupnost struktur jednotlivých podjednotek alespoň umožňuje předpověď (výpočet) mechanismu tvorby komplexu a další experimentální práci, která by v budoucnu mohla být terapeuticky využita.

Veškerá analýza a diskuze jsou zaměřeny převážně na výsledky strukturních studií, neboť se jedná právě o oblast specializace habilitanta. Práce je doplněna seznamem krystalových struktur, na jejichž řešení se habilitant podílel (viz. Přílohy).

Kapitola 3

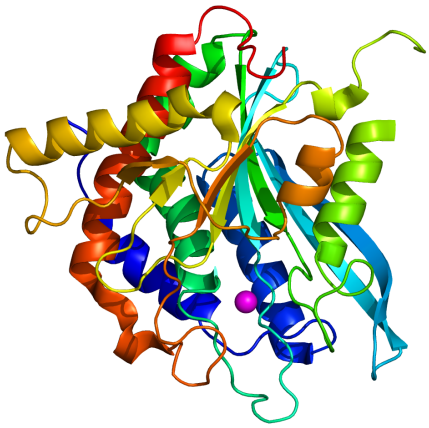
Interakce makromolekul s ligandy

3.1 Glutaminylcyklasa z organismu *Drosophila melanogaster*

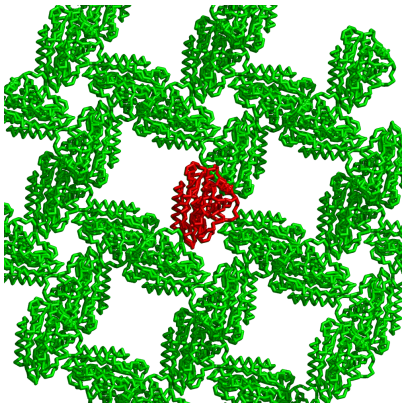
Glutaminylcyclasa (QC) katalyzuje tvorbu kyseliny pyroglutamové (pGlu) na N-konci peptidového řetězce. Předpokládá se, že hraje roli v rozvoji Alzheimerovy choroby [5]. Proto lidská QC představuje potenciální cíl pro návrh nových léčiv proti Alzheimerově chorobě na základě znalosti atomární struktury.

Dostupnost širokého spektra krystalizačních podmínek, ve kterých lze pěstovat krystaly různých krystalových forem, je zásadní pro detailní strukturní analýzu a analýzu vazby ligandů na cílový protein. V případě savčích QC se ukázalo, že krystalové kontakty mohou významně ovlivňovat způsob vazby inhibitoru PBD150 do aktivního místa enzymu [6]. Proto byly určeny struktury dvou isoforem QC z organismu *Drosophila melanogaster* (*Dm*) a jejich mutovaných forem v komplexu s inhibitorem PBD150 [1].

Enzym *Dm*QC původně krystalizoval s prostorovými grupami $P2_1$ a $P6_5$ [1]. Byly nalezeny nové krystalizační podmínky, které vedly k růstu nové formy krystalů s prostorovou grupou $I4$ [2]. Uspořádání molekul v krystalu z této krystalizační podmínky sice zabraňuje vazbě inhibitoru PBD150 do aktivního místa (aktivní místo v oblasti rezidua Phe292 je obsazeno řetězcem symetricky sdružené molekuly), přesto protokol zůstává



(a)



(b)

Obrázek 3.1: Struktura glutaminylcyklasy z organismu *Drosophila melanogaster*. (a) *Dm*QC je znázorněna v reprezentaci sekundární struktury (duhové zabarvení: N-konec modře, C-konec červeně), kationt zinku v aktivním místě je znázorněn fialovou koulí. (b) Uspořádání molekul *Dm*QC v krystalu s prostorovou grupou $I4$. Molekula *Dm*QC je reprezentována polohami atomů C^{α}. Původní molekula je zabarvena červeně, symetricky sdružené molekuly jsou zabarveny zeleně.

užitečný pro testování nových typů inhibitorů o jiném složení s odlišnými vazebnými módy. Výhodou této krystalizační podmínky je vysoká vnitřní symetrie a relativně dobrá difrakční kvalita vypěstovaných krystalů ve srovnání s předchozími publikovanými podmínkami [2].

Struktura, aktivní místo a mezimolekulární interakce uvnitř krystalu jsou znázorněny na obrázku 3.1. Krystalografická data jsou uvedena v publikaci [2].

3.2 Vývoj nových léčiv a interakce mezi enzymy a ligandy

Moderní metody vývoje nových léčiv velmi závisí na dostupnosti struktury cílové bílkoviny a struktury jejího komplexu s ligandem. Většina současných léčiv jsou velmi specifické inhibitory cílových enzymů. Krystalová struktura komplexu enzym:ligand však neposkytuje ucelenou informaci nutnou k návrhu či vylepšení léčiva. Potenciál inhibovat cílový enzym musí být vždy ověřen dodatečnými biochemickými experimenty. Výpočet inhibičních parametrů zvoleného inhibitoru pouze na základě znalosti struktury není doposud přesně zpracován. Optimalizace výpočetních parametrů stále představuje aktuální

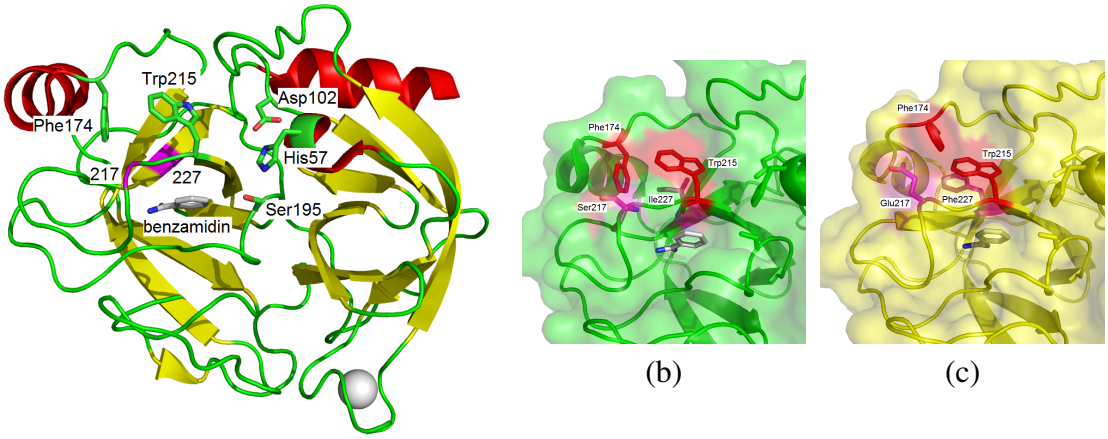
problém.

Nadrodina trypsinu podobných proteas je vhodný modelový systém pro zkoumání vazby ligandů a substrátové selektivity. Obsahuje velikou škálu funkčně odlišných enzymů, jejichž struktury se od sebe liší, s výjimkou mutovaných míst, jen nepatrně. Navíc obsahuje mnoho terapeuticky využitelných cílů, například inhibitory koagulačního faktoru Xa reprezentují léčiva regulující systém krevní hemostázy [3].

Pro účely naší studie byla nejdříve mutována rezidua aktivního místa hovězího trypsinu podle faktoru Xa. Následně byly provedeny dvě další mutace reziduí 217 a 227. Reziduum 217 se nachází na okraji substrát vázajícího místa, reziduum 227 je skryto pod reziduem Trp215, které interaguje s většinou inhibitorů v naší studii [3]. Celková struktura hovězího trypsinu se zvýrazněním poloh zkoumaných mutací je ukázána na obrázku 3.2.

Bylo vyřešeno 21 struktur (v publikaci uvedeno pouze 20) různých variant čtyř mutovaných forem a devíti inhibitorů. Rozlišení těchto struktur se pohybuje v rozmezí od 1,18 do 2,9 Å. Tyto struktury byly dále porovnány s předchozími variantami [7, 8]. Celkem bylo pozorováno devět krystalových forem, čtyři z nich již byly pozorovány dříve [8].

Všechna strukturní pozorování byla posléze konfrontována s potenciálem inhibovat mutovanou formu hovězího trypsinu. Přes výrazné rozdíly v afinitách inhibitorů k odlišným mutovaným formám byly pozorovány minimální strukturní změny. Již dříve bylo pozorováno, že k nejvýraznějším strukturním rozdílům dochází na smyčce obsahující reziduum Phe174, která nabývá dvou různých konformací [8]. Tyto konformace velmi významně ovlivňují schopnost inhibovat konkrétní mutovanou formu. Bylo prokázáno, že reziduum 217 má rovněž významný vliv na tuto oblast. Může hrát stabilizační i destabilizační roli. Reziduum 227 (původně Val227) se nachází pod reziduem Trp215, které poskytuje hydrofobní oblast pro vazbu substrátu/ligandu. Trp215 se nachází v těsné blízkosti Phe174. Záměna rezidua 227 tedy také ovlivňuje inhibiční vlastnosti inhibitorů vůči zvoleným mutovaným formám, přestože reziduum 227 se vůbec nepodílí na interakcích enzym:inhibitor.



(a) (b) (c)

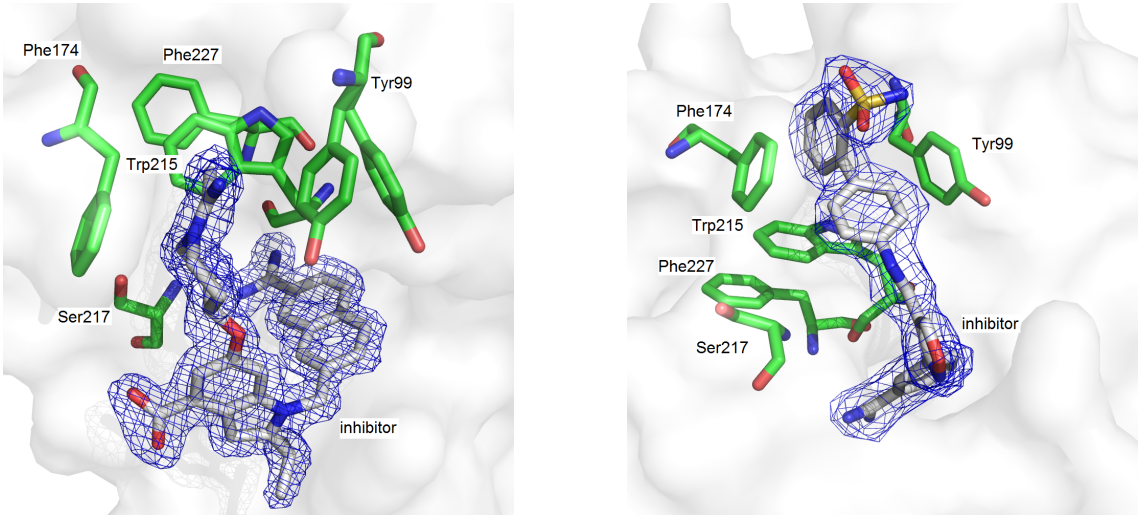
Obrázek 3.2: Struktura mutované formy hovězího trypsinu v komplexu s inhibitorem benzamidin. (a) Enzym je reprezentován zabarvenými prvky sekundární struktury, kationt vápníku je znázorněn bílou koulí, rezidua aktivního místa (His57, Asp102, Ser195) a další významná rezidua (Phe174 a Trp215) jsou znázorněna tyčinkovým modelem se zeleně zabarvenými uhlíkovými atomy, molekula benzamidinu je reprezentována tyčinkovým modelem se šedými uhlíkovými atomy, pozice mutovaných reziduí 217 a 227 jsou označeny fialovou barvou. (b) a (c) Detailní pohled na okolí aktivního místa struktury se spodní konformací rezidua Phe174 (panel b, PDB kód 3PLB) a horní konformací rezidua Phe174 (panel c, PDB kód 3UY9).

Přes velkou strukturní podobnost zkoumaných variant byly pozorovány velké rozdíly mezi experimentálními afinitami inhibitorů faktoru Xa a námi zkoumanými mutovanými variantami hovězího trypsinu. Přes naprostou nezbytnost strukturních analýz při návrhu nových léčiv demonstruje tento fakt omezení současných výpočetních přístupů a nutnost veškerá strukturní pozorování doplněná o předpovědi potenciálu navrhovaných inhibitorů konfrontovat s experimentálními přístupy.

Detailní pohled na vazbu dvou inhibiorů včetně pozorované mapy elektronové hustoty je ukázán na obrázku 3.3.

3.3 Biotechnologicky využitelné nukleasy

Nukleasy jsou enzymy, které lze průmyslově i vědecky využít v široké škále aplikací. Jsou využívány v laboratorních protokolech pro degradaci a odstraňování nukleových kyselin z různých druhů roztoků a rovněž jako analytický nástroj. Znalost jejich vlastností



(a) (b)

Obrázek 3.3: Detailní pohled na interakce inhibitorů s mutovanými formami hovězího trypsinu. Enzym je znázorněn průhlednou povrchovou reprezentací, významná rezidua (zelené atomy uhlíku) a inhibitory (šedé atomy uhlíku) jsou znázorněna tyčinkovou reprezentací. 2mFo-DFc mapa elektronové hustoty je znázorněna na hladině 1 σ modrou sítí. (a) PDB kód 3UPE. (b) PDB kód 3UQV.

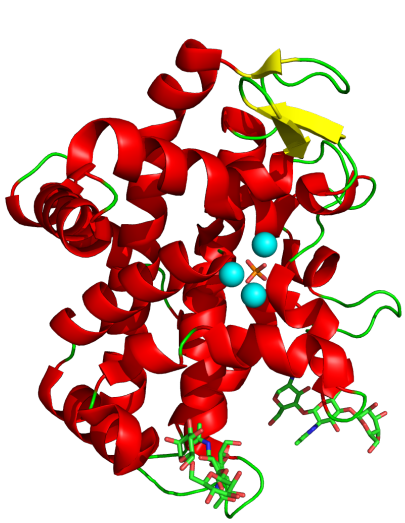
a struktur je také důležitá pro lidské zdraví.

3.3.1 Nukleasa TBN1 z organismu *Solanum lycopersicum*

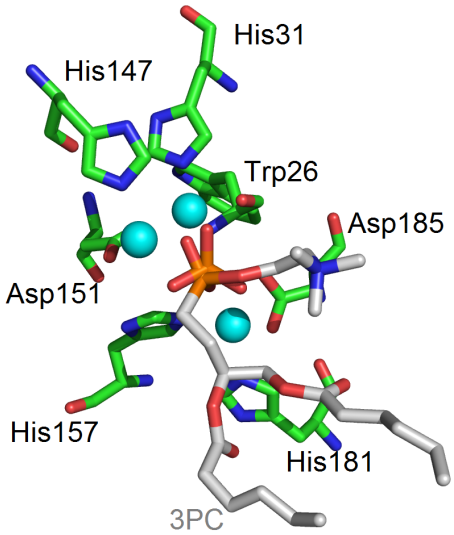
Multifunkční nukleasa TBN1 z organismu *Solanum lycopersicum* (rajče) hraje důležitou roli v apoptotických procesech a buněčném stárnutí v rostlinách, a proto se jedná o atraktivní molekulu s potenciálním využitím v nádorových terapiích [9].

Původní krystalová struktura TBN1 nukleasy neumožňovala nasáknutí nízkomolekulárních ligandů do aktivního místa z důvodu uspořádání molekul v krystalu. Aktivní místo je blokováno symetricky sdruženou molekulou. Optimalizovaná krystalizace mutované formy enzymu však umožnila pozorování fosfátových aniontů v aktivním místě enzymu - viz. obrázek 3.4 [10].

Enzym TBN1 se vyskytuje v roztoku převážně ve formě monomerů [9]. Vykazuje velmi širokou nukleasovou aktivitu na jednovláknové a dvouvláknové RNA, DNA i na strukturované RNA. Produkty reakce jsou mono- a oligonukleotidy [10]. Přesto je strukturní pozorování interakce enzymu s různými ligandy zásadně omezeno současnými ex-



(a)



(b)

Obrázek 3.4: Struktura mutované formy N211D enzymu TBN1 nukleasy v komplexu s fosfátovým aniontem a detailní pohled na aktivní místo enzymu. (a) TBN1 nukleasa v reprezentaci sekundární struktury, kationty zinku znázorněny světle modrými koulemi, fosfátový aniont a kovalentně navázané sacharidové podjednotky v tyčinkové reprezentaci. (b) Aktivní místo TBN1 nukleasy s pozorovaným fosfátovým aniontem a modelem (3s)-3,4-Di-N-Hexanoyloxybutyl-1-fosfocholinu (3PC) v tyčinkové reprezentaci. Rezidua koordinující koordinačně kovalentní vazbu kationtů zinku s atomy uhlíku zelené, uhlíkové atomy modelovaného ligandu 3PC šedé. 3PC byl modelován podle vazby na enzym fosfolipáza C z organismu *Bacillus cereus* (PDB kód 1P6D) díky vysoké strukturní podobnosti mezi oběma enzymy [11]. Tento model vysvětluje nízkou fosfolipázovou aktivitu pozorovanou rovněž u enzymu TBN1 [9, 10].

perimentálními přístupy. Krystalizační podmínka vedoucí k uvolnění aktivního místa pro studium vazby ligandů stále není známa. Mechanismus katalýzy může být tedy vydedukován pouze na základě srovnávacích analýz podobných molekul s obdobnými katalytickými vlastnostmi a pomocí *in silico* studií.

Přes rozdílné uspořádání molekul v různých krystalových formách je vazba smyčky jedné molekuly do aktivního místa druhé molekuly pozorována ve všech známých krystalových strukturách TBN1 nukleasy. Potvrzení následující hypotézy by vyžadovalo další experimentální studii, ale z předložených výsledků lze odvodit, že tato mezimolekulární interakce představuje důležitý regulační mechanismus enzymu [10].

3.3.2 S1 nukleasa z organismu *Aspergillus sp.*

S1 nukleasa z organismu *Aspergillus sp.* je extracelulární enzym, který hraje roli požírače nukleotidů a fosfátů [12]. Předpokládá se, že se podílí na výživě *Aspergillus sp*. Díky své unikátní specificitě je využíván jako analytický nástroj ke stanovení sekundární struktury nukleových kyselin. Enzym je aktivní na jednovláknových nukleových kyselinách, ale neaktivní na dvouvláknové DNA.

Navzdory tomu, že je S1 nukleasa hojně využívána v analytických procedurách a je mnoho let nabízena na trhu, nebyla doposud krystalová struktura enzymu známa a publikována. Hypotéza mechanismu katalysy byla navržena na velmi obecných základech [13] a nebyla plně pochopena ve strukturních detailech.

Enzym byl krystalizován v nezvykle širokém rozsahu pH (4.0 - 6.5). Bylo vyřešeno sedm struktur S1 nukleasy, z čehož šest struktur jsou struktury komplexu enzymu s produkty a inhibitory. Jedna z těchto struktur je struktura mutované formy D65N S1 nukleasy (obrázek 3.5). Tato pozorování jsou umožněna díky příznivému uspořádání molekul v krystalu a díky dostupnosti aktivního místa krystalizovaného enzymu pro navázání ligandu.

Struktury komplexů S1 nukleasy s různými ligandy ukazují, že S1 nukleasa váže nukleotidy a nukleosidy několika různými způsoby. Báze nukleosidů mohou být vázány

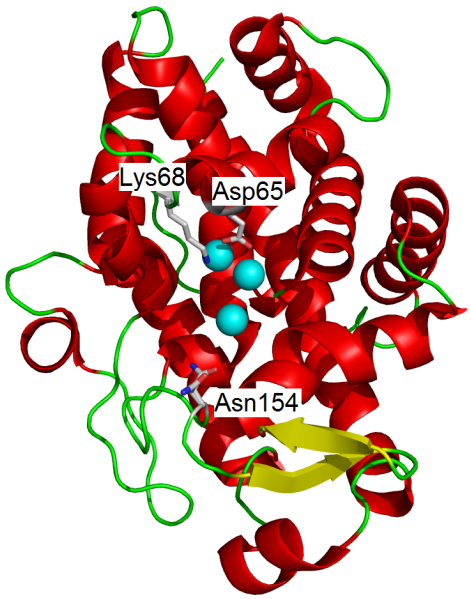
dvěma módy: mělkým a hlubokým. Vazebné místo rovněž podléhá strukturním změnám v závislosti na vazbě ligandu.

Ke strukturním změnám S1 nukleasy při vazbě ligandů dochází především u reziduí Asn154 a Phe81. Tato plasticita hraje nejspíše roli v pestrosti vazeb a aktivitě enzymu vůči různým druhům substrátů. Jedná se o první takové pozorování u nukleas S1-P1 rodiny. Ligandy mají velkou míru volnosti ve vytváření různých sítí vodíkových vazeb na enzym.

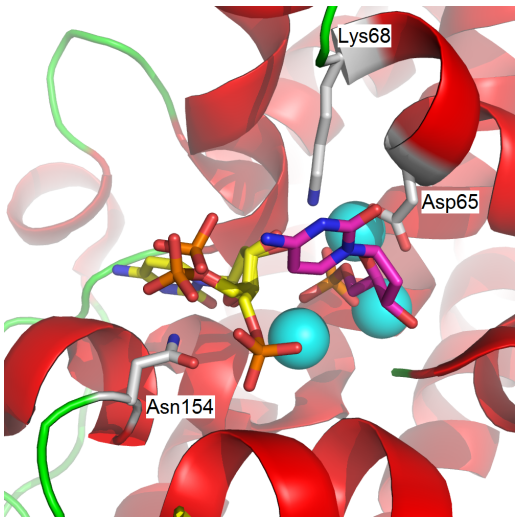
Ve strukturách komplexů bylo dále pozorováno další vazebné místo, díky němuž může enzym současně vázat dva různé ligandy. Toto vazebné místo však není sekvenčně zachováno v dalších nukleasach patřících do rodiny S1-P1, a proto nejspíš nebylo pozorováno u jiných enzymů.

Mutační studie odhalily vliv některých reziduí na aktivitu S1 nukleasy. Pro aktivitu enzymu je důležité reziduum zejména Asp65, které koordinuje zinečnaté ionty a interaguje s ligandy v aktivním místě. Reziduum Asn154 zajišťuje interakci enzymu se sacharidovými podjednotkami a Lys68 zprostředkovává vazbu enzymu s fosfátovými skupinami ligandů. Obě výše zmíněná rezidua (Asn154 a Lys68) nemají tak zásadní vliv na enzymatickou aktivitu S1 nukleasy jako Asp65.

Krystalová struktura S1 nukleasy je velmi podobná struktuře multifunkční nukleasy TBN1 z rajčete [9]. Náš současný cíl je (podobně jako u hovězího trypsinu a faktoru Xa) využít struktury S1 nukleasy, její snadné krystalizace a příznivých krystalových kontaktů, vytvořit mutované formy podle aktivního místa TBN1 nukleasy, provést strukturní analýzu a objasnit tak specificitu TBN1 nukleasy na základě tohoto modelového enzymu. Bez těchto znalostí nelze zodpovědět otázku proč S1 nukleasa není schopna odbourávat dvouvláknové nukleové kyseliny, zatímco TBN1 ano. Detaily těchto interakcí rovněž pomohou navrhnout další mutované formy TBN1 nukleasy s vylepšenými vlastnostmi pro nádorové terapie, neboť TBN1 nukeasa byla při *in vivo* testech aktivní proti lidským nádorům transplantovaným do myší.



(a)



(b)

Obrázek 3.5: Struktura enzymu S1 nukleasy a pohled na vazbu ligandů do aktivního místa enzymu. (a) Většina sekundární struktury S1 nukleasy je tvořena α-šroubovicemi. Aktivní místo je doplněno třemi ionty zinku. Postranní řetězce významných aminokyselin aktivního místa jsou znázorněny tyčinkovou reprezentací. (b) Detailní pohled na aktivní místo enzymu s navázanými ligandy v tyčinkové reprezentaci (barevně odlišené atomy uhlíku).

3.4 Receptor 2 lidského interferonu γ

Lidský interferon γ (hIFNγ) je důležitá molekula signalizační dráhy ovlivňující aktivitu vrozené i adaptabilní části imunitního systému proti patogenům a nádorovému bujení [4]. Signalizace je iniciována tvorbou ternárního komplexu tvořeného homodimerem hIFNγ, dvěma molekulami receptoru 1 liského interferonu γ (hIFNγ-R1) a molekulou, případně molekulami, receptoru 2 liského interferonu γ (hIFNγ-R2). Přesný poměr mezi molekulami jednotlivých podsložek (obzvlášvě hIFNγ-R2) není znám, neboť struktura celého komplexu ještě nebyla vyřešena. Publikován byl pouze komplex hIFNγ:hIFNγ-R1 (PDB kód 1FG9) [14]. Strukturní charakterizace hIFNγ-R2 chyběla úplně.

hIFNγ-R2 byl rekombinantně produkován pomocí hmyzích Schneider S2 buněk a následně biochemicky charakterizován. Plně glykosylovaná forma receptoru krystalizovala s prostorovou grupou $P2_12_12_1$, difrakční kvalita krystalů dosahovala rozlišení 2.8 Å. Řešení fázového problému bylo velmi komplikované, neboť nejvyšší podobnost již známé

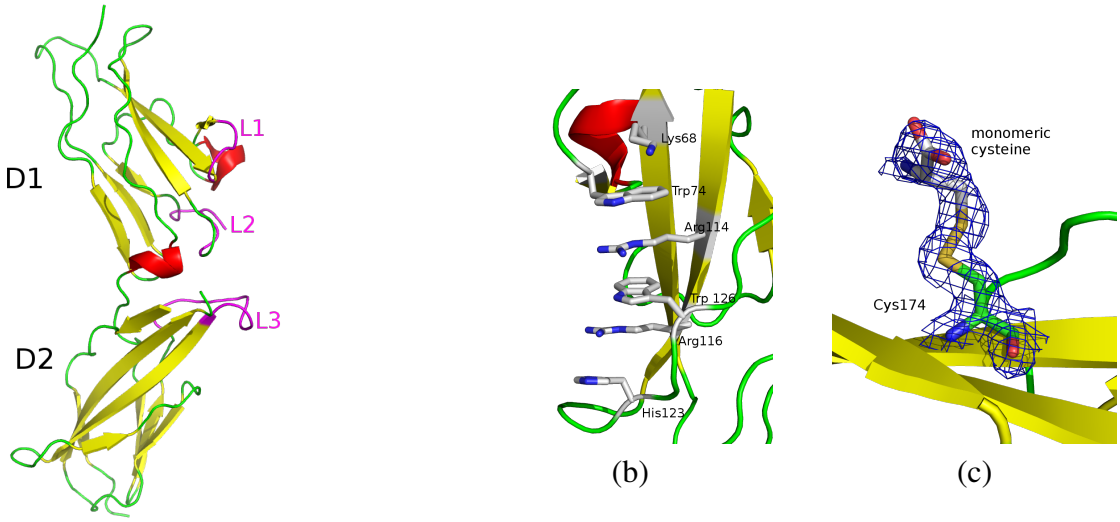
struktury na úrovni aminokyselinového složení byla méně než 25 %. Proto byl receptor částečně deglykosylován a tato forma byla znovu použita při hledání nové krystalizační podmínky. Nově získané krystaly deglykosylované formy se vyznačovaly významně lepší difrakční kvalitou - rozlišení 1.8 Å [4].

Trochu delší popis si nyní zaslouží samotné řešení struktury. Z pouhého zpracování difrakčních dat je v některých případech nemožné rozhodnout mezi správnou prostorovou grupou a všemi operátory symetrie uvnitř krystalu. Ze zpracování dat bylo zřejmé, že krystal náleží do hexagonální krystalové soustavy. Bylo však těžké rozhodnout o šroubových osách. Vyhasínání reflexí naznačovalo, že prostorová grupa krystalu je $P6_122$ nebo $P6_522$. Řešení fázového problému metodou molekulárního nahrazení pomocí struktury receptoru 1 lidského IFNγ (hIFNγ-R1) naznačovalo řešení $P6_522$. Avšak řešení struktury se zastavilo na vysokých R faktorech ($R_{\mathrm{work}}/R_{\mathrm{free}} \sim 0.41/0.45$). Přestože počáteční řešení v prostorové grupě $P6_122$ se zdálo být horší (podařilo se umístit pouze doménu D2 - viz níže) a R faktory modelu po molekulárním nahrazení dosahovaly vyšších hodnot ($R_{\mathrm{work}}/R_{\mathrm{free}} \sim 0.53/0.55$), bylo překvapivě toto řešení správné a drobná úprava modelu vedla k výraznému vylepšení fází a umožnění automatické stavby a upřesnění modelu pomocí dalších programů.

Struktura hIFNγ-R2 se skládá ze dvou domén fibronektinového typu - doména D1 (rezidua Ser28 – Thr139) a doména D2 (rezidua Val140 – Asp240). Struktura vykazuje několik následujících zajímavostí:

- Glykosylační místa Asn110 a Asn137 zaštiťuje hydrofobní postranní řetězec Trp131 před solventem,
- motiv sestavených π-interakcí sekvenčně vzdálených reziduí KWRWRH,
- kovalentní vázání monomerního cysteinu disulfidovou vazbou na reziduum Cys174.

Monomerní cystein vázaný na reziduum Cys174 pravděpodobně pochází z produkčního média, jehož byl složkou. Jeho pozorování nebylo předem očekáváno a přes jasnou lo-



(a) (b) (c)

Obrázek 3.6: Struktura a pohled na detaily receptoru 2 lidského interferonu γ. (a) hIFNγ-R2 v reprezentaci sekundární struktury se skládá z N-terminální domény C1 a C-terminální domény D2. Smyčky L1 – L3 jsou zabarveny fialově a označeny šipkami. (b) Detailní pohled na π motiv na povrchu domény D1. Rezidua podílející se na tvorbě motivu zobrazena tyčinkami, uhlíkové atomy šedou barvou. (c) 2mFo-DFc mapa elektronové hustoty (hladina 1σ, modrá síť) v okolí rezidua Cys174 a kovalentně vázaného monomerního cysteinu. Tato modifikace byla posléze potvrzena metodou hmotnostní spektrometrie.

kalizaci v mapě elektronové hustoty byla identita ligandu ověřena metodou hmotnostní spektroskopie (viz. obrázek 3.6) [4].

Navzdory tisícům provedených experimentů se dodnes nepodařilo vypěstovat krystaly ternárního komplexu [4]. Přes nedosažitelnost experimentálního pozorování lze alespoň vydedukovat oblasti hIFNγ-R2, které se na tvorbě tohoto komplexu podílí. Jedná se o velmi variabilní oblasti tří smyček L1 – L3 (smyčky L1 a L2 z domény D1 a smyčka L3 z domény D2). Lze tedy alespoň určitým způsobem zjednodušit výpočetní předpovědi výsledného modelu interakcí ve výsledném ternárním komplexu.

Kapitola 4

Shrnutí

Krystalografické studie komplexů makromolekul a jejich ligandů umožnují analýzu mechanismů aktivity makromolekul na atomární úrovni. Tyto studie jsou důležité pro celkové pochopení molekulárních základů života a jejich výsledky lze rovněž využít při vývoji nových biotechnologických aplikací či při vylepšování již známých technologií, které mohou nacházet uplatnění v široké škále průmyslových odvětví.

V této habilitační práci jsou prezentovány výsledky několika nezávislých projektů:

1. Nová krystalová forma glutaminyl cyklasy z organismu *Drosophila melanogaster* umožňuje rychlou a snadnou krystalografickou analýzu vazby a interakcí nových inhibitorů do aktivního místa enzymu [2].

   Příspěvek habilitanta: Veškerá experimentální práce je příspěvkem habilitanta, včetně návrhu samotného experimentu. Dále se habilitant nejvýznamněji podílel na tvorbě textu pro publikaci.

2. Analýza mnoha struktur komplexů čtyř mutovaných forem hovězího trypsinu s devíti různými inhibitory objasnila vliv reziduí z druhé vrstvy okolí aktivního místa na schopnost inhibitorů inhibovat danou mutovanou formu enzymu [3]. V souladu s předešlými výsledky byly pozorovány dvě konformace reziduí z okolí aktivního místa enzymu.

Příspěvek habilitanta: Habilitant byl odpovědný za konečné zpracování všech difrakčních experimentů a uložení výsledků do databáze PDB. Podílel se také na širší analýze v souvislosti s dalšími výsledky, příspěvek ke tvorbě publikovaného textu je však málo významný.

3. Poprvé byl pozorován ligand navázaný do aktivního místa multifunkční nukleasy TBN1 z organismu *Solanum lycopersicum* [10]. Pozorování objemnějších ligandů (zejména inhibitorů) není v této krystalové formě možné díky síti mezimolekulárních interakcí uvnitř krystalu.

   Příspěvek habilitanta: Habilitant se podílel na provedení a organizaci experimentální práce. Příspěvek k analýze výsledků a tvorbě textu byl málo významný.

4. Krystalové struktury S1 nukleasy z organismu *Aspergillus sp.* v komplexu s různými ligandy odhalily způsob vazby nukleotidů a významně tak přispěly k celkovému pochopení mechanismu katalytické aktivity enzymu [12].

   Příspěvek habilitanta: Habilitant se podílel na provedení a organizaci experimentální práce. Příspěvek k analýze výsledků a tvorbě textu byl málo významný.

5. Byla vyřešena krystalová struktura receptoru 2 lidského interferonu γ s rozlišením 1.8 Å. Bylo pozorováno kovalentní navázání monomerního cysteinu z produkčního média formou disulfidového můstku na volné reziduum cystein. Na základě sekvenční analýzy a analýzy komplexů s podobnou strukturou a funkcí lze předpovědět vazebné místo pro lidský interferon γ. Pokusy o krystalizaci a řešení struktury funkčního ternárního komplexu jsou nadále prováděny.

   Příspěvek habilitanta: Habilitant plně zajišťoval difrakční experiment, jeho vyhodnocení a uložení struktury do PDB databáze. Významně se podílel na tvorbě textu k publikaci.

Tabulka 4.1: Seznam krystalových struktur.

Protein	PDB kód	DOI
*Dm*QC	4FWU	10.2210/pdb4fwu/pdb
trypsin	3V0X	10.2210/pdb3v0x/pdb
	3PLB	10.2210/pdb3plb/pdb
	3PLK	10.2210/pdb3plk/pdb
	3PLP	10.2210/pdb3plp/pdb
	3PM3	10.2210/pdb3mp3/pdb
	3PMJ	10.2210/pdb3pmj/pdb
	3PWB	10.2210/pdb3pwb/pdb
	3PWC	10.2210/pdb3pwc/pdb
	3Q00	10.2210/pdb3q00/pdb
	3PYH	10.2210/pdb3pyh/pdb
	3UNQ	10.2210/pdb3unq/pdb
	3UNS	10.2210/pdb3uns/pdb
	3UOP	10.2210/pdb3uop/pdb
	3UPE	10.2210/pdb3upe/pdb
	3UUZ	10.2210/pdb3uuz/pdb
	3UQV	10.2210/pdb3uqv/pdb
	3UQO	10.2210/pdb3uqo/pdb
	3UY9	10.2210/pdb3uy9/pdb
	3YWI	10.2210/pdb3ywi/pdb
	3V12	10.2210/pdb3v12/pdb
	3V13	10.2210/pdb3v13/pdb
TBN1	4JDG	10.2210/pdb4jdg/pdb
S1	5FBA	10.2210/pdb5fba/pdb
	5FB9	10.2210/pdb5fb9/pdb
	5FBB	10.2210/pdb5fbb/pdb
	5FBC	10.2210/pdb5fbc/pdb
	5FBD	10.2210/pdb5fbd/pdb
	5FBF	10.2210/pdb5fbf/pdb
	5FBG	10.2210/pdb5fbg/pdb
hIFNγ-R2	5EH1	10.2210/pdb5eh1/pdb

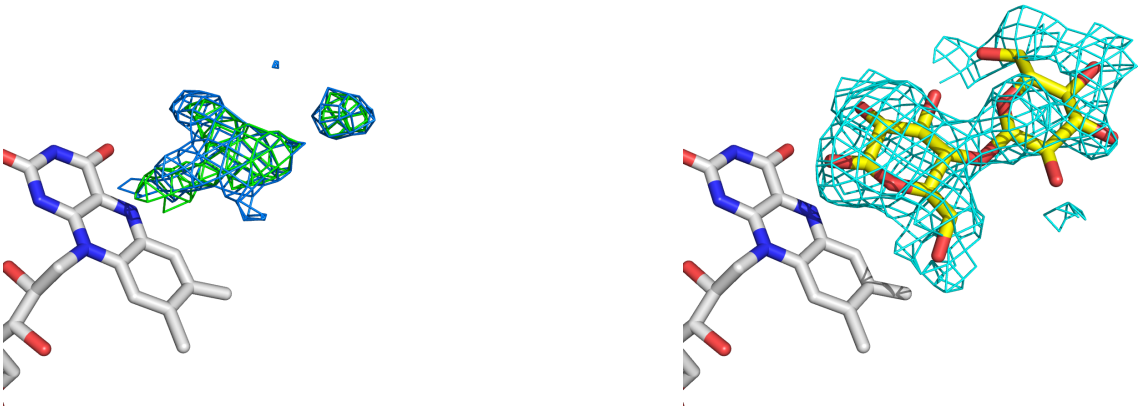
Kapitola 5

Perspektiva rozvoje oboru a metod

V blízké budoucnosti bude monokrystalová rentgenová strukturní analýza nadále hrát klíčovou roli ve strukturní biologii, především ve strukturní enzymologii. Žádná jiná metoda strukturní biologie v současnosti nedosahuje takové přesnosti lokalizace jednotlivých atomů. Neustále navíc probíhá vývoj experimentálních aparatur a experimentálních metod. To umožňuje přesnější a citlivější pozorování ligandů cílových makromolekul. Spolu s rozvojem krystalografických metod se rozvíjí metody přípravy vzorků a jejich krystalizace, díky nimž lze studovat stále širší množství makromolekul a jejich interakcí.

Rozvoj metodiky nezasahuje pouze experimentální metody. Vyvíjeny jsou rovněž způsoby zpracování difrakčních dat a jejich vyhodnocování [15, 16]. V současnosti je velmi zajímavá otázka využití dříve přehlížených dat s nízkou hodnotou signálu oproti šumu [17]. Dnešní technologie uchovávání původních difrakčních dat a jejich volné poskytování dalším uživatelům nabízí nové příležitosti k opětovnému vyhodnocení experimentu [18]. V neposlední řadě je očekáván další rozvoj v automatizaci zpracování dat a především v kontrole kvality výsledků a finálních modelů makromolekul.

V souvislosti s ligandy jsou rozvíjeny nové výpočetní metody umožňující pozorování slabě vázaných či částečně okupovaných nízkomolekulárních ligandů - například tzv. *polder* mapa [19]. Příklad rozdílu mezi standardním pozorováním vazby ligandu a *polder* mapou elektronové hustoty je ukázán na obrázku 5.1 (*zatím nepublikováno*). Vyhodnocení mapy elektronové hustoty vypočítané pomocí standardní metody by vedlo pozoro-



(a) (b)

Obrázek 5.1: Rozdíl mezi standardně užívanou 2mFo-DFc (modrá) a mFo-DFc (zelená) mapou elektronové hustoty (a) a *polder* (světle modrá) mapou (b). Pozorování bylo provedeno na krystalu enzymu s kofaktorem flavinadenindinukleotid (FAD) v aktivním místě nasáknutém v roztoku ligandu. Ze standardní mapy nevyplývá přítomnost ligandu navázaného v blízkosti FAD, zatímco *polder* mapa pokrývá takřka všechny atomy konečného modelu. *Polder* mapa ukazuje zdokonalení metodiky pozorování slabě vázaných ligandů. FAD (atomy uhlíku šedou barvou) a ligand (atomy uhlíku žlutou barvou) - model po několika cyklech upřesňování - jsou znázorněny v tyčinkové reprezentaci.

vatele k závěru, že ligand není navázán. Výpočet *polder* mapy však zjevně ukazuje na slabou vazbu částečně okupovaného ligandu. Přítomnost ligandu a správnost jeho určení byla později prokázána stabilním průběhem strukturního upřesňování a následnou kontrolou výsledného modelu. Existují i další metody zvýrazňující slabý experimentálně pozorovaný signál (*feature-enhanced map*, *composit omit map* atd.).

Pozorování slabě vázaných ligandů nadále představuje výzvu pro makromolekulární krystalografii. V případě získání prostředků na rozvoj této metodiky navrhuji sestavení výpočtu tzv. *kompozitní polder* mapy, která by rozšířila posílení slabého signálu navázaných ligandů na celý povrch a okolí molekuly, nejen na uživatelem zadanou oblast, jak je to dnes omezeně dostupné v současných výpočetních nástrojích.

Literatura

[1] B. Koch, P. Kolenko, M. Buchholz, D. Carrillo, C. Parthier, M. Wermann, J.-U. Rahfeld, G. Reuter, S. Schilling, M. Stubbs, and H.-U. Demuth, "Crystal structures of glutaminyl cyclases (qcs) from *Drosophilla melanogaster* reveal active site conservation between insect and mammalian qcs," *Biochemistry*, vol. 51, pp. 7383–7392, 2012.

[2] P. Kolenko, B. Koch, J.-U. Rahfeld, S. Schilling, H.-U. Demuth, and M. Stubbs, "Structure of glutaminyl cyclase from *Drosophila melanogaster* in space group $i4$," *Acta Cryst. F*, vol. 69, pp. 358–361, 2013.

[3] A. Tziridis, D. Rauh, P. Neumann, P. Kolenko, A. Menzel, U. Bräuer, C. Ursel, P. steinmetzer, J. Stürzebecher, A. Schweinitz, T. Steinmetzer, and M. Stubbs, "Correlating structure and ligand affinity in drug discovery: a cautionary tale involving second shell residues," *Biol. Chem.*, vol. 395, pp. 891–903, 2014.

[4] P. Mikulecký, J. Zahradník, P. Kolenko, J. Černý, T. Charnavets, L. Kolářová, I. Nečasová, P. Pham, and B. Schneider, "Crystal structure of human interferon-γ receptor 2 reveals the structural basis for receptor specificity," *Acta Cryst. D*, vol. 72, pp. 1017–1025, 2016.

[5] K.-F. Huang, Y.-L. Liu, W.-J. Cheng, T.-P. Ko, and A. Wang, "Crystal structures of human glutaminyl cyclase, an enzyme responsible for protein n-terminal pyroglutamate formation," *Proc. Natl. Acad. Sci.*, vol. 102, pp. 13117–13122, 2005.

[6] D. Ruiz-Carrillo, B. Koch, C. Parthier, M. Wermann, T. Dambe, M. Buchholz, H.-H. Ludwig, U. Heiser, J.-U. Rahfeld, M. Stubbs, S. Schilling, and H.-U. Demuth, "Structures of glycosylated mammalian glutaminyl cyclases reveal conformational variability near the active center," *Biochemistry*, vol. 50, pp. 6280–6288, 2011.

[7] D. Rauh, G. Klebe, J. Stürzebecher, and M. T. Stubbs, "Zz made ez: Influence of inhibitor configuration on enzyme selectivity," *Journal of Molecular Biology*, vol. 330, p. 761–770, Jul 2003.

[8] D. Rauh, G. Klebe, and M. T. Stubbs, "Understanding protein–ligand interactions: The price of protein flexibility," *Journal of Molecular Biology*, vol. 335, p. 1325–1341, Jan 2004.

[9] T. Koval', P. Lipovová, T. Podzimek, J. Matousek, J. Dusková, T. Skálová, A. Stepánková, J. Hasek, and J. Dohnálek, "Plant multifunctional nuclease tbn1 with unexpected phospholipase activity: structural study and reaction-mechanism analysis," *Acta Cryst. D*, vol. 69, pp. 213–226, 2013.

[10] J. Stránský, T. Koval', T. Podzimek, A. Týcová, P. Lipovová, J. Matoušek, P. Kolenko, K. Fejfarová, J. Dušková, T. Skálová, J. Hašek, and J. Dohnálek, "Phosphate binding in the active centre of tomato multifunctional nuclease tbn1 and analysis of superhelix formation by the enzyme," *Acta Cryst. F*, vol. 71, pp. 1408–1415, 2015.

[11] N. Antikainen, A. Monzingo, C. Franklin, J. Robertus, and S. Martin, "Using x-ray crystallography of the asp55asn mutant of the phosphatidylcholine-preferring phospholipase c from bacillus cereus to support the mechanistic role of asp55 as the general base," *Arch. Biochem. Biophys.*, vol. 417, pp. 81–86, 2003.

[12] T. Kovaľ, L. Oestergaard, J. Lehmbeck, A. Noergaard, P. Lipovová, J. Dušková, T. Skálová, M. Trundová, P. Kolenko, K. Fejfarová, J. Stránský, L. Švecová, J. Hašek, and J. Dohnálek, "Structural and catalytic properties of s1 nuclease from

Aspergillus oryzae responsible for substrate recognition, cleavage, non-specificity , and inhibition," *PLos ONE*, vol. 11, p. e0168832, 2016.

[13] N. Desai and V. Shankar, "Single-strand-specific nucleases," *FEMS Microbiol. Rev.*, vol. 26, pp. 457–491, 2003.

[14] D. Thiel, M. le Du, R. Walter, A. D'Arcy, C. Chene, M. Fountoulakis, G. Garotta, F. Winkler, and S. Ealick, "Observation of an unexpected third receptor molecule in the crystal structure of human interferon-gamma receptor complex," *Structure Fold.Des.*, vol. 8, pp. 927–936, 2000.

[15] W. Kabsch, "Xds," *Acta Cryst. D*, vol. 66, pp. 125–132, 2010.

[16] N. Sauter, J. Hattne, R. Grosse-Kunstleve, and N. Echols, "New python-based methods for data processing," *Acta Cryst. D*, vol. 69, pp. 1274–1282, 2013.

[17] P. Karplus and K. Diedrichs, "Linking crystallographic model and data quality," *Science*, vol. 336, pp. 1030–1033, 2012.

[18] Editorial, "Data sharing comes to structural biology," *Nature Methods*, vol. 13, p. 381, 2016.

[19] P. A. D. Liebschner, N. Moriarty, B. Poon, O. Sobolev, T. Terwilliger, and P. Adams, "Polder maps: improving omit maps by excluding bulk solvent," *Acta Cryst. D*, vol. 73, pp. 148–157, 2017.

Přílohy

- *Curriculum vitae* - Ing. Petr Kolenko, PhD., 2 strany.
- Publikace v impaktovaných časopisech v letech 2013 – 2017:

1. Kolenko P, Koch B, Rahfeld J-U, Schilling S, Demuth H-U, Stubbs MT. (2013). Crystal structure of the glutaminyl cyclase from Drosophila melanogaster in space group I4. *Acta Cryst.*, *F***69**, 358-361.
2. Tziridis A, Rauh D, Neumann P, Kolenko P, Menzel A, Bräuer U, Ursel C, Steinmetzer P, Stürzebecher J, Schweinitz A, Steinmetzer T, Stubbs MT. (2014). Correlating structure and ligand affinity in drug discovery: A cautionary tale involving second shell residues. *Biological chemistry*, **395**, 891-903.
3. Stránský J, Kovaľ T, Podzimek T, Týcová A, Lipovová P, Matoušek J, Kolenko P, Fejfarová K, Dušková J, Skálová T, Hašek J, Dohnálek J. (2015). Phosphate binding in the active centre of tomato multifunctional nuclease TBN1 and analysis of superhelix formation by the enzyme. *Acta Cryst.*, *F***71**, 1408-1415.
4. Kovaľ T, Oestergaard LH, Lehmbeck J, Noergaard A, Lipovová P, Dušková J, Skálová T, Trundová M, Kolenko P, Fejfarová K, Stránský J, Švecová L, Hašek J, Dohnálek J. (2016). Structural and Catalytic Properties of S1 Nuclease from Aspergillus oryzae Responsible for Substrate Recognition, Cleavage, Non-Specificity , and Inhibition. *PLoS ONE*, **11** (12), e0168832.
5. Mikulecký P, Zahradník J, Kolenko P, Černý J, Charnavets T, Kolářová L, Nečasová I., Pham P.N., Schneider B. (2016). Crystal structure of human

interferon-γ receptor 2 reveals the structural basis for receptor specificity. *Acta Cryst.*, *D***72**, 1017-1025.

Ing. Petr Kolenko, Ph.D.

Curriculum Vitae

OSOBNÍ ÚDAJE

Narozen	18. července 1982, Hodonín
Adresa	Na Lukách 1620/23, 268 01 Hořovice
Telefon	+420 224 358 606
Mobil	+420 723 489 379
E-Mail	`petr.kolenko@fjfi.cvut.cz`
WWW	`people.fjfi.cvut.cz/kolenpe1`
ORCID ID	0000-0002-4619-9276

VZDĚLÁNÍ

Doktorské: obor Fyzikální inženýrství 2006-2009

Fakulta jaderná a fyzikálně inženýrská ČVUT, Praha

Disertační práce: *Studie makromolekulárních komplexů v imunitním systému - vztah struktury a funkce.*

Magisterské: obor Fyzikální inženýrství 2001-2006

Fakulta jaderná a fyzikálně inženýrská ČVUT, Praha

Diplomová práce: *Stanovení struktury Fc fragmentu monoklonální protilátky IgG2b pomocí difrakčních metod.*

PRACOVNÍ ZKUŠENOSTI

2015-nyní

Fakulta jaderná a fyzikálně inženýrská ČVUT, částečný úvazek

Katedra inženýrství pevných látek - výuka (dva předměty) a vedení studentských závěrečných prací v bakalářském i magisterském oboru. Kontakt: Ladislav Kalvoda (doporučující dopis přiložen).

2015-nyní

Biotechnologický ústav AV ČR, v.v.i., částečný úvazek

Strukturní a biochemické studie biotechnologických enzymů, návrhy interakcí rozpoznávajících molekul, synchrotronová měření. Kontakt: Jan Dohnálek.

2013

Aston University, Birmingham, Anglie

Stáž zaměřená na přípravu, modifikaci a velkoobjemovou produkci biotechnologických proteinů pro průmyslové aplikace. Kontakt: Prof. Roslyn M. Bill.

2010-2012

Institute of Biochemistry and Biotechnology, MLU Halle-Saale, Německo

Biochemické a strukturní analýzy proteinů - převážně enzymů - pro účely cíleného návrhu léčiv - spolupráce s několika farmaceutickými firmami. Částečně podíl na výukových aktivitách.

Kontakt: Prof. Milton T. Stubbs (doporučující dopis přiložen).

2006-2015

Ústav makromolekulární chemie AV ČR, v.v.i.

Řešení struktur technologicky důležitých proteinů pomocí makromolekulární krystalografie.
Kontakt: Jindřich Hašek (současně na BTÚ AV ČR, v.v.i.).

ZNALOSTI

Odborné: Makromolekulární strukturní analýza, molekulární biologie, bakteriální a kvasinková produkce proteinů, široké spektrum biofyzikálních charakterizací proteinů.
Jazyky: Angličtina, Němčina

VÝBĚR PUBLIKACÍ (CELKEM 17)

- P. Mikulecký, J. Zahradník, P. Kolenko, *et al.* (2016). Crystal structure of human interferon-γ receptor 2 reveals the structural basis for receptor specificity. *Acta Cryst.* D**72**, 1017-1025. Impakt faktor: 2,5; citováno 1x.
- B. Koch, P. Kolenko, *et al.* (2012). Crystal structures of glutaminyl cyclases (QCs) from *Drosophila melanogaster* reveal active site conservation between insect and mammalian QCs. *Biochemistry* **51**, 7383-7392. Impakt faktor: 3,0; citováno 4x.
- P. Kolenko *et al.* (2009). New insights into intra- and intermolecular interactions of immunoglobulins: Crystal structure of mouse IgG2b-Fc at 2.1-Å resolution. *Immunology* **126**, 378-385. Impakt faktor: 3,7; citováno 15x.
- T. Skálová, ..., P. Kolenko, *et al.* (2009). The Structure of the Small Laccase from Streptomyces coelicolor Reveals a Link between Laccases and Nitrite Reductases. *J. Mol. Biol.* **385**, 1165-1178. Impakt faktor: 4,3; citováno 69x.

DALŠÍ TVŮRČÍ ČINNOST

- Více než 70 příspěvků na konferencích.
- Celkem 52 struktur v PDB databázi.

DALŠÍ

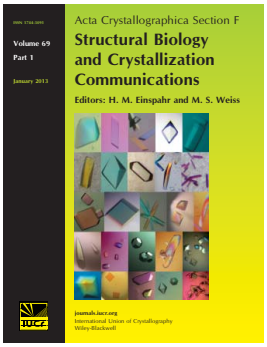
- Zkušenosti z pořádání několika kursů pro doktorandy v ČR i zahraničí.
- Vedoucí nově založené Laboratoře strukturní biologie na Katedře inženýrství pevných látek FJFI ČVUT.

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091
Editors: **H. M. Einspahr and M. S. Weiss**

Structure of glutaminyl cyclase from *Drosophila melanogaster* in space group *I*4

Petr Kolenko, Birgit Koch, Jens-Ulrich Rahfeld, Stephan Schilling, Hans-Ulrich Demuth and Milton T. Stubbs

Acta Cryst. (2013). F**69**, 358–361



Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of glutaminyl cyclase from *Drosophila melanogaster* in space group *I*4

Petr Kolenko,[a,b]* Birgit Koch,[c] Jens-Ulrich Rahfeld,[c] Stephan Schilling,[c] Hans-Ulrich Demuth[c] and Milton T. Stubbs[a]

[a]Department of Physical Biochemistry, Institute of Biochemistry and Biotechnology, MLU, Kurt-Mothes-Strasse 3, 06120 Halle, Germany, [b]Department of Structural Analysis of Biomacromolecules, Institute of Macromolecular Chemistry AS CR, Heyrovskeho nam. 2/1888, 162 06 Prague 6, Czech Republic, and [c]Probiodrug AG, Weinbergweg 22, 06120 Halle, Germany

Correspondence e-mail: petr.kolenko@gmail.com

Received 2 January 2013
Accepted 26 February 2013

PDB Reference: glutaminyl cyclase, 4fwu

The structure of ligand-free glutaminyl cyclase (QC) from *Drosophila melanogaster* (*Dm*QC) has been determined in a novel crystal form. The protein crystallized in space group *I*4, with unit-cell parameters $a = b = 122.3$, $c = 72.7$ Å. The crystal diffracted to a resolution of 2 Å at the home source. The structure was solved by molecular replacement and was refined to an *R* factor of 0.169. *Dm*QC exhibits a typical α/β-hydrolase fold. The electron density of three monosaccharides could be localized. The accessibility of the active site will facilitate structural studies of novel inhibitor-binding modes.

1. Introduction

Glutaminyl cyclases (QCs; EC 2.3.2.5) are enzymes that catalyze the conversion of N-terminal glutamine or glutamate into pyroglutamate and have been identified in many forms of life (>100 entries in the UniProt database; The UniProt Consortium, 2012). Whereas plant and bacterial QCs adopt a five-bladed β-propeller fold (Wintjens *et al.*, 2006; Ruiz-Carrillo *et al.*, 2010), mammalian (human and mouse; hQC and mQC, respectively) and insect QCs are zinc-dependent enzymes (Schilling *et al.*, 2003) with an α/β-hydrolase topology (Huang *et al.*, 2005, 2008, 2011; Ruiz-Carrillo *et al.*, 2011; Koch *et al.*, 2012). The main core of the tertiary structure is formed by a central six-stranded β-sheet surrounded by several α-helices (Huang *et al.*, 2005). The zinc cation at the active site is coordinated by the side chains of aspartate, histidine and glutamate residues.

We have recently published the crystal structures of wild-type *Drosophila melanogaster* QC (*Dm*QC) and a cysteine-lacking variant (Koch *et al.*, 2012) in complex with the highly potent competitive inhibitor PBD150 (Buchholz *et al.*, 2006, 2009), which are the first crystal structures of a secreted invertebrate QC. All known crystal structures of mammalian glutaminyl cyclases in complex with PBD150 show the influence of crystal contacts on the binding mode of the inhibitor (Ruiz-Carrillo *et al.*, 2011). In our previous study, inhibited *Dm*QC crystallized in space group $P2_1$, with the active site exposed to the solvent, and in space group $P6_5$, in which the crystal contacts significantly affected the binding mode of PBD150 (Koch *et al.*, 2012). In this study, we searched for a new crystal form for the purpose of inhibitor-soaking experiments.

2. Materials and methods

2.1. Macromolecule production

The protein (UniProtKB code Q9VRQ9; residues 29–340) was prepared according to the previously published protocol (Schilling *et al.*, 2007). Briefly, the protein was heterologously expressed in the yeast *Pichia pastoris*, purified by cation-exchange chromatography (Streamline SP XL column) and hydrophobic interaction chromatography (Butyl Sepharose FF column) and concentrated for crystallization (Table 1).

2.2. Crystallization

We performed extensive screening for novel crystallization conditions (roughly 900 conditions). Prism-like protein crystals of a novel crystal form (Fig. 1) appeared within two weeks. Further details

Table 1
Details of protein production.

DNA source	*D. melanogaster*
Forward primer	ATATAT**CTCGAG**AAAAGAAACATTGGATCCCAGTGGCGC
Reverse primer	ATATAT**GCGGCCGC**CTATGTTCGGTAAAAACGCAAATT
Expression vector	pPICαB
Expression host, strain	*P. pastoris*, X33 (AOX1, AOX2)

Table 2
Crystallization details.

Method	Hanging-drop vapour diffusion
Plate type	24-well NeXtal crystallization plate (Qiagen)
Temperature (K)	293
Protein concentration (g l^{-1})	15
Buffer composition of protein solution	0.1 *M* NaCl, 25 m*M* bis-tris pH 6.8
Composition of reservoir solution	1.6 *M* ammonium sulfate, 10%(*v*/*v*) 1,4-dioxane, 0.1 *M* HEPES pH 7.5
Volume of drop (µl)	2
Protein:reservoir ratio in drop	1:1
Volume of reservoir (µl)	1000

of the crystallization conditions are summarized in Table 2. Prior to the diffraction experiments, the crystal was flash-cooled in a nitrogen stream at 100 K with 20% ethylene glycol as a cryoprotectant.

2.3. Data collection and processing

The diffraction data set was collected using a Rigaku RU-300 rotating-anode generator equipped with focusing mirrors (MSC, USA) and a Saturn 940+ CCD detector. Integration of the images was performed using the *XDS* program package (Kabsch, 2010). Scaling was performed using *SCALA* from the *CCP*4 program package (Winn *et al.*, 2011). Statistics are given in Table 3.

2.4. Structure solution and refinement

The structure was solved by molecular replacement with the program *MOLREP* (Vagin & Teplyakov, 2010) using the *Dm*QC monomer in space group $P2_1$ (PDB entry 4f9u; Koch *et al.*, 2012) excluding water molecules and inhibitor as a search model. The process resulted in the localization of one monomer in the asymmetric unit (contrast = 30.6; score = 0.71; R_{work}/R_{free} = 0.225/0.264). Refinement was carried out with *REFMAC*5 (Murshudov *et al.*, 2011) and manual corrections of the model were performed using *Coot* (Emsley & Cowtan, 2004). Initially, water molecules were built with the help of *ARP*/*wARP* (Langer *et al.*, 2008). Refinement was continuously monitored using the R_{free} statistic (which was calculated using 5% of the reflections). The last refinement cycle was performed using all measured reflections. Statistics are given in Table 4.

Structure validation was carried out with *MolProbity* (Lovell *et al.*, 2003; Chen *et al.*, 2010) and *SFCHECK* (Vaguine *et al.*, 1999). The coordinates and structure factors have been deposited in the PDB (Berman *et al.*, 2000) with accession code 4fwu.

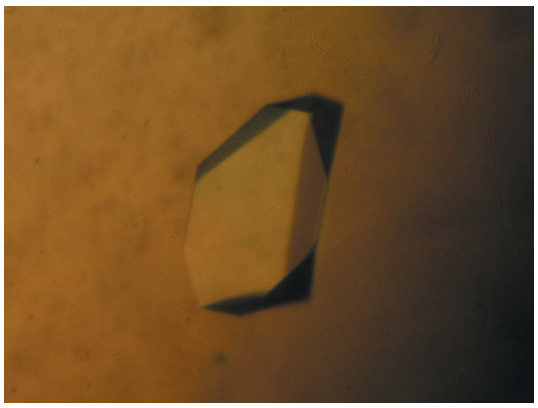


Figure 1
Crystal of *Dm*QC (~400 µm in one direction) in space group *I*4 after four weeks of growth. The photograph was taken shortly before cryoprotection and the diffraction experiment.

Table 3
Data-collection and processing statistics.

Values in parentheses are for the outer shell.

Wavelength (Å)	1.541
Temperature (K)	100
Crystal-to-detector distance (mm)	50
Rotation range per image (°)	1
Total rotation range (°)	180
Exposure time per image (s)	15
Space group	*I*4
Unit-cell parameters (Å, °)	$a = b = 122.3$, $c = 72.7$, $\alpha = \beta = \gamma = 90$
Mosaicity (°)	0.2
Resolution range (Å)	45.0–2.0 (2.11–2.00)
Total No. of reflections	125968 (11796)
No. of unique reflections	35645 (4846)
Completeness (%)	98 (93)
Multiplicity	3.5 (2.4)
$\langle I/\sigma(I)\rangle$	11.9 (2.0)
R_{merge}	0.075 (0.460)
Overall *B* factor from Wilson plot (Å^2)	24

Table 4
Structure-refinement statistics.

Values in parentheses are for the outer shell (2.05–2.00 Å).

No. of reflections, working set	32085
No. of reflections, test set	1778 [5%]
Final R_{work}	0.166 (0.276)
Final R_{free}	0.208 (0.318)
Final R_{all}	0.169
R.m.s. deviations	
Bonds (Å)	0.015
Angles (°)	1.448
No. of non-H atoms	
Protein	2427
Zinc ion	1
Saccharides	39
Ethylene glycol	12
Solvent	381
Total	2860
Average *B* factors (Å^2)	
Protein	23
Zinc ion	16
Saccharides	45
Ethylene glycol	31
Solvent	38
Total	25
Ramachandran plot	
Residues in favoured region	285
Residues in additionally allowed regions	10
Outliers	0

3. Results and discussion

We have found a new crystallization condition for *Dm*QC, leading to crystals that belonged to space group *I*4 and diffracted to 2.0 Å resolution. The structure could be solved by molecular replacement, and the enzyme adopts the α/β-hydrolase fold common to all known mammalian and insect glutaminyl cyclases (hQC, hisoQC, mQC, *Dm*QC and *Dm*isoQC; Koch *et al.*, 2012). The asymmetric unit contains one protein chain (Gln33–Phe337); localization of amino acids Leu195–Gln200 was not possible owing to the low quality of the electron-density map in this region.

The C-terminal region of *Dm*QC shows high flexibility and can cover distinct regions of the protein depending on the crystal contacts. Two possible conformations of the ten terminal residues have been observed (Koch *et al.*, 2012). The C-terminal conformation of the structure presented here is similar to that of chain *B* of wild-type *Dm*QC in space group $P2_1$ (PDB entry 4f9u), in which the C-terminal residues are directed away from the N-terminus of the protein.

In previously determined structures of *Dm*QC (PDB entries 4f9u and 4f9v), the crystal packing stabilized the conformation of the oligosaccharides attached to Asn42, allowing the localization of seven

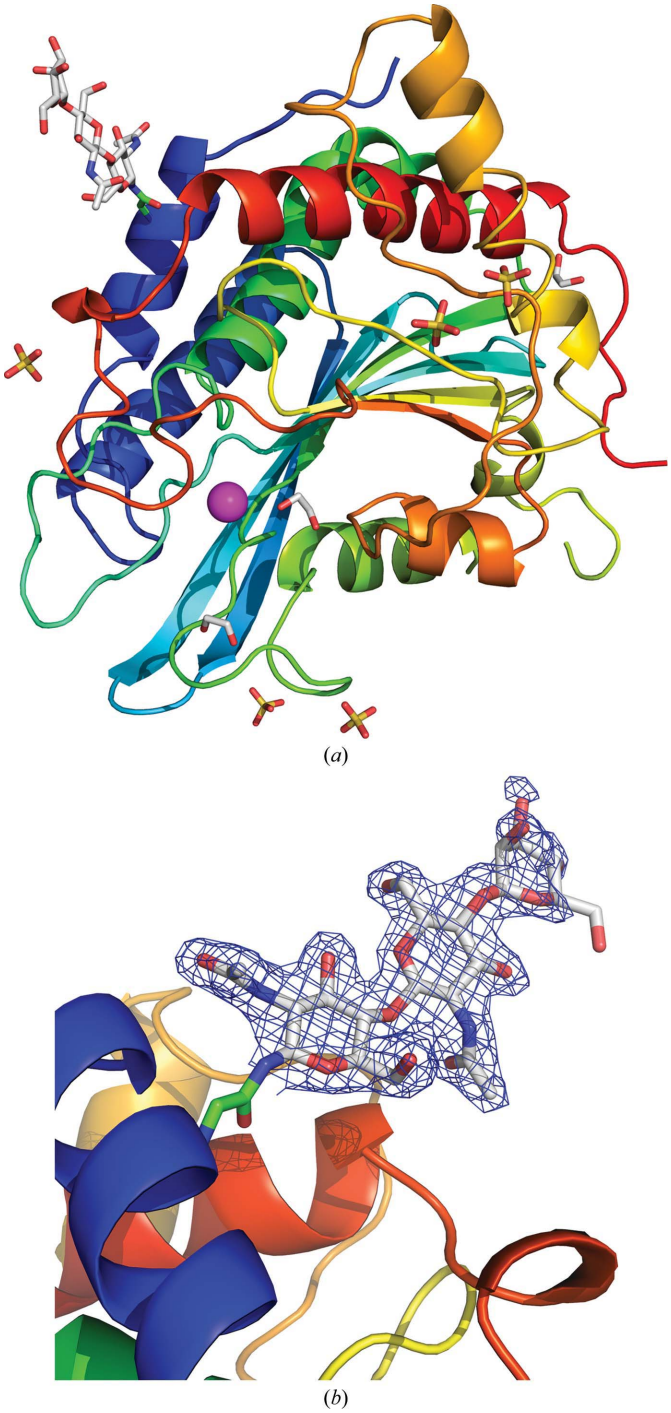


Figure 2
The structure of unliganded glutaminyl cyclase from *D. melanogaster* represented by secondary-structure elements (N-terminus in blue; C-terminus in red), and detail of the glycosylation site. (*a*) The oligosaccharides, sulfate anions and ethylene glycol molecules are represented as sticks. The zinc cation at the active site is represented as magenta sphere. (*b*) The $2F_o - F_c$ map around the oligosaccharide contoured at the 1σ level (blue).

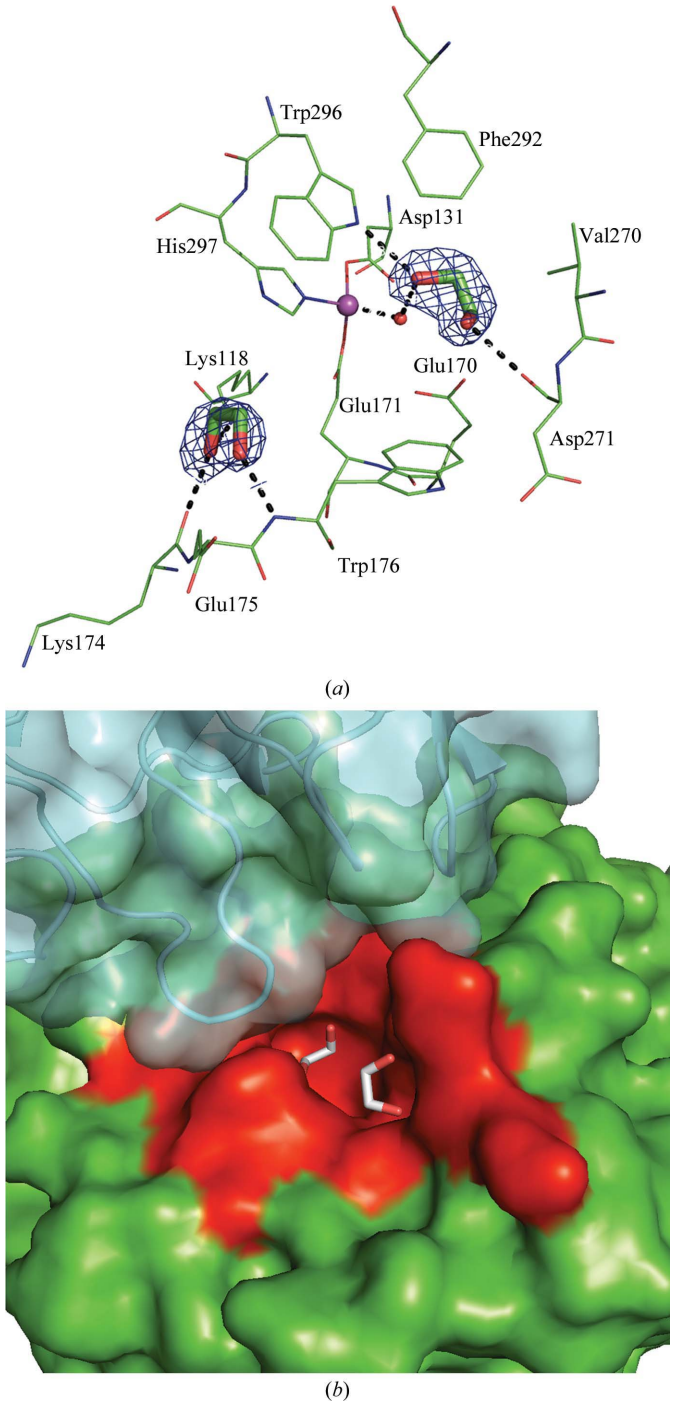


Figure 3
The active site and its solvent accessibility. (*a*) The two ethylene glycol molecules near the *Dm*QC active site. Amino acids are represented by lines, the zinc ion is shown as a sphere and ethylene glycol molecules are represented by sticks; hydrogen bonds are depicted by black dashes. The $2F_o - F_c$ map around the ethylene glycol molecules is contoured at 1σ (blue). (*b*) The *Dm*QC active site is accessible in the new *I*4 crystal form. *Dm*QC (green) and its active site (red) are represented by a surface; a symmetry-related molecule (cyan) is represented by a transparent surface and secondary-structure elements.

Table 5
Available crystallization conditions for glutaminyl cyclase from *D. melanogaster*.

Reservoir composition	pH	Protein concentration (g l^{-1})	Space group	Radiation source	Resolution (Å)	PDB code
10% PEG 35 000, 0.1 *M* Tris, 2 m*M* PBD150	8.0	7.5	$P2_1$	BESSY II	1.8	4f9u
12% PEG 8000, 0.1 *M* MES, 2 m*M* PBD150	6.5	10	$P6_5$	BESSY II	2.1	4f9v
1.6 *M* ammonium sulfate, 10% 1,4-dioxane, 40 m*M* MEGA-9, 0.1 *M* HEPES	7.5	15	*I*4	Rigaku RU-300	2.0	4fwu

monosaccharides in the electron density (Koch *et al.*, 2012). In the present structure (PDB entry 4fwu) only three solvent-exposed monosaccharide units could be detected (see Fig. 2); the sugars do not participate in crystal contacts.

Three ethylene glycol molecules, one chloride anion and five sulfate anions originating from the reservoir solution and cryoprotection could be localized on the surface of the molecule (see Fig. 2). One molecule of ethylene glycol binds directly to the active site and a second molecule binds close to the active site. Superposition of all five individual monomers of *Dm*QC deposited in the PDB (PDB entries 4f9u, 4f9v and 4fwu) revealed a high level of structural conservation, which included all of the residues in the active site: the r.m.s.d. between C^{α} atoms of pairwise compared monomers (∼298 atoms) does not exceed 0.45 Å (McNicholas *et al.*, 2011; Krissinel & Henrick, 2007). There are no significant conformational changes upon ligand binding.

The symmetry-related molecule in space group *I*4 interacts closely with the region around Phe292, preventing binding of the inhibitor PBD150 or similar derivatives in the active site. Our inhibition studies have previously shown that the potency of PBD150 is significantly lower for *Dm*QC than for hQC or mQC (Koch *et al.*, 2012). However, *Dm*QC remains a useful tool for the design and testing of new inhibitors with novel scaffolds and different binding modes. Soaking of these inhibitors would not be restricted by the crystal contacts observed in space group *I*4 (Fig. 3*b*).

4. Conclusion

We have succeeded in finding a novel crystallization condition for *Dm*QC that led to the growth of crystals in the new space group *I*4. This is the third space group found for crystals of *Dm*QC (a brief summary is given in Table 5). The new crystals will facilitate soaking experiments for novel inhibitors with alternative scaffolds and binding modes.

This work was supported by the ProNet-T^3 (Protein-Competence-Network-Halle: Tools, Targets and Therapeutics) of the German Federal Ministry for Education and Research (BMBF) (grant No. 03IS2211F/Inc-01) and in part by the Ministry of Education, Youth and Sports of the Czech Republic (grant No. CZ.1.07/2.3.00/30.0029).

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Buchholz, M., Hamann, A., Aust, S., Brandt, W., Böhme, L., Hoffmann, T., Schilling, S., Demuth, H. U. & Heiser, U. (2009). *J. Med. Chem.* **52**, 7069–7080.
Buchholz, M., Heiser, U., Schilling, S., Niestroj, A. J., Zunkel, K. & Demuth, H. U. (2006). *J. Med. Chem.* **49**, 664–677.
Carrillo, D. R., Parthier, C., Jänckel, N., Grandke, J., Stelter, M., Schilling, S., Boehme, M., Neumann, P., Wolf, R., Demuth, H. U., Stubbs, M. T. & Rahfeld, J. U. (2010). *Biol. Chem.* **391**, 1419–1428.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Huang, K.-F., Liaw, S.-S., Huang, W.-L., Chia, C.-Y., Lo, Y.-C., Chen, Y.-L. & Wang, A. H.-J. (2011). *J. Biol. Chem.* **286**, 12439–12449.
Huang, K.-F., Liu, Y.-L., Cheng, W.-J., Ko, T.-P. & Wang, A. H.-J. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 13117–13122.
Huang, K.-F., Wang, Y.-R., Chang, E.-C., Chou, T.-L. & Wang, A. H.-J. (2008). *Biochem. J.* **411**, 181–190.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Koch, B., Kolenko, P., Buchholz, M., Carrillo, D. R., Parthier, C., Wermann, M., Rahfeld, J. U., Reuter, G., Schilling, S., Stubbs, M. T. & Demuth, H. U. (2012). *Biochemistry*, **51**, 7383–7392.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Lovell, S., Davis, I., Arendall, W. B. III, de Bakker, P., Word, J., Prisant, M., Richardson, J. & Richardson, D. (2003). *Proteins*, **50**, 437–450.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Ruiz-Carrillo, D., Koch, B., Parthier, C., Wermann, M., Dambe, T., Buchholz, M., Ludwig, H. H., Heiser, U., Rahfeld, J. U., Stubbs, M. T., Schilling, S. & Demuth, H. U. (2011). *Biochemistry*, **50**, 6280–6288.
Schilling, S., Lindner, C., Koch, B., Wermann, M., Rahfeld, J. U., von Bohlen, A., Rudolph, T., Reuter, G. & Demuth, H. U. (2007). *Biochemistry*, **46**, 10921–10930.
Schilling, S., Niestroj, A. J., Rahfeld, J. U., Hoffmann, T., Wermann, M., Zunkel, K., Wasternack, C. & Demuth, H. U. (2003). *J. Biol. Chem.* **278**, 49773–49779.
The UniProt Consortium (2012). *Nucleic Acids Res.* **40**, D71–D75.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wintjens, R., Belrhali, H., Clantin, B., Azarkan, M., Bompard, C., Baeyens-Volant, D., Looze, Y. & Villeret, V. (2006). *J. Mol. Biol.* **357**, 457–470.

Anastasia Tziridis[a], Daniel Rauh[b], Piotr Neumann[c], Petr Kolenko[d], Anja Menzel[e], Ulrike Bräuer[f], Christian Ursel, Peter Steinmetzer, Jörg Stürzebecher, Andrea Schweinitz[g], Torsten Steinmetzer[h] and Milton T. Stubbs*

Correlating structure and ligand affinity in drug discovery: a cautionary tale involving second shell residues

Abstract: A high-resolution crystallographic structure determination of a protein–ligand complex is generally accepted as the 'gold standard' for structure-based drug design, yet the relationship between structure and affinity is neither obvious nor straightforward. Here we analyze the interactions of a series of serine proteinase inhibitors with trypsin variants onto which the ligand-binding site of factor Xa has been grafted. Despite conservative mutations of only two residues not immediately in contact with ligands (second shell residues), significant differences in the affinity profiles of the variants are observed. Structural analyses demonstrate that these are due to multiple effects, including differences in the structure of the binding site, differences in target flexibility and differences in inhibitor binding modes. The data presented here highlight the myriad competing microscopic processes that contribute to protein–ligand interactions and emphasize the difficulties in predicting affinity from structure.

Keywords: crystal structure; factor Xa; ligand affinity; protein flexibility; selectivity; structure-based drug design.

DOI 10.1515/hsz-2014-0158
Received March 3, 2014; accepted April 24, 2014

[a]Present address: Sanofi-Aventis Deutschland GmbH, Industriepark Hoechst, D-65926 Frankfurt/Main, Germany.
[b]Present address: Fakultät für Chemie und Chemische Biologie, Technische Universität Dortmund, Otto-Hahn-Strasse 6, D-44227 Dortmund, Germany.
[c]Present address: Institut für Mikrobiologie und Genetik, Georg-August-Universität, Justus-von-Liebig-Weg 11, D-37077 Göttingen, Germany.
[d]Present address: Institute of Macromolecular Chemistry, Heyrovského nám. 2, CZ-162 06 Praha 6, Czech Republic.
[e]Present address: Institut für Rechtsmedizin, Otto-von-Guerike-Universität, Leipziger Str. 44, D-39120 Magdeburg, Germany.
[f]Present address: School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.
[g]Present address: Institut für Physiologie II, Universitätsklinikum Jena, Kollegiengasse 9, D-07743 Jena, Germany.
[h]Present address: Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, Marbacher Weg 6, D-35032 Marburg, Germany.
***Corresponding author: Milton T. Stubbs,** Institut für Biochemie und Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, Kurt-Mothes-Straße 3, D-06120 Halle/Saale, Germany, e-mail: stubbs@biochemtech.uni-halle.de
Anastasia Tziridis, Daniel Rauh, Piotr Neumann, Petr Kolenko, Anja Menzel, Ulrike Bräuer and Christian Ursel: Institut für Biochemie und Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, Kurt-Mothes-Straße 3, D-06120 Halle/Saale, Germany
Peter Steinmetzer and Jörg Stürzebecher: Institut für Vaskuläre Medizin und Biologie, Friedrich-Schiller-Universität Jena, Nordhäuser Straße 78, D-99089 Erfurt, Germany
Andrea Schweinitz and Torsten Steinmetzer: Curacyte Chemistry GmbH, Winzerlaer Str. 2, D-07745 Jena, Germany

Dedicated to the memory of our friend and colleague Professor Dr. Jörg Stürzebecher, whose untimely death was a tragic loss to us and to the scientific community.

Introduction

The availability of a protein–ligand complex crystal structure at high resolution is commonly considered to be a 'gold standard' in modern drug discovery. The ability to analyze the interaction at (near) atomic resolution allows a precise characterization of structural elements and ligand binding geometries, providing a framework for rationalizing structure–affinity relationships and for guiding further modifications in chemical space exploration. Despite considerable progress in computer-aided drug discovery (Sliwoski et al., 2014), however, a robust correlation between structure and affinity (the 'scoring function') remains elusive (Huang et al., 2010; Martin and Clements, 2013).

An implicit assumption of most scoring functions in molecular docking is that the binding energy arising from interactions between atoms of the protein ligand complex that are in direct contact are the dominant contributors to affinity (with the obvious exception of electrostatic interactions, which are *per se* long-range). Yet any measure of ligand affinity must also incorporate any changes in

protein, ligand and solvent organization, effects that are not readily accessible at an atomic level and so must be estimated on an empirical basis. In particular, plasticity or flexibility of the target protein, the ligand or both continues to present a challenge to structure-guided drug design (Teague, 2003; Feixas et al., 2014).

To gain a better understanding of what information structures can deliver, there is a need for suitable model systems that allow dissection of protein–ligand interactions into their component contributions at the structural level. Due to a common structural scaffold and availability of a wide range of potential therapeutic targets, the trypsin-like serine proteinase superfamily has long provided a valuable resource for the analysis of ligand binding, affinity and selectivity (Turk et al., 1991; Stubbs et al., 1995; Hedstrom et al., 1996; Renatus et al., 1998; Dullweber et al., 2001). As an example, the selective inhibition of coagulation factor Xa (which occupies a strategic position at the intersection of the intrinsic and extrinsic coagulation pathways) represents a validated and attractive target for anticoagulant therapy to regulate the blood homeostasis system (Straub et al., 2011), and there is now a substantial database of factor Xa inhibitors that include orally-available compounds suitable for clinical use (Yeh et al., 2012; Steinberg and Becker, 2013).

Three sequential regions in the catalytic domain of human factor Xa have been implicated in inhibitor binding: the '190'-, the '99'- and the '175'-loops [numbering according to the chymotrypsin(ogen) system for trypsin-like serine proteases; Shotton and Hartley, 1970] (Figure 1), which flank the S1-, S2- and S3/S4-binding pockets of the enzyme. The primary specificity pocket S1, including the Asp189 necessary for binding to the P1-Arg/Lys residue of substrates, is more hydrophobic than that of trypsin through the presence of Ala190. The S2-site is effectively closed by the side-chain of Tyr99, explaining the strong preference for Gly as P2 residue in substrates. The S3/S4-pocket, formed by the surface-exposed aromatic side chains of Tyr99, Trp215 and Phe174, has a largely hydrophobic character and possesses an additional distal electronegative region formed by the carbonyl groups of Glu97, Thr98 and Ile176 (Stubbs et al., 1995; Brandstetter et al., 1996). Structural information derived from these binding site properties has played a key role in the development of orally available anticoagulants (Perzborn et al., 2011; Nar, 2012), and the structures of a large number of high affinity low molecular weight inhibitors have been described in complex with factor Xa and/or trypsin, although not all have been deposited in the Protein Data Bank (Rose et al., 2013). In particular, complexes are available for the inhibitors **(3)** (Stürzebecher et al., 1989; Renatus et al., 1998; Rauh et al., 2002, 2003), **(4)** (Faull et al., 1996; Rauh et al., 2002; Stubbs et al., 2002; Reyda et al., 2003), **(5)** (Schweinitz

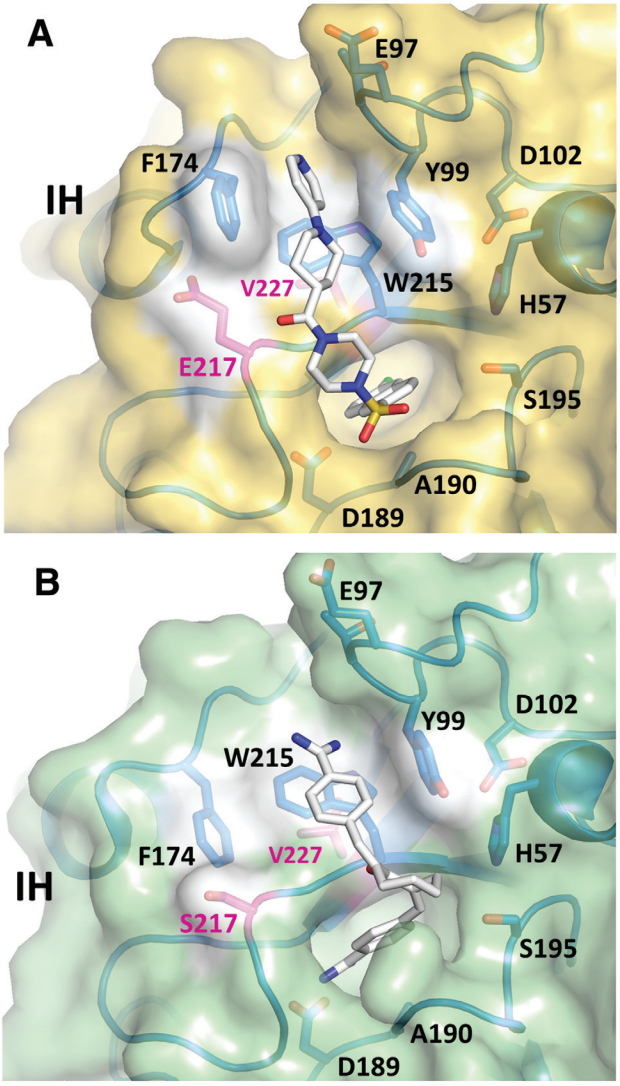


Figure 1 Surface representations of the binding sites of (A) *TripleGlu217*Val227 in complex with inhibitor **(4)** ('up'-conformation, PDB code 1v2k) (Rauh et al., 2004) and (B) *Triple*Ser217Val227 in complex with inhibitor **(3)** ('down'-conformation, Protein Data Bank code 1v2m) (Rauh et al., 2003).
Note the differences in orientation of the intermediate helix (IH) and the degree of exposure of Phe174 (see also Supplementary Movie 1). Whereas variant *Triple*Ser217Val227 appears to only adopt the 'down'-conformation, *TripleGlu217*Val227 can adopt both dependent upon ligand binding; the 'up'-conformation is indistinguishable in structure to the corresponding region in factor Xa. Inhibitors **(3)** and **(4)** would bind to factor Xa in an extended conformation to occupy both the primary specificity pocket and the aromatic box, formed by Tyr99, Trp215 and Phe174 in the up conformation. The more hydrophobic S1 pocket of factor Xa-like variants allows occupation by the chloronaphthyl group of **(4)**. Throughout this paper, the color coding is used to distinguish between 'up' and 'down' conformations (yellow and green surfaces, respectively).

et al., 2006; Stürzebecher et al., 2007), **(7)** (Whitlow et al., 1999; Arnaiz et al., 2000), **(8)** (Hirayama et al., 2002) and **(9)** (Pruitt et al., 2000) (Figure 2) studied here. Factor Xa active site inhibitors display an overall L-shape, allowing them to occupy both the S1 and the S3/S4 binding sites of the proteinase. As with other trypsin-like proteinases, a basic moiety (often amidine-based) may occupy the primary specificity pocket to form both ionic and hydrophilic interactions with Asp189, but due to the presence of Ala190, hydrophobic moieties are also accepted. The distal aromatic S3/S4 box binds large hydrophobic residues, but also readily accepts additional basic moieties through cation–π interactions (Schärer et al., 2005).

For the most part, issues of affinity and selectivity are investigated via chemical variation of the ligand(s) and/or comparison of closely-related but distinct enzymes. In previous studies, we have sought to explore elements of selectivity and specificity in protein–ligand interactions through grafting of the ligand binding pocket of human factor Xa on to the structurally-related rat and bovine trypsins using site-directed mutagenesis (Rauh et al., 2002, 2004; Reyda et al., 2003). In the three loops, trypsin Ser190 ('190-loop') was mutated to an Ala, yielding a larger hydrophobic cavity; the sequence Ser172–Ser173–Phe174–Ile175 was exchanged in the so-called '175-loop' (Tyr172–Pro173–Gly174–Gln175 in trypsin) and Glu97 and Tyr99 in the '99-loop' (Asn97 and Leu99 in trypsin) were replaced, yielding the so-called *Triple* variant that contains all three loops of factor Xa involved in inhibitor binding. Unexpectedly, these variants revealed strikingly different conformations to those of factor Xa, with a structural plasticity that had significant effects on ligand affinity.

The most prominent differences in the factor Xa binding site were found in the '175-loop' and the intermediate helix (IH in Figure 1, residues 168–182). In the initial rat trypsin *Triple* variant in complex with inhibitor **(1)** (Reyda et al., 2003), the side chain of Phe174 was observed to be buried in the core of the enzyme (the 'down' conformation, Figure 1B), resulting in the complete absence of one edge of the S3/S4-binding site and a partial unwinding of the α-helix from residue Asp165 to Ser170. Binding of the factor Xa-specific inhibitor **(4)**, however caused this variant to adopt the factor Xa-like 'up' conformation, with surface exposure of Phe174. Surprisingly, analogous mutations in the three loops in bovine trypsin generated a corresponding *Triple* variant (termed *Triple*Ser217Val227 here, see below; residues mutated from trypsin are denoted in italics) that exhibited only the 'down'-conformation (Rauh et al., 2004). Attempts to prevent this burial by further mutation of Phe174 in bovine trypsin variants resulted in extreme variation in the flexibility of this region, however. Inhibition studies revealed that the affinity of these variants for **(4)** was some three to six orders of magnitude lower than that for factor Xa, suggesting that conformational plasticity and/or structural reorganization of the binding site comes at an energetic cost and can severely compromise ligand affinity.

Residue 217 at the periphery of the ligand-binding site was identified as exerting a significant effect on this region (Rauh et al., 2004). In factor Xa, this residue is a glutamic acid, the side-chain carboxylate group of which is involved in hydrogen bonding to Ser172 O^{γ} and presumably stabilizes the '175'-loop. Substitution of Ser217 with Glu in the chimeras (*TripleGlu217*Val227), however, resulted in variants that showed both the 'up'- and 'down'- conformations (see Figure 1; Supplementary Movie 1) as a function of ligand binding. This serendipitous finding makes this system ideal for studying the influence of protein flexibility on ligand binding.

In our current studies, we have endeavored to influence the flexibility of the intermediate helix through manipulation of the packing of the hydrophobic core by replacing Val227 immediately below Trp215 (Figure 1). In factor Xa, this residue is an Ile; we hypothesized that steric hindrance caused by the more bulky hydrophobic side chain may disfavor burial of Phe174 and force the factor Xa-like 'up'-conformation. To accentuate this effect, we also introduced Phe in position 227, which is found in an equivalent location in the related serine proteinase thrombin. Both mutations were analyzed with respect to a range of inhibitors (Figure 2) in the context of Ser217 or Glu217, resulting in the four variants *Triple*Ser217*Ile227*, *Triple*Ser217*Phe227*, *TripleGlu217Ile227* and *TripleGlu217Phe227*.

Results and discussion

Activity and inhibitory profiles of the variants

All variants could be expressed and refolded in satisfactory yields, and displayed catalytic activity towards the commonly used chromogenic substrate Pefachrome® tPA (Table 1). Variants containing Ser217 exhibited a lower specific activity than their Glu217-containing counterparts, an effect that appears to be due to increased K_m values. Inhibitory profiles for compounds **(1–9)**, which include the weak non-selective inhibitors benzamidine **(1)** and **(2)**, were measured for each of the four variants, as well as the parent proteinases bovine trypsin and *Triple*Ser217Val227 and the target enzyme factor Xa (Table 1; Supplementary Figure 1). For the most part, variants *Triple*Ser217*Ile227*

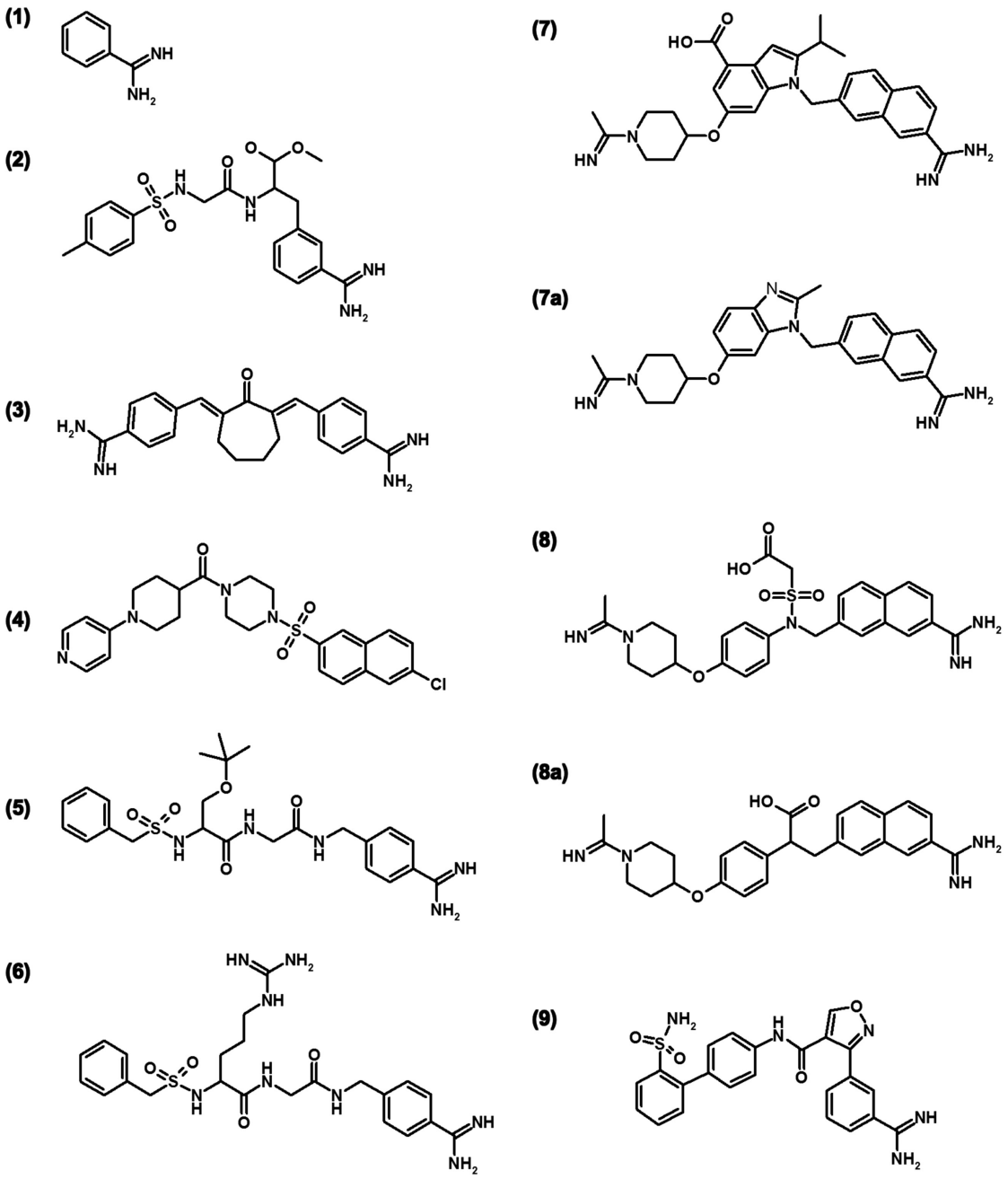


Figure 2 Inhibitors used in this study.
(1) Benzamidine (unspecific, weak inhibition of both trypsin and factor Xa, better for trypsin); **(2)** Nα-tosyl glycyl 3-amidino-D,L-phenylalanine-methylester (unspecific, weak inhibition of both enzymes, marginally better for factor Xa); **(3)** 2,7-bis(4-amidinobenzylidene)-cycloheptan-1-one (unspecific, strong inhibition of both enzymes, marginally better for factor Xa) (Stürzebecher et al., 1989); **(4)** [4-(6-chloro-naphthalene-2-sulfonyl)-piperazin-1-yl]-([3,4,5,6-tetrahydro-2H-1,4']bipyridinyl4-yl)-methanone (Zeneca, specific for factor Xa) (Faull et al., 1996); **(5)** benzylsulfonyl-D-Ser(*tert*butyl)-Gly-4-amidinobenzylamide (Curacyte Chemistry GmbH, Jena, Germany) and **(6)** benzylsulfonyl-D-Arg-Gly-4-amidinobenzylamide (Curacyte, Munich, Germany) (Schweinitz et al., 2006); **(7)** 1-[(7-carbamimidoyl naphthalen-2-yl)methyl]-6-(1-ethanimidoyl piperidin-4-yl)oxy-2-propan-2-yl-indole-4-carboxylic acid (Berlex, Montville, NJ, USA) (Arnaiz et al., 2000); **(8)** 2-[(7-carbamimidoyl naphthalen-2-yl)methyl-[4-(1-ethanimidoyl piperidin-4-yl)oxyphenyl]sulfamoyl] ethanoic acid (Yamanouchi, Nagano, Japan) (Hirayama et al., 2002); **(9)** 3-(3-carbamimidoylphenyl)-N-[4-(2-sulfamoylphenyl)phenyl]-1,2-oxazole-4-carboxamide (DuPont, Wilmington, DE, USA) (Pruitt et al., 2000).

and *Triple*Ser217*Phe227* showed small but significant decreases in affinity for the inhibitors compared to *Triple*Ser217Val227. All Glu217-containing variants, however, exhibited higher affinity than their corresponding Ser217 counterparts, with the greatest increases observed on going from *Triple*Ser217*Phe227* to *TripleGlu217Phe227*.

Table 1 Enzymatic activity and corresponding inhibition profiles of the variants.

Variant	Kinetic constants			K_i-values (μM)								
	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (10^6 $M^{-1}s^{-1}$)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Bovine trypsin	11.9	36.2	3.04	39[b]	4.4[c]	0.40[b]	13[b]	0.013	0.01	0.014	0.010	0.04
*Triple*Ser217Val227	104	n.d.[a]	n.d.[a]	69[b]	22[c]	1.25[b]	2.00[b]	0.054	0.050	0.102	0.039	0.60
*Triple*Ser217*Ile227*	988	247	0.25	227	110	1.80	3.60	0.360	0.300	0.447	0.140	1.95
*Triple*Ser217*Phe227*	1610	227	0.14	695	290	2.50	0.90	1.030	0.200	0.148	0.156	1.22
*TripleGlu217*Val227	61.6	176	2.86	51[c]	32[c]	0.05[c]	0.70[c]	0.022	0.007	0.008	0.006	0.04
TripleGlu217Ile227	241.0	250	1.04	96	23	1.20	0.50	0.064	0.020	0.023	0.008	0.07
TripleGlu217Phe227	88.1	165	1.87	186	24	0.08	0.024	0.037	0.003	0.004	0.003	0.02
Factor Xa	71.3	147	2.06	160	0.84	0.025	0.020	0.007	0.006	0.002	0.0009	0.0003

Values are averages of at least three measurements.
[a]n.d.=not determined.
[b]Rauh et al., 2003.
[c]Rauh et al., 2004.

Crystal structures of variant–ligand complexes

Co-crystals of each of the variants were obtained with several inhibitors under a narrow range of crystallization conditions (Table 2; Supplementary Table 1). Nine different crystal forms (**A–I**) were observed, of which four have been reported previously for bovine trypsin or variants thereof (Stubbs et al., 1995; Renatus et al., 1998; Rauh et al., 2004). Variant *Triple*Ser217*Ile227* crystallized in the commonly observed trigonal form **A**, regardless of the bound inhibitor, and in each case exhibited

Table 2 Crystallization conditions and data deposition.

Structure[a]	Crystallization conditions[b]	Ligand[c]	Crystal form[d]	Space group	Maximum resolution (Å)	PDB entry
*Triple*Ser217*Ile227*.A1	20% PEG 8000; 0.1 M Imi; 0.3 M AmS; pH 8.0	1	A	$P3_121$	1.18	3PLB
*Triple*Ser217*Ile227*.A2	20% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 7.0	2	A	$P3_121$	1.52	3PLK
*Triple*Ser217*Ile227*.A3	20% PEG 8000; 0.1 M Imi; 0.2 M AmS; pH 8.0	3	A	$P3_121$	1.63	3PLP
*Triple*Ser217*Ile227*.A5	30% PEG 8000; 0.1 M Imi; 0.2 M AmS; pH 8.0	5	A	$P3_121$	1.52	3PM3
*Triple*Ser217*Ile227*.A6	20% PEG 8000; 0.1 M Imi; 0.3 M AmS; pH 7.0	6	A	$P3_121$	1.45	3PMJ
TripleGlu217Ile227.A1	30% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 8.0	1	A	$P3_121$	1.63	3PWB
TripleGlu217Ile227.A2	20% PEG 8000; 0.1 M Imi; 0.2 M AmS; pH 7.0	2	A	$P3_121$	1.60	3PWC
TripleGlu217Ile227.D6	30% PEG 8000; 0.1 M Imi; 0.2 M AmS; pH 8.0	6	D	$P3_221$	1.70	3Q00
TripleGlu217Ile227.E9	30% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 7.0	9	E	$P6_5$	2.00	3PYH
*Triple*Ser217*Phe227*.C1	30% PEG 8000; 0.1 M Imi; 0.2 M AmS; pH 7.0	1	C	$P2_12_12_1$	1.62	3UNQ
*Triple*Ser217*Phe227*.B2	20% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 7.0	2	B	$P2_12_12_1$	1.80	3UNS
*Triple*Ser217*Phe227*.F6	20% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 8.0	6	F	$P3_221$	1.69	3UOP
*Triple*Ser217*Phe227*.F7	20% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 8.0	7	F	$P3_221$	1.54	3UPE
*Triple*Ser217*Phe227*.G8	20% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 8.0	8	G	$C222_1$	2.10	3UUZ
*Triple*Ser217*Phe227*.E9	30% PEG 8000; 0.1 M Imi; 0.2 M AmS; pH 7.0	9	E	$P6_5$	2.40	3UQV
*Triple*Ser217*Phe227*.H9	20% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 8.0	9	H	$P3_221$	1.80	3UQO
TripleGlu217Phe227.I1	30% PEG 8000; 0.1 M Imi; 0.3 M AmS; pH 7.0	1	I	P1	2.90	3UY9
TripleGlu217Phe227.B2	20% PEG 8000; 0.1 M Imi; 0.1 M AmS; pH 8.0	2	B	$P2_12_12_1$	1.42	3UWI
TripleGlu217Phe227.D6	30% PEG 8000; 0.1 M Imi; 0.2 M AmS; pH 8.0	6	D	$P3_221$	1.80	3V12
TripleGlu217Phe227.E9	20% PEG 8000; 0.1 M Imi; 0.2 M AmS; pH 7.0	9	E	$P6_5$	1.63	3V13

[a]Structures are distinguished as follows: {variant}.{crystal form} {inhibitor}; inhibitors are shown in Figure 1. Crystal forms A–D are the designations described previously (Rauh et al., 2004).
[b]Imi, imidazole; AmS, ammonium sulfate.
[c]See Figure 2.
[d]See Supplementary Table 1 for further details and data statistics.
PDB: Protein Data Bank.

the 'down'-conformation (Figure 3). Binding modes for **(1)**, **(2)** and **(3)** (Supplementary Figure 2) are as described previously for trypsin and/or *Triple*Ser217Val227 (Renatus et al., 1998; Rauh et al., 2002, 2003, 2004). In light of the complexity of the structural comparisons presented in this paper, solvent structures and their differences will not be discussed. Briefly, **(1)** occupies the S1 pocket as characteristic for benzamidine-based inhibitors, with the amidino moiety making a salt bridge to Asp189 and a hydrogen bond to the Gly219 carbonyl group. Similarly, the benzamidino group of **(2)** occupies the S1 pocket, while the tosyl moiety reaches towards what in factor Xa would be the aromatic box; the lack of further contacts is consistent with the rather weak affinity and the strong influence of crystal packing observed previously for this inhibitor (Renatus et al., 1998; Rauh et al., 2004). The chemically symmetric compound **(3)** binds in a (Z,Z) configuration, with the proximal amidinobenzyl residue occupying the S1 site and the distal group in the S3/S4 binding site as observed in *Triple*Ser217Val227 (Rauh et al., 2003).

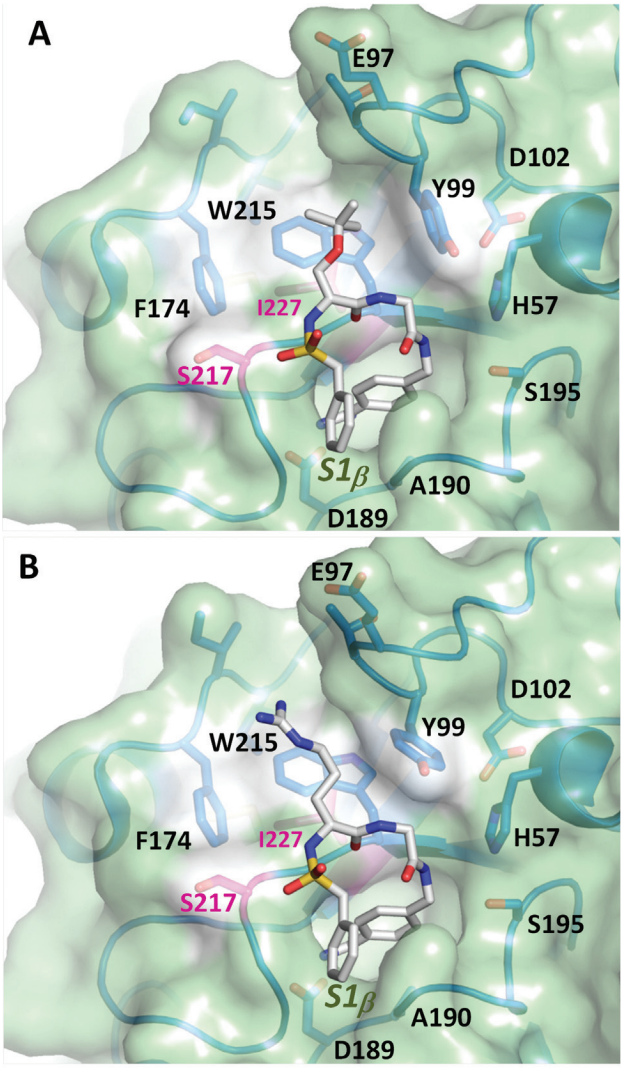


Figure 3 Crystal structures of variant *Triple*Ser217*Ile227* in complex with inhibitors **(5)** (A) and **(6)** (B).
In both cases, the enzyme is found in the 'down' conformation; the P3 D-amino acid residues reach out towards what would be the aromatic box in the Factor Xa 'up' conformation, whereas the P4 benzene ring nestles into the S1β pocket.

Inhibitor **(5)** exhibits a similar binding mode to that observed in factor Xa (Schweinitz et al., 2006), with the benzamidine-like S1 contacts described above supplemented by a hydrogen bond between the amide proton of the P1 4-amidinobenzylamide and the carbonyl oxygen of Ser214. The P3 D-Ser(*tert*-butyl) residue makes antiparallel main chain–main chain-like hydrogen bonds to Gly216, whereas a hydrogen bond is formed between one of the sulfonyl oxygens and Gly219 NH, and the P4 benzylsulfonyl ring nestles against the so-called $S1_{\beta}$ site near the mouth of the S1 pocket above the Cys220–Cys191 disulfide bridge. The *tert*-butyl group of **(5)** approaches the side chains of Tyr99 and Trp215, so that it would be ideally positioned to occupy the aromatic box of factor Xa. Inhibitor **(6)** binds in a similar fashion, with the side chain of Tyr99 rotating ~60° about χ_2 towards the P2 glycyl spacer and the C^{γ} and C^{δ} atoms of the P3 D-Arg; the guanidinium group points away from the enzyme, however, apparently stabilized by contacts to a symmetry related molecule.

Co-crystals of *TripleGlu217Ile227* with **(1)** and **(2)** also belonged to crystal form **A**, and apart from the presence of the Glu217 side chain, the structures are indistinguishable from the corresponding *Triple*Ser217*Ile227* complexes (Figure 4), i.e. the 'down'-conformation is found. Crystallization with **(6)**, however, resulted in the $P3_221$ crystal form **D** described previously (Rauh et al., 2004), and the enzyme is found in the 'up'-conformation. As a result, the D-Arg guanidinium moiety reorients with respect to its position in *Triple*Ser217*Ile227*–**(6)** to occupy the aromatic box, parallel to the side chain of Phe174, and is probably stabilized by cation–π interactions. All other interactions are comparable, including the side chain rotation of Tyr99, although no density is observed for the P4 benzene ring. Crystals were also obtained with **(9)** in a novel hexagonal $P6_5$ crystal form **E**. Once again, *TripleGlu217Ile227* is found in the 'up'-conformation, with the benzamidine group in the S1 pocket and the terminal aromatic rings occupying the aromatic box in an arrangement reminiscent of benzene crystal packing, as described for compound **(4)** in complex with the rat *Triple* variant (Reyda et al., 2003). In **(9)**, the carbonyl group of the amide linkage, which is co-planar with the isoxazole ring, is hydrogen bonded to Gly216; in trypsin (Pruitt et al., 2000), the amide bond

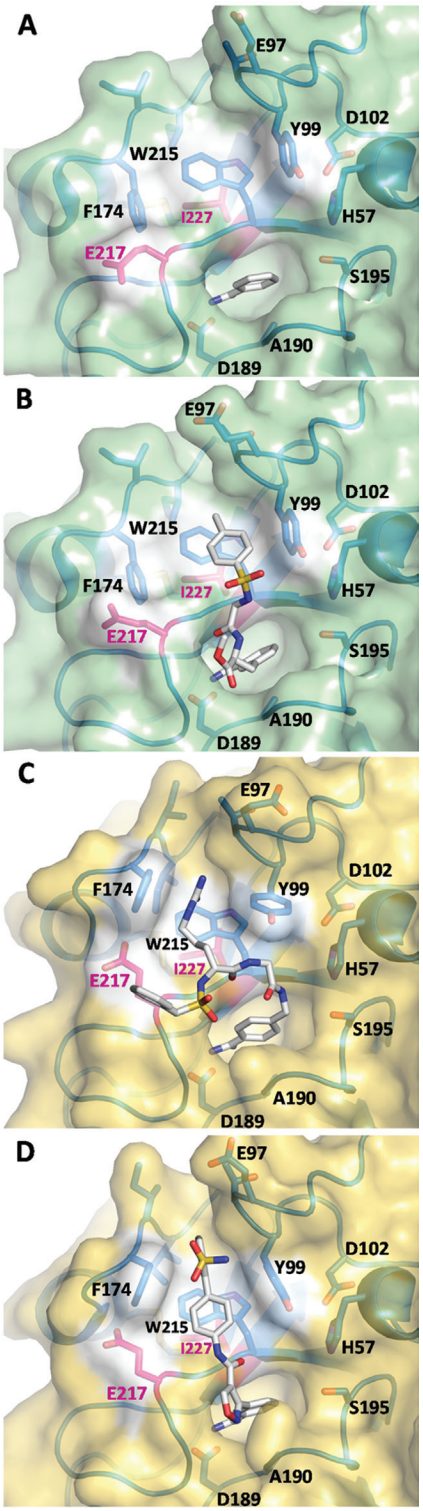


Figure 4 Structures of *TripleGlu217Ile227* inhibitor complexes. In the presence of the low affinity inhibitors **(1)** (A) and **(2)** (B), the proteinase adopts the 'down' conformation, whereas inhibitors **(6)** (C) and **(9)** (D) nestle into the aromatic box of the 'up' conformation. Note in panel (C) the different direction of the P3 D-Arg side chain of **(6)** compared to that of the same inhibitor in *Triple*Ser217*Ile227* (Figure 3B; see also Figure 7C), and that the P4 benzene ring is not well defined.

is perpendicular to the five-ring and is directed towards His57. The sulfonamide nitrogen makes hydrogen bonds with the Glu97 carbonyl oxygen.

Variant *Triple*Ser217*Phe227* crystallized in the presence of **(1)** in the orthorhombic crystal form **C**. Despite clear density for most of the protein, including the active site residues and the benzamidine inhibitor, residues Lys169–Ile175 of the intermediate helix (including Phe174 of the aromatic box) exhibit significant disorder (Figure 5), as does the side chain of Trp215 (although that of Phe227 on which it stacks is well defined). This disorder is consistent with a ~2°C decrease in apparent melting temperature measured by differential scanning calorimetry (Supplementary Figure 3). On the other hand, in crystal form **B** in the presence of **(2)**, the enzyme is observed in the 'up'-conformation. Although well defined by density, the inhibitor is strongly influenced by crystal packing as observed previously (Renatus et al., 1998; Rauh et al., 2004) and makes no contacts with the S3/S4 binding site. As in the complex with **(1)**, the Trp215 side chain density is weak. This variant is also observed in the 'up'-conformation in the presence of **(6)** in the trigonal crystal form **F**, where the binding mode of the inhibitor is equivalent to that described above for *TripleGlu217Ile227*, with the exception that the P4 benzene ring is well defined in the $S1_{\beta}$ position.

The naphthamidine-based inhibitor **(7)** yielded the same crystal form **F** and 'up'-conformation for the *Triple*Ser217*Phe227* variant, with the naphthamidine moiety occupying the S1 pocket. The acetimidoyl group is superimposable on the guanidinium moiety of **(6)**, and is sandwiched on one side by Phe174 and on the other by Tyr99, the side chain of which rotates to make a parallel stacking interaction. Yet another orthorhombic form **G** was obtained for *Triple*Ser217*Phe227* in the presence of **(8)**, once again in the 'up'-conformation. The naphthamidine group is superimposable upon that of **(7)**, but due to a different path of the aromatic linker, the acetimidoyl moiety is slightly displaced from the corresponding position in **(7)**, reaching a little further towards the 'cation hole'. This moiety is also sandwiched by the aromatic box side chains of Phe174 and Tyr99, but due to the different approach of the linker, these both adopt a more 'factor Xa'-like position. Finally, two crystal forms (**E** and **H**) were obtained in the presence of inhibitor **(9)**; in both cases, *Triple*Ser217*Phe227* was found in the 'up'-conformation, with ligand binding identical to that observed in *TripleGlu217Ile227* (crystal form **E**).

Co-crystals of benzamidine **(1)** with *TripleGlu217Phe227* belonged to the triclinic form **I**, with 16 complexes in the unit cell/asymmetric unit. Despite the relatively low resolution of 2.9 Å, it is clear that all protein monomers display

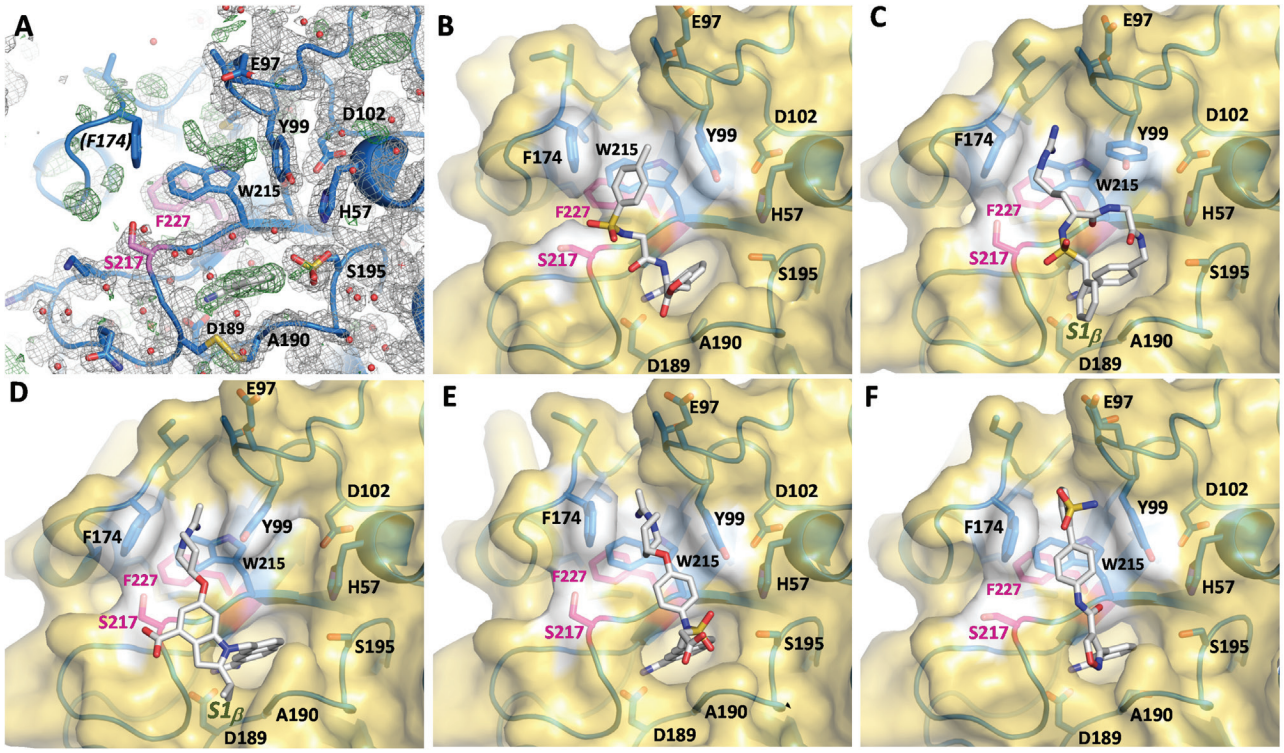


Figure 5 Structures of inhibitor complexes with *Triple*Ser217*Phe227*.
(A) Experimental electron densities (grey: 2*Fo–Fc* at 1σ; green *Fo–Fc* at 3σ) for the complex with **(1)**; both the inhibitor and residues Lys169–Ile175 were excluded from phase calculations ('omit map'). The missing density for the intermediate helix demonstrates disorder in this region. In complex with inhibitors **(2)**, **(6)**, **(7)**, **(8)** and **(9)** (panels B–F, respectively), the variant adopts the 'up' conformation.

the 'up'-conformation (Figure 6). Each of the monomers, arranged as four dimers of dimers, exhibit similar contacts within the crystals, with Phe174 approaching the bound S1 benzamidine of a neighboring molecule, so crystal packing might play a role in stabilizing this conformation. As noted previously (Stubbs et al., 2002; Rauh et al., 2004), bovine trypsin and variants thereof crystallize preferentially under the conditions used here in crystal forms **A**, **B** or **C** – or not at all. Thus the appearance of the new triclinic crystal form **I** suggests that variant *TripleGlu217Phe227* adopts the 'up'-conformation even in the absence of S3/S4 occupation. The same variant is also found in the 'up' conformation in complexes with **(2)**, **(6)** and **(9)** (crystal forms **B**, **D** and **E** respectively; Supplementary Figure 4), with binding characteristics as described above for *Triple*Ser217*Phe227*–**(2)** *TripleGlu217Ile227*–**(6)** and *TripleGlu217Ile227*–**(9)**.

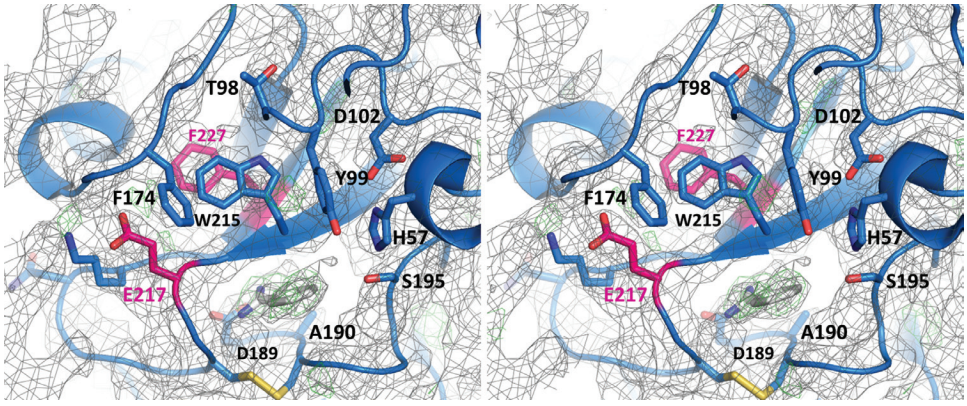


Figure 6 Stereo view of experimental electron densities (grey: 2*Fo–Fc* at 1σ; green *Fo–Fc* at 3σ) for *TripleGlu217Phe227* in complex with **(1)**. The inhibitor was not included in the phasing ('omit map'). The variant adopts the 'up' conformation, even though the aromatic box is unoccupied.

Taken together, correlation of the structural and inhibition data indicate that each of the six variants analyzed responds to inhibitor binding in a different manner. Despite the general drop in affinities compared to *Triple*Ser217Val227, the regions of *Triple*Ser217*Ile227* in contact with the inhibitors show no obvious differences in structure. The only structural variation appears to be for Ile175 and Ile176 (Supplementary Figure 2D), which make no contact with any of the inhibitors. As we can rule out an effect of crystal packing (as this variant only crystallizes in one form), the variation in position of Ile175 and Ile176 may be due to the additional methyl group of Ile227 compared to Val227 [although it should be noted that there is only one structure for *Triple*Ser217Val227 (Rauh et al., 2003, 2004), in complex with **(3)**]. We postulate that the overall reduced affinities for this variant may be due to microscopic structural fluctuations in the neighborhood of the ligand binding site (as observed for Ile175–Ile176, Supplementary Figure 2D). Conversely, the sole appearance of crystal form **A** suggests that, like its parent molecule *Triple*Ser217Val227 (Rauh et al., 2003, 2004), this variant prefers the 'down' conformation. Thus, the lack of aromatic box formation provides a possible explanation for the lack of discrimination between the P3 side chains D-Ser(*tert*butyl) in **(5)** and D-Arg in **(6)** observed for *Triple*Ser217Val227 and *Triple*Ser217*Ile227*.

It appears that the 'down'-conformation is strongly destabilized in *Triple*Ser217*Phe227* (in part corroborated by the absence of crystal form **A** for this variant), but that the intermediate helix can adopt a variety of conformations in addition to the 'up' conformation. In this variant, the weak side chain density of Trp215 (in contrast to that of Phe227) suggests that destabilization can be transmitted to the substrate binding site, providing an explanation for the high K_m for the substrate and low affinity for **(1)** – even though neither Phe227 nor Trp215 are involved in inhibitor binding – and the benzamidine-based inhibitor **(2)**. The 'up' conformation, however, can be readily accessed in this variant, allowing preferential binding of inhibitors that occupy the aromatic box.

In each of the *Triple* variants, replacement of Ser217 for Glu can facilitate a transition from the 'down' to the 'up' conformation. The effect is most pronounced for *TripleGlu217Phe227*: the cooperative effects of destabilization of the 'down' conformation by Phe227 and consolidation of the 'up' conformation by Glu217. That this variant prefers the 'up' conformation is corroborated by the differential scanning calorimetry data (Supplementary Figure 3). The inhibitor profile of this variant most closely resembles that of factor Xa (see Supplementary Figure 1), presumably due to fortification of the aromatic box. Nevertheless, it is unlikely to be a perfect model for factor Xa: substitution of the aromatic Phe227 for the aliphatic Ile227 results in small but potentially significant changes in the orientation of the Trp215 indole side chain, which may provide an additional rationalization for the remaining differences in affinity for the most potent factor Xa inhibitors. Nevertheless, a clear explanation for the marked discrepancy in affinity of the highly potent factor Xa inhibitor **(9)** for each of the variants remains elusive.

The importance of the aromatic box is most clearly seen for the closely related inhibitor pair **(5)** and **(6)**. No discrimination is seen between them for *Triple*Ser217Val227 and *Triple*Ser217*Ile227* (supporting the conclusion that these variants adopt only the 'down' conformation), but D-Arg is preferred by a factor of 3–12 over D-Ser(*tert*-butyl) in P3 for *Triple*Ser217*Phe227* and all *TripleGlu217* variants (which can all adopt the 'up' conformation), suggesting that occupation of the aromatic box by a basic group is energetically favorable for these variants. Interestingly, comparison of the structure of *Triple*Ser217*Phe227*–**(7)** with that of the closely related inhibitor **(7a)** in complex with trypsin (Whitlow et al., 1999) reveals significant differences in the binding mode (Figure 7), the most obvious being the positioning of the central indole/benzimidazole linker. Due to almost 180° dihedral rotations about each of the two bonds linking the naphthamidine and indole/benzimidazole rings, the central aromatic moiety is positioned over the Gly219 edge of the S1 pocket, so that the carboxylate group is exposed to solvent and the isopropyl group occupies the $S1_\beta$-site. This alternative location of the rigid linker is accompanied by a small but significant tilt of the naphthamidine group within the primary specificity pocket, as well as a ~3 Å repositioning of the acetimidoyl function away from the 'cation hole' and further into the aromatic box. Similar subtle differences in binding mode are observed between *Triple*Ser217*Phe227*–**(8)** and the closely related inhibitor **(8a)** in trypsin (Stubbs et al., 1995), namely a tilt of the naphthamidino group and a displacement of the basic acetimidoyl piperidine function ~2 Å further into the aromatic box, suggesting that this is due to specific differences in the binding sites of trypsin and the factor Xa-like variant. Thus cation-π interactions in the aromatic pocket seem to be preferred over hydrophobic burial, as discussed by the Diederich group (Schärer et al., 2005; Salonen et al., 2009, 2012); indeed, the guanidinium group of **(6)** and the acetimidoyl groups of **(7)** and **(8)** occupy similar positions to the tertiary amines in their structures. It therefore appears that the docking model of **(8a)** to factor Xa (Lin and

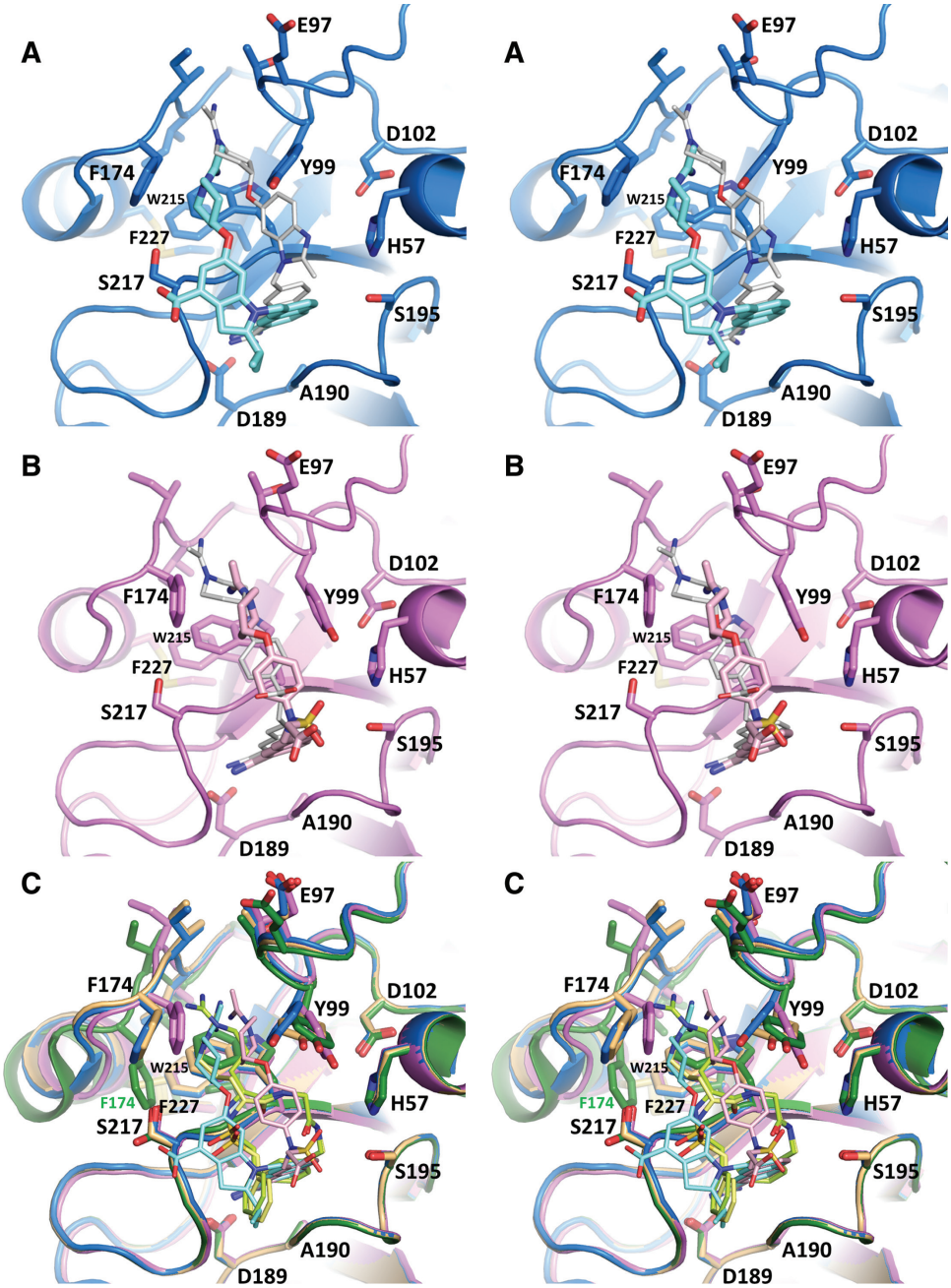


Figure 7 Comparison of 'factor Xa' and trypsin binding modes for naphthamidine-based inhibitors (stereo representations). (A) Overlay of *Triple*Ser217*Phe227*–**7** (blue/cyan) with **7a** (thin white sticks) from trypsin (Whitlow et al., 1999). Note the alternative orientations of the indole/benzimidazole spacers, the different tilts of the naphthamidine moieties, and the relative displacements of the acetimidoyl functions. (B) Overlay of *Triple*Ser217*Phe227*-**8** (violet/pink) with **8a** (DX9065a, Daiichi Corporation, Japan; thin white sticks) from trypsin (Stubbs et al., 1995). Corresponding subtle differences in binding mode are observed to those shown in (A). (C) Superposition of *Triple*Ser217*Ile227*-**6** ('down' conformation, green/light green), *Triple*Ser217*Phe227*-**6** (light orange/yellow), *Triple*Ser217*Phe227*-**7** (blue/cyan) and *Triple*Ser217*Phe227*-**8** (violet/pink). In the absence of the aromatic box (green structure), the terminal guanidinium group of **(6)** reaches out towards the bulk solvent. In each of the other three structures, the aromatic box is occupied by the respective terminal basic group (guanidinium in **6**, acetimidoyl in **7** and **8**) that we presume to be positioned in such a way as to optimize cation–π interactions. Note the alternative orientations of the Phe174 and Tyr99 side chains, which shift to accommodate the differing approaches of the inhibitors.

Johnson, 1995) is closer to the genuine binding mode than that suggested by structural studies using the surrogate trypsin (Stubbs et al., 1995; Whitlow et al., 1999).

In conclusion, the data presented here highlight many of the potential problems in correlating structural, mutational and affinity data. Even the introduction of a methyl

group (as in *Triple*Ser217*Ile227*) can have a significant influence on affinity, even though no obvious changes in structure or binding mode are observed. Destabilization of the target enzyme (as seen for *Triple*Ser217*Phe227*) has a more predictable deleterious effect on ligand binding, whereas the availability of alternative states (as for example with *TripleGlu217*Val227 and *TripleGlu217Ile227*) alters the inhibitory profile in a ligand-dependent fashion. A combination of destabilization of the 'down' and stabilization of the 'up' conformations in variant *TripleGlu217Phe227* results in a chimeric enzyme that appears to mimic the structural properties of factor Xa; nevertheless, there remain significant discrepancies in affinity for the high potency factor Xa inhibitors, suggesting that factor Xa itself possesses very different properties to these chimeric proteinases and probably harbors a rigid substrate binding site. On the other hand, plasticity in the neighborhood of the intermediate helix has been reported in other trypsin-like serine proteinases, where it has been linked to cofactor dependent activity modulation (see Reyda et al., 2003).

Obviously, our understanding of the binding process remains incomplete. A better understanding of the type of protein–ligand interactions observed here – especially those involving conformational transitions – may ultimately permit factorization of the multitudinous competing processes that contribute to affinity, in turn facilitating a fine-tuning of computational methods in drug design optimization.

Materials and methods

Protein preparation and purification

Desired mutants were designed by site directed mutagenesis using the Quick-Change™ kit (Weiner et al., 1994) and a pET-vector system. Mutations were introduced using pairs of complementary primers (Ile227: 5′-CT CAC GTA GTT GCA GAC CTT TGT GTA GAT ACC AGG CTT G-3′ sense, 5′-C AAG CCT GGT ATC TAC ACA AAG GTC TGC AAC TAC GTG AG-3′ anti-sense; Phe227: 5′-CT CAC GTA GTT GCA GAG CTT AGT GTA GAA ACC AGG CTT G-3′ sense, 5′-C AAG CCT GGT TTC TAC ACT AAG CTC TGC AAC TAC GTG AG-3′ anti-sense). For calorimetric experiments, it was necessary to use proteolytically inactive variants; the active site Ser195 in each of the variants was mutated to Ala using the primers 5′-CAC AGG GCC ACC TGC GTC ACC CTG GC-3′ (forward) and 5′-GC CAG GGT GAC GCA GGT GGC CCT GTG-3′ (reverse). *Escherichia coli* BL21 (DE3) strain was transformed with the mutated DNA and the protein expressed in zymogen form. The accumulated inclusion bodies were isolated and folded using established procedures (Rauh et al., 2002). After activation with enterokinase (Roche Diagnostics, Penzberg, Germany), the trypsin variants were purified by affinity chromatography on soybean trypsin inhibitor (Hedstrom et al., 1996) covalently bound to agarose (Sigma-Aldrich, Steinheim, Germany) as described in Rauh et al. (2002). The yields of active proteinase ranged from 0.4 mg/l cultivation media (*TripleGlu217Phe227*) to 2.3 mg/l (*Triple*Ser217*Phe227*).

K_i-value and kinetic characterization

Inhibitor **(1)** was purchased from Sigma and inhibitor **(4)** was kindly provided by Hans-Dieter Gerber, Marburg. All other inhibitors were provided courtesy of Jörg Stürzebecher, Erfurt and Curacyte Chemistry GmbH, Jena, Germany. All substrates were supplied by Pentapharm Ltd., Basel, Swizerland. K_m-, k_{cat}- and K_i- values were determined photometrically using the chromogenic substrate methylsulfonyl-D-cyclohexylalanyl-glycyl-arginyl-*p*-nitroanilide (Pefachrome®tPA) and the fluorogenic substrate methylumbelliferyl-*p*-guanidinobenzoate (for active site titration) as described by Stürzebecher et al. (1997). K_i-values were determined in a microplate reader (VersaMax, Molecular Devices, Biberach, Germany) at 405 nm at 25°C under the following conditions: 0.05 M Tris-HCl (pH 8.0), 0.154 M NaCl and 0.01 M $CaCl_2$ and 5% (v/v) ethanol at 25°C and obtained from Dixon plots (Dixon, 1972).

Crystallization, data collection, structure determination and refinement

Variants were incubated at a concentration of 20 mg/ml with inhibitor solutions (1–10 mg/ml) for 30 min and screened for crystallization using 12 conditions [0.1, 0.2 or 0.3 M ammonium sulfate; 0.1 M imidazole at pH 7.0 or 8.0; and 20% or 30% (w/v) PEG 8000]. Crystals grew at room temperature within one to 21 days in a variety of crystal forms and space groups (Table 1) and were measured under cryogenic conditions at -180°C in-house using an R-AxisIV^{++} image plate system (MSC, TX, USA) installed on a Rigaku rotating anode generator or at the synchrotron beamlines at DESY (Hamburg) or BESSY (Berlin). Data were processed and scaled using the HKL package (Otwinowski and Minor, 1997) or Crystal Clear™ 1.3.6 (MSC, TX, USA).

Conventional crystallographic molecular replacement and refinement were performed with CCP4 (Bailey, 1994) or CNS (Brunger et al., 1998) programs; starting coordinates were taken from Rauh et al. (2004). Model building was carried out using O (Jones et al., 1991) or COOT (Emsley and Cowtan, 2004). Data collection and refinement statistics are given in Supplementary Table 1. The figures were created in PyMol v0.98 (DeLano Scientific, San Carlos, CA, USA).

Differential scanning microcalorimetry

Experiments were performed using a VP-differential scanning microcalorimetry instrument (Microcal, LLC, Northampton, MA, USA). Extensive buffer screening was performed using the inactive variant *TripleGlu217*Val227*Ala195* to establish conditions for reversible thermal unfolding transitions; unfortunately, these conditions were not transferable to any of the other variants, so that only the apparent melting temperatures T_m^{app} could be measured (Supplementary Figure 3). Variants at a concentration of 3–35 μM were dialyzed against 5 mM Borat/NaOH pH 8.0 with 1 mM $CaCl_2$. The degassed dialysis buffer was used to fill the instrument and to record an instrument baseline. Temperature scans of the degassed samples were run from

20°C to 80°C with a scan rate of 1.5 K min^{-1}. At least two scans were made for each variant to ascertain whether unfolding was reversible.

Acknowledgments: We thank Diana Lieber for assistance with protein preparation and crystallization. Hans-Dieter Gerber (Marburg) and Curacyte Chemistry GmbH (Jena) kindly provided additional inhibitors. We also thank Gerhard Klebe (Marburg) and Norbert Sträter (Leipzig) for use of their X-ray facilities during the initial stages of this project. The diffraction data for structures *Triple*Ser217*Ile227*.A1 and *TripleGlu217Phe227*.F1 variants were measured on the MPG/GBF beamline BW6 at DESY, while the diffraction experiments for *Triple*Ser217*Ile227*.A4 and *TripleGlu217Phe227*.B2 variants were carried out at the Protein Structure Factory beamline BL14.1 of BESSY and Free University Berlin at BESSY. This work was supported in part by the DFG Graduiertenkolleg 1026 'Conformational transitions in macromolecular interactions' to MTS.

References

Arnaiz, D.O., Zhao, Z., Liang, A., Trinh, L., Witlow, M., Koovakkat, S.K., and Shaw, K.J. (2000). Design, synthesis, and *in vitro* biological activity of indole-based factor Xa inhibitors. Bioorg. Med. Chem. Lett. *10*, 957–961.

Bailey, S. (1994). The CCP4 Suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Brandstetter, H., Kühne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K., and Engh, R.A. (1996). X-ray structure of active site-inhibited clotting factor Xa – implications for drug design and substrate recognition. J. Biol. Chem. *271*, 29988–29992.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Dixon, M. (1972). Graphical determination of K_m and K_i. Biochem. J. *129*, 197.

Dullweber, F., Stubbs, M.T., Musil, D., Stürzebecher, J., and Klebe, G. (2001). Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. J. Mol. Biol. *313*, 593–614.

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

Faull, A.W., Mayo, C.M., Preston, J., and Stocker, A. (1996). Aminoheterocyclic derivatives as antithrombotic or anticoagulant agents. (Patent, WO, 9610022).

Feixas, F., Lindert, S., Sinko, W., and McCammon, J.A. (2014). Exploring the role of receptor flexibility in structure-based drug discovery. Biophys. Chem. *186*, 31–45.

Hedstrom, L., Lin, T.Y., and Fast, W. (1996). Hydrophobic interactions control zymogen activation in the trypsin family of serine proteases. Biochemistry *35*, 4515–4523.

Hirayama, F., Koshio, H., Katayama, N., Kurihara, H., Taniuchi, Y., Sato, K., Hisamichi, N., Sakai-Moritani, Y., Kawasaki, T., Matsumoto, Y., et al. (2002). The discovery of YM-60828: a potent, selective and orally-bioavailable Factor Xa inhibitor. Bioorg. Med. Chem. *10*, 1509–1523.

Huang, S.Y., Grinter, S.Z., and Zou, X. (2010). Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. Phys. Chem. Chem. Phys. *12*, 12899–12908.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. Acta Crystallog. Sect. A *47*, 110–119.

Lin, Z. and Johnson, M.E. (1995). Proposed cation-π mediated binding by factor Xa: a novel enzymatic mechanism for molecular recognition. FEBS Lett. *370*, 1–5.

Martin, S.F. and Clements, J.H. (2013). Correlating structure and energetics in protein-ligand interactions: paradigms and paradoxes. Annu. Rev. Biochem. *82*, 267–293.

Nar, H. (2012). The role of structural information in the discovery of direct thrombin and factor Xa inhibitors. Trends Pharmacol. Sci. *33*, 279–288.

Otwinowski, Z. and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Macromol. Crystallog. A *276*, 307–326.

Perzborn, E., Roehrig, S., Straub, A., Kubitza, D., and Misselwitz, F. (2011). The discovery and development of rivaroxaban, an oral, direct factor Xa inhibitor. Nat. Rev. Drug Discov. *10*, 61–75.

Pruitt, J.R., Pinto, D.J., Estrella, M.J., Bostrom, L.L., Knabb, R.M., Wong, P.C., Wright, M.R., and Wexler, R.R. (2000). Isoxazolines and isoxazoles as Factor Xa inhibitors. Bioorg. Med. Chem. Lett. *10*, 685–689.

Rauh, D., Reyda, S., Klebe, G., and Stubbs, M.T. (2002). Trypsin mutants for structure-based drug design: expression, refolding and crystallisation. Biol. Chem. *383*, 1309–1314.

Rauh, D., Klebe, G., Stürzebecher, J., and Stubbs, M.T. (2003). ZZ made EZ: influence of inhibitor configuration on enzyme selectivity. J. Mol. Biol. *330*, 761–770.

Rauh, D., Klebe, G., and Stubbs, M.T. (2004). Understanding protein-ligand interactions: the price of protein flexibility. J. Mol. Biol. *335*, 1325–1341.

Renatus, M., Bode, W., Huber, R., Stürzebecher, J., and Stubbs, M.T. (1998). Structural and functional analyses of benzamidine-based inhibitors in complex with trypsin: implications for the inhibition of factor Xa, tPA, and urokinase. J. Med. Chem. *41*, 5445–5456.

Reyda, S., Sohn, C., Klebe, G., Rall, K., Ullmann, D., Jakubke, H.D., and Stubbs, M.T. (2003). Reconstructing the binding site of factor Xa in trypsin reveals ligand-induced structural plasticity. J. Mol. Biol. *325*, 963–977.

Rose, P.W., Bi, C., Bluhm, W.F., Christie, C.H., Dimitropoulos, D., Dutta, S., Green, R.K., Goodsell, D.S., Prlic, A., Quesada, M., et al. (2013). The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res. *41*, D475–D482.

Salonen, L.M., Bucher, C., Banner, D.W., Haap, W., Mary, J.L., Benz, J., Kuster, O., Seiler, P., Schweizer, W.B., and Diederich, F. (2009). Cation-π interactions at the active site of factor Xa: dramatic enhancement upon stepwise N-alkylation of ammonium ions. Angew. Chem. Int. Ed. *48*, 811–814.

Salonen, L.M., Holland, M.C., Kaib, P.S., Haap, W., Benz, J., Mary, J.L., Kuster, O., Schweizer, W.B., Banner, D.W., and Diederich,

F. (2012). Molecular recognition at the active site of factor Xa: cation-pi interactions, stacking on planar peptide surfaces, and replacement of structural water. Chemistry *18*, 213–222.

Schärer, K., Morgenthaler, M., Paulini, R., Obst-Sander, U., Banner, D.W., Schlatter, D., Benz, J., Stihle, M., and Diederich, F. (2005). Quantification of cation-pi interactions in protein-ligand complexes: crystal-structure analysis of Factor Xa bound to a quaternary ammonium ion ligand. Angew. Chem Int. Ed. *44*, 4400–4404.

Schweinitz, A., Stürzebecher, A., Stürzebecher, U., Schuster, O., Stürzebecher, J., and Steinmetzer, T. (2006). New substrate analogue inhibitors of factor Xa containing 4-amidinobenzylamide as P1 residue: part 1. Med. Chem. *2*, 349–361.

Shotton, D.M. and Hartley, B.S. (1970). Amino-acid sequence of porcine pancreatic elastase and its homologies with other serine proteinases. Nature *225*, 802–806.

Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E.W., Jr. (2014). Computational methods in drug discovery. Pharmacol. Rev. *66*, 334–395.

Steinberg, B.A. and Becker, R.C. (2013). Structure-function relationships of factor Xa inhibitors: implications for the practicing clinician. J Thromb. Thrombolysis. *37*, 234–241.

Straub, A., Roehrig, S., and Hillisch, A. (2011). Oral, direct thrombin and factor Xa inhibitors: the replacement for warfarin, leeches, and pig intestines? Angew. Chem. Int. Ed. *50*, 4574–4590.

Stubbs, M.T., Huber, R., and Bode, W. (1995). Crystal-structures of factor Xa specific inhibitors in complex with trypsin – structural grounds for inhibition of factor Xa and selectivity against thrombin. FEBS Lett. *375*, 103–107.

Stubbs, M.T., Reyda, S., Dullweber, F., Möller, M., Klebe, G., Dorsch, D., Mederski, W.W.K.R., and Wurziger, H. (2002). pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. Chembiochem *3*, 246–249.

Stürzebecher, J., Stürzebecher, U., Vieweg, H., Wagner, G., Hauptmann, J., and Markwardt, F. (1989). Synthetic inhibitors of bovine factor Xa and thrombin comparison of their anticoagulant efficiency. Thromb. Res. *54*, 245–252.

Stürzebecher, J., Prasa, D., Hauptmann, J., Vieweg, H., and Wikström, P. (1997). Synthesis and structure-activity relationships of potent thrombin inhibitors: piperazides of 3-amidinophenylalanine. J. Med. Chem. *40*, 3091–3099.

Stürzebecher, A., Dönnecke, D., Schweinitz, A., Schuster, O., Steinmetzer, P., Stürzebecher, U., Kotthaus, J., Clement, B., Stürzebecher, J., and Steinmetzer, T. (2007). Highly potent and selective substrate analogue factor Xa inhibitors containing D-homophenylalanine analogues as P3 residue: part 2. ChemMedChem. *2*, 1043–1053.

Teague, S.J. (2003). Implications of protein flexibility for drug discovery. Nat. Rev. Drug Discov. *2*, 527–541.

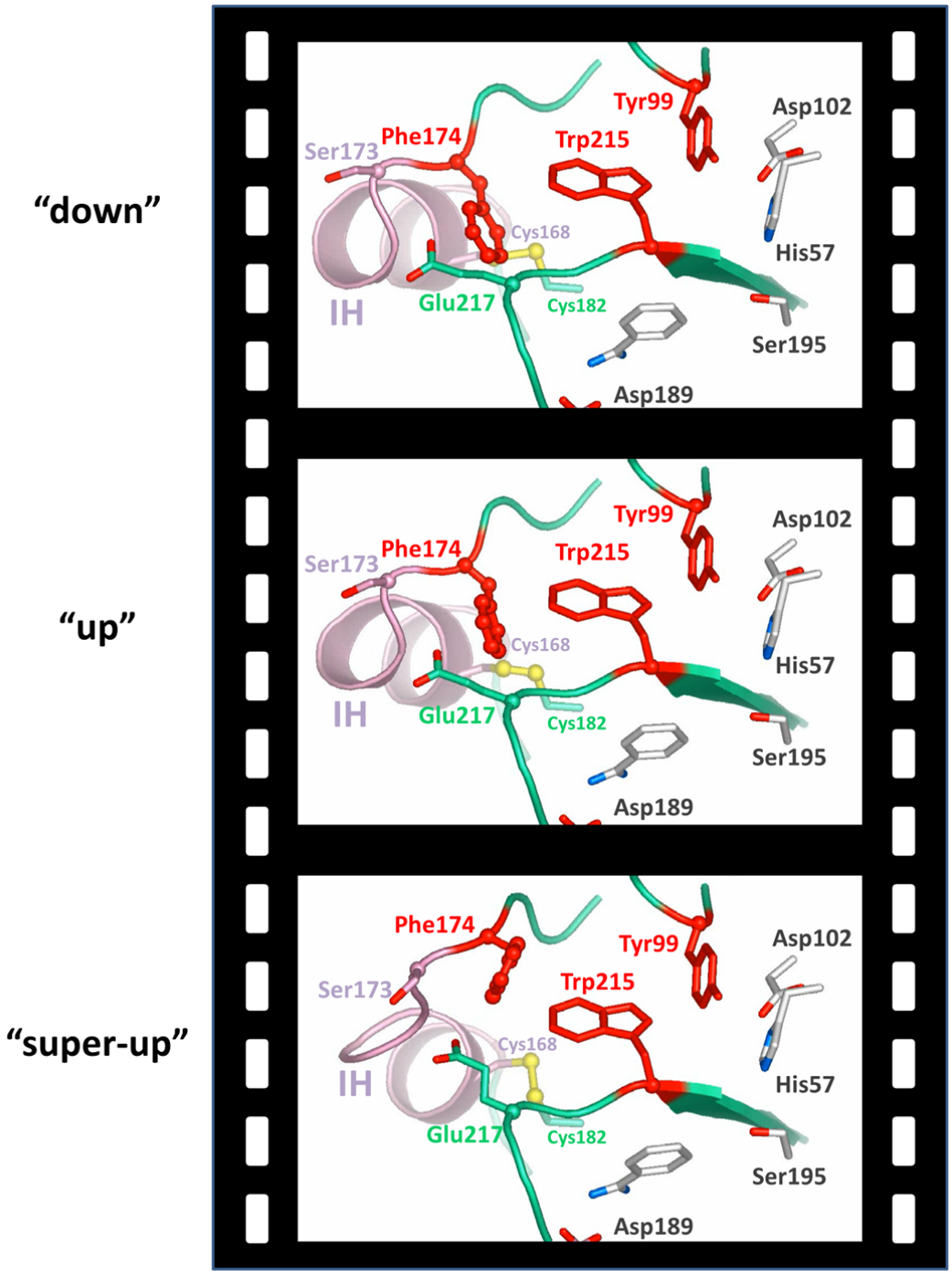
Turk, D., Stürzebecher, J., and Bode, W. (1991). Geometry of binding of the Nα-tosylated piperidides of *m*-amidino-, *p*-amidino- and *p*-guanidino phenylalanine to thrombin and trypsin. X-ray crystal structures of their trypsin complexes and modeling of their thrombin complexes. FEBS Lett. *287*, 133–138.

Weiner, M.P., Costa, G.L., Schoettlin, W., Cline, J., Mathur, E., and Bauer, J.C. (1994). Site-directed mutagenesis of double-stranded DNA by the polymerase chain-reaction. Gene *151*, 119–123.

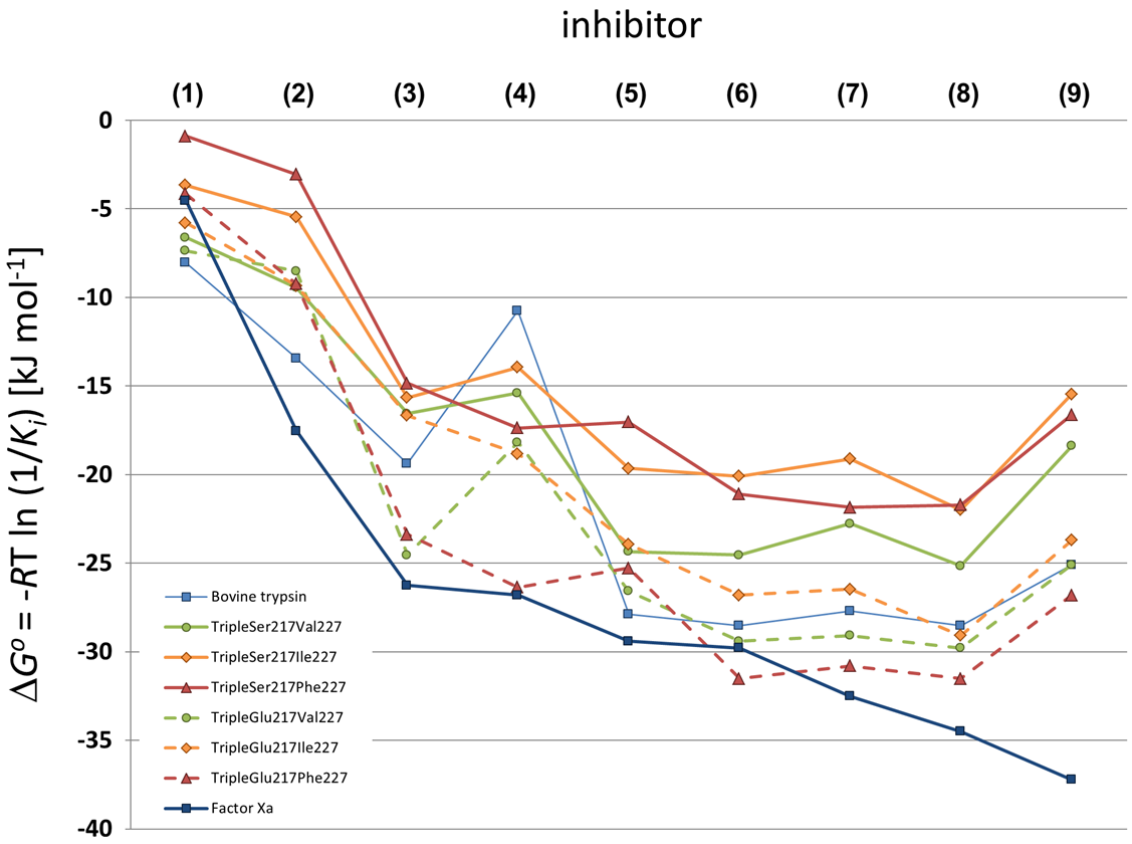
Whitlow, M., Arnaiz, D.O., Buckman, B.O., Davey, D.D., Griedel, B., Guilford, W.J., Koovakkat, S.K., Liang, A., Mohan, R., Phillips, G.B., et al. (1999). Crystallographic analysis of potent and selective factor Xa inhibitors complexed to bovine trypsin. Acta Crystallogr. D Biol. Crystallogr. *55*, 1395–1404.

Yeh, C.H., Fredenburgh, J.C., and Weitz, J.I. (2012). Oral direct factor Xa inhibitors. Circ. Res. *111*, 1069–1078.

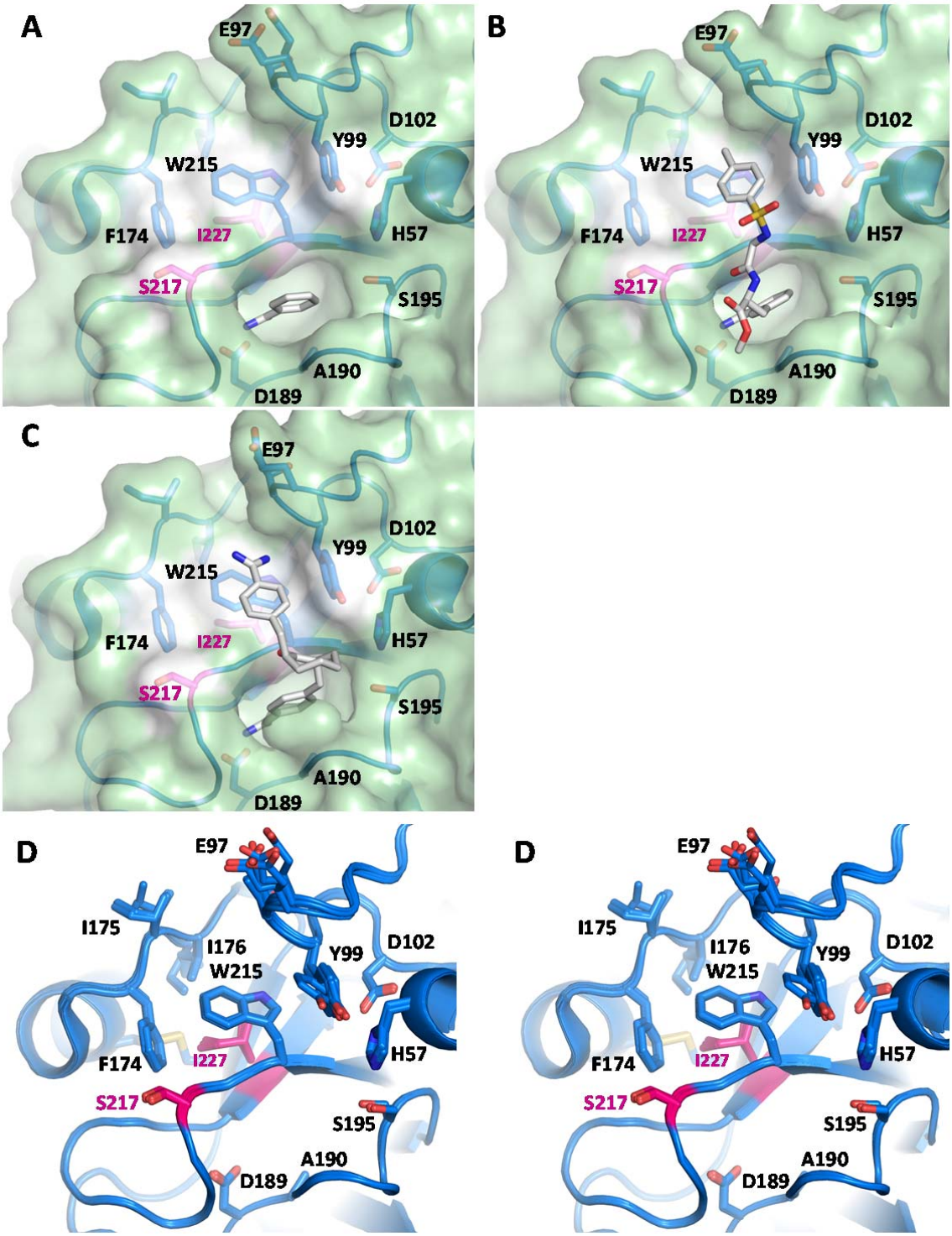
Supplemental Material: The online version of this article (DOI 10.1515/hsz-2014-0158) offers supplementary material, available to authorized users.



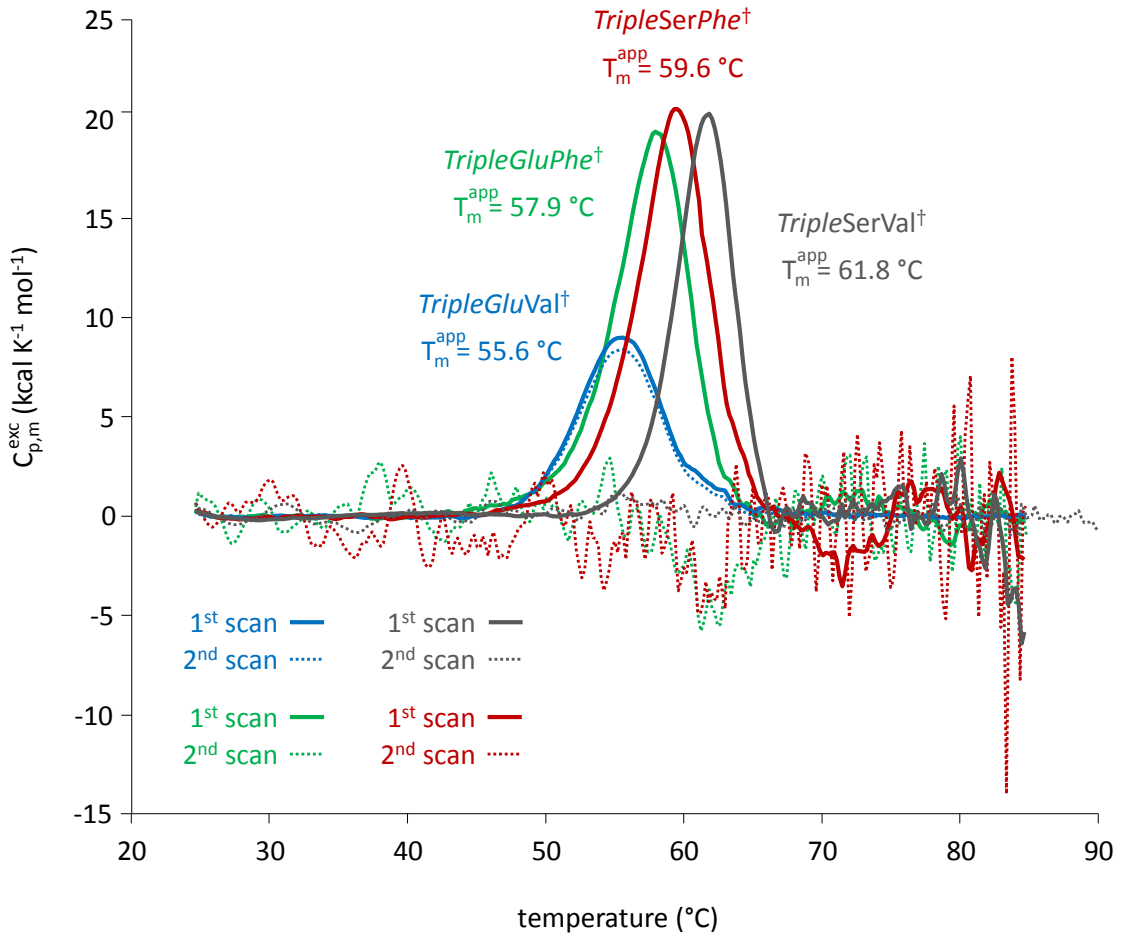
Supplementary Movie 1: Morphing between 'down' and 'super-up' conformations of the factor Xa-like bovine trypsin variant *TripleGlu217*Val227 in complex with benzamidine **(1)** *via* the 'up' conformation (Rauh et al., 2004) demonstrates the changes in position of the side chain of Phe174 of the aromatic box (red), as well as in the intermediate helix and the disulphide bond Cys168 – Cys182.



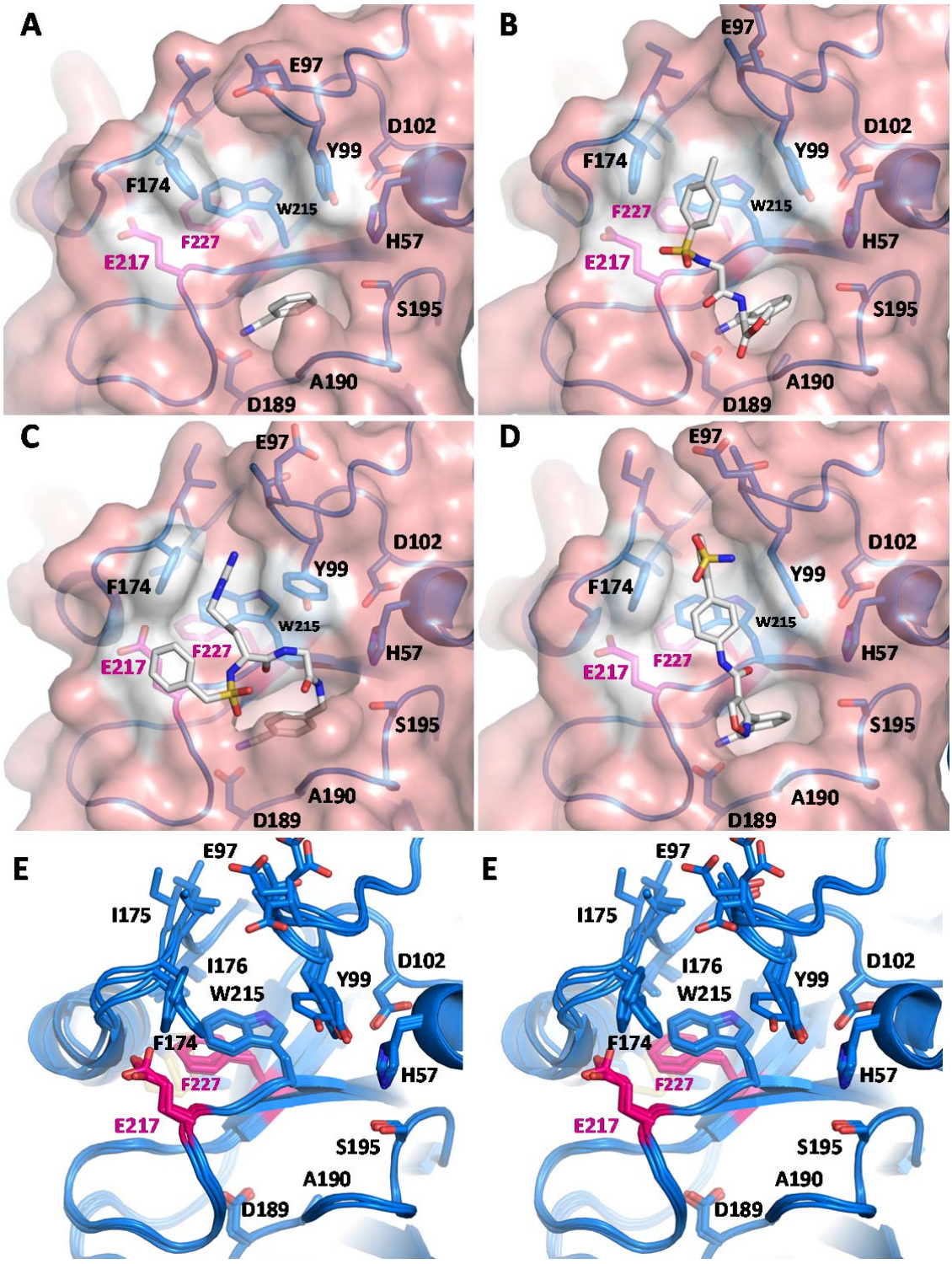
Supplementary Figure 1: Assuming K_i can be equated to the dissociation constant K_D (Dullweber et al., 2001), the inhibition constants of the compounds in this study can be converted into standard Gibbs free energies of binding ΔG^o using the equation $\Delta G^0 = -RT \ln (1/K_D)$, where R is the universal gas constant and T the absolute temperature. A free energy difference $\Delta\Delta G^o$ of 1.7 kJ/mol represents a two-fold difference in K_i at 298K, while a difference of 5.7 kJ/mol corresponds to a ten-fold difference in K_i.



Supplementary Figure 2: Binding modes of inhibitors **(1)**, **(2)** and **(3)** in complex with *Triple*Ser217*Ile227* (**A, B** and **C** respectively) are as those found in *Triple*Ser217Val227 in crystal form A (Rauh et al., 2004). Note that *Triple*Ser217*Ile227* is found solely in the 'down' conformation. **(D)** Residues Ile175 and Ile176 show a variability in position in each of the structures (which may be due to the replacement Val227Ile), as do the side chains of Glu97-Thr98-Tyr99 (which is most likely due to contacts with the inhibitors).



Supplementary Figure 3: Thermal unfolding transitions of the inactive (denoted by the † superscript) variants *Triple*Ser217Val227*Ala195*, *Triple*Ser217*Phe227Ala195*, *TripleGlu217*Val227*Ala195* and *TripleGlu217Phe227Ala195*, measured using differential scanning microcalorimetry. Under the elucidated buffer conditions, reversible thermal unfolding was observed for variant *TripleGlu217*Val227*Ala195* only (demonstrated by the overlap of the first and second scans, full and dotted lines respectively). *Triple*Ser217*Phe227Ala195* shows a 2.2 °C lower T_m^{app} compared to *Triple*Ser217Val227*Ala195*, whereas both *Glu217* variants exhibit a significant reduction in thermal stability. In contrast to the Ser217-containing variants however, the T_m^{app} of *TripleGlu217Phe227Ala195* is 2.3 °C higher than that of the corresponding Val227-containing variant, reflecting the plasticity of *TripleGlu217*Val227*Ala195* compared to the 'up' conformation of *TripleGlu217Phe227Ala195*.



Supplementary Figure 4: Binding modes of inhibitors **(1)**, **(2)**, **(6)** and **(9)** in complex with *TripleGlu217Phe227* (**A, B, C** and **D** respectively) are as those found in *Triple*Ser217*Phe227* in the corresponding crystal forms. Note that the P4 benzene ring of **(6)** is not well defined in **(C)**. The variant *TripleGlu217Phe227* is only observed in the 'up' conformation, although there remains significant structural variability in the ligand binding site **(E)**.

Supplementary Table 1: Data collection and refinement statistics

Structure[a]	crystal form	space group	cell constants a α	b β	c γ	observed / unique reflections	R_{symm} [b] (%)	resolution (Å)	non-hydrogen atoms per asymmetric unit: Protein	Ions	Solvent	Inhibitor	rms deviation: Bond(Å)	Angle(deg.)	R_{work} (%)	R_{free} (%)	PDB
*Triple*Ser217*Ile227*.A1	A	$P3_121$	54.7 90	54.8 90	107.9 120	140865/ 62088	4.7 (44.0)	99.0-1.18 1.20-1.18	1633	1 Ca^{2+}; 1 SO_4^{2-}	219	9	0.0100	1.569	20.7	22.7	3PLB
*Triple*Ser217*Ile227*.A2	A	$P3_121$	54.6 90	54.6 90	109.0 120	344918/ 29785	4.5 (12.9)	50.0-1.52 1.57-1.52	1633	1 Ca^{2+}; 1 SO_4^{2-}	227	30	0.0090	1.290	15.9	18.0	3PLK
*Triple*Ser*217Ile227*.A3	A	$P3_121$	55.1 90	55.1 90	108.9 120	210585/ 24576	6.9 (26.9)	28.9-1.63 1.69-1.63	1633	1 Ca^{2+}	208	28	0.0110	1.410	17.8	21.8	3PLP
*Triple*Ser217*Ile227*.A5	A	$P3_121$	54.5 90	54.5 90	111.7 120	350479/ 30348	3.0 (6.5)	50.0-1.52 1.57-1.52	1633	1 Ca^{2+}	246	35	0.0130	1.500	16.6	19.5	3PM3
*Triple*Ser217*Ile227*.A6	A	$P3_121$	54.7 90	54.7 90	112.0 120	189250/ 32308	6.8 (31.9)	21.8-1.45 1.50-1.45	1633	1 Ca^{2+}	153	35	0.0095	1.579	22.9	25.7	3PMJ
TripleGlu217Ile227.A1	A	$P3_121$	54.6 90	54.6 90	106.5 120	167485/ 23624	10.7 (42.9)	19.7-1.63 1.69-1.63	1633	1 Ca^{2+}; 1 SO_4^{2-}	197	9	0.0088	1.546	19.9	24.4	3PWB
TripleGlu217Ile227.A2	A	$P3_121$	54.4 90	54.4 90	109.1 120	131550/ 24403	5.8 (24.6)	28.8-1.60 1.66-1.60	1633	1 Ca^{2+}; 1 SO_4^{2-}	184	30	0.0099	1.615	19.5	22.8	3PWC
TripleGlu217Ile227.D6	D	$P3_221$	54.0 90	54.0 90	137.1 120	172054/ 26369	10.2 (38.5)	50.0-1.70 1.75-1.70	1633	1 Ca^{2+}; 1 SO_4^{2-}	102	35	0.0092	1.62	22.4	25.1	3Q00
TripleGlu217Ile227.E9	E	$P6_5$	81.7 90	81.7 90	63.6 120	148593/ 16366	16.8 (39.8)	19.6-2.00 2.07-2.00	1633	1 Ca^{2+}; 1 SO_4^{2-}	126	33	0.0080	1.501	19.6	23.9	3PYH
*Triple*Ser217*Phe227*.C1	C	$P2_12_12_1$	57.9 90	64.1 90	70.4 90	417418/ 48215	6.7 (39.0)	50.0-1.62 1.71-1.62	1638	1 Ca^{2+}; 1 SO_4^{2-}	170	9	0.0112	1.706	20.6	22.2	3UNQ
*Triple*Ser217*Phe227*.B2	B	$P2_12_12_1$	54.3 90	57.2 90	66.9 90	112316/ 19879	8.3 (41.4)	24.5-1.80 1.86-1.80	1637	1 Ca^{2+}; 1 SO_4^{2-}	105	30	0.0092	1.567	21.6	23.7	3UNS
*Triple*Ser217*Phe227*.F6	F	$P3_221$	55.1 90	55.1 90	154.9 120	156002/ 28059	8.6 (28.2)	35.0-1.69 1.75-1.69	1637	1 Ca^{2+}; 1 SO_4^{2-}	123	35	0.0096	1.555	22.8	24.7	3UOP
*Triple*Ser217*Phe227*.F7	F	$P3_221$	54.6 90	54.6 90	154.9 120	168812/ 40804	5.0 (19.5)	47.3-1.54 1.59-1.54	1638	1 Ca^{2+}	205	39	0.0090	2.250	21.3	23.9	3UPE
*Triple*Ser217*Phe227*.G8	G	$C222_1$	74.5 90	77.8 90	171.8 90	532407/ 29135	8.0 (14.7)	30.0-2.10 2.23-2.10	3276	2 Ca^{2+}	237	74	0.0150	1.631	17.1	21.2	3UUZ
*Triple*Ser217*Phe227*.E9	E	$P6_5$	84.5 90	84.5 90	51.4 120	80694/ 10755	16.7 (38.8)	29.8-2.40 2.49-2.40	1637	1 Ca^{2+}; 1 SO_4^{2-}	87	33	0.0099	1.615	18.9	25.8	3UQV
*Triple*Ser*217Phe227*.H9	H	$P3_221$	54.7 90	54.7 90	113.0 120	107032/ 18782	9.5 (36.5)	27.6-1.80 1.86-1.80	1637	1 Ca^{2+}; 1 SO_4^{2-}	86	33	0.0102	1.580	24.1	27.3	3UQO
TripleGlu217Phe227.I1	I	P1	88.7 89.9	98.6 88.9	120.8 104.0	370469/ 78880	11.3 (39.2)	50.0-2.90 3.04-2.90	26240	1 Ca^{2+}; 1 SO_4^{2-}	n.d.	144	0.008	1.400	25.0	27.5	3UY9
TripleGlu217Phe227.B2	B	$P2_12_12_1$	53.6 90	57.3 90	65.7 90	434909/ 38823	4.4 (21.7)	50.0-1.42 1.47-1.42	1640	1 Ca^{2+}	175	30	0.0110	1.450	18.6	20.4	3UWI
TripleGlu217Phe227.D6	D	$P3_221$	54.3 90	54.3 90	136.9 120	146771/ 22494	9.5 (36.1)	34.2-1.80 1.86-1.80	1640	1 Ca^{2+}; 1 SO_4^{2-}	148	35	0.0095	1.609	21.1	24.1	3V12
TripleGlu217Phe227.E9	E	$P6_5$	81.8 90	81.8 90	63.9 120	270243/ 30439	11.8 (39.5)	35.4-1.63 1.69-1.63	1640	1 Ca^{2+}; 1 SO_4^{2-}	212	33	0.0100	1.638	18.8	20.4	3V13

[a] Structures are distinguished as follows: {variant}.{crystal form} {inhibitor}; inhibitors are shown in Figure 1. Crystal forms A-D are the designations as described in (Rauh et al., 2004).

[b] $R_{sym} = \sum(|I-\langle I\rangle|)/\sum(I)$; values in parentheses are for the highest resolution shell

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN: 2053-230X
journals.iucr.org/f

Phosphate binding in the active centre of tomato multifunctional nuclease TBN1 and analysis of superhelix formation by the enzyme

Jan Stránský, Tomáš Koval', Tomáš Podzimek, Anna Týcová, Petra Lipovová, Jaroslav Matoušek, Petr Kolenko, Karla Fejfarová, Jarmila Dušková, Tereza Skálová, Jindřich Hašek and Jan Dohnálek

Acta Cryst. (2015). F**71**, 1408–1415

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Phosphate binding in the active centre of tomato multifunctional nuclease TBN1 and analysis of superhelix formation by the enzyme

Jan Stránský,[a,b]* Tomáš Koval',[a,c] Tomáš Podzimek,[d] Anna Týcová,[e,f] Petra Lipovová,[d] Jaroslav Matoušek,[e] Petr Kolenko,[a,b,c] Karla Fejfarová,[a,c] Jarmila Dušková,[a] Tereza Skálová,[a] Jindřich Hašek[a] and Jan Dohnálek[a,c]*

[a]Institute of Biotechnology CAS, v.v.i., Vídeňská 1083, 142 20 Praha 4, Czech Republic, [b]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Břehová 7, 115 19 Praha 1, Czech Republic, [c]Institute of Macromolecular Chemistry CAS, v.v.i., Heyrovského nám. 2, 162 06 Praha 6, Czech Republic, [d]University of Chemistry and Technology, Technická 5, 166 28 Praha 6, Czech Republic, [e]Institute of Plant Molecular Biology, Biology Centre, CAS, v.v.i., Branišovská 31, 370 05 České Budějovice, Czech Republic, and [f]Faculty of Science, University of South Bohemia in České Budějovice, Branišovská 1760, 370 05 České Budějovice, Czech Republic. *Correspondence e-mail: stransky@ibt.cas.cz, dohnalek@ibt.cas.cz

Received 15 July 2015
Accepted 30 September 2015

Keywords: tomato multifunctional nuclease; TBN1; type I nuclease; superhelix.

PDB reference: tomato multifunctional nuclease, N211D mutant, 4jdg

Supporting information: this article has supporting information at journals.iucr.org/f

Tomato multifunctional nuclease TBN1 belongs to the type I nuclease family, which plays an important role in apoptotic processes and cell senescence in plants. The newly solved structure of the N211D mutant is reported. Although the main crystal-packing motif (the formation of superhelices) is conserved, the details differ among the known structures. A phosphate ion was localized in the active site of the enzyme. The binding of the surface loop to the active centre is stabilized by the phosphate ion, which correlates with the observed aggregation of TBN1 in phosphate buffer. The conserved binding of the surface loop to the active centre suggests biological relevance of the contact in a regulatory function or in the formation of oligomers.

1. Introduction

Tomato multifunctional nuclease 1 (TBN1; UniProt accession No. Q0KFV0) is a 31.6 kDa, mainly α-helical glycoprotein which plays an important role in specific apoptotic functions, vascular-system development, stress response and tissue differentiation in plants (Matoušek *et al.*, 2007). Furthermore, TBN1 exhibits anticancerogenic properties (Matoušek *et al.*, 2008, 2010). The enzyme possesses endonuclease-like and exonuclease-like activity on double-stranded (ds) and single-stranded (ss) RNA and DNA and on structured RNA, with the production of 5′-mononucleotides and oligonucleotides, and specific 3′-nucleotidase activity (Podzimek *et al.*, 2011).

The structures of wild-type TBN1 (PDB entry 3sng) and the N211D mutant (PDB entry 4dj4) have recently been solved and a reaction mechanism has been suggested (Koval' *et al.*, 2013; Dohnálek *et al.*, 2011). The structural similarity to phospholipase C from *Bacillus cereus* (Antikainen *et al.*, 2003) led to our discovery of the phospholipase activity of TBN1 (Koval' *et al.*, 2013).

TBN1 is related to fungal P1 nucleases. The enzymes share a fold and active centre. In contrast to the multifunctionality of TBN1, P1 nuclease cleaves only single-stranded DNA and RNA (Romier *et al.*, 1998). The major differences between the P1 and TBN1 structures are in the overall shape and the electrostatic potential distribution (Koval' *et al.*, 2013).

As part of our effort to confirm the reaction mechanism and the binding modes of products and substrates, TBN1 was co-crystallized with 5′-mononucleotides. We solved a crystal

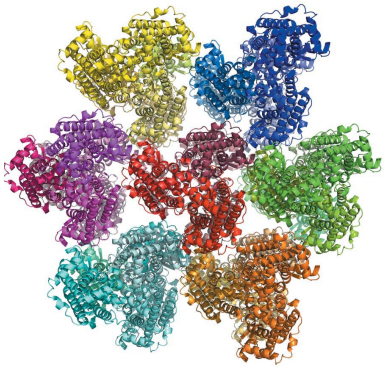


Table 1
Enzyme expression.

Source organism	*Solanum lycopersicum*
Forward primer	5′-GACATTGAAGGGGACTTCACTGACGGAATTT-GG-3′
Reverse primer	5′-CCAAATTCCGTCAGTGAAGTCCCCTTCAATG-TC-3′
Expression vector	pLV07
Expression host	*N. benthamiana*
Complete amino-acid sequence of the construct produced	WSKEGHVMTCRIAQGLLNDEAAHAVKMLLPEYVN-GDLSALCVWPDQVRHWYKYKWTSPLHFIDTPD-KACNFDYERDCHDQHGVKDMCVAGAIQNFTTQ-LSHYREGTSDRRYNMTEALLFLSHFMGDIHQP-MHVGFTSDAGGNSIDLRWFRHKSNLHHVWDRE-IILTAAKDYYAKDINLLEEDIEGNFTDGIWSD-DLASWRECGNVFSCVNKFATESINIACKWGYK-GVEAGETLSDDYFNSRLPIVMKRVAQQGIRLA-MLLNNVFGASQQEDSVV

Table 2
Enzyme crystallization.

Method	Vapour diffusion, hanging drop
Plate type	Qiagen EasyXtal 24-well
Temperature (K)	291
Protein concentration (mg ml^{-1})	2
Buffer composition of protein solution	10 m*M* 5′-AMP, 50 m*M* Tris–HCl pH 7.5, 0.3 *M* NaCl, 1%(*v*/*v*) glycerol, 10 m*M* 3-(1-pyridino)-1-propanesulfonate (NDSB-201)
Composition of reservoir solution	0.1 *M* bis-tris pH 6.5, 10%(*w*/*v*) PEG 8000
Volume and ratio of drop	1 µl, 1:1
Volume of reservoir (µl)	500

structure of the N211D mutant (N-glycosylation site removed) from a different crystallization condition in a new crystal form. Mononucleotide was not observed in the structure; however, an unexpected phosphate ion was localized in the active centre. The change in the crystallization conditions and the presence of phosphate ion did not change the packing of TBN1 in the crystals.

2. Materials and methods

2.1. Protein preparation

For crystallization and diffraction experiments, a recombinant N211D mutant variant of tomato plant nuclease type I (R-TBN1-N211D) was used. The gene for wild-type TBN1 (UniProt sequence accession code No. Q0KFV0) was mutated in the plant expression vector pLV07 by PCR site-directed mutagenesis to replace Asn211 with Asp. The mutation leads to the hypoglycosylated version of the protein, with only two N-glycosylated sites. *Agrobacterium tumefaciens* LBA4404 was transformed by electroporation with the vector pLV07 and the protein was produced using leaf disc infiltration and by expression in *Nicotiana benthamiana* leaves (Matoušek *et al.*, 2009). Further, the protein was purified by precipitation in ammonium sulfate, ion-exchange chromatography (HiTrap Q FF, Amersham) and affinity chromatography (HiTrap Heparin HP, Amersham), and was desalted on a PD10 (Amersham) column (Lipovova *et al.*, 2008; Matoušek *et al.*, 2008). Protein expression details are summarized in Table 1.

The storage buffer for R-TBN1-N211D consisted of 50 m*M* Tris–HCl pH 7.5, 0.3 *M* NaCl, 1%(*v*/*v*) glycerol with 10 m*M* 3-(1-pyridino)-1-propanesulfonate (NDSB-201) to improve the protein solubility.

Table 3
Data and structure-refinement statistics for TBN1-N211D-PO_4.

Values in parentheses are for the highest resolution shell.

X-ray source	BL14.1, BESSY II
Detector	MAR CCD 225
Wavelength (Å)	0.918
Crystal-to-detector distance (mm)	339.5
No. of frames	120
Exposure per image (s)	3
Oscillation angle (°)	1
Space group	*H*3
Unit-cell parameters (Å)	$a = b = 115.9$. $c = 74.7$
Resolution (Å)	35.0–2.8 (2.89–2.80)
No. of measured reflections	35101
No. of unique reflections	9202 (850)
Average multiplicity	3.8 (3.2)
Completeness (%)	99.8 (98.7)
Average $I/\sigma(I)$	17.2 (2.0)
R_{merge}	0.08 (0.50)
R_{work}	0.181
R_{free}	0.250
R_{all}	0.187
Average ADP (Å^2)	55.5
R.m.s.d., bond lengths (Å)	0.012
R.m.s.d., bond angles (°)	1.805
Solvent content (%)	52.9
Matthews coefficient (Å^3 Da^{-1})	2.61
No. of non-H atoms	2284
No. of monomers in asymmetric unit	1
No. of localized amino acids	265
No. of water molecules	13
Other localized ions	3 Zn^{2+}, PO_4^{3-}
Localized glycosylation	4 GlcNAc, 2 β-D-mannose, 1 α-D-mannose

2.2. Crystallization

For crystallization, R-TBN1-N211D was concentrated to 2 mg ml^{-1} in the storage buffer and 5′-AMP (Sigma–Aldrich) was added to a final concentration of 10 m*M* for co-crystallization. The hanging-drop vapour-diffusion method was used at 291 K with a 1:1 volume ratio of protein solution to reservoir solution in the drop (0.5 µl + 0.5 µl). The original crystallization reservoir solution, condition No. 2 from Crystal Screen 2 [0.1 *M* HEPES pH 7.5, 8%(*v*/*v*) ethylene glycol, 10%(*w*/*v*) PEG 8000; Hampton Research], was optimized to a final solution consisting of 0.1 *M* bis-tris pH 6.5, 10%(*w*/*v*) PEG 8000. Crystallization information is summarized in Table 2.

2.3. Structure solution and refinement

Crystals were equilibrated for 15 s in mother solution with the addition of glycerol to a concentration of 20%(*v*/*v*) as a cryoprotectant. The crystals were mounted in nylon Cryo-Loops (Hampton Research) and vitrified in liquid nitrogen.

Data collection was carried out at 100 K on beamline 14.1 at the BESSY II synchrotron-radiation source, Helmholtz-

Zentrum Berlin using a MAR CCD 225 detector and a mini-kappa goniometer (Mueller *et al.*, 2012). The data were processed using *XDS* (Kabsch, 2010*a*,*b*); the data statistics are summarized in Table 3. The structure was solved by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010) in the *CCP*4 suite (Winn *et al.*, 2011) with PDB entry 4dj4 (Koval' *et al.*, 2013) chain *A* as a search model. The procedure resulted in a solution with an $R_{\rm cryst}$ of 0.395 and a contrast of 10.65 as defined in *MOLREP*.

Structure refinement involved iterative cycles of manual model building using *Coot* (Emsley *et al.*, 2010) according to $2mF_{\rm o} - DF_{\rm c}$ and $mF_{\rm o} - DF_{\rm c}$ Fourier maps and restrained refinement using *REFMAC*5 (Murshudov *et al.*, 2011). $R_{\rm free}$ was used as a cross-validation statistic. The structure was validated using *MolProbity* (Chen *et al.*, 2010) and the set of validation tools in *Coot* (Emsley *et al.*, 2010). All figures were created using *PyMOL* (DeLano, 2004).

The structure was deposited in the PDB with accession code 4jdg. The structure-refinement statistics are summarized in Table 3. The structure is referred to as TBN1-N211D-PO_4; the structure of the wild type (PDB entry 3sng) is referred to as TBN1-wt and a previous structure of the N211D mutant (PDB entry 4dj4) is referred to as TBN1-N211D.

3. Results and discussion

3.1. Structure of a mutant form of TBN1

TBN1 belongs to the phospholipase C/P1 nuclease superfamily (Pfam ID CL0368) and the S1/P1 nuclease family (Pfam ID PF02265) and shares its fold with other members of the family with known structure (Koval' *et al.*, 2013; Yu *et al.*, 2014; Romier *et al.*, 1998). The enzyme molecule consists of 11 α-helices, a 3_{10}-helix and a small β-sheet, and is stabilized by four disulfide bridges. Asn119 and Asn137 are N-glycosylated (Fig. 1). The modelled oligosaccharide chain connected to Asn119 consists of two units of *N*-acetylglucosamine (GlcNAc) and one unit of β-D-mannose; in the case of Asn137

Table 4
Zinc-cluster parameters of TBN1 (TBN1-N211D-PO_4, TBN1-N211D and TBN1-wt) and AtBFN2 from *A. thaliana* (AtBFN2-wt and AtBFN2-PO_4).

The PDB code is given for each structure.

	TBN1-N211D-PO_4	TBN1-wt	TBN1-N211D	AtBFN2-wt	AtBFN2-PO_4
	4jdg	3sng	4dj4 (chain *A*, *B*)	4cwm (chain *A*, *B*)	4cxv (chain *A*, *B*)
Zn–Zn2 (Å)	3.4	3.3	3.3, 3.3	3.4, 3.5	3.5, 3.6
Zn1–Zn3 (Å)	4.8	4.8	4.9, 4.9	5.0, 5.0	4.9, 4.9
Zn2–Zn3 (Å)	5.7	5.9	5.7, 6.0	5.7, 5.7	5.8, 5.8
Zn2–Zn1–Zn3 angle (°)	86.1	91.2	86.5, 92.2	84.1, 82.9	84.1, 84.9

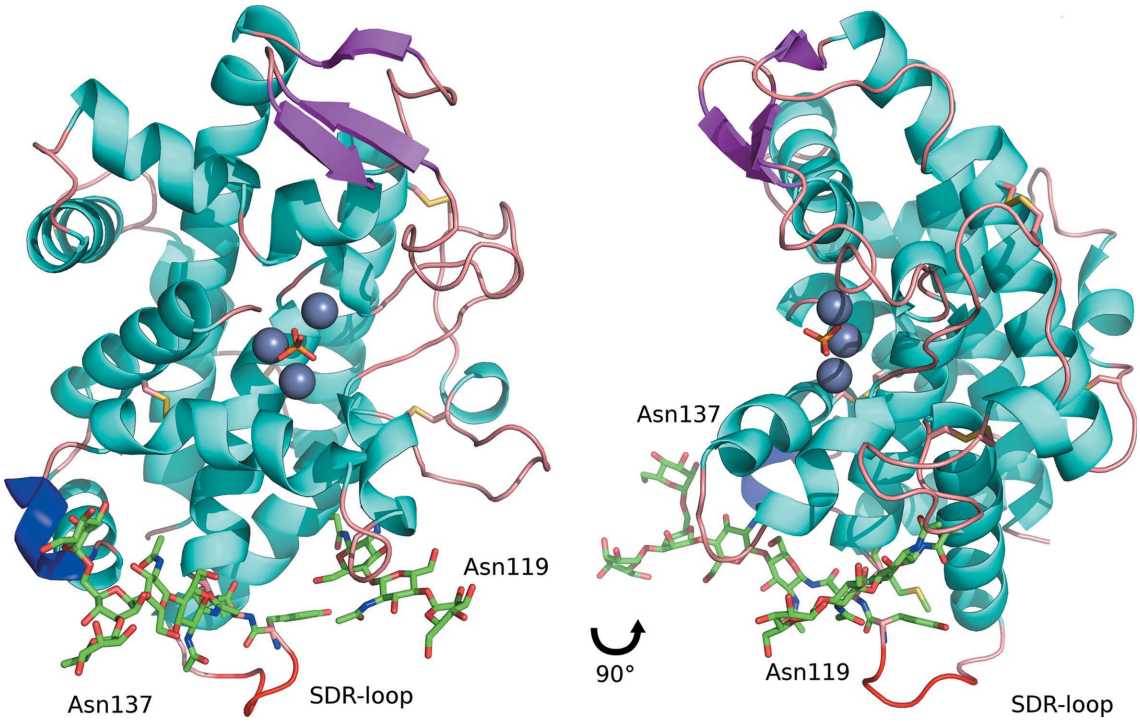


Figure 1
Overall structure of TBN1-N211D-PO_4, shown in secondary-structure representation. α-Helices are in cyan, 3_{10}-helices in blue, β-sheets in magenta, unstructured loops in salmon, oligosaccharides (named by modified amino-acid residues) in green, phosphorus in orange and zinc ions in grey. The SDR-loop (residues 131–135) is shown in red.

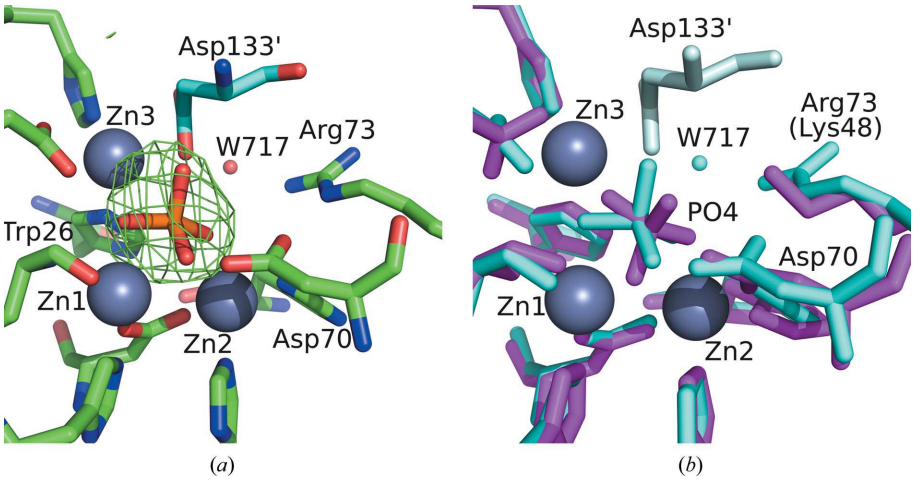


Figure 2
(*a*) The active centre of TBN1-N211D-PO_4 with bound phosphate ion (phosphorus in orange). The omit $mF_o - DF_c$ Fourier map at the 3.5 r.m.s.d. level is shown in green. The zinc-coordinating residues are shown in green and Asp133′ of the neighbouring molecule is shown in cyan. (*b*) The active centre of TBN1-N211D-PO_4 (cyan) superimposed (*LSQ* using zinc ions in *Coot*) with that of AtBFN2-PO_4 (magenta; PDB entry 4cxv, Yu *et al.*, 2014). Asp133′ of the neighbouring TBN1-N211D-PO_4 molecule is shown in pale cyan. The phosphate ion is rotated by 48° around the P—O4 bond and the position of the P—O4 bond is shifted by 0.5 Å.

two units of *N*-acetylglucosamine (GlcNAc), one unit of β-D-mannose and one unit of α-D-mannose were modelled.

The active centre of TBN1 is formed by three Zn^{2+} ions. The zinc cluster forms an almost right-angled triangle (cluster parameters and a comparison are given in Table 4). The positions of the Zn^{2+} ions were confirmed by the anomalous signal. The zinc cluster is conserved in this group of zinc-dependent nucleases (Table 4), including the multifunctional nuclease TBN1 (Koval' *et al.*, 2013), the multifunctional nuclease AtBFN2 (Yu *et al.*, 2014) and the single-strand-specific P1 nuclease (Romier *et al.*, 1998).

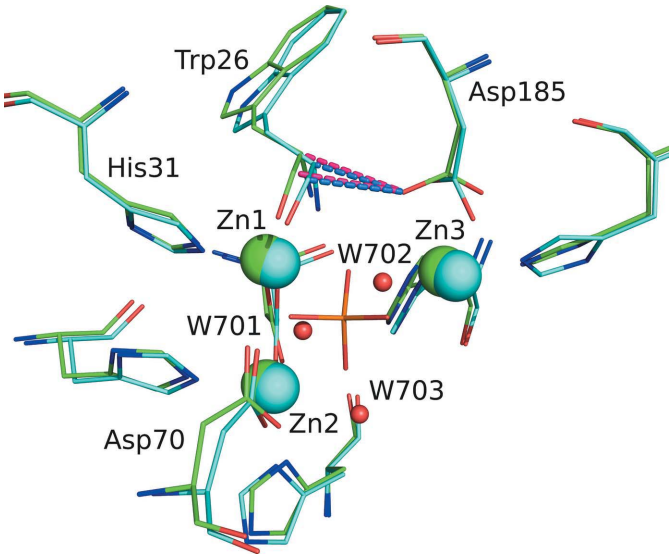


Figure 3
Superimposed active centres of TBN1 structures: TBN1-N211D (cyan) and TBN1-N211D-PO_4 (green) (*LSQ* using C^{α} atoms of all residues in *Coot*). Zinc ions are shown as spheres, the zinc-binding residues and phosphate ion are shown as sticks and water molecules of TBN1-N211D are shown as small red spheres. Blue dashed lines show the distances between the Trp26 main chain (C and C^{α}) of TBN1-N211D-PO_4 and Asp185 OD1 of TBN1-N211D-PO_4 (3.0 Å) and magenta dashed lines those between the Trp26 main chain (C and C^{α}) of TBN1-N211D and Asp185 OD1 of TBN1-N211D-PO_4 (2.6 Å). The Trp26 main-chain atoms are shifted by 0.5 Å between TBN1-N211D-PO_4 and TBN1-N211D. The zinc ions are shifted by 0.25 Å for Zn1, 0.26 Å for Zn2 and 0.4 Å for Zn3.

Compared with the previous structures of TBN1 (Koval' *et al.*, 2013), coordination of the zinc ions is conserved except for residue Asp70 (Zn2) and the newly occurring coordination by Asp133′ (Zn3) from a symmetry-related molecule (symmetry operator $-y + 2/3, x - y + 1/3, z + 1/3$). While in the previous structures TBN1-wt and TBN1-N211D the $O^{\delta1}$ atom of Asp70 formed a weak coordination to Zn2 with a distance of 2.7 Å, in the reported structure this coordination is disappearing, with a distance of 3.1 Å. The side chain of Asp70 is the most flexible part of the active centre according to its high atomic displacement parameter. The $O^{\delta1}$ atom of Asp133′ coordinates Zn3 with a distance of 2.3 Å and participates in the binding of the phosphate ion in the zinc cluster.

A peak in the difference $mF_o - DF_c$ Fourier map within the zinc cluster was interpreted as a phosphate ion (Fig. 2*a*). It is most likely that the phosphate ion originates from impurities in the added 5′-AMP (purity 97%, Sigma–Aldrich, St Louis, USA). However, 5′-AMP was not observed in the structure. It was not possible to confirm the presence of the phosphate ion from the anomalous signal. There were also no peaks of anomalous density observed at the sulfur sites.

The active centre of TBN1-N211D-PO_4 exhibits differences when compared with that of TBN1-N211D (Fig. 3). The distance between the carboxyl group of Asp185 and C^{α} of Trp26 is increased, which opens space for the free rotation of the carboxyl group (both shown positions are stable in the refinement of TBN1-N211D-PO_4). Moreover, the distance of 3.0 Å suggests the possibility of the formation of a hydrogen bond between the carboxyl group of Asp185 and delocalized electrons of the 26–27 peptide bond and the N-terminal amine of Trp26.

A phosphate ion in the zinc cluster was observed in the structure of multifunctional nuclease type 2 from *Arabidopsis*

thaliana (AtBFN2; Yu *et al.*, 2014; PDB entry 4cxv; Fig. 2*b*; referred to as AtBFN2-PO_4). Here, the phosphate ion is rotated towards Zn3. Two O atoms of the phosphate ion are located in the plane defined by the zinc ions in TBN1-N211D-PO_4. The bonding distances of the phosphate ion in both structures are listed in Table 5 and the zinc-cluster parameters are compared in Table 4. A similar position of a phosphate ion in the zinc cluster was observed in a complex of phospholipase C from *B. cereus* with the substrate analogue (3*S*)-3,4-di-*N*-hexaoyloxybutyl-1-phosphocholine (Antikainen *et al.*, 2003; PDB entry 1p6d; Fig. 4).

3.2. Crystal composition of TBN1 in known structures

In the reported structure of the deglycosylated mutant TBN1-N211D-PO_4, the molecules of TBN1 are packed into a new crystal form (Table 6) with some conserved crystal motifs in comparison to the previously solved structures of the wild type (PDB entry 3sng; referred to as TBN1-wt) and the deglycosylated N211D mutant (PDB entry 4dj4; referred to as TBN1-N211D). The molecules form superhelices in the crystals in all of the structures of TBN1 known to date (Fig. 5), which is caused by an interaction of the active centre of one molecule with a surface loop Ser132′-Asp133′-Arg134′ (SDR-loop) of a neighbouring molecule (residues of the SDR-loop in symmetry-related molecules are marked with a prime). The superhelices are generated by crystal symmetry around the 3_1 axis.

However, the manner in which the superhelices pack with each other to form the crystal varies among the known structures. In TBN1-wt the orientation of two interacting superhelices is antiparallel and they are related by a crystallographic twofold. In TBN1-N211D and TBN1-N211D-PO_4 the orientation of the superhelices is parallel and there is no additional symmetry involved besides crystallographic translational symmetry; however, the mutual position of the superhelices differs (Fig. 6).

Table 5
Bond distances (in Å) of phosphate-ion O atoms to zinc-cluster atoms and active-centre residues in TBN1-N211D-PO_4 and AtBFN2-PO_4 (PDB entry 4cxv; Yu *et al.*, 2014).

Residue numbers according to TBN1-N211D-PO_4 are in parentheses. Atom labelling is according to TBN1-N211D-PO_4.

	TBN1-N211D-PO_4				AtBFN2-PO_4			
	PO_4^{3-} (601)				PO_4^{3-} (401-A)			
	O1	O2	O3	O4	O1	O2	O3	O4
Zn1 (401)	2.3		2.3		3.3		1.8	
Zn2 (402)			2.7	2.2			2.2	2.1
Zn3 (403)	2.6				2.1			
Asp OD1 (70)		4.0	3.0	3.5		2.4	3.3	3.9
Asp OD2 (70)		3.8	2.2	4.3		3.2	3.3	3.0
Arg NE (73)		5.2		4.3				
Arg NH (73)		5.4		4.3				
Lys NZ (48)†						3.6		3.1

† The residue in AtBFN2-PO_4 in the place of Arg73 in TBN1-N211D-PO_4.

The putative asymmetrical enzyme dimers formed by the interaction of the SDR-loop with the active centre show only minor differences in their arrangement across the known structures. The mutual orientation of the two molecules within a dimer varies by less than 10°. Details of the interaction interfaces of the dimers and their variations are shown in Table 7.

Thus, in contrast to the conservation of the superhelix motif, the contact of the SDR-loop with the active centre differs in the structures (Fig. 7). Coordination of Zn3 is completed by Ser132′ in TBN1-wt; however, Asp133′ takes this role in TBN1-N211D and TBN1-N211D-PO_4. The interaction of Arg134′ with Phe86 and Asp88 in the base-binding site

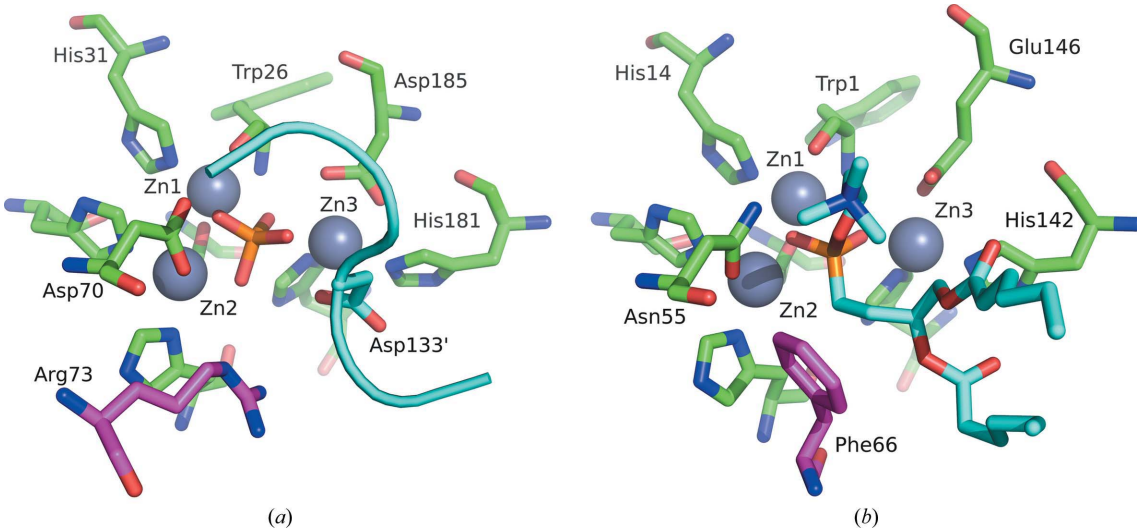


Figure 4
Comparison of the active centres of (*a*) TBN1-N211D-PO_4 with phosphate ion and the SDR-loop of a symmetry-related molecule (cyan) and (*b*) phospholipase C from *B. cereus* (Antikainen *et al.*, 2003) with (3*S*)-3,4-di-*N*-hexaoyloxybutyl-1-phosphocholine (cyan; PDB entry 1p6d). Zinc-coordinating residues are shown as green sticks. P atoms are shown in orange. All of the zinc-binding residues are conserved except Asp70, which is substituted by Asn55 in phospholipase C. Arg73, which is involved in the catalytic mechanism (Koval' *et al.*, 2013), in TBN1-N211D-PO_4 and Phe66 in the same position in phospholipase C are shown in magenta. Notice the similarity in phosphate ion position and orientation.

Table 6
Parameters of the crystal structures TBN1-wt, TBN1-N211D and TBN1-N211D-PO_4.

	TBN1-wt	TBN1-N211D	TBN1-N211D-PO_4
PDB code	3sng	4dj4	4jdg
Protein form	Wild type	N211D mutant	N211D mutant
Glycosylation	Full	Hypoglycosylated	Hypoglycosylated
Additional ligands	SO_4^{2-}, bis-tris, Cl^-	Na^+, Cl^-	PO_4^{3-}
Crystallization solution	1.0 *M* $(NH_4)_2SO_4$, 0.1 *M* bis-tris pH 5.5, 1%(*w*/*v*) PEG 3350	0.16 *M* $MgCl_2$, 0.08 *M* Tris pH 8.5, 19%(*w*/*v*) PEG 3350, 3.3%(*w*/*v*) 1,4-butanediol	0.1 *M* bis-tris pH 6.5, 10%(*w*/*v*) PEG 8000; additive 10 m*M* 5′-AMP
Resolution (Å)	35–2.16	30–2.48	35–2.80
Space group	$P3_121$	$H3$	$H3$
Unit-cell parameters (Å)	$a = b = 101.0$, $c = 71.5$	$a = b = 113.4$, $c = 138.4$	$a = b = 115.9$, $c = 74.7$
No. of monomers per asymmetric unit	1	2	1

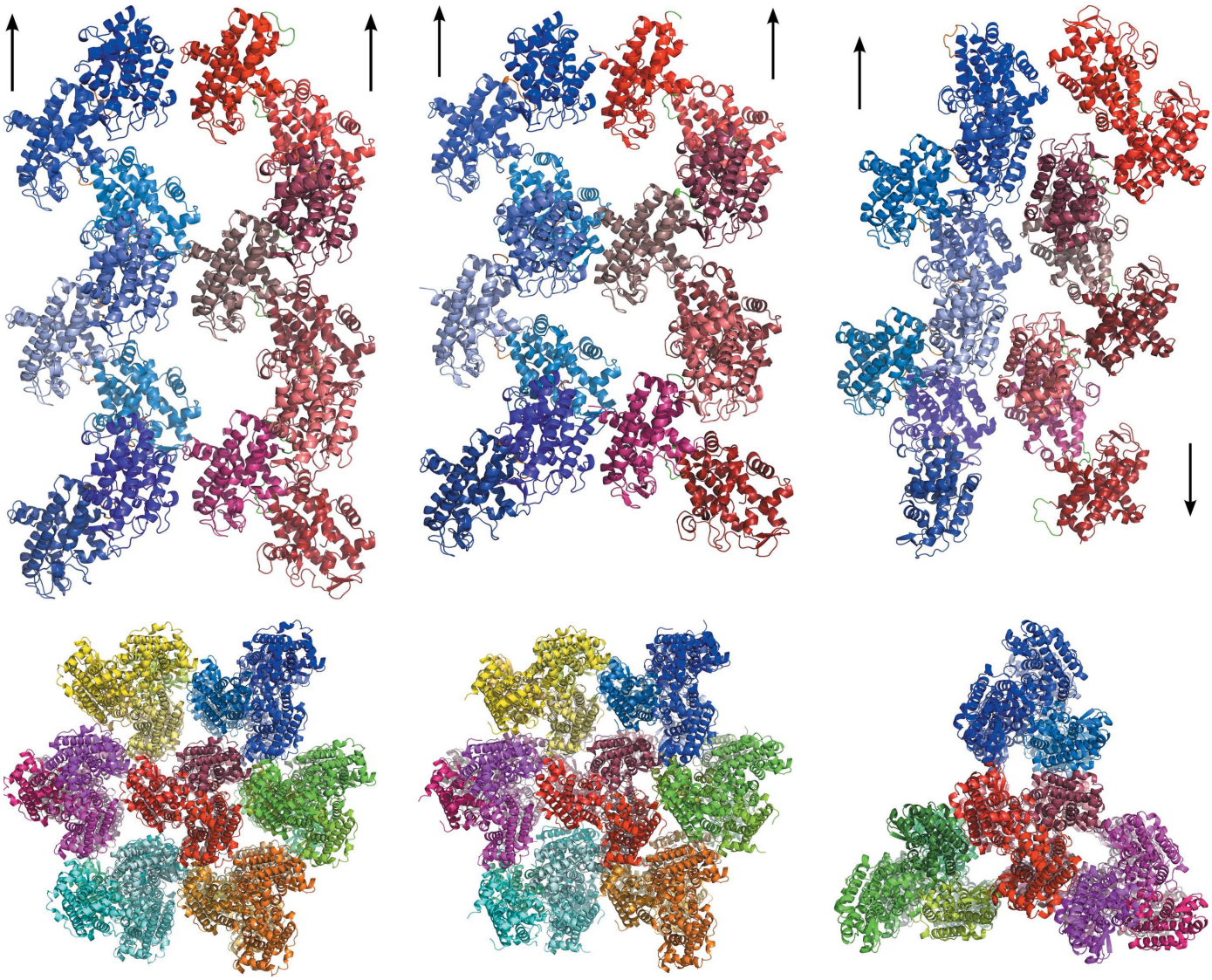


Figure 5
Superhelices of TBN1 in (*a*) TBN1-N211D-PO_4, (*b*) TBN1-N211D and (*c*) TBN1-wt. Top row, two neighbouring helices viewed perpendicular to the superhelix axes; arrows indicate the direction of the superhelices. Bottom row, packing of all superhelices in contact with the central (red) superhelix viewed along the superhelix axes. Colours distinguish different superhelices; shades of one colour distinguish the molecules in a superhelix.

Table 7
Variation of the dimer interface of TBN1, the key intermolecular interaction leading to the formation of superhelices, in crystals in known structures calculated by *PISA* (Krissinel & Henrick, 2007).

The percentage with respect to the whole TBN1 molecule is given in parentheses. BASA stands for buried accessible surface area.

	TBN1-N211D-PO_4		TBN1-N211D		TBN1-wt	
	Active site	SDR-loop	Active site	SDR-loop	Active site	SDR-loop
No. of residues	28 (10.5%)	21 (7.9%)	26 (10.4%)	20 (7.5%)	30 (11.2%)	21 (7.9%)
No. of non-H atoms	95 (4.4%)	89 (4.1%)	99 (4.9%)	87 (4.1%)	100 (4.7%)	79 (3.7%)
BASA (Å^2)	792 (6.3%)	837 (6.6%)	762 (6.4%)	833 (6.7%)	715 (5.6%)	802 (6.3%)

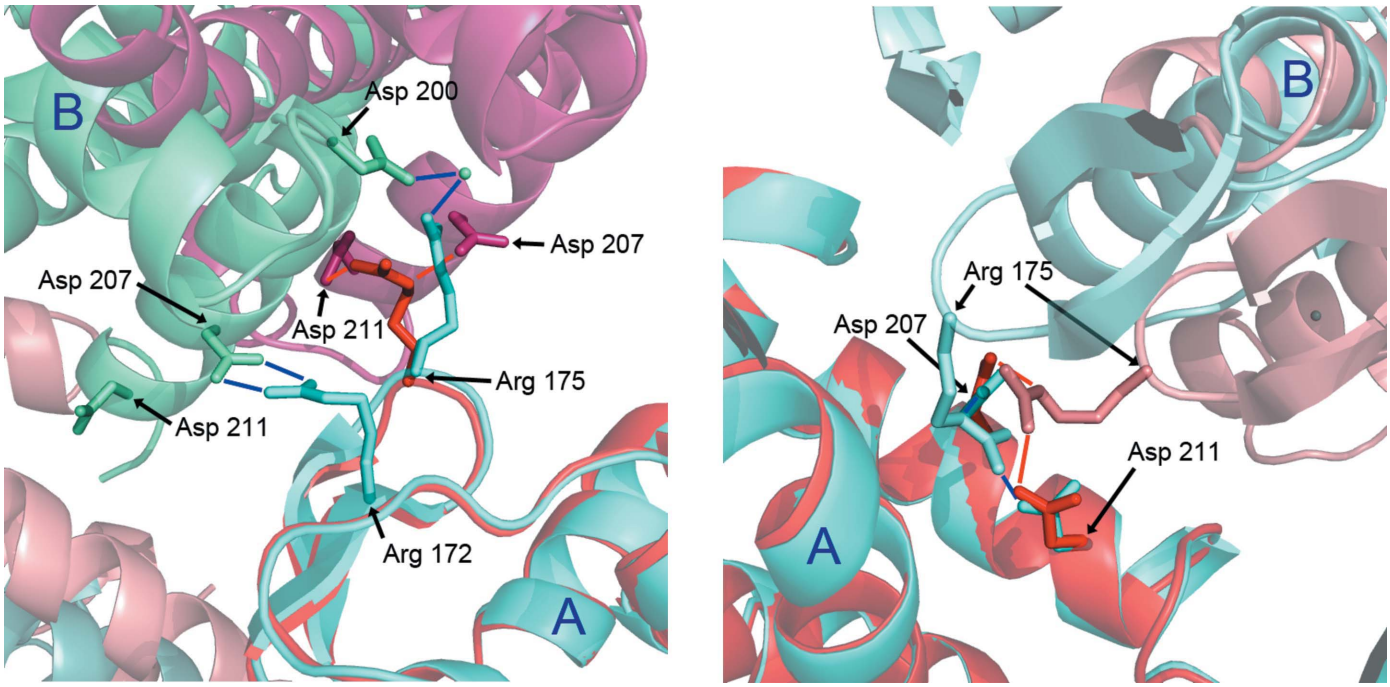


Figure 6
Ilustration of differences in superhelix packing details: comparison of the crystal contacts in TBN1-N211D (cyan) and TBN1-N211D-PO_4 (red) which involve Arg175. Crystals are superimposed by the molecules labelled A. A and B refer to chains in the TBN1-N211D structure. The distance between the C^{α} atoms of Asp207 in the left panel is 11.6 Å. The distance between the C^{α} atoms of Arg175 in the right panel is 9.7 Å.

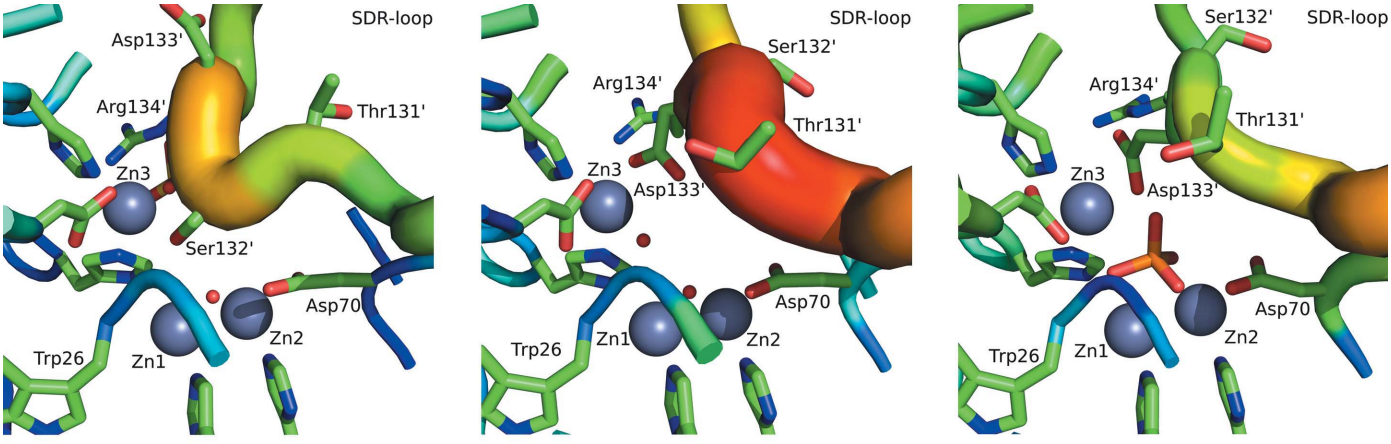


Figure 7
Comparison of the interaction of the SDR-loop (right, in sticks, residues denoted by primes) with the active centre (left) in the TBN1-wt (*a*), TBN1-N211D (*b*) and TBN1-N211D-PO_4 (*c*) structures. The main chain is shown in cartoon representation with colours and diameters according to the values of the atomic displacement parameters of the C^{α} atoms (low, blue and thin; high, red and thick).

remains the same as in the previously reported structures of TBN1. The currently reported structure provides an observation of the most orderly localized SDR-loop. In the TBN1-wt and TBN1-N211D structures the SDR-loop has high atomic displacement parameters and is even not localized in chain *B* of TBN1-N211D. Hence, we propose that the presence of the phosphate ion in the active centre may stabilize the interaction, but it is not crucial for the formation of the contact and the superhelices.

Every molecule in TBN1-wt is surrounded by four other protein molecules, whereas every molecule in TBN1-N211D and TBN1-N211D-PO_4 is surrounded by six other molecules. However, the contact causing the formation of the superhelices, *i.e.* the contact involving the SDR-loop, is the most significant (buried accessible surface area of about 800 Å as calculated by *PISA*; Krissinel & Henrick, 2007), in contrast to other relatively weaker contacts (buried accessible surface area of about 100 $Å^2$); these weaker contacts mediate the mutual packing of superhelices into the crystal.

4. Conclusion

The third reported crystal structure of the enzyme TBN1 confirms that the superhelices, which are the main motif in the crystal composition, persist in all of the different crystal forms. The binding of the surface SDR-loop of one molecule to the active centre of a neighbouring molecule, present in all known structures of this enzyme, suggests that this intermolecular contact is not a crystallization artefact, as might be perceived. Instead, it may have a biological role. In the presented structure, a phosphate anion is bound to the zinc cluster in the active centre, which leads to better localization (lower atomic displacement parameters) of the SDR-loop. Phosphate binding to the zinc cluster has been observed for other related enzymes, but here it takes a special place. The previous results (Koval' *et al.*, 2013) are now corroborated by the crystallographic data. The superhelices can also form without phosphate, but we propose that the presence and binding of phosphate stabilizes their formation, which is most likely to lie behind enzyme aggregation in certain conditions. The observed interactions may represent an important regulatory process controlled by phosphate abundance and phosphate starvation, which is known to be correlated with senescence processes.

Acknowledgements

This publication is supported by the project BIOCEV – Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University (CZ.1.05/1.1.00/02.0109) from the European Regional Development Fund, by the Ministry of Education, Youth and Sports of the Czech Republic (grant Nos. EE2.3.30.0029 and LG14009), by the Grant Agency of the Czech Technical University in Prague (grant No. SGS13/219/OHK4/3T/14), by institutional support of IBT CAS, v.v.i. (RVO 86652036) and by institutional support of BC CAS, v.v.i. (IPMB RVO 60077344). The authors wish to thank Dr Uwe Müller and Dr Manfred Weiss of the Helmholtz-Zentrum Berlin for support at beamline BL14.1 of BESSY II.

References

Antikainen, N., Monzingo, A., Franklin, C., Robertus, J. & Martin, S. (2003). *Arch. Biochem. Biophys.* **417**, 81–86.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
DeLano, W. (2004). *Abstr. Pap. Am. Chem. Soc.* **228**, 030-CHED.
Dohnálek, J., Koval', T., Lipovová, P., Podzimek, T. & Matoušek, J. (2011). *J. Synchrotron Rad.* **18**, 29–30.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Kabsch, W. (2010*a*). *Acta Cryst.* D**66**, 133–144.
Kabsch, W. (2010*b*). *Acta Cryst.* D**66**, 125–132.
Koval', T., Lipovová, P., Podzimek, T., Matoušek, J., Dušková, J., Skálová, T., Štěpánková, A., Hašek, J. & Dohnálek, J. (2013). *Acta Cryst.* D**69**, 213–226.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Lipovova, P., Podzimek, T., Orctova, L., Matousek, J., Pouckova, P., Soucek, J. & Matousek, J. (2008). *Neoplasma*, **55**, 158–164.
Matoušek, J., Kozlová, P., Orctová, L., Schmitz, A., Pešina, K., Bannach, O., Diermann, N., Steger, G. & Riesner, D. (2007). *Biol. Chem.* **388**, 1–13.
Matoušek, J., Podzimek, T., Poučková, P., Orctová, L., Souček, J., Škvor, J., Lipovová, P. & Matoušek, J. (2008). *Anticancer Res.* **28**, 3399.
Matousek, J., Podzimek, T., Pouckova, P., Stehlik, J., Škvor, J., Lipovova, P. & Matousek, J. (2010). *Neoplasma*, **57**, 339–348.
Matoušek, J., Podzimek, T., Poučková, P., Stehlík, J., Škvor, J., Souček, J. & Matoušek, J. (2009). *Oncol. Res.* **18**, 163–171.
Mueller, U., Darowski, N., Fuchs, M. R., Förster, R., Hellmig, M., Paithankar, K. S., Pühringer, S., Steffien, M., Zocher, G. & Weiss, M. S. (2012). *J. Synchrotron Rad.* **19**, 442–449.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Podzimek, T., Matoušek, J., Lipovová, P., Poučková, P., Spiwok, V. & Šantrůček, J. (2011). *Plant Sci.* **180**, 343–351.
Romier, C., Dominguez, R., Lahm, A., Dahl, O. & Suck, D. (1998). *Proteins*, **32**, 414–424.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yu, T.-F., Maestre-Reyna, M., Ko, C.-Y., Ko, T.-P., Sun, Y.-J., Lin, T.-Y., Shaw, J.-F. & Wang, A. H.-J. (2014). *PLoS One*, **9**, e105821.

PLOS ONE

RESEARCH ARTICLE

Structural and Catalytic Properties of S1 Nuclease from *Aspergillus oryzae* Responsible for Substrate Recognition, Cleavage, Non–Specificity, and Inhibition

Tomáš Kovaľ[1]*, Lars H. Østergaard[2], Jan Lehmbeck[3], Allan Nørgaard[4], Petra Lipovová[5], Jarmila Dušková[1], Tereza Skálová[1], Mária Trundová[1], Petr Kolenko[1], Karla Fejfarová[1], Jan Stránský[1], Leona Švecová[1], Jindřich Hašek[1], Jan Dohnálek[1]*

1 Laboratory of Structure and Function of Biomolecules, Institute of Biotechnology CAS, v. v. i., Biocev, Vestec, Czech Republic, **2** Department of Agile Protein Screening, Novozymes A/S, Bagsvaerd, Denmark, **3** Department of Fungal Strain Technology, Novozymes A/S, Bagsvaerd, Denmark, **4** Department of Protein Biochemistry and Stability, Novozymes A/S, Bagsvaerd, Denmark, **5** Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague, Czech Republic

* koval.tomas@gmail.com (TK); jan.dohnalek@ibt.cas.cz (JD)

Citation: Kovaľ T, Østergaard LH, Lehmbeck J, Nørgaard A, Lipovová P, Dušková J, et al. (2016) Structural and Catalytic Properties of S1 Nuclease from *Aspergillus oryzae* Responsible for Substrate Recognition, Cleavage, Non–Specificity, and Inhibition. PLoS ONE 11(12): e0168832. doi:10.1371/journal.pone.0168832

Editor: Andreas Hofmann, Griffith University, AUSTRALIA

Received: August 18, 2016

Accepted: December 7, 2016

Published: December 30, 2016

Data Availability Statement: Data is available from the Protein Data Bank using the accession codes 5FB9 (S1-free), 5FBA (S1-PO4pH5.5), 5FBB (S1-PO4/5'AMP), 5FBC (S1-5'dAMP(S)), 5FBD (S1-PO4/dCyt), 5FBF (S1-5'dCMP), and 5FBG (S1D65N-PO4/dCyt/dGua).

Funding: This work was supported by the project BIOCEV Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University (CZ.1.05/1.1.00/02.0109) from the European

Abstract

The single–strand–specific S1 nuclease from *Aspergillus oryzae* is an archetypal enzyme of the S1–P1 family of nucleases with a widespread use for biochemical analyses of nucleic acids. We present the first X–ray structure of this nuclease along with a thorough analysis of the reaction and inhibition mechanisms and of its properties responsible for identification and binding of ligands. Seven structures of S1 nuclease, six of which are complexes with products and inhibitors, and characterization of catalytic properties of a wild type and mutants reveal unknown attributes of the S1–P1 family. The active site can bind phosphate, nucleosides, and nucleotides in several distinguished ways. The nucleoside binding site accepts bases in two binding modes–shallow and deep. It can also undergo remodeling and so adapt to different ligands. The amino acid residue Asp65 is critical for activity while Asn154 secures interaction with the sugar moiety, and Lys68 is involved in interactions with the phosphate and sugar moieties of ligands. An additional nucleobase binding site was identified on the surface, which explains the absence of the Tyr site known from P1 nuclease. For the first time ternary complexes with ligands enable modeling of ssDNA binding in the active site cleft. Interpretation of the results in the context of the whole S1–P1 nuclease family significantly broadens our knowledge regarding ligand interaction modes and the strategies of adjustment of the enzyme surface and binding sites to achieve particular specificity.

Regional Development Fund; by the Ministry of Education, Youth and Sports of the Czech Republic grant No. LG14009 (to JD); support of large infrastructures by the Ministry of Education, Youth and Sports CR, CIISB - LM2015043; by the Czech Science Foundation, project 15-05228S (to JD); and by BioStruct-X (EC FP7 project 283570). Novozymes provided support in the form of salaries for authors LHØ, JL, and AN. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of all authors are articulated in the "Author contributions" section.

Competing Interests: The authors declare that no competing interests exist. Authors affiliated to Novozymes A/S declare no competing interests related to employment, consultancy, patents, products in development or marketed products. All authors adhere to PLOS ONE policies on sharing data and materials.

Introduction

Nucleases of the S1–P1 family [1, 2] can be found in fungi, plants, protozoan parasites and, interestingly, in some bacteria. They are zinc dependent nucleases/3'nucleotidases active on both RNA and DNA with acidic or close to neutral pH optima. They act as phosphoesterases cleaving the P–O3' bond and producing 5'mononucleotides as end products. A cluster of three zinc ions coordinates the substrate/product scissile phosphate and the reaction mechanism utilizes water activated by Zn^{2+} ions as a nucleophile [2]. Their native role usually lies in nucleotides/nucleosides scavenging [3], specific apoptotic functions [4, 5] and in symbiont (pathogen)–host interactions [6]. This enzyme class is generally substrate sequence–nonspecific and representatives from different types of organisms realize substrate binding via a variation of the same basic approach. Even if several studies already addressed ligand binding in this nuclease family exhaustive experimental evidence regarding ligand binding principles in this family is still missing.

S1 nuclease from *Aspergillus oryzae* (EC 3.1.30.1; NCBI sequence ID: XP_001818636; S1–P1 nuclease family in Pfam, PF02265) is an extracellular, single–strand–specific, sugar non–specific, Zn^{2+}–dependent, fungal nuclease with 3'–mononucleotidase activity [7] with a pH optimum in range 4.0–4.3 [8]. Mature S1 nuclease is a glycoprotein with two N–glycosylation sites. It is composed of 267 amino acids with a molecular mass of 29.1 kDa (about 35 kDa when fully glycosylated). It can be utilized in many biochemical methods; in one of its applications it is used as an analytical tool for determination of the secondary structure of nucleic acids [9]. The most likely natural role of S1 lies in scavenging of phosphate and nucleotides [2].

Structures of three S1 nuclease homologs are known: P1 from *Penicillium citrinum* [10, 11], TBN1 from *Solanum lycopersicum* [12] and AtBFN2 from *Arabidopsis thaliana* [13, 14]. The fungal single–strand–specific nuclease P1 shares the highest sequence identity with S1 (51%). Despite close sequence and fold similarity, there are several important differences between these two nucleases. P1 has slightly more basic pH optimum (around 5.5). It prefers 3'AMP over RNA and single–stranded DNA (ssDNA) and can cleave also double–stranded DNA (dsDNA) although at a significantly reduced level [2]. S1 is more active on single–stranded nucleic acids with preference for ssDNA over RNA and 3'AMP [2]. Both nucleases prefer 3'–ribomononucleotides over 3'–deoxyribomononucleotides, but with slightly different base preference [2, 7]. Tomato bifunctional nuclease TBN1 (27% sequence identity with S1) and bifunctional nuclease AtBFN2 (31% sequence identity with S1) belong to the plant nuclease I family, in which they form a sub–family called plant S1–like nucleases [5]. Interestingly, despite the same fold and active site composition there are important differences in the catalytic activity of S1, TBN1, and AtBFN2. Along with differences in pH optima and base and sugar preferences the main difference lies in the fact that TBN1 is active against multiple types of nucleic acids (NAs) including structured DNA and highly stable viroid RNA [15]. AtBFN2 is also capable of cleaving dsDNA although it prefers single–stranded NAs. S1 and P1 on the other hand highly prefer single–stranded NAs. This raises questions regarding the key factors causing such major distinction between plant and fungal nucleases. Finally, the role of amino acids adjacent to the active site in the hydrolysis of NAs has not been fully explained either.

Nuclease and nucleotidase products are competitive inhibitors of the members of the S1–P1 nuclease family, however, no potent selective inhibitors for this family are known. In this study, we focus on the structure–function relationship in S1 nuclease, as a model system of the family, in complexes with varied ligands, in order to investigate substrate recognition, draw conclusions regarding its catalytic properties and provide structure–based background for design of inhibitors with potential applications in biotechnology and medicine.

Along with the wild type of S1 nuclease (S1wt), biochemical properties of four structure–based mutants (S1D65N, S1K68N, S1N154A, and S1N154S) were studied in this work. Analyses

of the S1 reaction mechanism and substrate/inhibitor interactions are based on six structures of S1wt and one structure of S1D65N in complex with ligands obtained at pH 4.2, 5.5, and 6.5.

Results

Seven structures of S1 nuclease in complexes with various ligands and at different pH were obtained. The following paragraphs describe the structural features of S1 nuclease common for all the currently reported structures and the most important details in each of them.

General features of S1 nuclease structure

Crystal packing, protein fold, and post–translational modifications. Crystals of S1 nuclease were obtained under four distinct crystallization conditions with the main difference being in their pH (4.2, 5.5, and 6.5). Depending on the particular condition and the type of ligand in co–crystallization, S1 crystallized with five different unit cells and in four different space groups (Table 1). In six of the seven structures reported here at least one N–linked N–acetyl–D–glucosamine moiety (GlcNAc; the first saccharide unit left after deglycosylation using Endo F1) participates in crystal contacts.

Table 1. Structures of S1 nuclease–crystallization conditions, ligands, selected data collection and structure refinement statistics.

Structure title	5FB9 – unoccupied	5FBA– phosphate	5FBB– inhibitors	5FBC– remodeled	5FBD– nucleotidase products	5FBF–nuclease products	5FBG–mutant with products
PDB ID; figure	5FB9; Fig 3A	5FBA	5FBB; Fig 3B and 3D	5FBC; Fig 3E	5FBD; Fig 3C and 3F	5FBF; Fig 3H and 3I	5FBG; Fig 3G
Crystallization condition	Optimized Index No. 70, pH 5.5	Index No. 70, pH 5.5	Index No. 54, pH 6.5	Index No. 70, pH 5.5	Optimized Index No. 40, pH 4.2	Optimized Index No. 40, pH 4.2	Index No. 70, pH 5.5
Co–crystallization partner	dGua	None	5'AMP	dA(pS)dA	dC(pS)dC	5'dCMP	$d(GC)_6$
Important ligands	None (water)	Pi	2x Pi, 2x 5'AMP, 2x Na	5'dAMP(S)	Pi, dCyt	2 x 5'dCMP	2x Pi, 2x dCyt, dGua
Space group	$P1$	$P2_1$	$P1$	$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P3_121$
Unit–cell *a*, *b c*, (Å)	43.2, 48.6, 65.5	41.8, 62.3, 48.0	42.8, 47.6, 62.6	41.9, 62.6, 48.2	43.0, 62.4, 84.1	53.7, 62.4, 62.8	106.8, 106.8, 127.9
Unit–cell α, β, γ (°)	107.4, 90.1, 105.7°	90, 106.7, 90	106.4, 90.1, 106.3	90, 107.0, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution range (Å)	43.05–1.50 (1.53–1.50)	45.97–1.80 (1.84–1.80)	35.51–1.75 (1.78–1.75)	46.14–1.75 (1.78–1.75)	30.08–1.75 (1.78–1.75)	44.25–1.04 (1.06–1.04)	37.47–1.97 (2.02–1.97)
R_{meas}	0.084 (0.264)	0.090 (0.504)	0.137 (0.791)	0.120 (0.785)	0.072 (0.504)	0.097 (0.394)	0.144 (0.750)
Mean $I/\sigma(I)$	9.2 (3.8)	11.7 (2.2)	8.1 (2.6)	10.7 (2.0)	13.6 (2.1)	10.7 (2.7)	9.3 (2.0)
Completeness (%)	88.2 (54.6)	95.3 (69.0)	95.8 (94.7)	99.8 (99.8)	95.2 (72.2)	94.5 (54.5)	99.6 (97.6)
R_{work}	0.149	0.152	0.145	0.131	0.152	0.111	0.158
R_{free}	0.177	0.199	0.188	0.181	0.212	0.135	0.187
R.m.s.d. bonds (Å)	0.015	0.016	0.017	0.017	0.017	0.013	0.016
R.m.s.d. angles (°)	1.647	1.688	1.728	1.678	1.706	1.695	1.598
Ramachandran plot (%)[a]							
Favored	98.37	96.99	97.01	98.17	97.74	97.93	97.77
Outliers	0	0	0	0	0	0	0

Values in parentheses are for the highest resolution shell. dGua stands for 2'–deoxyguanosine, dCyt for 2'–deoxycytidine and Pi for phosphate ion. Details regarding crystallization are summarized in Table A in S1 File. Data processing and refinement statistics are summarized in Table B in S1 File.

[a]As calculated by MolProbity [17].

doi:10.1371/journal.pone.0168832.t001

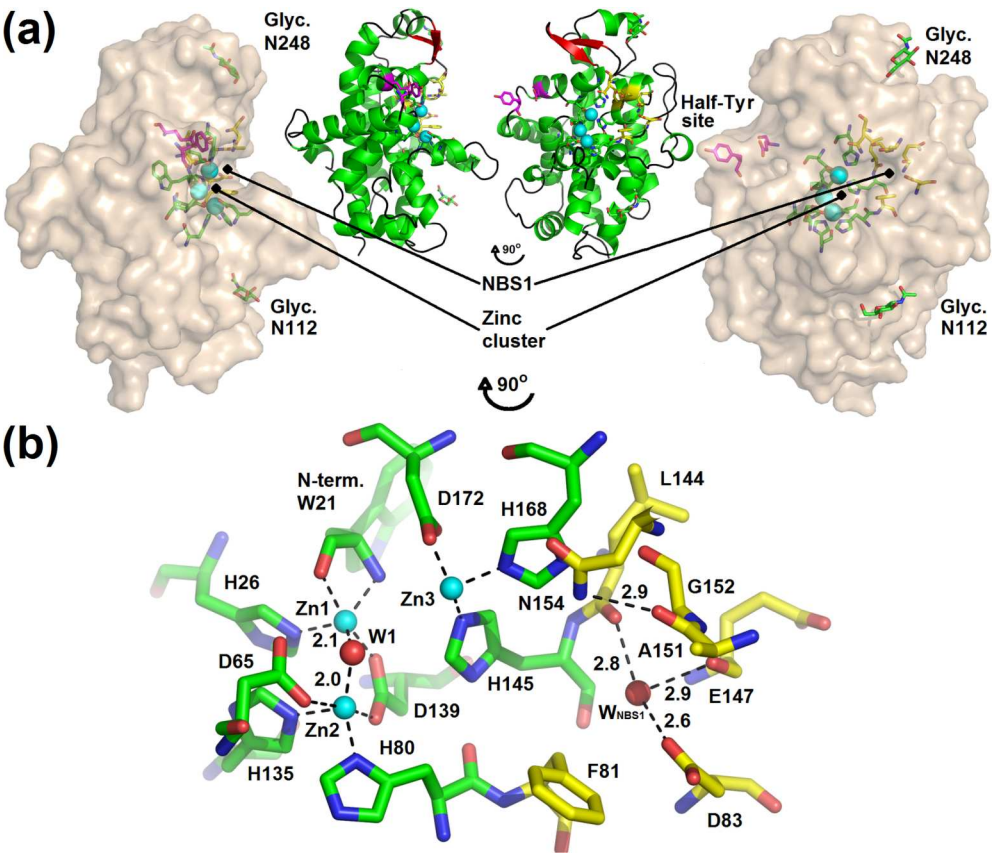


Fig 1. Overview of the structure of S1 nuclease. (a) S1 nuclease represented by its surface with important sites in sticks. The secondary structure representation is shown in the inset (helices are colored green, β–strands red, loops black). The catalytic zinc ions are shown as light blue spheres. Glycans are shown as sticks (carbon–green) and labelled according to the modified residues. Residues involved in the zinc cluster coordination are shown as sticks (carbon–green), residues forming NBS1, (also known as Phe–site) as sticks with carbon in yellow. Residues forming the Half–Tyr site are shown as sticks (carbon–magenta). NBS1 and the Half–Tyr site delimit the extent of the active site cleft. The graphics was created based on the structure 5FBF–nuclease products with removed 5'dCMP. **(b)** The active site of S1 nuclease. The color scheme is the same as in (a). Residues involved in the zinc cluster coordination and formation of NBS1 are labelled. The activated water molecule is labelled as W1, the water molecule present inside NBS1 as W_{NBS1}. Important interaction distances of W1, W_{NBS1} and the distance of Asn154$N^{\delta 2}$ to Gly152O are given in Å. The graphics was created based on the structure 5FBA–phosphate, alternative without PO_4. Phe81 is shown in alternative A (occupancy factor 0.7). Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

doi:10.1371/journal.pone.0168832.g001

S1 nuclease has the phospholipase C/P1 nuclease–like fold stabilized by two disulfide bridges. The mature protein chain starts with Trp21 (the first twenty residues form a signal sequence and are cleaved off in the process of maturation) and ends with Ser287. Two N–glycosylation sites were identified at Asn112 and Asn248. All samples used for crystallization were deglycosylated using Endo F1 (Figure A in S1 File) and so only one GlcNAc can be present at each glycosylation site in the structures reported here. The main features of the structure of mature S1 nuclease are shown in Fig 1A and important structural attributes are also marked in the amino acid sequence alignment in Fig 2.

Active site. The main features of the active site follow all the other members of this family with known structure. The active site is located in a surface cleft (Fig 1) and composed of the obligatory catalytic trinuclear zinc cluster supplemented by Lys68 and the nucleoside binding

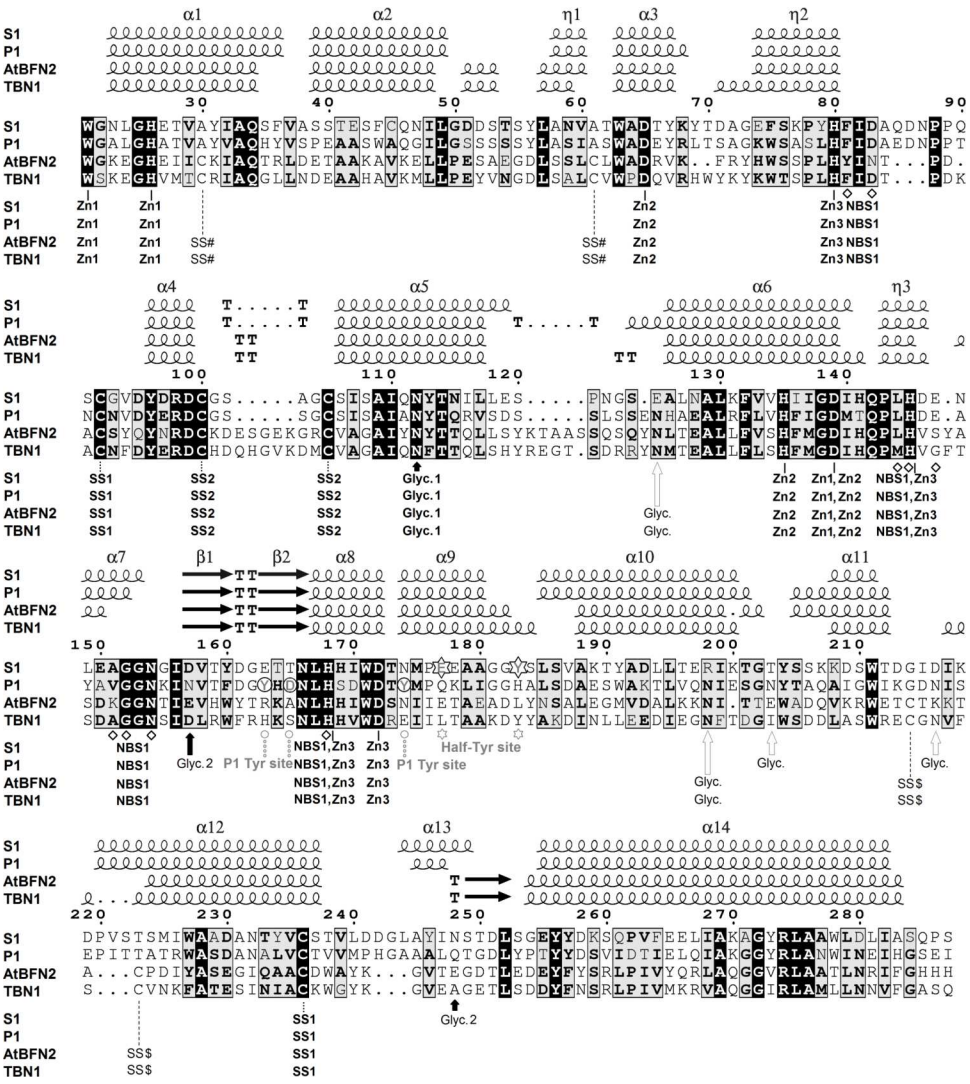


Fig 2. Comparison of amino–acid sequences and secondary structures of S1, P1, TBN1, and AtBFN2 nucleases. Secondary–structure labeling for S1 nuclease is shown: α–helices as α, 3_{10}–helices as η, β–sheets as β, β–turns as T. Residues coordinating Zn^{2+} ions are marked by solid line, cysteine residues forming disulfide bridges by dotted line, S1 nuclease glycosylation sites conserved in P1 nuclease (Glyc. 2 is conserved in position in space, not in sequence) are marked by black arrows. Glycosylation sites not present in S1 are marked by empty arrows. Residues forming the Nucleoside binding site 1 (NBS1) are marked by diamonds. Labels of features conserved in all four nucleases are in bold letters. The Half–Tyr site of S1 nuclease is marked by stars. The Tyr site of P1 nuclease is marked by circles. Secondary structure was assigned by ENDscript [16] based on the structures of S1 5FBF–nuclease products, P1 nuclease–PDB ID: 1AK0 [11], AtBFN2 –PDB ID: 3W52 [13] and TBN1 –PDB ID: 3SNG [12]. The figure was created using ESPript [16] and manually edited.

doi:10.1371/journal.pone.0168832.g002

site 1 (NBS1; also called Phe site in the previous studies [11, 12]). The zinc ions are coordinated by the N–terminus main chain (Trp21) and side chains of several histidine and aspartic acid residues. The cluster connects distant parts of the protein chain (Figs 1 and 2). Organization of the zinc cluster was described in detail in our previous study of TBN1 [12]. NBS1 is an open pocket on the enzyme surface in close proximity to the zinc cluster. NBS1 in S1 nuclease is composed of the Phe81 side chain and the Ala151–Gly152 peptide bond, both providing stacking interactions to a nucleobase, and of Asp83 providing hydrogen bonding. These four amino

acids are exposed to the solvent and create the NBS1 opening. The bottom of the pocket is composed of the main chain carbonyls of Leu144 and Glu147 and residue His145. In vacant NBS1 and in some complexes there is a water molecule (W_{NBS1}) linking together the two main chain oxygen atoms and the side chain oxygen of Asp83. NBS1 is completed by the side chain of Asn154, which stabilizes the pocket through a hydrogen bond to the main chain oxygen of Ala151. $N^{\delta 2}$ of Asn154 can act as a donor in hydrogen bonding to the sugar moiety of a ligand. His168 (involved in the coordination of Zn3) can also be considered a part of NBS1. Residues involved in formation of the active site are shown in Fig 1B and marked in Fig 2.

Structures of complexes

A structure of S1 nuclease with unoccupied active site, two binary complexes, three ternary complexes, and one quaternary complex are further described. To provide evidence for binding of the discussed ligands composite omit maps were calculated for each important ligand and are included in the supplementary data (Figure B to Figure H in S1 File).

Structure with unoccupied zinc cluster. The crystal of the structure 5FB9 –unoccupied was obtained at pH 5.5. The structure contains two protein chains in the asymmetric unit. Coordination of the zinc ions is completed by four water molecules W1–W4 (Fig 3A and Figure B in S1 File). This structure represents a possible resting state of the enzyme.

Complex with phosphate ion. The crystal of the structure 5FBA–phosphate was obtained at pH 5.5. A phosphate ion inside the zinc cluster in the first binding mode is modeled with occupancy 0.5 (Figure C in S1 File). Phosphate ion is a product of 3'–mononucleotidase activity and simultaneously an inhibitor of S1 nuclease. This structure brings evidence for one of its binding modes and for its capability to bind inside the active site. It binds in the same mode as in the structure 5FBB–inhibitors (the next paragraph).

Complex with two independent inhibitors: phosphate ion and 5'AMP. This complex brings clear evidence for the fact that two inhibitors can bind to the active site at the same time and without direct competition. Each of the two protein chains in the asymmetric unit binds one phosphate and one molecule of adenosine 5'–monophosphate (5'AMP) in the same binding mode. The crystal of the structure 5FBB–inhibitors was obtained at pH 6.5. Phosphate is one product of the 3'–mononucleotidase activity, 5'AMP is a product of nuclease activity. Both ligands are also inhibitors. Phosphate is present inside the zinc cluster, 5'AMP occupies the NBS1 site in an inverted orientation with respect to the standard substrate/product binding (the ribose moiety is far from the zinc cluster, Fig 3D).

One of the phosphate oxygen atoms replaces the water molecule W1 and binds almost symmetrically between Zn1 and Zn2 (Fig 3B). Another oxygen atom binds asymmetrically between Zn1 and Zn3 replacing W2. W3 stays in the same position as in the unoccupied zinc cluster and W4 is displaced (compare Fig 3A and 3B for water replacement/displacement). The phosphate ion also interacts with Asp65 and Lys68 (Fig 3B and Figure D in S1 File). We further refer to this binding mode as the first binding mode of the phosphate ion, similar to the previous observations in TBN1 (PDB ID: 4JDG [18]) and AtBFN2 (PDB ID: 4CXV [14]).

The phosphate moiety of 5'AMP interacts not with the zinc cluster but instead with $Asp83O^{\delta 2}$ of NBS1 through a sodium ion (Fig 3D). The placement of the phosphate moiety is slightly affected by a crystal contact–a direct interaction with symmetry–related $Glu42O^{\delta 1}$ (Figure D in S1 File). There is no direct interaction of the ribose moiety with the protein. The adenine base binds inside NBS1 in a position almost exactly above and parallel to the side chain of Phe81. The water molecule W_{NBS1} is substituted by the amino group of the adenine base, which enables a deeper penetration of the ligand, and therefore we denote this binding mode as deep. Hydrogen bonding to Asp83 is realized through the Hoogsteen face of adenine (Fig 3D).

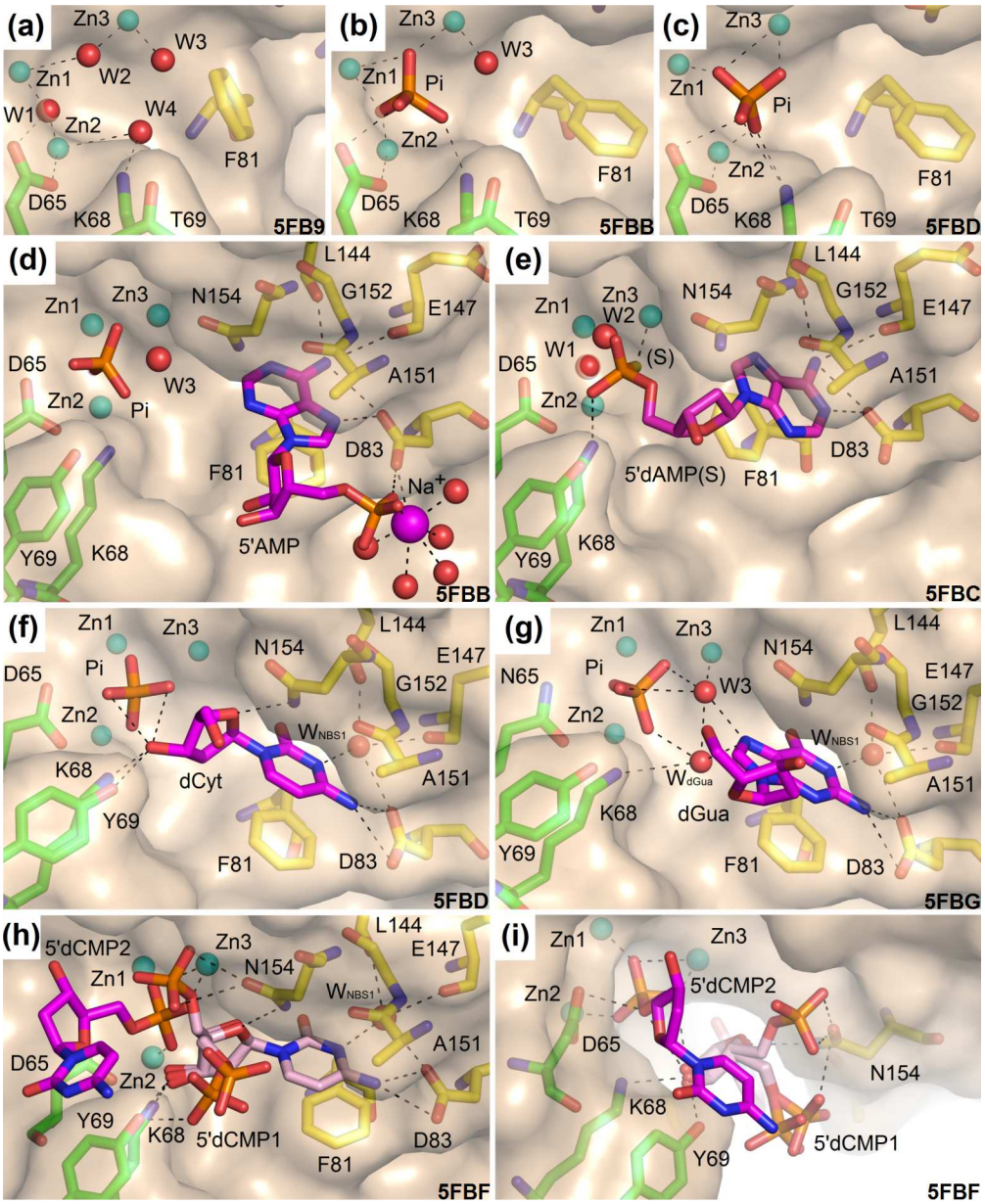


Fig 3. Observed binding of the ligands in S1 nuclease structures. The catalytic zinc ions are shown as light blue spheres. Asp65 (Asn65 in the case of the mutant) is shown in sticks (carbon–green). Other residues involved in the zinc cluster coordination are not shown. Lys68 and Tyr69 interact with ligands and are shown as sticks (carbon–green). Residues forming NBS1 are shown as sticks (carbon–yellow). Important water molecules are shown as red spheres and phosphate ions as orange/red sticks. Selected interactions are shown as black dashed lines. Molecular graphics were created using *PyMOL* (Schrödinger, LLC). PDB ID of each structure is shown. **(a)** Binding of water molecules in the unoccupied zinc cluster in the structure 5FB9 – unoccupied (pH 5.5). **(b)** The first binding mode of the phosphate ion in the structure 5FBB–inhibitors (pH 6.5). 5'AMP is excluded from the graphics for clarity. **(c)** The second binding mode of phosphate ion in the structure 5FBD–nucleotidase products (pH 4.2). All water molecules are displaced and none of the oxygen atoms of the phosphate ion occupies the original positions of water molecules. dCyt is excluded from the graphics for clarity. **(d)** The inverted deep binding mode of 5'AMP in the complex 5FBB–inhibitors (pH 6.5). Water W_{NBS1} is replaced by the inhibitor. The zinc cluster is occupied by a phosphate ion. **(e)** The binding mode of 2'–deoxyadenosine 5'–thio–monophosphate in the complex 5FBC–remodeled (pH 5.5). The thiophosphate moiety binds outside the zinc cluster interacting only with Zn3 using its sulfur atom (marked as S), with Lys68 by the oxygen atom and through a water network with Asn154. In this deep binding mode W_{NBS1} is replaced. Notice the same position of the adenine amino group, similar to the position described in (d), but an entirely inverted orientation of the base. NBS1 is remodeled to the extended form. **(f)** The observed binding mode of

2'–deoxycytidine (carbon–magenta) in the complex 5FBD–nucleotidase products. The cytosine moiety interacts with the protein using several types of interactions. Its π–conjugated system interacts with Phe81 and the peptide bond between Ala151 and Gly152. It has a direct polar interaction with the side chain of Asp83 and several water–mediated interactions, including involvement of water inside the NBS1 site. The 2'–deoxyribose moiety binds close to the zinc cluster and interacts directly with Lys68N$^{\zeta}$, Tyr69O$^{\eta}$, Asn154N$^{\delta 2}$, and the phosphate ion. A similar binding occurs in the complex 5FBF–nuclease products (for the dCyt moiety of 5'dCMP) and in the complex 5FBG–mutant with products. **(g)** The observed binding mode of 2'–deoxyguanosine (carbon–magenta) in the complex 5FBG–mutant with products, chain B. Binding of the pyrimidine–like part of guanine mimics the orientation of cytosine inside NBS1 as shown in panel (f). N7 of the imidazole–like part is involved in the water network (W3 and W_{dGua}) connecting this atom to Zn3, the phosphate ion, and Lys68N$^{\zeta}$. Asp154N$^{\delta 2}$ can interact with the π–conjugated electrons of the guanine moiety. **(h)** The observed binding mode of two molecules of 5'dCMP in the complex 5FBF–nuclease products (pH 4.2). The binding mode of the first 5'dCMP (carbon–pale pink) is almost identical with binding of 2'–deoxycytidine in the case of the complex 5FBD–nucleotidase products (panel f). The phosphate moiety is disordered and interacts either with Asn154 or with Lys68 and Tyr69. **(i)** The phosphate moiety of the second 5'dCMP in the complex 5FBF–nuclease products (carbon–magenta) binds in the zinc cluster in the second binding mode (as phosphate ion in 5FBD–nucleotidase products, panel f). The cytosine moiety likely interacts with Tyr69O$^{\eta}$ (hydroxyl group) through its π–conjugated system.

doi:10.1371/journal.pone.0168832.g003

Complex with inhibitor 5'dAMP(S) in remodeled NBS1. The crystal of the structure 5FBC–remodeled was obtained at pH 5.5. The structure contains one molecule of 2'–deoxyadenosine–5'–thiomonophosphate (5'dAMP(S)) in the active site (Fig 3E). Here, one molecule of the inhibitor binds simultaneously to both NBS1 and the zinc cluster. This complex also shows an alternative binding of the adenosine moiety (when compared with the binding mode present in 5FBB–inhibitors) in which NBS1 undergoes remodeling.

The thiophosphate moiety of 5'dAMP(S) interacts with the zinc cluster from the side of NBS1, binding only to Zn3 by its thiol group and to Lys68N$^{\zeta}$ by an oxygen atom. The deoxyribose moiety does not interact directly with protein and is disordered. The adenine moiety inside NBS1 is in the deep binding mode but different from that in the structure 5FBB–inhibitors (compare Fig 3D and 3E). The amino group still substitutes W_{NBS1} as in 5FBB–inhibitors but the adenine moiety interacts with Asp83 using the Watson–Crick face (Fig 3E and Figure E in S1 File).

Here NBS1 is remodeled when compared to all the previously published structures of the S1–P1 nuclease family members. The phenyl ring of Phe81 is rotated closer to Zn3 and makes an almost parallel stacking interaction with the adenine moiety of the ligand. The H–bond between the main chain of Ala151 and the side chain of Asn154 is broken, and the side chain of Asn154 is displaced in order to avoid clashes with the base and the 2'–deoxyribose moiety of the ligand. Remodeling of NBS1 will be described further and is also shown in Fig 4.

Complex with products of 3'–nucleotidase activity: phosphate and 2'–deoxycytidine (dCyt). The structure contains one phosphate ion and one molecule of dCyt in the active site (Fig 3C and 3F). This ternary complex (further called 5FBD–nucleotidase products) with the phosphate ion and dCyt represents binding of products of 3'nucleotidase activity and was obtained at pH 4.2. Both, the phosphate ion and the deoxyribose moiety of dCyt, are found in novel binding modes, observed here for the first time.

Two of the phosphate oxygen atoms bind directly to Zn1 and Zn2 (Fig 3C) and the ion also interacts with Asp65 and Lys68. All water molecules of the previously unoccupied zinc cluster (Fig 3A) are displaced but none of them is substituted by any oxygen atom of the phosphate ion (Fig 3C). We further refer to this interaction as the second binding mode of the phosphate ion.

The cytosine moiety of dCyt binds in NBS1 via stacking interactions with the side chain of Phe81 and the peptide bond Ala151–Gly152 (Fig 3F). The pyrimidine ring of dCyt is not placed exactly above the Phe81 side chain but slightly shifted out of NBS1. The deep position in NBS1 remains occupied by W_{NBS1}. Therefore, this nucleoside binding mode is denoted as

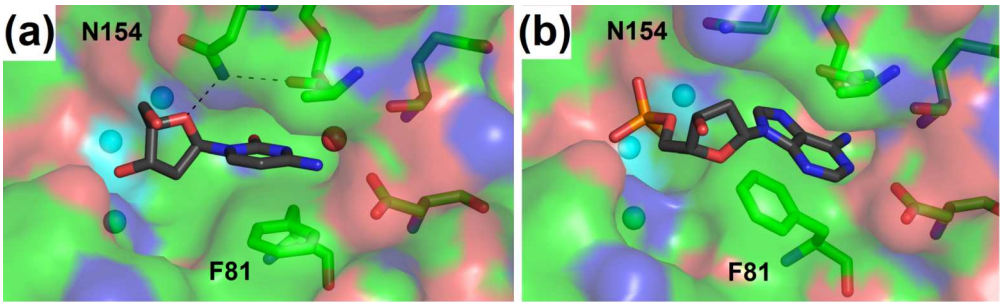


Fig 4. Remodeling of NBS1. The Nucleoside binding site 1 of S1 nuclease is represented as sticks (carbon–green). The solvent accessible surface of protein, color–coded by atom types, is shown. Zinc ions are shown as light blue spheres. Ligands are shown as sticks (carbon–grey). Selected interactions are shown as black dashed lines. The protein orientation in both panels is identical; changes of Phe81 and Asn154 can be seen. **(a)** The compact form of NBS1 with shallow base binding observed in the structure 5FBD–nucleotidase products. **(b)** The extended form of NBS1 with deep base binding observed in the structure 5FBC–remodeled. Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

doi:10.1371/journal.pone.0168832.g004

shallow. The difference between the deep and shallow binding modes of nucleobase can be seen in Fig 3D and 3F or in Fig 4A and 4B. Asp83 along with W_{NBS1} provide hydrogen bonding interactions to the base. The position of O1 of the cytosine moiety inside NBS1 allows interaction with the π–system of His145 (His145 is not shown in Fig 3F). The deoxyribose moiety binds via O4' to Asn154$N^{\delta 2}$. Its O3' interacts with Lys68N^{ζ} and Tyr69O^{η} and is placed approximately between the two atoms (Fig 3F and Figure F in S1 File).

Complex with products of dinucleotide cleavage: two molecules of 5'dCMP. The complex 5FBF–nuclease products was crystallized at pH 4.2 and contains two molecules of 2'–deoxycytidine 5'–monophosphate (5'dCMP) bound in the –1 and +1 positions with respect to the cleaved P–O3' bond (Fig 3H and 3I, and Figure G in S1 File). This ternary complex represents one of the possible binding states of single–stranded NA in the active site after cleavage.

One molecule of 5'dCMP binds to the active site in a similar way as 2'–deoxycytidine in the case of the complex 5FBD–nucleotidase products described above (compare Fig 3F and 3H). Its phosphate moiety is disordered and present in two main positions (Fig 3H and Figure G in S1 File). The phosphate moiety of the second molecule of 5'dCMP is bound inside the zinc cluster in the second phosphate binding mode (Fig 3H and 3I). Its deoxyribose moiety has no direct contact with the protein but is indirectly connected via a network of water molecules. Its oxygen atom O3' is positioned in the active site cleft and oriented towards the Half–Tyr site (described further) at a distance of about 12 Å from the side chain of Tyr183. The cytosine moiety is placed close to Tyr69 and it is likely that the hydroxyl group Tyr69O^{η} and the π–conjugated system of Tyr69 interact with the cytosine π–conjugated electrons; its position is slightly affected by a crystal contact (Figure G in S1 File).

Quaternary complex of mutant D65N with products of nucleotidase activity in the active site and a nucleoside in the secondary binding site. The complex 5FBG–mutant with products crystallized at pH 5.5 with two protein chains in the asymmetric unit shows the active site after mutation of the key catalytic residue Asp65 to Asn. Products of 3'monunucleotidase activity are bound in the active site: phosphate with 2'–deoxycytidine or 2'–deoxyguanosine. The third important feature is the identification of the secondary binding site in the active site groove based on binding of dCyt.

Asp65 is a part of the active site (Fig 1B) of wt S1 nuclease. The mutant D65N has the same tertiary structure as the wild type. The active site organization is also unchanged, with the only difference being the change of Asp65 to Asn.

A phosphate ion is bound inside the zinc cluster in the first binding mode in protein chain A (Figure H in S1 File). One molecule of dCyt binds in the active site as in the complex 5FBD–nucleotidase products (Figure H in S1 File). The second molecule of dCyt binds in the secondary binding site (Fig 5A and Figure H in S1 File) composed of Tyr183 providing a stacking interaction for the nucleobase and Glu177 providing hydrogen bonding. We refer to this site as "Half–Tyr site" with respect to the P1 nuclease Tyr site which is composed of two tyrosine residues [11]. The Half–Tyr site is a part of the active site groove (Fig 1A), about 13 Å distant from the catalytic zinc cluster, and exposed to solvent (Figs 1A and 5).

Chain B also contains a phosphate ion in the first binding mode. Simultaneously, one molecule of 2'–deoxyguanosine (dGua) is bound in the active site, interacting with NBS1 and the zinc cluster (Fig 3G and Figure H in S1 File). Interestingly, the placement of the pyrimidine part of the pyrimidine–imidazole ring system of dGua is the same as that of the cytosine moiety in the previous cases (compare Fig 3F and 3G). Additionally, the imidazole part interacts with a water network involving W3 and W_{dGua}, which is only present in this structure. These two water molecules mediate its interaction with Zn3, Lys68N^{ζ}, Asn154$O^{\delta1}$, and the phosphate ion within the zinc cluster. Asn154$N^{\delta2}$ likely interacts with the π–system of the guanine moiety. In this binding mode deoxyribose is oriented out of the active site without any direct interaction with the protein.

This complex was crystallized in the presence of self–complementary DNA $d(GC)_6$, however, the structure contains single nucleosides. The sample of S1D65N used in crystallization had a small residual activity. Self–complementary $d(GC)_6$ easily adopts Z–DNA conformation and can form double helical structures with single–stranded overhangs due to possible shifts in the hybridization process. Such DNA form is then susceptible to cleavage by S1 nuclease.

Catalytic activity

S1wt under the given reaction conditions (see Materials and Methods) digests RNA with a V_{lim} of 28.0 ± 3.0 ΔA_{260}/min/μg and a K_m of 0.16 ± 0.04 mg/ml. It digests ssDNA with a V_{lim} of 45.0 ± 3.0 ΔA_{260}/min/μg and a K_m of 0.14 ± 0.03 mg/ml. Activity towards dsDNA under standard reaction conditions is about ten times lower than towards ssDNA.

Inhibition by phosphate. Phosphate ion is often present in the active site of the structures reported here. The inhibition effect of inorganic phosphate was tested on ssDNase activity of S1wt under standard reaction conditions. Phosphate inhibits activity by 70% at 10 mM concentration and by 91% at 100 mM concentration.

Catalytic activity of the structure–based mutants S1D65N, S1K68N, S1N154A, and S1N154S. Residues Asp65, Lys68, and Asn154 are located in the active site and interact with ligands (Figs 1 and 3). Therefore they were identified as targets for structure–based mutational studies. Catalytic properties of the variants of S1 nuclease are reported here, in Fig 6 and Table 2; raw data can be found in Figure I in S1 File.

Under the standard reaction conditions, mutant S1D65N digests both single–stranded substrates roughly at a five hundred–fold lower rate than S1wt (0.16% of S1wt cleavage rate for RNA and 0.19% for ssDNA). Mutant S1K68N shows 27% of the S1wt cleavage rate for RNA and 41% for ssDNA. Mutant S1N154A has 60% of the S1wt cleavage rate for RNA and 52% for ssDNA, and a different mutant of the same residue, S1N154S shows 62% of the S1wt cleavage rate for RNA and 34% for ssDNA. The changes of activity are presented in Fig 6; for clarity all values in Fig 6 are related to the level of the ssDNAse activity being equal 100%.

Kinetics measured for both mutants S1N154A and S1N154S confirm a significant decrease of the catalytic activity expressed as V_{lim}/K_m, with the residue mutated to serine the effect being more pronounced. The mutant S1N154A shows inhibition by ssDNA (K_s = 0.3 ± 0.1

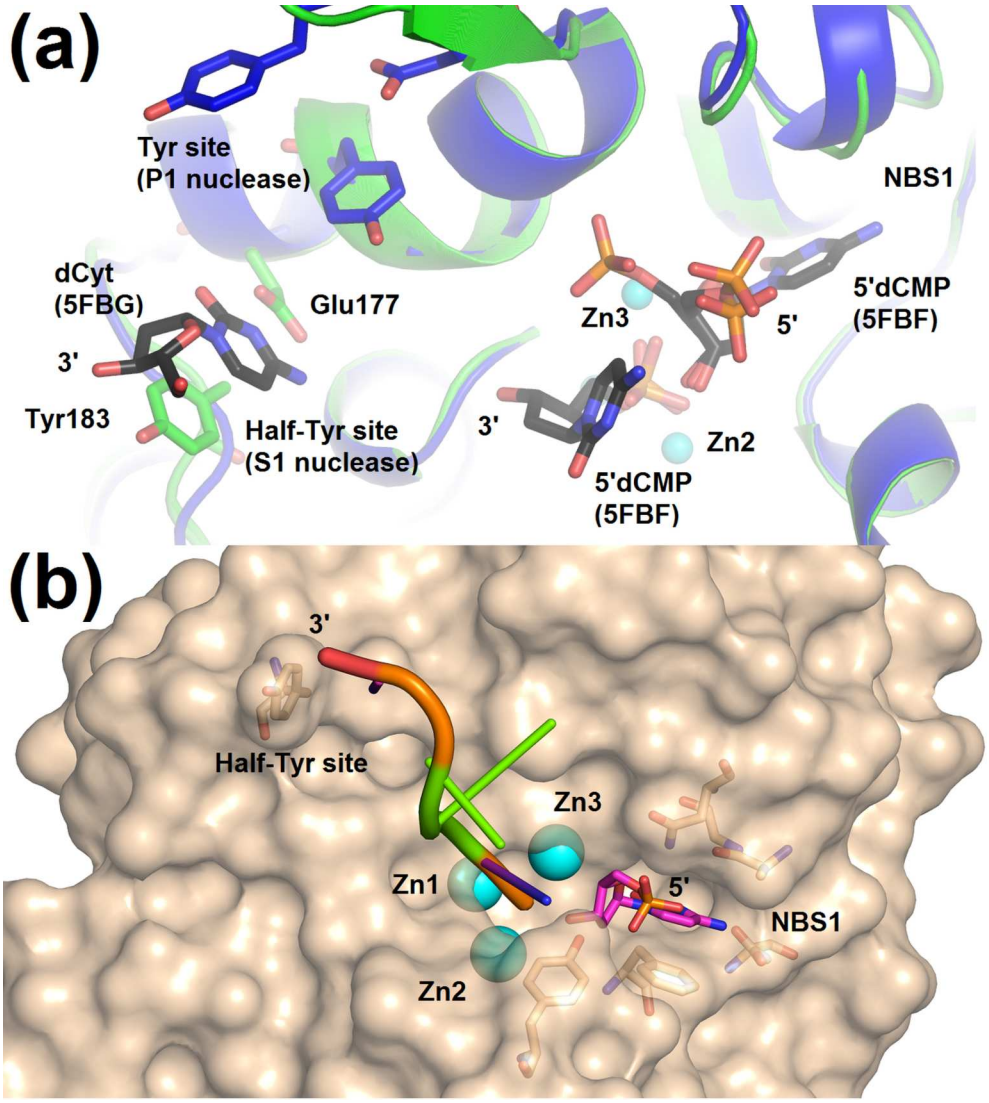


Fig 5. The Half–Tyr site of S1 nuclease and the proposed binding of ssDNA. (a) 2'–deoxycytidine (shown as sticks, carbon–magenta) bound in the Half–Tyr site and comparison of the position of the Half–Tyr site of S1 nuclease (carbon–green) with the position of the Tyr site of P1 nuclease (carbon–dark blue) and with respect to the active site. Both secondary sites are positioned at the same end of the active site cleft, albeit on the opposite "banks". Superposition of the complexes 5FBG–mutant with products, chain A, 5FBF–nuclease products, and of P1 nuclease (PDB ID: 1AK0 [11]) was calculated using the SSM Superpose tool in *Coot* [19]. Residues forming the Half–Tyr site are shown as sticks (carbon–green) and labelled. Residues of P1 nuclease involved in the formation of the Tyr site are shown as sticks (carbon–dark blue) and not labelled. The catalytic zinc ions (light blue spheres) with the phosphate ion (red/orange sticks) inside the zinc cluster are shown only for S1 nuclease. NBS1 is not shown but its position is marked. 2'–deoxycytidine and 5'dCMP binding in the active site cleft are shown (carbon–grey). **(b)** The proposed binding mode of ssDNA in the active site cleft based on the observed interactions of nucleotides and nucleosides in the S1 nuclease structures. The placement of the five nucleotides is based on x–ray structure coordinates except for the two nucleotides shown in green, which were positioned manually with optimized chain geometry in order to demonstrate the possible direction of ssDNA binding. Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

doi:10.1371/journal.pone.0168832.g005

mg/ml), as opposed to no inhibition in the mutant S1N154S or in the wild type. Changes of the kinetic parameters are summarized in Table 2.

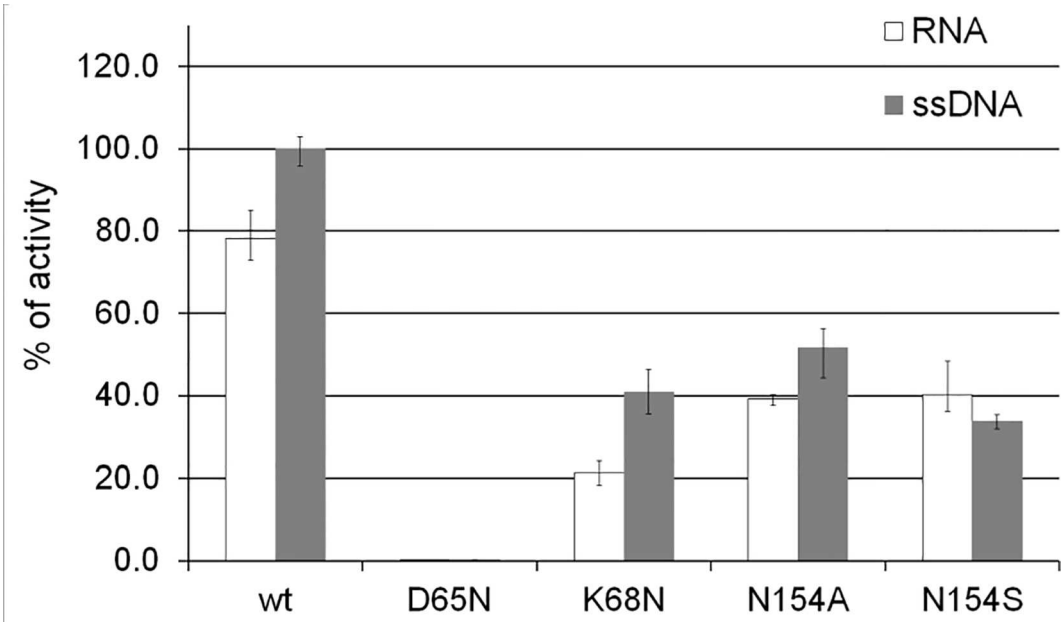


Fig 6. Comparison of the catalytic activity of S1 wild type and mutants. The catalytic activity of S1wt, S1D65N, S1K68N, S1N154A, and S1N154S is shown as a percentage of the S1wt activity on ssDNA.

doi:10.1371/journal.pone.0168832.g006

Discussion

Asp65 is critical for the reaction mechanism

The zinc cluster plays a major role in the catalytic mechanism [1, 2, 20]. Residues involved in coordination of the zinc ions are highly conserved in the S1–P1 nuclease family (Fig 2). Asp65 is involved in the coordination of Zn2 and in the reaction mechanism itself [11, 12]. It can play the role of a general base [21] or can assist in proper orientation of the activated water for the nucleophilic attack [11].

For the first time we provide a structure of a mutant at this position. In order not to severely influence the active site organization, Asp was replaced by Asn. The structure of the protein and organization of the active site are unaffected by this mutation and the enzyme can still bind ligands (see structure 5FBG–mutant with products, Fig 3G and Figure H in S1 File). This mutation basically abolishes the enzymatic activity (see Results and Fig 6), from which it follows that Asp65 is critical for the reaction mechanism. This observation supports the mechanism proposed previously [11, 12, 21], in which the water molecule is activated between Zn1 and Zn2 using Asp65 as a general base. The activated water molecule/hydroxide ion (labeled

Table 2. Comparison of kinetic parameters of nuclease S1 wild type and its mutants N154S and N154A using ssDNA and RNA as a substrate. Standard deviations are given.

	ssDNA				RNA		
Nuclease S1	V_{lim} [ΔA_{260nm}/min/µg]	K_m [µg/µL]	V_{lim}/K_m [ΔA_{260nm}µL/min/µg^2]	K_S [µg/µL]	V_{lim} [ΔA_{260nm}/min/µg]	K_m [µg/µL]	V_{lim}/K_m [ΔA_{260nm}µL/min/µg^2]
Wild type	45 ± 3	0.14 ± 0.03	321	–	28 ± 3	0.16 ± 0.04	175
N154A	55 ± 7	0.32 ± 0.06	172	0.3 ± 0.1	17 ± 2	0.14 ± 0.03	121
N154S	16 ± 1	0.12 ± 0.02	133	–	23 ± 1	0.39 ± 0.04	59

doi:10.1371/journal.pone.0168832.t002

W1 in Fig 3A) acts as a nucleophile in the in–line attack of the phosphorous atom with a consequential inversion of configuration of the scissile phosphate after cleavage.

Lys68 has multiple roles but is not critical for the reaction mechanism

The role of the residue at position 68 (Lys68 in S1 nuclease, Lys or Arg in the S1–P1 nuclease family, Fig 3) in the reaction mechanism was proposed previously [11, 12, 22] and investigated using quantum chemical calculations [21]. Lys68 interacts with ligands in six of the seven structures of S1 nuclease reported here (Fig 3), therefore we decided to investigate its importance by mutating to Asn. This mutation causes a decrease of activity to about one third of the original level of S1wt (Fig 6).

Based on the observed interactions of Lys68 with the phosphate ion in the active site (Fig 3B, 3C and 3E) and with products (Fig 3B, 3C, 3F, 3H and 3I) and adhering to the previous studies [11, 21] we propose that Lys68 has multiple functions in the active site; it can stabilize binding of a substrate or of an intermediate state of the scissile phosphate and it can also assist in the recycling of the active site (discussed further). Thus the decrease but not an entire cessation of activity of the K68N mutant can be explained given the fact that its role in the mechanism is not central, albeit important in assisting the right positioning and orientation of the molecular moieties during the enzymatic cycle. Moreover, it is not excluded that the side chain of Asn68 in the variant can still fulfill some of the original roles. We propose that altering the nature of the residue 68 can fine–tune the level of activity of these 3'nucleotidases/nucleases.

Asn154 is important for proper substrate binding but not essential for catalysis

Asn154, as a part of NBS1, is highly conserved in the whole S1–P1 nuclease family (Figure J in S1 File). Its role in the reaction mechanism was proposed previously [12] but the hypothesis was never tested.

Binding of a nucleobase inside NBS1 can be accompanied by a hydrogen bond between Asn154$N^{\delta 2}$ and O4' of 2'–deoxyribose. This interaction was observed in the structures 5FBD–nucleotidase products, 5FBF–nuclease products and 5FBG–mutant with products (Fig 3F and 3H), and previously in P1 nuclease (PDB ID: 1AK0 [11]) and AtBFN2 (PDB ID: 4CXO [14]). Based on our structures we designed a mutation, first, to a short apolar side chain (Ala) to disable any possibility of hydrogen bonding and, second, to a shorter polar side chain (Ser). In a simple comparison of the level of activity under identical conditions both mutations (S1N154A and S1N154S) decrease the rate of cleavage to about 60% for RNA and 30–50% for ssDNA as a substrate (Fig 6). This confirms the significance of the Asn side chain for the catalytic mechanism in this position. To discriminate the detailed effect of these mutations, the full kinetics were also measured (Table 2). Despite all the experimental effort our structures of complexes with Asn154 exposed to contacts with the sugar moiety contain only 2'deoxynucleotides, and therefore direct structural interpretation of the kinetic data can be attempted only for ssDNA as a substrate. The decrease of the overall efficiency on ssDNA of the N154A mutant is caused by a higher K_m, which complies with the assumed role of Asn154 in the proper binding of the substrate and is in agreement with the observed interaction Asn154Nδ2 –O4' in several structures. Removal of the Asn154 side chain leads to a decrease of the affinity of ssDNA substrate in the productive position. The mutation N154S restores the affinity but the enzyme is not capable of the original turnover, which could be caused by an improper orientation of the substrate. The mutant N154A also shows inhibition by substrate in the case of ssDNA, presumably caused by the lack of the Asn154$N^{\delta 2}$ –O4' interaction, possibly also leading to non–productive substrate binding.

Changes of K_m in these two mutants with RNA as substrate follow an inverse pattern compared to ssDNA (Table 2). The lack of structural information for possible ribose–Asn154 interactions impairs any direct interpretation of the data. Clearly, Asn154 plays a role also in RNA substrate binding and in the catalytic mechanism. There are differences compared to the ssDNase activity but the underlying details remain to be determined.

Asn154 is conserved throughout the S1–P1 nuclease family and has an important role in the catalytic mechanism, however it is not essential for activity. In the case of ssDNA it is involved in binding of 2'–deoxyribose and contributes to the formation of a productive enzyme–substrate complex.

S1 nuclease non–specificity is primarily caused by NBS1 promiscuity

NBS1 (Fig 1B) is responsible for the nucleoside binding at position –1 with respect to the cleaved P–O3' bond. This site is always present in the S1–P1 nuclease family. NBS1 is capable of accommodating not only unmodified natural nucleobases (this study) but also unrelated compounds (*e.g.* an arginine side chain in the case of TBN1 [12]). Promiscuity of NBS1 can be attributed to its capability to undergo remodeling and to offer varied H–bonding patterns.

Remodeling of NBS1. The nucleoside binding site 1 in the structures of S1 nuclease occurs in two distinct conformations. We further refer to these conformations as compact NBS1 and extended NBS1 (Fig 4).

Compact NBS1 is present in every structure of the members of the S1–P1 nuclease family published up to date [11–14, 18]. In this form NBS1 maintains the Ala151 main chain–Asn154 side chain H–bond (Fig 3D, 3F, 3G and 3H). The phenyl ring of Phe81 is located near Asp83 and in an almost parallel orientation with the plane of the peptide bond Ala151 –Gly152 (Fig 4A). The compact form of NBS1 enables the formation of a direct H–bond of Asn154$N^{\delta 2}$ to deoxyribose O4' of a ligand (Figs 3F and 4A).

Remodeling of NBS1 was reported for P1 nuclease previously [11] but the respective structure was not published. Remodeled NBS1 is present in several structures reported in this study. The complex 5FBC–remodeled contains extended (remodeled) NBS1 (with occupancy factor 0.8 to correspond with the occupancy of the ligand). Partially remodeled NBS1 is also present in the structures 5FB9 –unoccupied and 5FBA–phosphate. To the best of our knowledge, these are the first published coordinates of an enzyme from the S1–P1 nuclease family with remodeled NBS1.

In the extended (remodeled) form of NBS1 (*e.g.* in 5FBC–remodeled) the Phe81 phenyl ring is rotated closer to Zn3 to make an almost parallel stacking interaction with the adenine moiety of a ligand (Figs 3E and 4B). The side chain of Asn154 is displaced from its usual position in order to avoid clashing with the ligand; the direct sugar–Asn154 contact is thus disabled (Fig 4B).

The capability of NBS1 to undergo remodeling between the compact and extended forms allows for binding of substrates/products with varying sugar–base conformations and therefore can significantly contribute to the ligand binding promiscuity (Fig 4). One could speculate that NBS1 remodeling could also play its role in the catalytic mechanism especially with respect to various types of substrates, however, structure–function data in support of such hypothesis are missing. The part of NBS1 formed by the Ala151 –Gly152 peptide bond is also flexible. In the presented structures its position changes by up to 1 Å (not shown; based on a superposition of all the reported structures).

The conformational state of unoccupied NBS1 is not clear. Based on the structure 5FBA–phosphate, without any ligand in NBS1, we conclude that in the unliganded state the Asn154 side chain maintains the H–bond with the main chain of Ala151 while the side chain of Phe81 can be found in either of the two observed conformations (Figure C in S1 File).

Variable hydrogen bonding patterns. The second important feature responsible for the ligand binding promiscuity of NBS1 is a high degree of freedom in the formation of H–bonds with ligands as observed in the current structures (see Fig 3D to 3H). One of the key features of NBS1 of S1 nuclease is the presence of water W_{NBS1} (Fig 2B and other). Based on the utilization of this water molecule the nucleobase binding mode can be distinguished as shallow (W_{NBS1} is employed, Fig 3F) or deep (W_{NBS1} is replaced by a part of the nucleobase, Fig 3D). In the structures presented here, cytosine and guanine are found in the shallow binding mode, whereas adenine in the deep mode. These observations do not rule out the possibility of different types of binding for all nucleobases, especially when considering the high flexibility of NBS1 and the fact that at least one observed binding mode of adenine (Fig 3D) and the observed interaction of guanine (Fig 3G) are clearly not consistent with the most likely reaction mechanism (the position of the sugar moiety is too far from the zinc cluster). W_{NBS1} is also present in the structures of P1 nuclease (PDB ID: 1AK0 [11]) and AtBFN2 (PDB ID: 4CXO [14]). Both contain thymidine in the shallow binding mode. W_{NBS1} is also present in the structure of TBN1, PDB ID: 3SNG [12], with an arginine side chain bound inside NBS1. Although W_{NBS1} is located almost in the same position in all the known structures of S1–P1 nucleases, slight changes in its position and coordination may result in the existence of different binding modes for different nucleobases in distant members of the S1–P1 nuclease family.

The side chain of Asp83 (Asn in some cases, see Figure J in S1 File) is also involved in hydrogen bonding of a nucleobase. The protonation state of the Asp side chain depends on pH. The possibility of increased pK_a of the Asp side chain inside NBS1 was discussed in previous studies [11, 12]. An estimated value of pK_a of Asp83 using the Karlsberg+ server [23] is about 7. The binding modes of ligands observed in the reported complexes support protonation of the side chain of Asp83 as derived from the possible combinations of H–bond patterns in NBS1.

The experimental data within this study clearly show that both the NBS1 remodeling and the variable H–bonding patterns within NBS1 lead to a whole spectrum of ligand binding states of the enzyme. The extended and compact forms of NBS1 were observed in the currently reported structures in these combinations with the deep and shallow binding modes of a nucleobase: compact/deep (5FBB–inhibitors, Fig 3D), compact/shallow (5FBD–nucleotidase products, Figs 3F and 4A; 5FBF–nuclease products, Fig 3H, and 5FBG–mutant with products, Fig 3G) and extended/deep (5FBC–remodeled, Figs 3E and 4B). The combination extended/shallow was not observed, but is not excluded.

ssDNA binding as implied by structural data

The Half–Tyr site. One of the puzzling enigmas of P1 nuclease is the role of its second nucleobase binding site, called the Tyr site [11]. The confusion was caused by the fact that this site was not conserved in the sequences of many P1 homologues including S1 nuclease (Fig 2 and Figure J in S1 File). Although the P1–like Tyr site is not conserved in the S1 amino acid sequence, its role is most likely substituted by an alternative site–the Half–Tyr site. We propose that the role of the Tyr site in substrate binding in P1 is in S1 performed by the Half–Tyr site based on the structure 5FBG–mutant with products, which contains a molecule of dCyt in the Half–Tyr site (Fig 5A and Figure H in S1 File). From the comparison of the P1 and S1 structures it is clear that both sites are located in a similar position with respect to the active site and the active site cleft (Fig 5A). Tyrosine has no base preference and can adopt a wide range of conformations when stacked to nucleobases [24]. Moreover, the Half–Tyr site is entirely exposed to the solvent (Fig 5A). All these properties make this site highly promiscuous and hence useful for an unspecific nuclease, such as S1.

Many of the S1–P1 nuclease family members from fungi contain either the P1–like Tyr site or the S1–like Half–Tyr site (sequence alignment, Figure J in S1 File). The tyrosine residue is in some cases substituted by phenylalanine or histidine.

The Half–Tyr site is also likely present in many plant S1–P1 nucleases (Figure J in S1 File), although in the cases of distant homologues such predictions must be interpreted with caution. In TBN1 Tyr183 is conserved in the sequence (while Glu177 is substituted by Leu) but it is not present in the same position in the enzyme structure and, instead, interacts with one of the glycosylation oligosaccharides. In TBN1 in the position corresponding to the Half–Tyr site of S1 there are positive residues (Lys, Arg) presumably involved in binding of ds substrates [12]. AtBFN2 employs completely different binding sites for single–stranded nucleic acids [14]. These differences underline the great variability of the S1–P1 nuclease family and clearly show that one cannot easily predict the binding sites in these versatile nucleases merely by sequence analysis.

Simultaneous binding of –1 and +1 nucleotides of ssDNA in the active site. The ternary complex with two molecules of 5'dCMP as products (5FBF–nuclease products; Fig 3H and 3I) brings the first experimental evidence for the capability of an S1–P1 nuclease to simultaneously bind the +1 and –1 nucleotides (with respect to the scissile phosphate) with clear localization in the enzyme active site (Fig 3H). In the previous studies the general orientation of the cleaved single–stranded substrate was assumed [11, 12, 14]. Here, for the first time the way of the nucleotide chain placement in the active site cleft can be inferred from the structural data. It implies a tight turn of the ssDNA chain bound in the active site with the phosphate moieties of nucleotides -1, +1 and +2 placed at an angle of about 90 degrees. Such geometry is governed by the interactions of the -1 nucleobase with NBS1 and of the phosphate moiety with the zinc cluster.

Proposed binding of ssDNA. To the best of our knowledge, the ternary complex of two 5'dCMP molecules with S1 nuclease (5FBF–nuclease products, Fig 3H) brings the first structural evidence about the directionality of ssDNA binding in the active site cleft. Together with the assumed role of the Half–Tyr site (Fig 5A) we can propose the expected binding of ssDNA in the active site cleft of S1 nuclease (Fig 5B). The predicted interaction is characterized by the 90˚ turn (measured by phosphate ions) in the catalytic center and most likely no contacts between nucleotide +2 and protein. The nucleotide in position +3 partially interacts with the active site cleft and in position +4 binds in the Half–Tyr site.

Deoxyribose moiety after cleavage can bind at least in two distinct positions

Localization of the O3' oxygen of the scissile phosphate during the catalytic cycle is a key element in understanding substrate binding, catalysis, and product removal.

In our study, in three cases (5FBD–nucleotidase products, 5FBF–nuclease products, and 5FBG–mutant with products) the O3' oxygen of deoxyribose of products occurs in the vicinity of the zinc cluster but without direct interaction. The deoxyribose moiety is rotated away from the zinc cluster and the O3' oxygen binds between Lys68N^{ζ} and Tyr69O^{η}. This is in contrast with the situation in the structures of P1 nuclease (PDB ID: 1AK0 [11]) and of AtBFN2 (4CXO [14]), where the O3' oxygen of the ligands interacts with Zn3 of the cluster. (Fig 7). The structures 5FBD–nucleotidase products (Fig 3F) and 5FBF–nuclease products (Fig 3I), were obtained at pH 4.2, close to the S1 nuclease pH optimum for ssDNAse activity (around pH 4). The structures of P1 nuclease and AtBFN2 with product–like ligands (Fig 7B), were obtained at pH 5.3 and 7.5, respectively. The preference for the second binding mode of phosphate in our two structures may be correlated with pH (compare other structures of the series, Table 1 and Fig 3) and this mode excludes a concurrent interaction of deoxyribose O3' with Zn3.

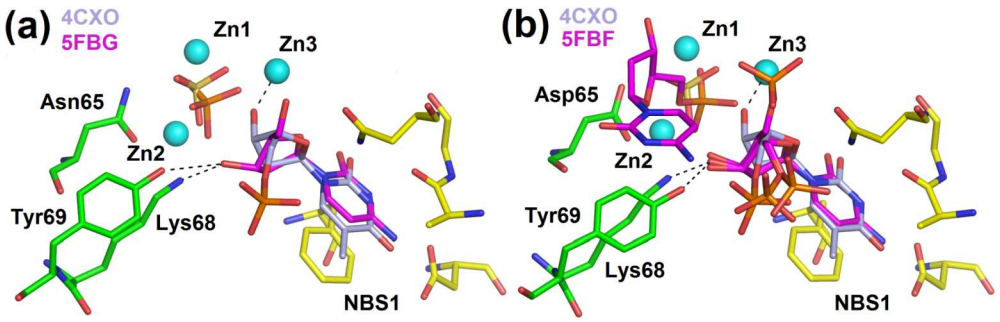


Fig 7. Two types of interactions of the O3' oxygen with the active site in the S1–P1 nuclease family. Zinc ions are shown as light blue spheres. Asp65 (Asn65 in the case of the mutant) is shown in sticks (carbon–green). Other residues involved in the zinc cluster coordination are not shown. Lys68 and Tyr69 interact with ligands and are shown as sticks (carbon–green). Residues forming NBS1 are shown as sticks (carbon–yellow). Phosphate ion is shown as orange/red sticks, sulfate ion as yellow/red sticks, and selected interactions as black dashed lines. Molecular graphics were created using *PyMOL* (Schrödinger, LLC). **(a)** Comparison of the deoxyribose binding and position of its O3' in the structure 5FBG–mutant with products with its position in the structure of AtBFN2 (PDB ID: 4CXO [14]). The zinc cluster, residues involved in the interactions and NBS1 are shown only for S1 nuclease. Ligands present in the structure of S1 nuclease are phosphate and dCyt (carbon–magenta). Ligands present in AtBFN2 are sulfate and thymidine 5'–monophosphate (carbon–silver). Notice the difference in the positions of the O3' oxygens and of the phosphate ion in S1 –excluding the interaction of O3' with Zn3. **(b)** Comparison of deoxyribose binding and the position of its O3' in the structures 5FBF–nuclease products and AtBFN2 (PDB ID: 4CXO [14]). The zinc cluster, residues involved in the interactions and NBS1 are shown only for S1 nuclease. Ligands present in the structure of S1 nuclease are two molecules of 5'dCMP (carbon–magenta). Ligands of AtBFN2 are displayed as in panel (a). Notice the difference in the positions of the O3' oxygens and of the phosphate moiety in the case of S1 which excludes binding of O3' to Zn3.

doi:10.1371/journal.pone.0168832.g007

Being the only available structures of ternary complexes with both products and at the optimal pH for the respective activity we propose that these can represent the situation in the active site right before the products of the nucleotidase (5FBD) and nuclease (5FBF) activity are released. The phosphate moiety already escaped the tight first binding mode complying with the cleavage step and O3' of deoxyribose of the nucleotide in position –1 left Zn3 and, instead, interacts with Lys68 and Tyr69.

Variability of phosphate binding

Inorganic phosphate is a product of 3'nucleotidase activity and, as shown in the Results section, also an inhibitor. Only the first binding mode of phosphate (Fig 3B) was observed in the previous structures of the members of the S1–P1 nuclease family till now [14, 18]. It follows from our results that two main binding modes of the phosphate ion in the active site of the wild type of S1 nuclease can be distinguished (Fig 3B and 3C). In both cases the ion interacts with all three zinc ions, Asp65 and Lys68. The difference lies in the position of its oxygen atoms with respect to the zinc ions (compare Fig 3B and 3C). The first binding mode is realized in this study in the structures crystallized at pH 5.5 and 6.5. The second binding mode is observed in this study only in the case of crystallization at pH 4.2 and is accompanied by binding of 2'–deoxycytidine or 5'dCMP in NBS1 (Fig 3F, 3H and 3I). This implies a certain role of pH in the formation of the interactions of the phosphate ion/moiety with the cluster, however, the influence of the nucleotides/nucleosides present in NBS1 on the phosphate binding mode cannot be neglected and the current results do not allow a simple interpretation.

To the best of our knowledge, an interaction of the phosphate moiety of a mononucleotide/oligonucleotide with the active site in the S1–P1 nuclease family was never observed before. In our structures we observe three different ways of binding of the phosphate moiety of 5'–

mononucleotide to S1 nuclease. First, it can mimic the interactions of the free phosphate ion (structure 5FBF–nuclease products, Fig 3I). Second, the thiophosphate moiety of 5'dAMP(S) can interact with the zinc cluster from an outer position (structure 5FBC–remodeled, Fig 3E) and, third, the phosphate moiety can also interact with NBS1 instead of the zinc cluster (structure 5FBB–inhibitors, Fig 3D).

5'–mononucleotides can inhibit S1 nuclease via several distinct binding modes

Three types of binding of a 5'–mononucleotide (product and simultaneously inhibitor [2]) in the active site were observed in the structures of S1 nuclease reported here.

In the first case two molecules of 5'dCMP mimic one of the possible arrangements of products after cleavage (structure 5FBF–nuclease products, Fig 3H and 3I, discussed above) and at the same time they show a two–molecule inhibition mechanism. This complex confirms the capability of a 5'mononucleotide to inhibit this nuclease not only as a ligand of NBS1 (–1 position) but also as a ligand of the zinc cluster (+1 position).

In the second case (5FBC–remodeled) the thiophosphate moiety of 5'dAMP(S) interacts with the zinc cluster from an outer position. Contrary to the complex with 5'dCMP, binding of 5'dAMP(S) is marked by the phosphate moiety missing from the zinc cluster with a simultaneous remodeling of NBS1 to the extended form (Figs 3F and 4).

In the third observed case a nucleotide/nucleoside binds in the way that its sugar and phosphate moieties are positioned too far from the zinc cluster to participate in interactions (structures 5FBG–mutant with products, chain B, Fig 3G and 5FBB–inhibitors, Fig 3D). The phosphate moiety either interacts with NBS1 (Fig 3D) or is disordered with no interaction with the protein (Fig 3G). In this type of complexes a phosphate ion can be present inside the zinc cluster at the same time.

We conclude that inhibitors of this nuclease family can utilize the zinc cluster but can also inhibit without direct interactions with the cluster. The most probable binding mode of an inhibitor depends on the nature of the nucleobase, on the presence of phosphate ions and pH. It is likely that a mixture of the different binding modes of a given inhibitor exists in solution and also that new, so far unobserved, binding modes are possible.

Specificity–related differences between S1 nuclease and other members of the S1–P1 family are defined by the enzyme surface

The fold of S1 nuclease is very similar to those of nucleases P1, TBN1, and AtBFN2 (Fig 2). Residues involved in the coordination of the zinc cluster, the glycosylation site Asn112, as well as two disulfide bridges present in S1 nuclease, are conserved (Fig 2). The main structural differences lie in the facilitation of the secondary substrate binding sites, which largely determines specificity, in the closely related charge distribution on the enzyme surface and in the enzyme shape (Figure K in S1 File). The overall structural features of mature S1 nuclease as well as a comparison with P1, TBN1 and AtBFN2 nucleases are shown in the amino acid sequence alignment in Fig 2.

Glycosylation has a minor role in S1 nuclease stabilization and does not affect activity

S1 nuclease has two N–glycosylation sites. The Asn112 site is conserved in all members of this family with known structure (Fig 2) and most probably also in the whole family (Figure J in S1 File). Under normal conditions, oligosaccharides account for about 18% of the mature protein

mass [2]. Elimination of the major part of oligosaccharides by Endo F1 neither decreases activity (Figure A in S1 File) nor causes the enzyme to unfold (presented 3D structures) or aggregate (verified by DLS, Figure A in S1 File). This is in contrast with the behavior of the so far studied plant homologues TBN1 and AtBFN2, losing stability and solubility upon deglycosylation [11, 25]. A slight decrease of its thermal stability is the only observed effect caused by S1 nuclease deglycosylation (Figure L in S1 File). It should be noted that the only so far studied bacterial member of the S1–P1 nuclease family, M1 nuclease from *Mesorhizobium loti*, naturally lacks glycosylation and yet is stable and active [6].

Conclusions

1. The three–dimensional structure of S1 nuclease revealed new specific features including the Half–Tyr site and NBS1 remodeling and confirmed conservation of some of the main features of the S1–P1 family, including the helical fold with the central cluster of three zinc ions involving the N–terminal Trp26.

2. Asp65, coordinating one of the zinc ions, is essential for activity. Its mutation to asparagine does not affect the active site organization but abolishes activity. Lys68 on the NBS1–distal side of the active site is not essential but its mutation decreases activity. Asn154 of NBS1 is not essential for activity but, amongst others, is involved in interactions with deoxyribose of ssDNA substrate. Kinetic data for ssDNA activity of the Asn154 mutants can be explained by changes in substrate affinity.

3. NBS1 can accommodate a nucleobase in different binding modes–shallow and deep. In the shallow mode a nucleobase interacts with water W_{NBS1}; in the deep mode this water molecule is replaced by a part of the nucleobase. In addition to the shallow and deep binding modes, upon ligand binding NBS1 can undergo substantial remodeling depending on the character of the ligand. The two most extreme observed states of NBS1 –the compact and the extended forms, distinguished by conformations of Phe81 and Asn154 –bind their ligands in positions differing in the orientation of the base plane by about 30 degrees and in the position of the (deoxy)ribose ring. NBS1 can thus bind substrate in two or more different conformations.

4. A new accessory nucleotide binding site, named Half–Tyr site, was identified by ligand binding. It utilizes Tyr183 and Glu177 and is located about 13 Å from the zinc cluster downstream along the active site cleft. Its existence explains the absence of the Tyr site, previously found in P1 nuclease, in sequences of many fungal members of the S1–P1 nuclease family. Most likely all nucleases of this family require a second ligand binding site apart from NBS1, which is located across the zinc cluster. Such site can be realized as a more specific nucleobase binding site (Tyr site, Half–Tyr site) preferred by single–strand–specific enzymes, or by positive surface patches for double strand activity (TBN1, [12]). Based on our structural data a placement of an ssDNA strand in the active site cleft of S1 nuclease is proposed with the implication of a 90˚ turn of the chain in the zinc cluster.

5. Four different ways of the interaction of phosphate with S1 nuclease were observed, three of them for the first time in the S1–P1 nuclease family. There are two different binding modes of phosphate inside the zinc cluster; the phosphate moiety of a nucleotide can interact with the zinc cluster via binding only to Zn3 and not inside the cluster; it was also observed to interact with NBS1. These phosphate binding modes are likely affected by pH and possibly also by the binding mode of the carrying or co–binding nucleotide or nucleoside in the active site.

6. Five different binding modes of nucleotides and nucleosides in the active site were observed distinguished by the position and orientation of the ligand base and by placement of the phosphate moiety. Two of them represent binding of both products after cleavage of a dinucleotide or of a mononucleotide. The remaining ones are considered inhibitor binding. In four cases two molecules, either of the same ligand or two different ligands, were bound in the active site. The wide range of the utilized interactions shows the true diversity of the nucleic acid binding patterns in the S1–P1 family, which explains its universality and non–specificity. It also broadens the set of interaction points which can be utilized in inhibitor design.

Materials and Methods

Expression and purification

Cloning of S1 nuclease wild type. Gene for S1 nuclease (*A. oryzae* RIB40 gene Ao09000 1000075) was amplified from *A.oryzae* strain IFO4177 by primers 5′–GACGCGGCCGCACC ATGCCGCGCTTACTCCC and 5′–GACGCGATCGCTCAAGAGGGCTGACTCG having overhangs with recognition for restriction endonuclease sites, NotI and SgfI, respectively. The amplified DNA (band of 988 base pairs) was digested with restriction endonucleases NotI and SgfI and the resulting 977–base–pair product was cloned into the corresponding restriction sites of an expression vector.

Construction of mutated variants of S1 nuclease. Genes of all variants were generated by spliced overlap extension (SOE) polymerase chain reaction (PCR) with flanking primers 5′–AA CTGGGGATCCACCATGCCGCGCTTACTCC (forward) and 5′–ACCAGGTCTTAAGTCAAGAGG GCTGACTCGCAATC (reverse) and hybrid primers. The resulting nuclease variant genes were cloned into an expression vector as a BamHI–AflII fragment using standard molecular biology techniques.

Expression and purification. After verification by DNA sequencing, constructs were transformed into protoplasts of *Aspergillus oryzae* for expression driven by the TAKA amylase promoter. The transformed strain of *A. oryzae* was typically grown for 4 days at 30˚C in DAP–4C medium (11 g $MgSO_4$–$7H_2O$, 1 g KH_2PO_4, 2 g citric acid, 20 g dextrose, 10 g maltose, 5.2 g K_3PO_4–H_2O, 0.5 g yeast extract, 0.5 ml trace metals, 0.5 g $CaCO_3$, 23 ml of a 50% solution of $(NH_4)_2HPO_4$, 33 ml of a 20% solution of lactic acid per liter). The fermentation broth was sterile filtered to remove fungal hyphae. Salts and other low molecular weight solutes were removed by ultrafiltration. 1 M Tris–HCl, pH 7.5 was added to the resulting retentate to a final concentration of 25 mM. pH and ionic strength were determined to be within the acceptable range for anion exchange chromatography. The chromatography was then conducted with an ÄKTA Prime instrument (Amersham Biosciences). Briefly, the protein was bound to a column with 20 ml Q Sepharose High–Performance pre–equilibrated with 25 mM Tris–HCl, pH 7.5. After a thorough wash with the equilibration buffer, the bound protein was eluted from the column with a linear NaCl gradient (0–0.5 M) in the equilibration buffer over ten column volumes. Collected fractions containing pure nuclease, as estimated by SDS–PAGE, were pooled. All purification steps were carried out at room temperature.

Deglycosylation

All samples were enzymatically deglycosylated in order to improve crystallizability. Two different deglycosylation enzymes were tested on S1wt: Endoglycosidase F1 (Endo F1) from *Elizabethkingia miricola* (EC 3.2.1.96) and α–Mannosidase (α–Mann) from *Canavalia ensiformis*

(EC 3.2.1.24). All reactions were done in 100 mM sodium acetate, pH 4.5 with 1 mM $ZnCl_2$. For 0.1 mg of S1 0.1 μg of Endo F1 or α–Mann was used. Reactions were carried out at 37˚C for 1 h. Cleavage was monitored by SDS–PAGE (Figure A in S1 File). Oligomeric state of both samples after deglycosylation was monitored by dynamic light scattering and isoelectric focusing (Figure A in S1 File). The activity of S1 nuclease deglycosylated by Endo F1 (S1–Endo F1) was compared to the activity of the fully glycosylated version (Figure A in S1 File). S1D65N was deglycosylated only with Endo F1 prior to crystallization. All chemicals were purchased from Sigma–Aldrich.

Nuclease activity

The activity of all S1 nuclease variants was measured towards commercially available isolated nucleic acids. The reaction mixtures contained 50 μl of native DNA from calf thymus (dsDNA), heat–denatured DNA from calf thymus (ssDNA) or RNA from torula yeast (concentration 1 mg/ml in 0.1 M sodium acetate buffer, pH 4.5 containing 50 mM NaCl), and 50 μl of the enzyme diluted in the same buffer. All reactions were carried out at 37˚C for 5 min. These assay settings are referred to as standard reaction conditions. Each reaction was stopped by adding 250 μl of 96% ethanol. The mixture was vortexed and incubated at –20˚C for 20 min. The precipitated undigested substrate was centrifuged (22 000 x g, 20 min, 4˚C) and the absorbance of the supernatant was measured at 260 nm. Each measurement was performed in triplicate. Separate background readings for individual concentration points of all substrates were used in all cases. The inhibitory effect of phosphate was tested by measuring activities towards ssDNA in the presence of inorganic phosphate at two different concentrations (10 mM and 100 mM). The reaction was performed as described above. All chemicals were purchased from Sigma–Aldrich.

Calculation of kinetic parameters. Specific activities of S1wt and mutants N154S and N154A were determined using ssDNA or RNA as a substrate. The values of K_m, and V_{lim} were calculated using the iterative method for statistical evaluation of deviations in the tool Solver (Microsoft Excel). The following equation was used to calculate the inhibition constant, where [*S*] represents substrate concentration, v_0 initial velocity, K_m is Michaelis constant, V_{lim} maximum velocity and K_S is substrate inhibition constant [26]:

$$v_0 = \frac{V_{lim}.[S]}{K_m + \left(1 + \frac{[S]}{K_S}\right).[S]}$$

Crystallization

For initial crystallization trials fully glycosylated S1wt as well as S1wt deglycosylated with Endo F1 and α–Mann were used. Prior to crystallization, all samples were transferred to 25 mM Bis–Tris pH 6.0 with the addition of 50 mM NaCl by several cycles of concentration/dilution using a Nanosep® centrifugal device with 10 kDa cut–off (Pall Corporation) and concentrated to 25 mg/ml. The Index crystallization screen (Hampton Research) and the hanging drop vapor diffusion method at 18˚C with the ratio of the protein to reservoir drop volume 1:1 (0.4 μl + 0.4 μl) were used as the screening setup. Crystals appeared only in the case of S1–Endo F1 and grew in several conditions in 3 to 21 days. The most promising crystals originated from conditions no. 40 (0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3350), no. 54 (0.05 M $CaCl_2$, 0.1 M Bis–Tris pH 6.5, 30% v/v Polyethylene glycol monomethyl ether 550) and no. 70 (0.2 M NaCl, 0.1 M Bis–Tris pH 5.5, 25% w/v Polyethylene glycol 3350). Crystals used in this study were obtained using the original conditions no. 54 and no. 70, the optimized condition no. 40 (0.1 M Citric acid pH 3.8, 25% w/v Polyethylene glycol 3350), and the optimized

condition no. 70 (0.05 M $CaCl_2$, 0.2 M NaCl, 0.1 M Bis–Tris pH 5.5, 25% w/v Polyethylene glycol 3350) as reservoir solutions. All crystals were obtained using the hanging drop or the sitting drop vapor diffusion methods at 18˚C with the ratio of protein to reservoir drop volume 1:1 (0.4 μl + 0.4 μl). For a more precise determination of pH in crystallization drop mixtures of storage buffer with reservoir solutions (without polymer) in ratio 1:1 were prepared. The measured values of pH were within 0.1 unit difference from the values of the corresponding reservoir buffers in the cases of nos. 54 and 70 whereas for the optimized condition no. 40, the resulting value was 4.2 compared to the original 3.8. S1D65N –Endo F1 was crystallized using the condition no. 70 and the same setup as in the case of S1wt.

Crystals of 5FBA–phosphate were obtained without any co–crystallization effort with protein concentration 25 mg/ml. For co–crystallization experiments solutions of ligands were mixed with S1 nuclease in v/v ratio 1:10. The resulting protein concentration was always 22.5 mg/ml. Mixtures were incubated for 1 h at room temperature and then crystallized as described above. Successful co–crystallization was achieved with adenosine 5'–monophosphate (5'AMP) and 2'–deoxycytidine 5'–monophosphate (5'dCMP) as a ligand. The structure 5FB9 –unoccupied was obtained from unsuccessful co–crystallization with 2'–deoxyguanosine (dGua). The final concentration of the above mentioned ligands in the mixtures was 10 mM. Three other complexes were obtained using thiophosphorylated dinucleotides and dsDNA. The final concentration of thiophosphorylated 2'–deoxyadenosine dinucleotide dA(pS)dA in the protein–ligand mixture was 1 mM, of thiophosphorylated 2'–deoxycytidine dinucleotide dC(pS)dC 1.8 mM, and of dsDNA $d(GC)_6$ 1.2 mM. The latter three nucleic acids were purchased from Generi Biotech (Czech Republic). Unless stated otherwise, all chemicals were purchased from Sigma–Aldrich. Selected crystallization parameters of the reported structures are given in Table 1. Further crystallization details are summarized in Table A in S1 File.

X–ray diffraction analyses, structure solution, and refinement

Single crystals obtained from the conditions based on Index no. 40 and no. 70 were equilibrated in reservoir solution containing 20% (v/v) glycerol as a cryoprotectant for 15 seconds. There was no need for the addition of cryoprotectant in the case of crystals from the condition no. 54. All crystals were mounted in round LithoLoops (Molecular Dimensions) of appropriate size and vitrified in liquid nitrogen. X–ray diffraction data for 5FBD–nucleotidase products were collected at 120 K using a Gemini Enhanced Ultra diffractometer with an Enhanced Ultra copper (Cu) source and an Atlas CCD detector (Agilent Technologies). Diffraction data were processed using CrysAlis^{Pro} (Agilent Technologies), and scaled and merged using Aimless [27] from the CCP4 suite [28]. Data for 5FB9 –unoccupied were collected at the beamline P13 of the synchrotron radiation source DESY: PETRA III, Hamburg using a Dectris Pilatus 6M–F detector and a Maatel MD2 micro–diffractometer with mini kappa goniometer at 100 K. All the remaining data were collected at the synchrotron radiation source BESSYII, Helmholtz Zentrum Berlin [29] at 100 K. Data for 5FBA–phosphate, and 5FBF–nuclease products were collected at beamline BL14.1 using a Dectris Pilatus 6M detector and a mini kappa goniometer. Data for 5FBB–inhibitors and 5FBG–mutant with products were collected at the beamline BL14.2 using a MAR Mosaic CCD 225 detector and a MAR Research Desktop Beamline goniometer. Data for 5FBC–remodeled were collected at the beamline BL14.3 using a MARmosaic CCD 225 detector and a MAR Research dtb goniometer. Data collected at DESY: PETRA III and BESSYII were processed either using XDSGui [30] or XDSapp [31], scaled using CORRECT and merged using Aimless [27], or alternatively processed using iMosflm [32] and scaled and merged using Aimless [27]. Selection of the processing programs was

based on the quality of data statistics and electron density. Data collection and processing statistics are given in detail in Table B in S1 File. Selected statistics are reported in Table 1.

The phase problem for the first structure (5FBA–phosphate) was solved by molecular replacement in MOLREP [33] using the structure of P1 nuclease as a model (PDB ID: 1AK0 [11]). The phase problem for all the other structures was solved using the protein chain of 5FBA–phosphate as a model. All structures were built using Coot [34] and refined in REFMAC5 [35] using restrained refinement. Restraints for nucleotides and nucleosides were generated using the Grade Web Server (grade.globalphasing.org). Standard restraints of the CCP4 library [28] were used for all the other ligands and moieties. R_{free} was used as a cross validation method in the refinement. Atomic displacement parameters (ADPs) were refined as isotropic for all structures except 5FBF–nuclease products, for which ADPs were refined as anisotropic due to the atomic resolution of the structure. Hydrogen atoms in riding positions generated by REFMAC5 [35] were used in all refinements. Automatically generated local NCS was used in the cases of 5FB9 –unoccupied, 5FBB–inhibitors and 5FBG–mutant with products. The quality of each structure was assessed using the set of the validation tools in Coot [34], the validation services provided by wwPDB [36] and MolProbity [17]. Detailed refinement statistics are given in Table B in S1 File. Selected statistics are reported in Table 1. Coordinates and structure factors were deposited in the Protein Data Bank under accession numbers: 5FB9 (5FB9 –unoccupied), 5FBA (5FBA–phosphate), 5FBB (5FBB–inhibitors), 5FBC (5FBC–remodeled), 5FBD (5FBD–nucleotidase products), 5FBF (5FBF–nuclease products), and 5FBG (5FBG–mutant with products).

Supporting Information

S1 File. Supporting information with additional Materials and Methods, Figures, and Tables.
(PDF)

S1 PDBvalidationreports. Validation reports for the reported PDB entries.
(ZIP)

Acknowledgments

We would like to thank Dr. Michal Dušek for assistance with crystal testing and data collection on the in–house diffractometer at the Institute of Physics CAS CZ, and Dr. Müller and Dr. Weiss of the Helmholtz–Zentrum Berlin for support at BESSY II.

Author Contributions

Conceptualization: TK J. Dohnálek.

Formal analysis: TK PL J. Dušková TS MT PK KF JS LS J. Dohnálek.

Funding acquisition: J. Dohnálek.

Investigation: TK LHO JL AN PL J. Dohnálek.

Project administration: TK J. Dohnálek.

Resources: TK LHO JL AN PL J. Dušková TS MT PK KF JS LS JH J. Dohnálek.

Supervision: TK J. Dohnálek.

Validation: TK J. Dušková TS KF PK J. Dohnálek.

Visualization: TK J. Dohnálek.

Writing – original draft: TK J. Dohnálek.

Writing – review & editing: TK LHO AN PL J. Dušková TS MT PK KF JS LS JH J. Dohnálek.

References

1. Yang W. Nucleases: diversity of structure, function and mechanism. Q Rev Biophys. 2011; 44(1):1–93. doi: 10.1017/S0033583510000181 PMID: 20854710
2. Desai NA, Shankae V. Single–strand–specific nucleases. FEMS Microbiol Rev. 2003; 26(5):457–491. PMID: 12586391
3. Balabanova LA, Pivkin MV, Rasskazov VA. The Distribution and Substrate Specificity of Extracellular Nuclease Activity in Marine Fungi. Open Journal of Marine Science. 2012; 2:188–195.
4. Matousek J, Podzimek T, Pouckova P, Stehlik J, Skvor J, Lipovova P, et al. Antitumor activity of apoptotic nuclease TBN1 from *L. esculentum*. Neoplasma. 2010; 57:339–348. PMID: 20429625
5. Lesniewicz K, Karlowski WM, Pienkowska JR, Krzywkowski P, Poreba E. The plant S1–like nuclease family has evolved a highly diverse range of catalytic capabilities. Plant Cell Physiol. 2013; 54:1064–1078. doi: 10.1093/pcp/pct061 PMID: 23620482
6. Pimkin M, Miller CG, Blakesley L, Oleykowski CA, Kodali NS, Yeunq AT. Characterization of a periplasmic S1–like nuclease coded by the *Mesorhizobium loti* symbiosis island. Biochem Biophys Res Commun. 2006; 343:77–84. doi: 10.1016/j.bbrc.2006.02.117 PMID: 16529713
7. Oleson EA, Sasakuma M. S1 nuclease of *Aspergillus oryzae*: a glycoprotein with an associated nucleotidase activity. Arch Biochem Biophys. 1980; 204:361–370 PMID: 6252849
8. Vogt VM. Purification and further properties of single–strand–specific nuclease from *Aspergillus oryzae*. Eur J Biochem. 1973; 33:192–200 PMID: 4691350
9. Rushizky GW. S1 nuclease of *Aspergillus oryzae*. Gene Amplif Anal. 1981; 2:205–215 PMID: 6101052
10. Volbeda A, Lahm A, Sakiyama F, Suck D. Crystal structure of *Penicillium citrinum* P1 nuclease at 2.8 A resolution. EMBO J. 1991; 10:1607–1618 PMID: 1710977
11. Romier C, Dominguez R, Lahm A, Dahl O, Suck D. Recognition of single–stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. Proteins. 1998; 32:414–424 PMID: 9726413
12. Koval T, Lipovova P, Podzimek T, Matousek J, Duskova J, Skalova T, et al. Plant multifunctional nuclease TBN1 with unexpected phospholipase activity: structural study and reaction–mechanism analysis. Acta Crystallogr. 2013; D69:213–226.
13. Chou TL, Ko TP, Ko CY, Lin TY, Guo RT, Yu TF, et al. Mechanistic insights to catalysis by a zinc–dependent bi–functional nuclease from *Arabidopsis thaliana*. Biocatal Agric Biotechnol. 2013; 2:191–195
14. Yu TF, Maestre–Reyna M, Ko CY, Ko TP, Sun YJ, Lin TY, et al. Structural Insights of the ssDNA Binding Site in the Multifunctional Endonuclease AtBFN2 from *Arabidopsis thaliana*. PLoS ONE. 2014; 9(8): e105821. doi: 10.1371/journal.pone.0105821 PMID: 25157844
15. Podzimek T, Matousek J, Lipovova P, Pouckova P, Spiwok V, Santrucek J. Biochemical properties of three plant nucleases with anticancer potential. Plant Science. 2011; 180:343–351. doi: 10.1016/j.plantsci.2010.10.006 PMID: 21421379
16. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014; 42(W1):W320–W324.
17. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all–atom structure validation for macromolecular crystallography. Acta Crystallogr. 2010; D66:12–21.
18. Stransky J, Koval T, Podzimek T, Tycova A, Lipovova P, Matousek J, et al. Phosphate binding in the active centre of tomato multifunctional nuclease TBN1 and analysis of superhelix formation by the enzyme. Acta Crystallogr. 2006; 71: 1408–1415.
19. Krissinel E, Henrick K. Secondary–structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. 2004; D60:2256–2268.
20. Gite S, Shankar V. Characterisation of S1 nuclease. Eur J Biochem. 1992; 210:437–441. PMID: 1281097
21. Liao RZ, Yu JG, Himo F. Phosphate mono–and diesterase activities of the trinuclear zinc enzyme nuclease P1 –insights from quantum chemical calculations. Inorg Chem. 2010; 49:6883–6888. doi: 10.1021/ic100266n PMID: 20604512

22. Gite S, Reddy G, Shankar V. Active–site characterization of S1 nuclease I. Affinity purification and influence of amino–group modification. Biochemical Journal. 1992; 285:489–494 PMID: 1637340

23. Kieseritzky G, Knapp EW. Optimizing pKA computation in proteins with pH adapted conformations. Proteins. 2008; 71:1335–1348. doi: 10.1002/prot.21820 PMID: 18058906

24. Wilson KA, Kellie JL, Wetmore SD. DNA–protein π–interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. Nucleic Acids Res. 2014; 42:6726–6741. doi: 10.1093/nar/gku269 PMID: 24744240

25. Ko CY, Lai YL, Liu WY, Lin CH, Chen YT, Chen LF, et al. Arabidopsis ENDO2: its catalytic role and requirement of N–glycosylation for function. J Agric Food Chem. 2012; 60:5169–5179. doi: 10.1021/jf300945c PMID: 22506810

26. Copeland RA. Enzymes. Wiley–VCH Inc., New York; 2000.

27. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr. 2013; D69:1204–1214.

28. Winn MD, Ballard ChC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr. 2011; D67:235–242.

29. Mueller U, Darowski N, Fuchs MR, Forster R, Hellmig M, Paithankar KS, et al. Facilities for macromolecular crystallography at the Helmholtz–Zentrum Berlin. J Synchrotron Radiat. 2012; 19:442–449. doi: 10.1107/S0909049512006395 PMID: 22514183

30. Kabsch W. Integration, scaling, space–group assignment and post–refinement. Acta Crystallogr. 2010; D66:133–144.

31. Krug M, Weiss MS, Heinemann U, Mueller U. XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. J Appl Crystallogr. 2012; 45:568–572.

32. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction–image processing with MOSFLM. Acta Crystallogr. 2011; D67:271–281.

33. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallogr. 2010; D66:22–25.

34. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr. 2010; D66:486–501.

35. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. 2011; D67:355–367.

36. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nat Struct Biol. 2003; 10:980. doi: 10.1038/nsb1203-980 PMID: 14634627

Supporting information

Structural and Catalytic Properties of S1 Nuclease from *Aspergillus oryzae* Responsible for Substrate Recognition, Cleavage, Non–specificity, and Inhibition

Tomáš Koval'[1*], Lars H. Østergaard[2], Jan Lehmbeck[3], Allan Nørgaard[4], Petra Lipovová[5], Jarmila Dušková[1], Tereza Skálová[1], Mária Trundová[1], Petr Kolenko[1], Karla Fejfarová[1], Jan Stránský[1], Leona Švecová[1], Jindřich Hašek[1], and Jan Dohnálek[1*]

[1] Laboratory of Structure and Function of Biomolecules, Institute of Biotechnology CAS, v. v. i., Biocev, Vestec, Czech Republic
[2] Department of Agile Protein Screening, Novozymes A/S, Bagsvaerd, Denmark
[3] Department of Fungal Strain Technology, Novozymes A/S, Bagsvaerd, Denmark
[4] Department of Protein Biochemistry and Stability, Novozymes A/S, Bagsvaerd, Denmark
[5] Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague, Czech Republic

* Corresponding authors

E–mail: koval.tomas@gmail.com (TK), jan.dohnalek@ibt.cas.cz (JD)

Materials and methods

SDS–PAGE

The SDS–PAGE monitoring of S1 nuclease deglycosylation was performed under non–reducing conditions using an XCell SureLock™ mini–cell electrophoresis system, pre–cast polyacrylamide NuPAGE® Bis–Tris mini gels with 4–12% gradient, and the Mark12™ Unstained Standard (Life Technologies Corp.). Electrophoresis was performed according to the manufacturer's instructions. 5 μg of each sample was loaded on the gel.

Determination of experimental pI

The experimental pI of S1 nuclease wild type and of all deglycosylated versions was determined by isoelectric focusing (IEF) using an XCell SureLock™ mini–cell electrophoresis system, pre–cast 5% polyacrylamide Vertical Novex® IEF Mini Gels pH 3 – 10, and IEF Marker 3–10 (Life Technologies Corp.). Electrophoresis was performed according to the manufacturer's instructions.

Behavior of S1 nuclease in solution

Oligomerization state and behavior of S1 nuclease samples in solution were tested by dynamic light scattering (DLS). DLS experiments were performed using a Zetasizer Nano (Malvern Instruments) and a 45 μl quartz cuvette. All measurements were performed at 18 °C with a protein concentration of 1 mg/ml, in the storage buffer (25 mM Bis–Tris pH 6.0 with addition of 50 mM NaCl).

Nuclease activity

The reaction mixtures contained 50 µl of heat–denatured DNA from calf thymus (ssDNA) or RNA from torula yeast (concentration 1 mg/ml in 0.1 M sodium acetate buffer, pH 4.5 containing 50mM NaCl), and 50 µL of the enzyme diluted in the same buffer. After 5 min at 37 °C the reaction was stopped by adding 250 µl of 96% v/v ethanol. These assay settings are referred to as standard reaction conditions. The mixture was vortexed and incubated at −20 °C for 20 min. The precipitated undigested substrate was centrifuged (22000 x g, 20 min, 4 °C) and the absorbance of the supernatant was measured at 260 nm. Each measurement was performed in triplicate. Separate background readings for individual concentration points of all substrates were used in all cases.

Thermal unfolding using differential scanning fluorimetry

Thermal stability of fully glycosylated S1 nuclease and a sample treated with Endoglycosidase F1 (see deglycosylation details in the main article) was analyzed by differential scanning fluorimetry using a Prometheus NT.48 apparatus and Prometheus NT.48 Series nanoDSF Grade Standard Capillaries (NanoTemper Technologies GmbH). Samples were in the storage buffer (25 mM Bis–Tris pH 6.0 with addition of 50 mM NaCl). Concentration of both samples was about 0.5 mg/ml. Thermal unfolding was performed in the range from 20 °C to 95 °C at a scan rate of 2.5 °C per minute.

Surface electrostatic potential distribution

Surface electrostatic potential distribution was calculated for protonation states at pH 4, 6, and 8.5; pH 4 is close to the pH optimum for nuclease activity, pH 6 is close to the pH optimum for 3'–mononucleotidase activity, and pH 8.5 was chosen as a point of minimal catalytic activity of S1 nuclease. The calculations were done using APBS [1]. Parameter files were created by PDB2PQR 1.8 using the AMBER force field [2]. Protonation states were assigned by PropKa [3].

SUPPLEMENTARY TABLES

Table A. Crystallization of S1 nuclease.

Structure title	Composition of reservoir solution	Protein concentr. [mg/ml]	Co–crystallization partner	Co–cryst. partner concentr. [mM]	Setup	Reservoir volume [µl]
5FB9 – unoccupied	0.05 M CaCl2, 0.2 M NaCl, 0.1 M Bis–Tris pH 5.5, 25% w/v Polyethylene glycol 3350	22.5	2'–deoxyguanosine	10	Sitting drop vapour diffusion	70
5FBA – phosphate	0.2 M NaCl, 0.1 M Bis–Tris pH 5.5, 25% w/v Polyethylene glycol 3350	25	none	-	Hanging drop vapour diffusion	600
5FBB – inhibitors	0.05 M CaCl2, 0.1 M Bis–Tris pH 6.5, 30% v/v Polyethylene glycol monomethyl ether 550	22.5	adenosine 5'–monophosphate	10	Hanging drop vapour diffusion	600
5FBC – remodeled	0.2 M NaCl, 0.1 M Bis–Tris pH 5.5, 25% w/v Polyethylene glycol 3350	22.5	thiophosphorylated 2'–deoxyadenosine dinucleotide	1	Hanging drop vapour diffusion	600
5FBD – nucleotidase products	0.1 M Citric acid pH 3.8 (pH 4.2), 25% w/v Polyethylene glycol 3350	22.5	thiophosphorylated 2'–deoxycytidine dinucleotide	1.8	Hanging drop vapour diffusion	600
5FBF – nuclease products	0.1 M Citric acid pH 3.8 (pH 4.2), 25% w/v Polyethylene glycol 3350	22.5	2'–deoxycytidine 5'–monophosphate	10	Sitting drop vapour diffusion	70
5FBG – mutant with products	0.2 M NaCl, 0.1 M Bis–Tris pH 5.5, 25% w/v Polyethylene glycol 3350	22.5	dsDNA: $d(GC)_6$	1.2	Hanging drop vapour diffusion	600

All S1 nuclease samples were deglycosylated using Endoglycosidase F1 from *Elizabethkingia miricola* and transferred to 25 mM Bis–Tris pH 6.0, 50 mM NaCl, prior to crystallization. Crystallization temperature was 18 °C and the ratio of protein to reservoir drop volume was 1:1 (0.4 µl + 0.4 µl) in all cases. 5FB9 – unoccupied and 5FBF – nuclease products were obtained using CrystalQuick 96 Well Sitting Drop Plate (Greiner) sealed with ClearSeal Film™ (Hampton Research). The rest of the crystals were obtained using 24–well VDX Plates sealed with 18 mm x 0.22 mm siliconized circle cover slides and vacuum grease (Hampton Research).

Table B. Data collection and structure refinement statistics.

Structure title	**5FB9 – unoccupied**	**5FBA – phosphate**	**5FBB – inhibitors**
PDB ID	5FB9	5FBA	5FBB
Data collection statistics			
End station, detector	DESY: PETRA III P13, Pilatus 6M–F	BESSY II BL14.1, Pilatus 6M	BESSY II BL14.2, MAR CCD 225
Wavelength (Å)	1.00000	0.91841	0.91841
Crystal–detector distance (mm)	285.6	421.4	180.0
No. of oscillation images processed	3599	1798	400
Exposure per image (s)	0.05	1.0	2.6
Oscillation width (º)	0.05	0.1	0.5
Space group	*P*1	$P2_1$	*P*1
Unit–cell parameters (*a, b, c*, in Å; *α, β, γ* in °)	43.18, 48.59, 65.47; 107.4, 90.1, 105.7	41.76, 62.31, 48.00; 90.0, 106.7, 90.0	42.84, 47.64, 62.60; 106.4, 90.1, 106.3
Resolution range (Å)	43.05 – 1.50 (1.53–1.50)	45.97 – 1.80 (1.84–1.80)	46.14 –1.75 (1.78–1.75)
No. of observations	127078 (3881)	65030 (1936)	96158 (5195)
No. of unique reflections	68828 (2125)	20922 (889)	43898 (2371)
Multiplicity (I+= I-)	1.8 (1.8)	3.1 (2.2)	2.2 (2.2)
Data completeness (%)	88.2 (54.6)	95.3 (69.0)	95.8 (94.7)
Average mosaicity (°)	0.12	0.29	0.97
Mean *I/σ(I)*	9.2 (3.8)	11.7 (2.2)	8.1 (2.6)
R_{merge}	0.060 (0.187)	0.074 (0.386)	0.102 (0.594)
R_{meas}	0.084 (0.264)	0.090 (0.504)	0.137 (0.791)
$R_{p.i.m.}$	0.060 (0.187)	0.049 (0.318)	0.090 (0.519)
CC1/2	0.993 (0.907)	0.997 (0.381)	0.989 (0.454)
Wilson B	5.8	17.6	9.9
Refinement parameters			
No. of reflections (working set)	65463	19742	41645
No. of reflections (test set)	3364	1078	2252
R_{work}	0.149	0.152	0.145
R_{free}	0.177	0.199	0.188
R_{all}	0.150	0.153	0.146
Mean ADP (Å^2)	13.32	21.29	16.10
RMSD bond lengths (Å)	0.015	0.016	0.017
RMSD bond angles (°)	1.647	1.688	1.728
Solvent content (%)	43.5	40.6	39.2
No. of monomers in ASU	2	1	2
The last modelled residue (chain A / chain B)	S287 / Q285	S287	S287 / S287
No. of non–hydrogen atoms in ASU	5047	2284	4705
No. of water molecules in ASU	871	214	439
List of non–water ligands in ASU	6x Zn^{2+}, 2x Na^+, 2x Bis–Tris	3x Zn^{2+}, 1x Pi, 1x Glycerol	6x Zn^{2+}, 2x Pi, 2x 5'AMP, 3x Na^+, 4x Bis–Tris, 5x Ca^{2+}, 1x PEG MME
Modelled glycosylation (chain A / chain B)	N112, N248 / N112, N248	-, N248	N112, N248 / -, -

5FBC – remodeled	**5FBD – nucleotidase products**	**5FBF – nuclease products**	**5FBG – mutant with products**
5FBC	5FBD	5FBF	5FBG
BESSY II BL14.3, MAR CCD 225	Gemini Enhanced Ultra, Atlas CCD	BESSY II BL14.1, Pilatus 6M	BESSY II BL14.2, MAR CCD 225
0.89450	1.54056	0.91841	0.91841
180.0	55.2	200.5	230.0
166	283	2000	120
5.0	30.0	0.2	4.4
1.0	0.5	0.1	1
$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P3_121$
41.89, 62.59, 48.24; 90.0, 107.0 90.0	43.04, 62.43, 84.12; 90.0, 90.0, 90.0	53.74, 62.39, 62.76; 90.0, 90.0, 90.0	106.76, 106.76, 127.91; 90.0, 90.0, 120.0
46.14 –1.75 (1.78–1.75)	30.08 – 1.75 (1.78 – 1.75)	44.25 – 1.04 (1.06–1.04)	37.47–1.97 (2.02–1.97)
84531 (4544)	81964 (1979)	580610 (6260)	347503 (14849)
24085 (1299)	22221 (903)	96234 (2708)	59852 (4088)
3.5 (3.5)	3.7 (2.2)	6.0 (2.3)	5.8 (3.6)
99.8 (99.8)	95.2 (72.2)	94.5 (54.5)	99.6 (97.6)
0.26	1.22	0.11	1.01
10.7 (2.0)	13.6 (2.1)	10.7 (2.7)	9.3 (2.0)
0.101 (0.663)	0.061 (0.376)	0.089 (0.307)	0.131 (0.645)
0.120 (0.785)	0.072 (0.504)	0.097 (0.394)	0.144 (0.750)
0.064 (0.416)	0.036 (0.331)	0.037 (0.242)	0.058 (0.372)
0.995 (0.657)	0.996 (0.841)	0.996 (0.815)	0.993 (0.686)
11.1	5.1	4.8	10.4
22868	21053	91430	56785
1198	1138	4712	3024
0.131	0.152	0.111	0.158
0.181	0.212	0.135	0.187
0.133	0.153	0.113	0.159
15.05	11.74	8.32	21.93
0.017	0.017	0.013	0.016
1.678	1.706	1.695	1.598
41.1	36.9	32.3	66.2
1	1	1	2
S287	S287	S287	S287 / Q285
2376	2464	2606	4859
268	377	485	652
3x Zn^{2+}, 1x Glycerol, 1x 5’dAMP(S)	3x Zn^{2+}, 1x Pi, 1x dCyt	3x Zn^{2+}, 1x Na^{+}, 2x 5’dCMP	6x Zn^{2+}, 2x Pi, 2x dCyt, 1x dGua
N112, N248	N112, -	N112, N248	N112, N248 / N112, N248

Values in parentheses are for the highest resolution shell. ASU stands for asymmetric unit, Pi for phosphate ion, PEG MME for a fragment of polyethylene glycol monomethyl ether, 5’AMP for adenosine 5'–monophosphate, 5’dAMP(S) for 2'–deoxyadenosine 5'–thio–monophosphate, 5’dCMP for 2'–deoxycytidine 5'–monophosphate, dGua for 2'–deoxyguanosine, and dCyt for 2'–deoxycytidine.

SUPPLEMENTARY FIGURES

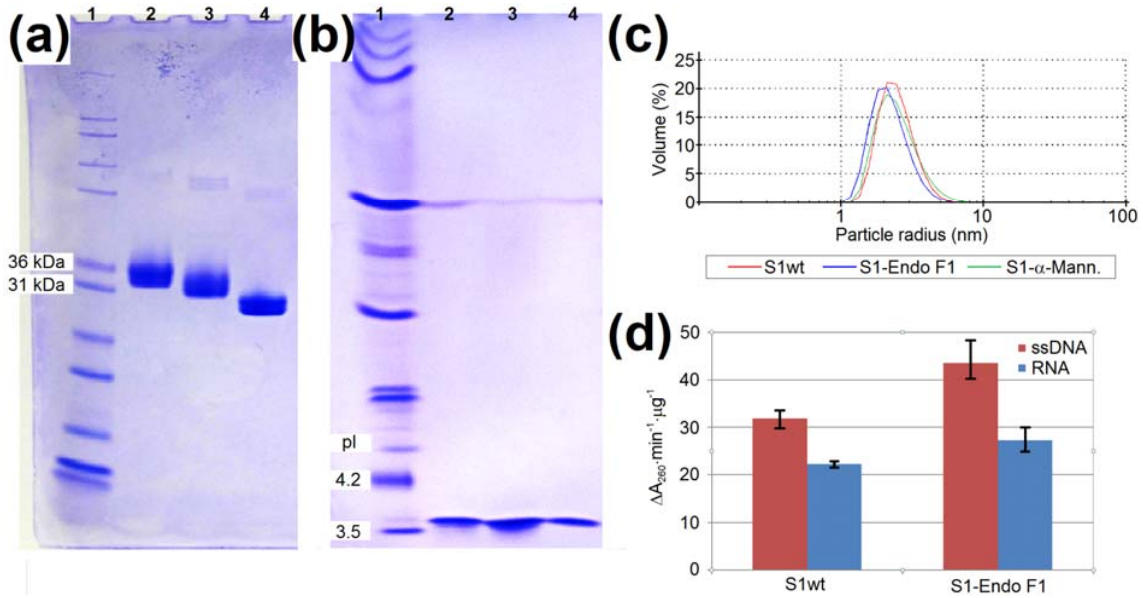


Figure A. Monitoring of S1 nuclease deglycosylation. (**a**) SDS–PAGE analysis. Lane 1: Mark12™ Unstained Standard, lane 2: fully glycosylated S1wt, lane 3: S1wt treated by α–Mannosidase from *Canavalia ensiformis*, lane 4: S1wt treated by Endoglycosidase F1 from *Elizabethkingia miricola.* α–Mannosidase leaves three or more carbohydrate units at every N–glycosylation site. Endo F1 leaves only one carbohydrate unit (N–acetyl–D–glucosamine) for every N–glycosylation site. Successful deglycosylation can be seen by the difference in the resulting S1 nuclease mass. (**b**) IEF analysis. Lane loading is the same as in (a) except for the marker lane: IEF Marker 3–10. Determined values of the isoelectric point (~3.6) for S1wt and both deglycosylated versions are the same. (**c**) DLS analysis. Both deglycosylated versions behave similar to S1wt in the storage buffer and they are monomeric. The measured hydrodynamic radius is 2.49 ± 0.70 nm for S1wt, 2.47 ± 0.89 nm for S1–α–mann, and 2.23 ± 0.69 nm for S1–Endo F1. The apparent trend of decrease of hydrodynamic radius by deglycosylation cannot be reliably interpreted due to the observed experimental errors. (**d**) The activity of S1wt and S1 treated with Endoglycosidase F1 against ssDNA and RNA. Activity is reported as a change of absorbance at 260 nm over time normalized to the amount of enzyme used. S1wt and S1–Endo F1 display similar activity taking into account the decrease of enzyme mass by about 18% by deglycosylation and also the experimental error.

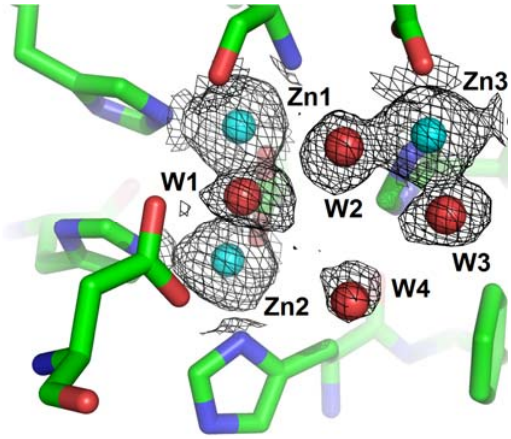


Figure B. The catalytic zinc cluster of the structure 5FB9 – unoccupied with bound water molecules. Zinc ions are shown as light blue spheres and water molecules as red spheres. The composite omit $2mF_o$-DF_c electron density map was calculated using *Phenix* [4] with the refinement mode. It is shown as a black mesh and contoured at a 1.5 σ level around the zinc cluster and water molecules present in the cluster. The four water molecules labelled W1–W4 are displaced and/or substituted upon binding of the various ligands in the other structures of this study. Their presence in the unoccupied active site is confirmed by the composite omit map. Molecular graphics were created using *PyMOL* (Schrödinger, LLC) and chain A of the structure 5FB9.

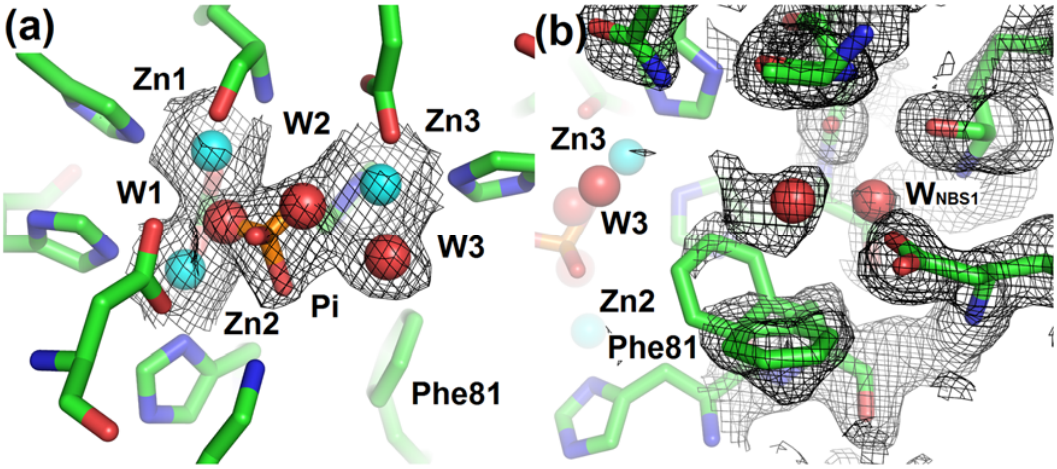


Figure C. The active site of the structure 5FBA – phosphate. Zinc ions are shown as light blue spheres, water molecules as red spheres, and phosphate as orange/red sticks. The composite omit $2mF_o$-DF_c electron density map was calculated using *Phenix* [4] with the refinement mode. It is shown as a black mesh and contoured at a 1.0 σ level around the zinc cluster, NBS1 and ligands of interest. (**a**) The presence of a phosphate ion (labelled Pi) in the catalytic zinc cluster is confirmed by the composite omit electron density. Based on behavior in the refinement, the phosphate ion was modelled with half occupancy as well as water molecules W1 and W2. (**b**) The nucleoside binding site (NBS1) with disordered Phe81. Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

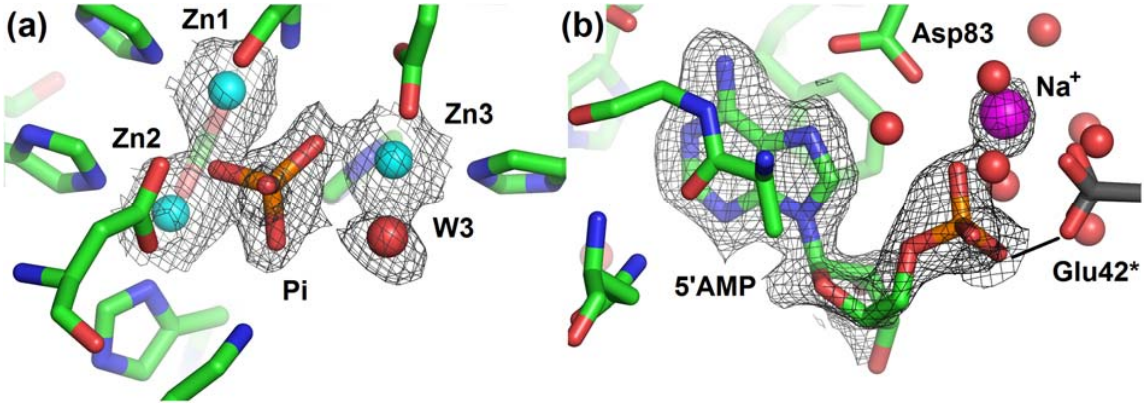


Figure D. The active site of the structure 5FBB – inhibitors with phosphate ion and adenosine 5'–monophosphate. Zinc ions are shown as light blue spheres and water molecules as red spheres. Phosphate is shown as orange/red sticks. The composite omit $2mF_o$-DF_c electron density map was calculated using *Phenix* [4] with the refinement mode. It is shown as a black mesh and contoured at a 1.0 σ level around the zinc cluster and ligands of interest. (**a**) The presence of the phosphate ion (labelled Pi) in the zinc cluster is confirmed by the composite omit electron density. (**b**) NBS1 with adenosine 5'–monophosphate. The presence of 5'AMP and a sodium ion is confirmed by the composite omit electron density. The presence of a sodium ion (magenta sphere) is supported by the coordination distances and behavior in the refinement. 5'AMP interacts with the sodium ion through the phosphate group and through this sodium ion also with Asp83 of NBS1. The sodium ion has no direct contact with symmetry–related protein molecules. The phosphate moiety of 5'AMP also interacts with Glu42 (shown in sticks, carbon – dark grey, marked by *) from a symmetry–related protein chain. This is the only direct contact of the ligand with a symmetry–related molecule. Molecular graphics were created using *PyMOL* (Schrödinger, LLC) and chain A of the structure 5FBB.

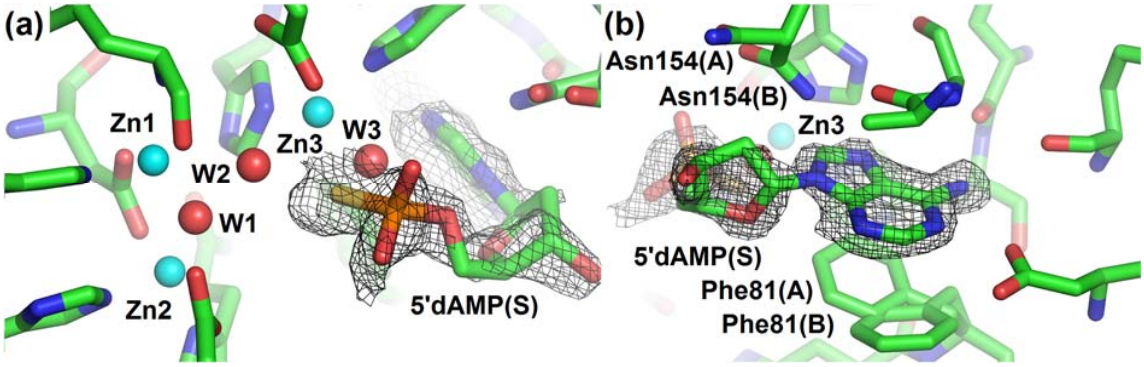


Figure E. The active site of the structure 5FBC – remodeled with 2'–deoxyadenosine 5'–thio–monophosphate. Zinc ions are shown as light blue spheres, water molecules as red spheres, and thiophosphate as yellow/orange/red sticks. The composite omit $2mF_o$-DF_c electron density map was calculated using *Phenix* [4] with the refinement mode. It is shown as a black mesh and contoured at a 1.0 σ level only around the ligand of interest. 5'dAMP(S) in the structure is a product of cleavage of thiophosphorylated 2'–deoxyadenosine dinucleotide present in the crystallization experiment. (**a**) Interaction of the thiophosphate moiety of 5'dAMP(S) (sulfur – yellow, phosphorus – orange) with the zinc cluster. The orientation of the thiophosphate moiety with the sulfur atom interacting with Zn3 is confirmed by the intensity of the electron density peak at this position and also by the distance of this peak to the peak for phosphorus. 5'dAMP(S) was modelled with occupancy 0.8 due to the observed disorder in NBS1 (see panel b). Water molecule W3 was modelled with occupancy 0.2. (**b**) The nucleobase binding site (NBS1) with 5'dAMP(S). The presence of 5'dAMP(S) is confirmed by the composite omit map. The observed binding mode of 5'dAMP(S) is only compatible with Asn154 and Phe81 in alternative A, so it was modelled with occupancy 0.8 to interpret the observed disorder. The adenine moiety is slightly deformed after refinement of the structure, which is caused by the disorder in NBS1. The refinement procedure itself forces this deformation. Nevertheless, deformation of the base is insignificant for the interpretation of the binding mode. Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

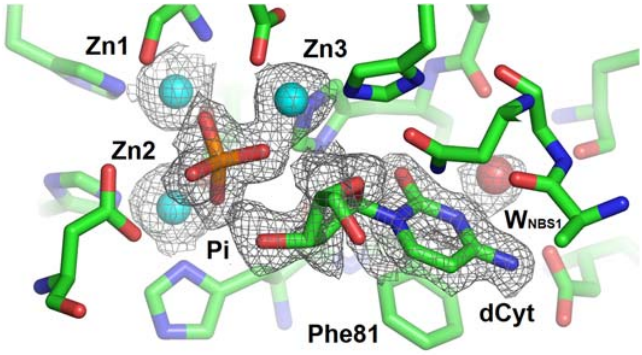


Figure F. The active site of the structure 5FBD – nucleotidase products with phosphate ion and 2'–deoxycytidine. The presence of the phosphate ion (orange/red sticks, labelled Pi) and 2'–deoxycytidine (labelled dCyt) is confirmed by the composite omit $2mF_o$-DF_c electron density map (black mesh). 2'–deoxycytidine is a product of cleavage of thiophosphorylated 2'–deoxycytidine dinucleotide which was present in the crystallization experiment. Zinc ions are shown as light blue spheres, water molecules as red spheres, and phosphate as orange/red sticks. The composite omit $2mF_o$-DF_c electron density map was calculated using *Phenix* [4] with the refinement mode. It is shown as a black mesh and contoured at a 1.0 σ level only around the ligands of interest and the zinc cluster. Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

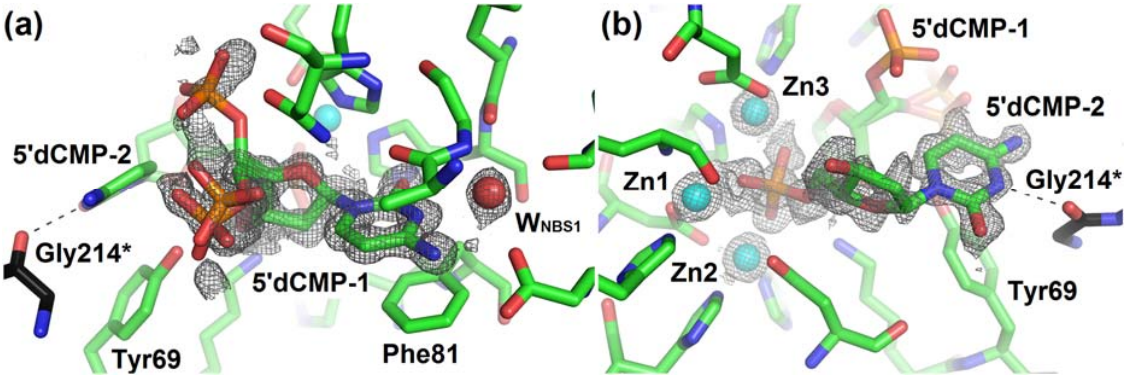


Figure G. The active site of the structure 5FBF – nuclease products with two molecules of 2'–deoxycytidine 5'–monophosphate. The catalytic zinc ions are shown as light blue spheres, water molecules as red spheres, and the phosphate moiety as orange/red sticks. The composite omit $2mF_o$-DF_c electron density map was calculated using *Phenix* [4] with the refinement mode. It is shown as a black mesh and contoured at a 1.0 σ level only around the zinc cluster and the ligands of interest. The presence of both molecules of 2'–deoxycytidine 5'–monophosphate (labelled 5'dCMP) is confirmed by the composite omit $2mF_o$-DF_c electron density map. (**a**) One molecule of 5'dCMP interacts with NBS1 (in the compact form) in the shallow binding mode, similar to 2'–deoxycytidine in the structure 5FBD – nucleotidase products (Figure F). The phosphate and 2'–deoxycytidine moieties of 5'dCMP in position -1 are modelled in three possible conformations (modelled based on mF_o-DF_c and refined) with the main differences in the positions of the phosphate moiety. (**b**) The second molecule of 5'dCMP binds inside the zinc cluster via its phosphate group. The cytosine moiety only interacts directly with the symmetry–related protein chain (Gly214*, shown as sticks, carbon black) and so participates in the crystal contact. The cytosine moiety is also potentially involved in an interaction with Tyr69. The phosphate group binds inside the zinc cluster in the

same mode as the phosphate ion in the case of 5FBD – nucleotidase products (Figure F). Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

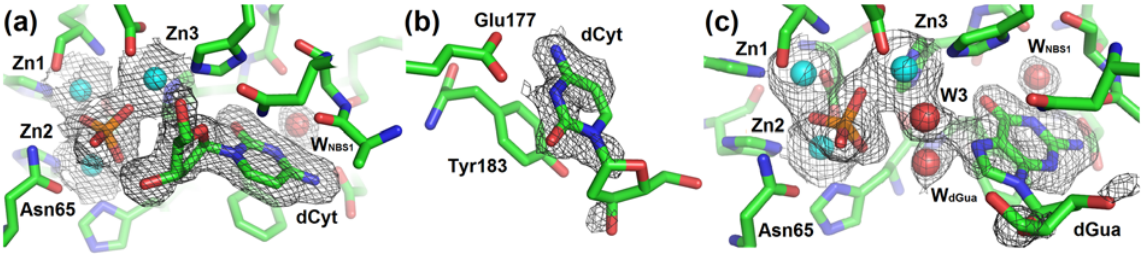


Figure H. The active sites and the Half–Tyr site of the structure 5FBG – mutant with products with the phosphate ion, 2'–deoxycytidine, and 2'–deoxyguanosine. Zinc ions are shown as light blue spheres, water molecules as red spheres, and phosphate ions as orange/red sticks. The composite omit $2mF_o$-DF_c electron density map was calculated using *Phenix* [4] with the refinement mode. It is shown as a black mesh and contoured at a 1.0 σ level only around the zinc cluster and the ligands of interest. Only inorganic phosphate, 2'–deoxycytidine (dCyt), and 2'–deoxyguanosine (dGua) could be clearly identified in electron density and built. Both protein chains of the asymmetric unit have inorganic phosphate inside the zinc cluster but in slightly different orientations with respect to the zinc ions. The active site of chain A contains one molecule of dCyt whereas chain B one molecule of dGua. One additional molecule of dCyt binds in a site on the surface of chain A, near the active site. **(a)** The active site of chain A. Presence of the phosphate ion (orange/red sticks) and 2'–deoxycytidine (labelled dCyt) is confirmed by the composite omit $2mF_o$-DF_c electron density map. **(b)** dCyt interacting with the Half–Tyr site. Electron density with a similar shape at this site can be observed also in chain B. It was not interpreted for its lower clarity. **(c)** The active site of chain B. Presence of the phosphate ion (orange/red sticks), 2'–deoxyguanosine (labelled dGua), and water molecules participating in the interaction of the guanine moiety with the zinc cluster (red spheres, labelled as W3 and W_{dGua}) is confirmed by the composite omit $2mF_o$-DF_c electron density map. It is not clear whether the observed ligand is 2'–deoxyguanosine or 2'–deoxyguanosine 5'–monophosphate due to the disorder of the 2'–deoxyribose moiety. We chose to model this ligand as 2'–deoxyguanosine because the presence of the phosphate moiety is not supported by electron density. Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

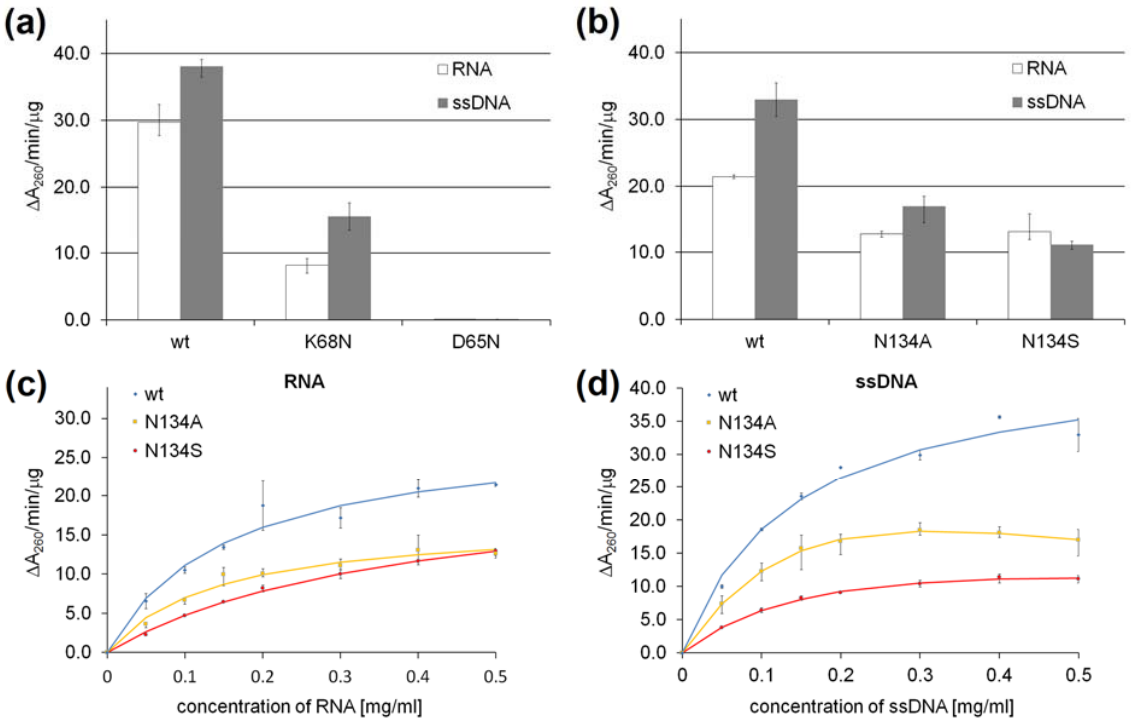


Figure I. Comparison of the catalytic activity of S1 wild type and mutants. (a) and **(b)** Activity was measured under standard reaction conditions and is reported as a change of absorbance at 260 nm over time normalized to the amount of enzyme used. Mutants reported in panels (a) and (b) were not measured at the same time. D65N and K68N were successfully expressed and purified earlier than N134A and N134S. **(c)** and **(d)** Comparison of kinetic parameters of nuclease S1 wild type and its mutants N154S and N154A using ssDNA and RNA as a substrate.

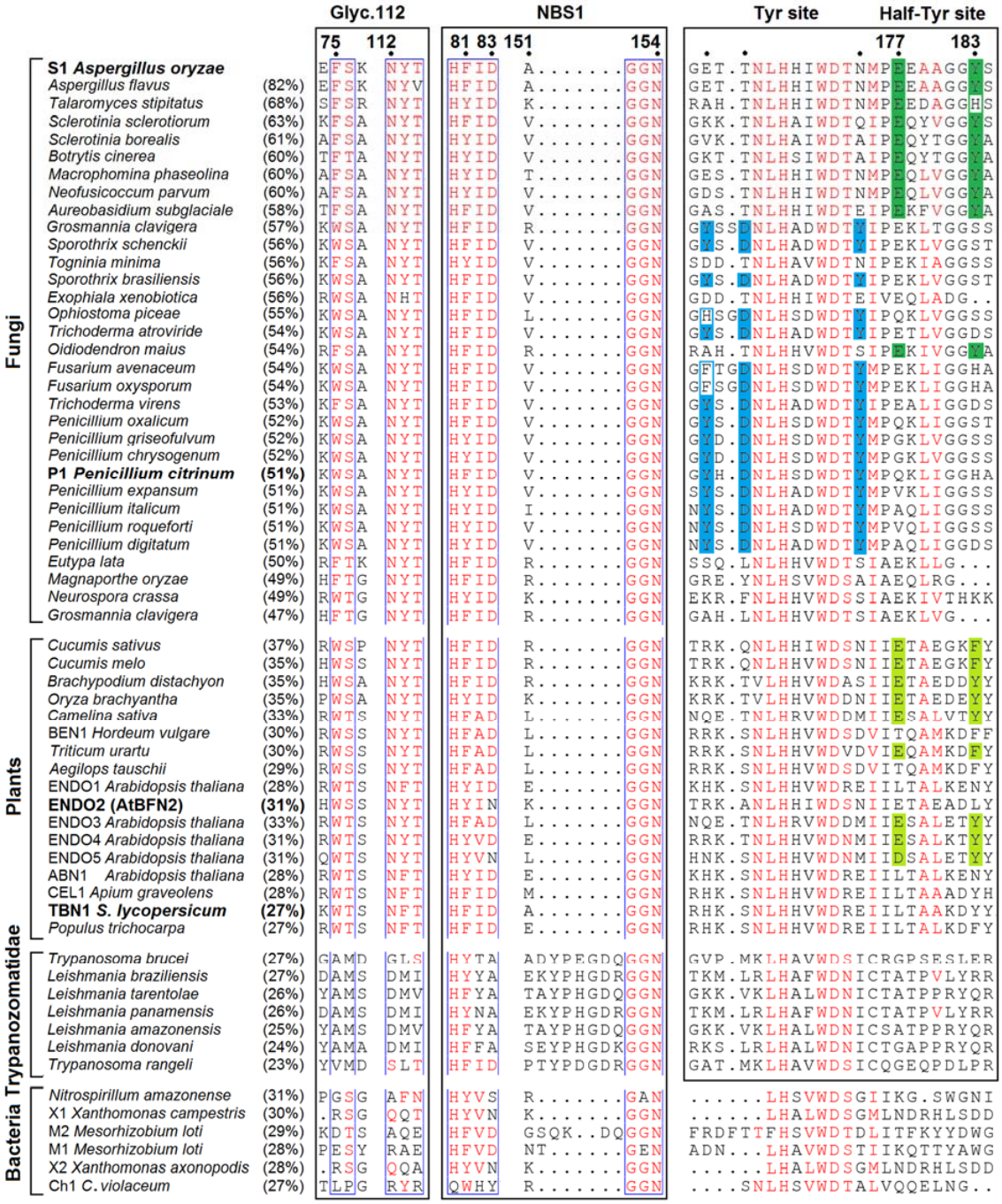


Figure J. Conservation of selected features in the S1–P1 nuclease family. Homologs of S1 nuclease were identified using an NCBI BLAST search [5]. Due to the high number of found homologs the list presented here was manually edited based on the intended demonstration of the selected features. Most of the bacterial homologs were selected based on Pimkin *et al.* [6]. The sequences were aligned using ClustalW2 [7]. Names of the enzymes with known structure are in bold characters. The column following the name shows sequence identity to S1 nuclease. The figure was created using ESPript [8] and edited. (**Glyc112**) Glycosylation site at position 112 along with its interacting partner, an aromatic amino acid at position 75. (**NBS1**) Alignment of selected residues involved in the formation of NBS1. The side chain of the residue at position 81 is involved in the stacking interaction with a nucleobase. The side

chain of the amino acid at position 83 provides hydrogen bonding to a nucleobase. Asp or Asn is conserved at position 83 in fungi and plants. In trypanozomatidae and bacteria this pattern is often broken. However, without the structure of such nuclease it is hard to estimate whether the function of the hydrogen bond partner is substituted by another amino acid, or nucleobase binding is facilitated in a different way. The second π–system donor is the peptide bond between the residue 151 and Gly152. Asn154 is always conserved. (**Tyr site and Half–Tyr site**) Residues involved in the formation of the P1 nuclease Tyr site are in the light blue box. Residues involved in the formation of the S1 nuclease Half–Tyr site are in the green box. These two sites are conserved only in fungi. Moreover, several S1–P1 like nucleases from fungi apparently do not possess either site. Residues with a possible role in the Half–Tyr site in plants are in the light green box. However, it is not possible to confirm the role of these residues in plants without the corresponding structures.

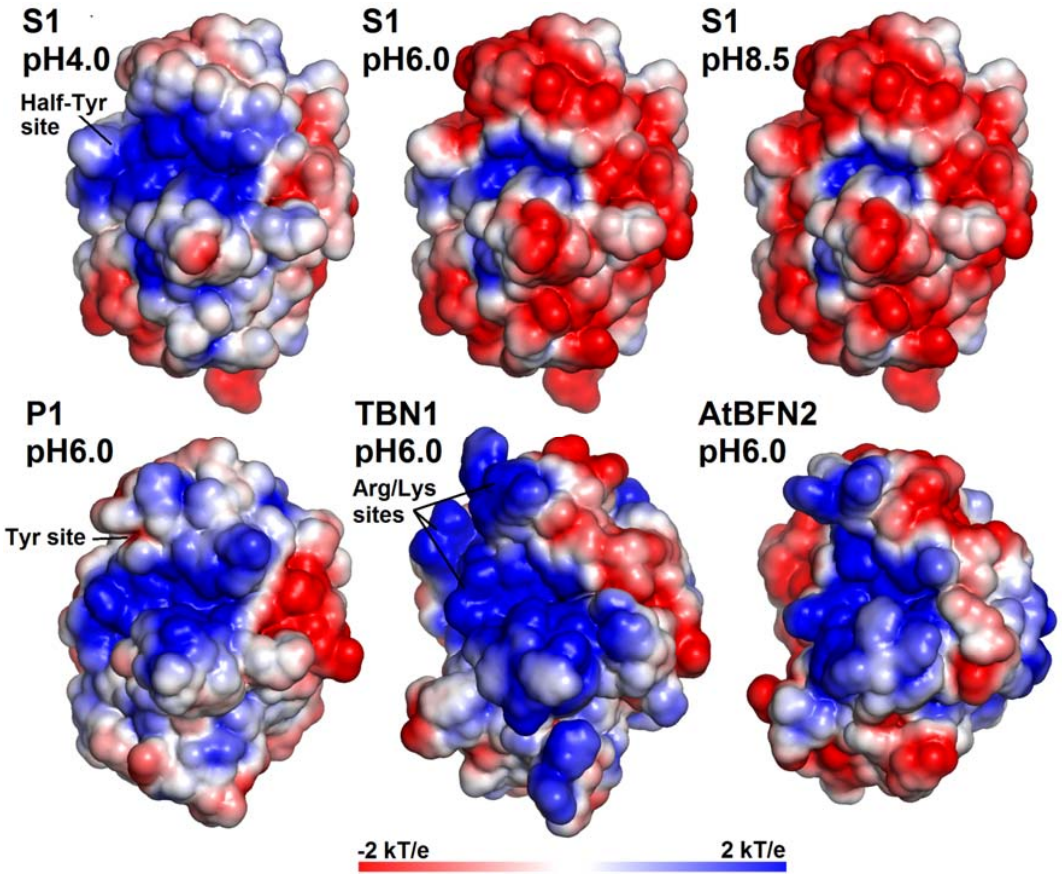


Figure K. Electrostatic potential distribution on the solvent accessible surface of the S1–P1 nuclease family members with known structure. Electrostatic potential distribution for S1 nuclease was calculated for protonation states at pH 4.0, pH 6.0, and pH 8.5 using the structure 5FBF – nuclease products. In the cases of P1 (PDB ID: 1AK0 [9]), TBN1 (PDB ID: 3SNG [10]), and AtBFN2 (PDB ID: 4CXO [11]) the electrostatic potential distribution was calculated only at pH 6.0 (close to their pH optimum of activity). All structures are shown in the same orientation with respect to the zinc cluster. Molecular graphics were created using *PyMOL* (Schrödinger, LLC).

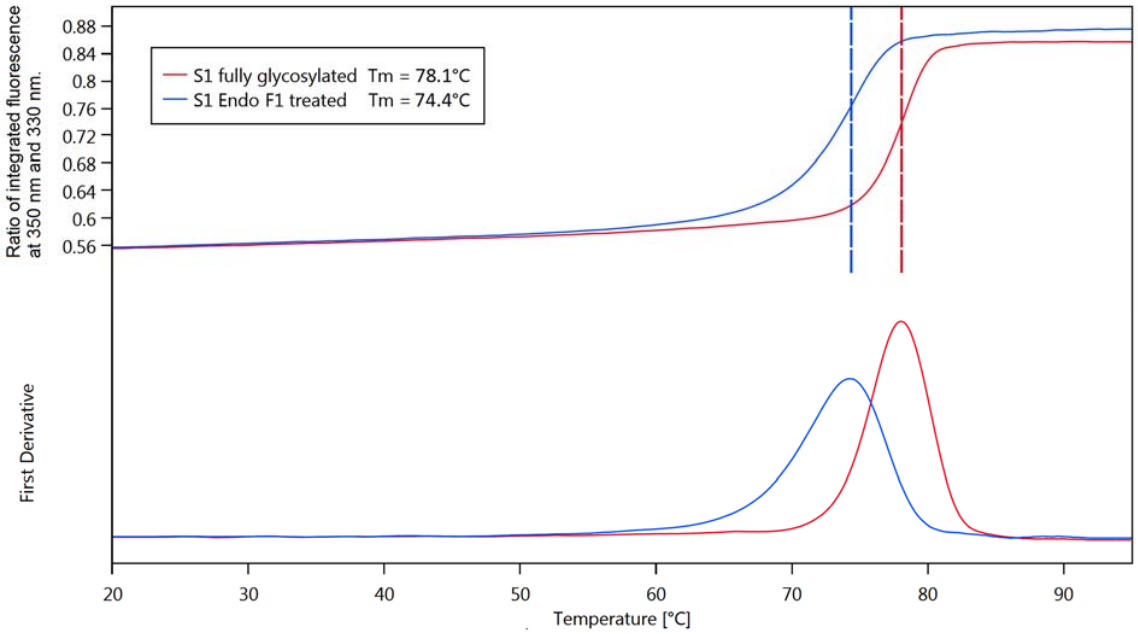


Figure L. Thermal stability of fully glycosylated S1 nuclease (red) and S1 nuclease treated with Endoglycosidase F1 (blue) measured by DSF. Measurements were performed in 25 mM Bis–Tris pH 6.0 with the addition of 50 mM NaCl, with protein concentration 0.5 mg/ml, and in temperature range 20 – 95 °C.

References

1. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U.S.A. 2001;98:10037–10041.
2. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 2007;35:W522–525.
3. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. Chem Theor Comput. 2001;7:2289–2295.
4. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. 2012 D68:352–367.
5. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a more efficient report with usability improvements. Nucleic Acids Res. 2013;41(W1):W29–W33.
6. Pimkin M, Miller CG, Blakesley L, Oleykowski CA, Kodali NS, Yeung AT. Characterization of a periplasmic S1–like nuclease coded by the *Mesorhizobium loti* symbiosis island. Biochem Biophys Res Commun. 2006;343:77–84.
7. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 2013;41(W1):W597–W600.
8. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014;42(W1):W320–W324.
9. Romier C, Dominguez R, Lahm A, Dahl O, Suck D. Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. Proteins. 1998;32:414–424.
10. Koval T, Lipovova P, Podzimek T, Matousek J, Duskova J, Skalova T, et al. Plant multifunctional nuclease TBN1 with unexpected phospholipase activity: structural study and reaction-mechanism analysis. Acta Crystallogr. 2013;D69:213–226.
11. Yu TF, Maestre-Reyna M, Ko CY, Ko TP, Sun YJ, Lin TY, et al. Structural Insights of the ssDNA Binding Site in the Multifunctional Endonuclease AtBFN2 from *Arabidopsis thaliana*. PLoS ONE. 2014;9(8):e105821.

STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

Crystal structure of human interferon-γ receptor 2 reveals the structural basis for receptor specificity

Pavel Mikulecký, Jirí Zahradník, Petr Kolenko, Jiří Černý, Tatsiana Charnavets, Lucie Kolářová, Iva Nečasová, Phuong Ngoc Pham and Bohdan Schneider

Acta Cryst. (2016). D**72**, 1017–1025

STRUCTURAL BIOLOGY

ISSN 2059-7983

Crystal structure of human interferon-γ receptor 2 reveals the structural basis for receptor specificity

Pavel Mikulecký,‡ Jirí Zahradník,‡ Petr Kolenko, Jiří Černý, Tatsiana Charnavets, Lucie Kolářová, Iva Nečasová, Phuong Ngoc Pham and Bohdan Schneider*

Institute of Biotechnology CAS, BIOCEV, Prumyslova 595, 252 50 Vestec, Czech Republic. *Correspondence e-mail: bohdan.schneider@gmail.com

Received 17 May 2016
Accepted 27 July 2016

Edited by Z. S. Derewenda, University of Virginia, USA

‡ These authors share first authorship.

Keywords: interferon-γ receptor 2; fibronectin type III domain; class 2 cytokine receptors.

PDB reference: interferon-γ receptor 2, 5eh1

Supporting information: this article has supporting information at journals.iucr.org/d

Interferon-γ receptor 2 is a cell-surface receptor that is required for interferon-γ signalling and therefore plays a critical immunoregulatory role in innate and adaptive immunity against viral and also bacterial and protozoal infections. A crystal structure of the extracellular part of human interferon-γ receptor 2 (IFNγR2) was solved by molecular replacement at 1.8 Å resolution. Similar to other class 2 receptors, IFNγR2 has two fibronectin type III domains. The characteristic structural features of IFNγR2 are concentrated in its N-terminal domain: an extensive π–cation motif of stacked residues KWRWRH, a NAG–W–NAG sandwich (where NAG stands for *N*-acetyl-D-glucosamine) and finally a helix formed by residues 78–85, which is unique among class 2 receptors. Mass spectrometry and mutational analyses showed the importance of N-linked glycosylation to the stability of the protein and confirmed the presence of two disulfide bonds. Structure-based bioinformatic analysis revealed independent evolutionary behaviour of both receptor domains and, together with multiple sequence alignment, identified putative binding sites for interferon-γ and receptor 1, the ligands of IFNγR2.

1. Introduction

Interferon-γ receptor 2 is a cell-surface receptor that represents a crucial molecule in the interferon-γ (IFNγ) signalization pathway, influencing innate and adaptive immunity against pathogens and tumours (Schoenborn & Wilson, 2007; Lin & Young, 2013). The signalling cascade is initiated by the binding of IFNγ to its high-affinity cell surface receptor 1, forming a binary complex with a structure that has already been determined [PDB entries 1fg9 (Thiel *et al.*, 2000) and 1fyh (Landar *et al.*, 2000)]. However, to activate this binary complex and activate the JAK/STAT signalization pathway (Jung *et al.*, 1987; Cook *et al.*, 1994; Hemmi *et al.*, 1994), IFNγ receptor 2 must participate in a ternary complex created by a homodimer of IFNγ, two molecules of receptor 1 and molecule(s) of receptor 2. To date, detailed structural and biophysical characterization of IFNγ receptor 2, the key molecule for proper IFNγ signalization, is lacking.

From its sequence similarity, IFNγ receptor 2 (also known as IFNγ receptor β chain or accessory factor 1, AF-1), has been classified as a member of the class 2 receptor family. This large group of cytokine receptors includes IFNγ receptor 1, receptors of interferon-α and interferon-β, receptors of interleukin-10 and interleukin-20, and receptors of other interleukins belonging to the IL-10 family (Langer *et al.*, 2004). The mature IFNγ receptor 2 protein comprises of 310 amino acids and has a predicted molecular mass of 35 kDa. It consists of a relatively short 69-amino-acid intracellular domain, a 21-amino-acid transmembrane domain and a 220-

OPEN ACCESS

amino-acid extracellular domain that is structured into two fibronectin type III domains. The extracellular domain contains five cysteine residues and six potential N-linked glycosylation sites. Such extensive glycosylation contributes to a significant size heterogeneity, which is observed even when the receptor molecule is isolated from the same cell type; the molecular weight of mature receptor 2 of human interferon-γ ranges from 61 to 67 kDa (Bach *et al.*, 1995).

Despite its biological significance, three-dimensional structural data on IFNγ receptor 2 are lacking. Here, we report a 1.8 Å resolution crystal structure of the extracellular portion of IFNγ receptor 2 (hereafter called IFNγR2); the structure has been deposited in the Protein Data Bank as entry 5eh1. The structure and sequence of IFNγR2 are discussed in the context of the structures and sequences of the related class 2 cytokine receptors, with emphasis on the sequentially closest receptors of interleukins from the IL-10 family. Structure-based and sequence-based alignments suggested regions securing binding specificity of these receptors for their cytokine ligands.

2. Materials and methods

2.1. Cloning, expression and purification of IFNγR2

The gene encoding the extracellular part of IFNγR2 (residues 28–247 of UniProt entry P38484) was cloned into a *Drosophila* pMT-BiP-V5-His_A vector using BglII and AgeI restriction enzymes in frame with an N-terminal BiP signal peptide and a C-terminal 6×His tag. This expression vector was co-transfected into insect Schneider S2 cells along with the pCoBlast selection plasmid using Effectene Transfection Reagent according to the manufacturer's instructions. Blasticidin-resistant S2 cells were selected by growing the cells in HyClone SFX Medium supplemented with 10% FBS and 25 µg ml^{-1} blasticidin S. Large-scale protein expression was achieved after expansion and substitution into HyClone SFX serum-free medium, and protein expression was induced by the addition of 0.75 m*M* $CuSO_4$ for 6 d (the cell concentration was approximately 35 million per millilitre) and 1.5 m*M* $CuSO_4$ for a further 2 d until the percentage of living cells did not decrease below 95%. After expression, the cells were discarded by centrifugation and the medium containing secreted glycosylated IFNγR2 protein was supplemented with the following additives at the following final concentrations: 5 m*M* $CaCl_2$, 1 m*M* $NiSO_4$, 250 m*M* NaCl and 50 m*M* Tris–HCl pH 8. The protein was purified on an IMAC HP column charged with $NiSO_4$ and equilibrated with EQ buffer (50 m*M* Tris–HCl pH 8, 500 m*M* NaCl). The column was washed with W buffer (50 m*M* Tris–HCl pH 8, 500 m*M* NaCl, 20 m*M* imidazole pH 8) and the protein was eluted with EL buffer (50 m*M* Tris–HCl pH 8, 500 m*M* NaCl, 250 m*M* imidazole pH 8). It was further purified to homogeneity by size-exclusion chromatography at room temperature on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with HN buffer (10 m*M* HEPES pH 7.5, 100 m*M* NaCl). Samples were analyzed by 12% SDS–PAGE.

IFNγR2 was produced in insect cells as a secreted protein bearing oligosaccharide moieties of approximately 10 kDa according to SDS–PAGE analysis. Deglycosylation by peptide: N-glycosidase F (PNGase F) or endoglycosidase H (Endo H) with a C-terminal *Strep*-tag (§S1, Supporting Information) was performed after purification of IFNγR2 on an IMAC column during dialysis against TN buffer (50 m*M* Tris buffer pH 8, 150 m*M* NaCl) or HN buffer, respectively. Endoglycosidases were removed on a *Strep*-Tactin column and the nonbound fraction containing IFNγR2 was further purified by size-exclusion chromatography in HN buffer.

The single IFNγR2 variants (N110Q, N137Q and N231Q, respectively) were introduced using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies). Primers are listed in Supplementary Table S1. The fully mutated IFNγR2 variant bearing N56Q, N110Q, N137Q and N231Q mutations was obtained as a GeneArt Strings DNA Fragment and was cloned with the same protocol as the wild type. The expression and purification of all IFNγR2 variants were performed in the same way as described above.

2.2. Biophysical measurements

Circular-dichroism (CD) spectra were recorded using a Chirascan-plus spectrometer (Applied Photophysics) in steps of 1 nm over the wavelength range 185–260 nm. Samples diluted with water to a concentration of 0.2 mg ml^{-1} were placed into the holder in a 0.05 cm path-length quartz cell and individual spectra were recorded at a temperature of 23°C. The CD signal was expressed as the ellipticity and the resulting spectra were buffer-subtracted. To analyze the ratio of secondary structures, we used the *CDNN* program (Böhm *et al.*, 1992) provided with the Chirascan CD spectrometer. CD melting measurements were performed using samples diluted with water to a protein concentration of 0.5 mg ml^{-1}. A 10 mm path-length quartz cell was placed into the thermostated holder and sample absorption was recorded at 280 nm in 1°C increments at a rate of 0.5°C min^{-1} over the temperature range 20–85°C with an averaging time of 12 s. Melting curves were normalized to relative values between 0.0 and 1.0 to visually magnify differences between the melting profiles, and the melting temperature (T_m) was estimated from the first derivative of the melting curves.

2.3. Glycosylation analysis and disulfide-bond determination

IFNγR2 glycosylation sites were determined by MALDI-MS analysis preceded by protein digestion as described previously (Plíhal *et al.*, 2004). Disulfide bonds in IFNγR2 were determined by SDS–PAGE and subsequent identification by mass spectrometry (MS) in analogy to the previously described procedure (Pompach *et al.*, 2009). 20 µg of sample in nonreducing conditions was loaded onto a 4–12% gradient gel (Life Technologies) in the presence of 200 µ*M* cystamine. Bands corresponding to highly glycosylated IFNγR2 were excised and subjected to in-gel deglycosylation and proteolysis. Deglycosylation using Endo H (New England Biolabs) was carried out for 4 h at 37°C and the resulting

Table 1
Data-collection statistics and structure-refinement parameters.

Values in parentheses are for the highest resolution shell.

X-ray source	MX 14.1, HZB
Wavelength (Å)	0.91841
Total oscillation angle (°)	180
Resolution range (Å)	62.88–1.80 (1.91–1.80)
Space group	$P6_122$
Unit-cell parameters (Å)	$a = b = 58.102$, $c = 377.266$
No. of measured reflections	688675 (110540)
No. of unique reflections	36723 (5733)
Average multiplicity	18.8 (19.3)
Completeness (%)	99.9 (99.8)
Average $I/\sigma(I)$	17.6 (2.5)
Overall B factor from Wilson plot (Å^2)	20
Average B factor (Å^2)	28
No. of non-H atoms	
Protein	1734
Saccharides	42
Waters	311
All	2121
R_{merge}	0.148 (1.395)
Half-data-set correlation coefficient $CC_{1/2}$	99.9 (81.2)
No. of reflections, test set	1820
Final $R_{work}/R_{free}/R_{all}$	0.190/0.222/0.191
Ramachandran plot	
Residues in favoured region	213 [96.3%]
Residues in allowed regions	219 [99.1%]
Outliers	2 [0.9%]

partly deglycosylated sample was digested with trypsin (sequencing grade, Promega) for 12 h at 37°C at a protein:enzyme ratio of 30:1(*w*/*w*). After digestion, the tryptic peptide mixture was desalted on a peptide MicroTrap column (Michrom Bioresources) and separated on a reversed-phase C18 column (Acclaim PepMap 100, 5 µm, 0.1 × 20 mm; Thermo Scientific). The mobile phases consisted of 0.1% formic acid in 2% acetonitrile (solvent *A*) and 0.1% formic acid in 98% acetonitrile (solvent *B*). Peptides were eluted under the following gradient conditions: 2–45% solvent *B* in 40 min, 45–95% solvent *B* in 5 min. The flow rate was 0.5 µl min^{-1} and the column was directly connected to the mass spectrometer. Mass spectra were acquired on a solariX XR FTMS instrument equipped with a 12 T superconducting magnet (Bruker Daltonics). For the identification of disulfide bonds, we used the *Links* algorithm, previously described as *ASAP* (*Automated Spectrum Assignment Program*; Schilling *et al.*, 2003). To generate deconvoluted spectra and export the *m*/*z* values, we used a script utilizing the *SNAP* 2.0 algorithm of the *DataAnalysis* 4.2 software suite (Bruker Daltonics).

2.4. Crystallization and diffraction data collection

Crystals of Endo H-deglycosylated IFNγR2 receptor were grown using the sitting-drop vapour-diffusion method in 96-3 three-well Intelli-Plate trays (Art Robbins Instruments). The reservoir solution consisted of 0.1 *M* MES pH 5.0, 10% PEG 6000 (final pH 6.0): condition No. 61 of The JCSG Core I Suite (Qiagen). Drops consisting of 0.2 µl protein sample (15 mg ml^{-1} protein in HN buffer) and 0.2 µl reservoir solution were prepared with a Gryphon liquid-pipette robot (Dunn Labortechnik) and were equilibrated against 100 µl reservoir solution. Crystals appeared after 30 d of incubation at 291 K. Crystals were mounted in Round LithoLoops (Molecular Dimensions) and flash-cooled in liquid nitrogen after cryoprotection in 20%(*v*/*v*) glycerol. X-ray diffraction data were collected at 100 K on beamline MX 14.1 of the BESSY II synchrotron-radiation source at the Helmholtz-Zentrum Berlin (HZB). A native data set was collected at a wavelength of 0.918 Å.

2.5. Data processing, structure determination and refinement

The diffraction and refinement statistics are summarized in Table 1. Diffraction data were processed and scaled using the *XDS* program package (Kabsch, 2010). The structure was solved with *BALBES* (Long *et al.*, 2008), but the structure model needed significant manual remodelling. Only the C-terminal domain of IFNγR2 was found and the initial *R* factors were about 0.49 and 0.52 for R_{work} and R_{free}, respectively. Residues missing from the initial model were built in with significant help from *ARP*/*wARP* (Langer *et al.*, 2008); manual corrections and building were performed using *Coot* (Emsley & Cowtan, 2004). Refinement was then carried out with *REFMAC*5 (Murshudov *et al.*, 2011) and the structure was validated by *MolProbity* (Chen *et al.*, 2010). The coordinates and structure factors have been deposited in the PDB with accession code 5eh1.

2.6. Sequence and structural bioinformatics

The UniProt database was searched with the *BLAST* tool (Camacho *et al.*, 2009) using the sequence of the extracellular part of IFNγR2 as the query sequence. The automated result was manually reviewed to select 90 sequences from different species. These sequences were used to calculate a multiple sequence alignment with *Clustal Omega* (Sievers *et al.*, 2011) as implemented in *UGENE* (Okonechnikov *et al.*, 2012). The *ConSurf* server (Glaser *et al.*, 2003) was used to estimate the evolutionary conservation of amino-acid positions in the protein structures. The calculations were based on the crystal structure of IFNγR2 (PDB entry 5eh1) and the alignment prepared by *Clustal Omega*. Structural comparison was prepared by *MatchMaker* as implemented in the *UCSF Chimera* software (Pettersen *et al.*, 2004).

Root-mean-square deviation (r.m.s.d.) values between the N- and C-terminal domains were calculated using *VMD* (Humphrey *et al.*, 1996). The backbone atoms of 34 sequentially conserved residues in each domain (listed in Supplementary Table S2) were used for the structure superposition of all possible pairs of N- and C-terminal domains of the 12 available crystal structures of class 2 cytokine receptors. The *VMD* commands `measure fit` and `move` were used for the structural overlay, followed by `measure rmsd` operating on the same selection of residues and backbone atoms for the calculation of r.m.s.d. values. The stability of the N-terminal domain of IFNγR2 was estimated by calculating the pairwise interaction energy at the DFT-D level as detailed in Supplementary Fig. S3.

3. Results and discussion

3.1. Summary

IFNγR2 was produced in insect S2 cells, purified, characterized by biophysical techniques (Fig. 1) and its crystal structure was solved at 1.8 Å resolution. Fig. 2 highlights some of the structural features of IFNγR2; Figs. 3 and 4 provide a comparison of IFNγR2 to the other cytokine receptors from the class 2 family, with the aim of correlating the sequences and structures of these proteins.

3.2. Glycosylation and overall fold stability

The IFNγR2 protein has six potential glycosylation sites (Asn56, Asn110, Asn137, Asn231, Asn85 and Asn219), of which the first four were confirmed by mass-spectrometric analysis as glycosylated in our construct; the crystal structure later revealed that position Asn85 was also glycosylated. The greatest heterogeneity was observed at position Asn137. To remove oligosaccharide moieties from these residues, we used two endoglycosidases, Endo H and PNGase F, but both enzymes left several forms of IFNγR2 with residual glycosylation as observed by a distribution of molecular mass on SDS–PAGE (Supplementary Fig. S1). Mass spectrometry identified α(1–3)-fucose at position Asn231. Because α(1–3)-fucose abolishes the activity of both endoglycosidases, its presence is a likely to be reason for the mass distribution of IFNγR2. Deglycosylation by Endo H caused an approximately 7 kDa shift in molecular weight on SDS–PAGE and, in contrast to deglycosylation by PNGase, did not induce protein oligomerization, as checked by size-exclusion chromatography. Measurements using CD spectroscopy (Fig. 1) and thermal shift assay (§S1 in Supporting Information and Supplementary Fig. S2) showed no significant difference in melting temperatures between the glycosylated and Endo H-deglycosylated forms of the IFNγR2 protein. Because the CD spectra of these two forms are also virtually identical, we believe that deglycosylation does not influence the secondary structure of IFNγR2. Although the CD spectra of IFNγR2 and IFNγR1 differ considerably (Černý *et al.*, 2015), both proteins belong to the same fold of the fibronectin type III domain family (Pfam PF00041).

Besides deglycosylation by the endoglycosidases, we designated asparagine-to-glutamine mutants to decrease the level of glycosylation. We prepared a fully mutated IFNγR2 variant bearing N56Q, N110Q, N137Q and N231Q mutations and single-point mutants N110Q, N137Q and N231Q, respectively. All of these constructs were transfected into insect S2 cells, but none of them were secreted into the cell-culture medium. This correlates with the earlier observation that IFNγR2 mutants with changed glycosylation patterns were located in the cytoplasmic fraction (Moncada-Vélez, 2013). Structurally significant is glycosylation at positions Asn110 and Asn137, where the bound *N*-acetyl-D-glucosamine (NAG) residues sandwich Trp131 (Fig. 2*b*), thus shielding its hydrophobic surfaces from solvent. As suggested by the failure to express and/or purify the N56Q, N110Q, N137Q and N231Q mutants, glycosylation is necessary for IFNγR2 production by stabilizing the fold and transport to and/or across the cellular membrane.

3.3. The IFNγR2 fold is stable without disulfide bonds

IFNγR2 contains five cysteine residues, and our mass-spectrometric analysis identified disulfide bonds linking Cys86 to Cys94 and Cys209 to Cys234. We observed the same protein mobility under nonreducing and reducing conditions during SDS–PAGE analysis. The melting temperature of both glycosylated and deglycosylated IFNγR2 measured by thermal shift assay (Supplementary Fig. S2) decreased by only $\sim$1°C in the presence of 5 m*M* TCEP (a reducing agent to break the disulfide bonds), so that the IFNγR2 fold is stable without S—S bonds. This contrasts with the behaviour of IFNγR1 (Fountoulakis *et al.*, 1990), in which the protein fold is

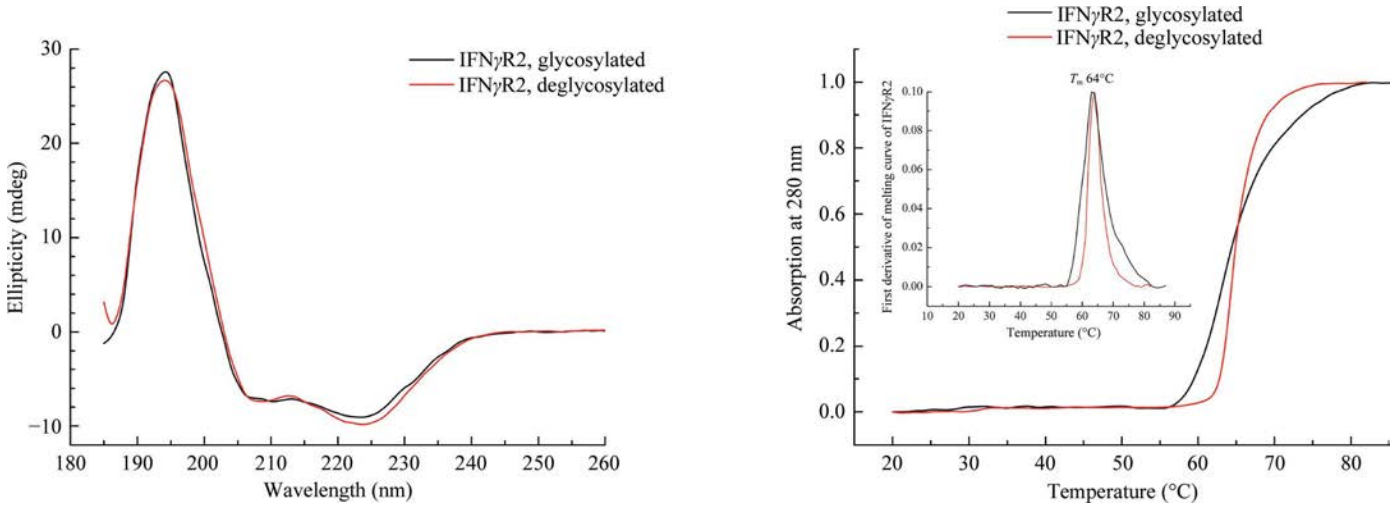


Figure 1
Left: circular-dichroism (CD) spectra of glycosylated and deglycosylated IFNγR2. The CD spectra of both proteins are highly similar, suggesting that the partial removal of the oligosaccharide moieties did not affect the overall structure of IFNγR2. Right: normalized melting curves measured from temperature-dependent CD spectra at 280 nm. The melting temperature was estimated as 64°C for both glycosylated and deglycosylated IFNγR2.

stabilized to a large extent by S—S bridges. The fifth IFNγR2 cysteine residue, Cys174, does not form an intramolecular S—S bridge but is bound to a monomeric cysteine. Binding of free cysteine to the sterically accessible Cys174 probably occurs after secreting IFNγR2 into the cell-culture medium, which contains free cysteine and stabilizes the monomeric form of IFNγR2.

3.4. The overall structure of the extracellular portion of IFNγR2

The structure of IFNγR2 was solved at 1.8 Å resolution and electron density was observed for amino-acid residues 28–240 of UniProt entry P38484, except for two two-residue loops. Data-collection and refinement parameters are shown in Table 1. The extracellular part of the IFNγR2 molecule consists of two domains (Fig. 2), the N-terminal D1 domain of UniProt residues 28–133 and the C-terminal D2 domain of residues 144–247. Both domains belong to the immunoglobulin fold with fibronectin type III topology, forming β-sandwiches (Pfam PF00041). The inter-domain torsion angle is approximately 120°, similar to those of IFNγR1 (Thiel *et al.*, 2000; Walter *et al.*, 1995) and human tissue factor (Harlos *et al.*, 1994); the D1–D2 torsion angle is defined in Supplementary Table S3. The D1 domain is composed of three β-strands stacked on a layer of four β-strands, and the D2 domain is created by four β-strands arranged against four other β-strands; both domains are connected by a short linker (residues 134–143 in IFNγR2) comprising a short helix that is also found in IFNγR1, human tissue factor and other receptors.

3.5. Structural motifs in D1 and D2

D1 contains a distinct structural motif of six stacked residues: Lys68, Trp74, Arg114, Trp126, Arg116 and His123. The average distance between the mean planes of the individual side chains of this extensive π–cation interaction is 3.65 Å.

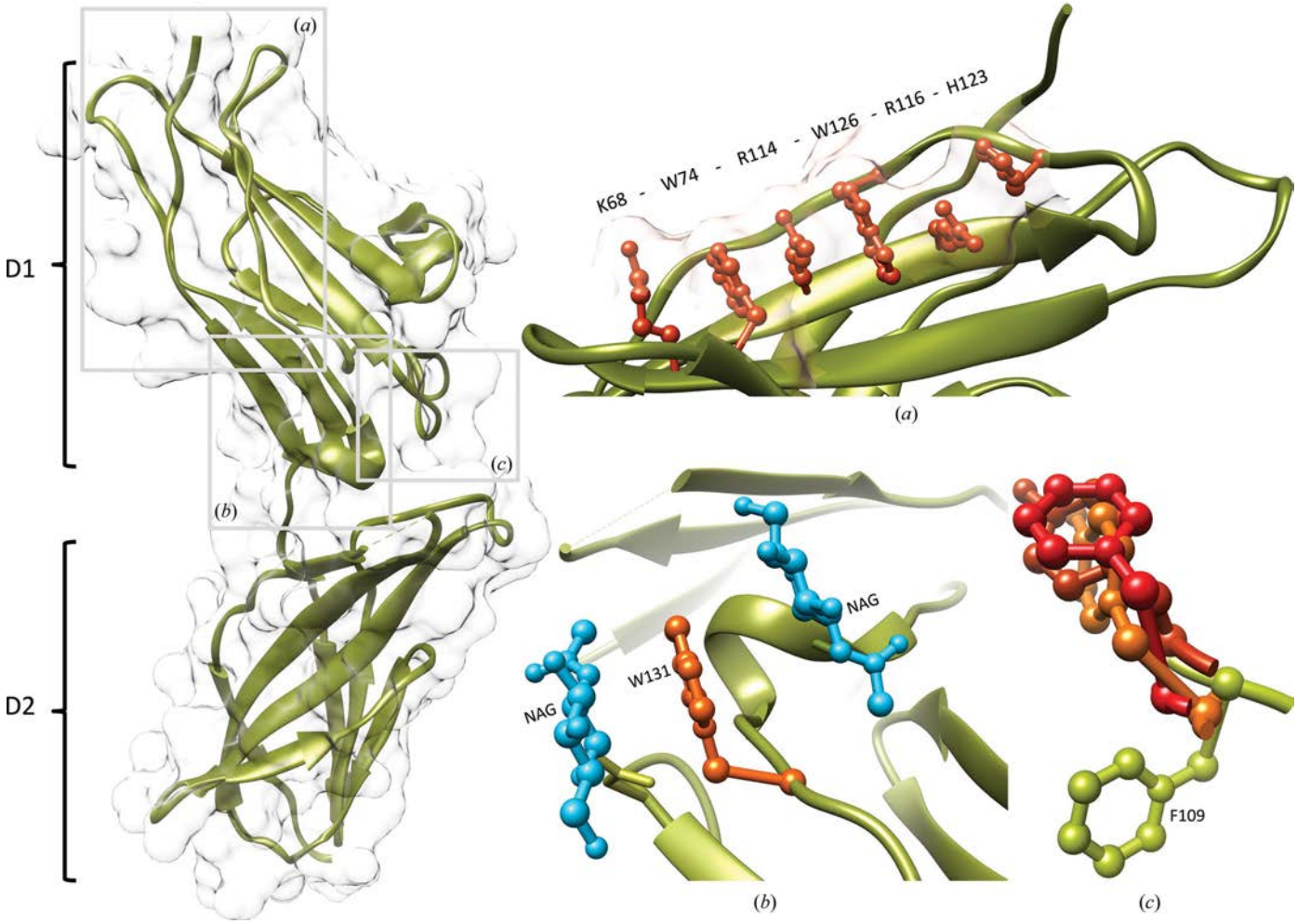


Figure 2
Left: ribbon and surface representations of the IFNγR2 structure. D1 and D2 indicate domains 1 and 2, respectively. Insets: (*a*) residues Lys68, Trp74, Arg114, Trp126, Arg116 and His123 of the D1 domain form a stacking motif on the IFNγR2 surface. (*b*) *N*-Acetyl-D-glucosamines (NAGs; blue) glycosylating Asn110 and Asn137 sandwich Trp131 (orange), reducing its hydrophobic character. (*c*) The superposition of aromatic binding epitopes shows differences between IFNγR2 (Phe109 in green) on one side and promiscuous shared cytokine receptors on the other [in red; Tyr82 of IL10R2 (PDB entry 3lqm; Yoon *et al.*, 2010), Phe169 of gp130 (PDB entry 1bqu; Bravo *et al.*, 1998) and Tyr103 of γ_c receptor (PDB entry 4gs7; Ring *et al.*, 2012)]. No corresponding aromatic residue is observed in IL20R2 (Logsdon *et al.*, 2012).

Analysis of the interaction energies in D1 revealed that the motif contributes significantly to the overall stability of the whole domain. These six surface residues are involved in interactions that are comparable in strength to the hydrophobic core of the domain and are likely to play an important role in the process of domain folding. The residues responsible for domain stability are depicted in Supplementary Fig. S3, in which the colour and thickness of the cartoon representation show the relative interaction energy per residue ranging from low (blue) to high (red) stabilizing values.

An analogous stacking motif with the consensus sequence WS*X*WS (Bazan, 1990) has been predicted by sequence alignments in D1 of the class 1 receptor family (McElroy *et al.*, 2009), but such a motif is missing in the D2 domains of both class 1 and class 2 receptors. Based on the presence of the KWRWRH motif in IFNγR2, we performed structural alignment of the class 2 receptor structures and discovered a similar but sequentially noncontinuous motif with the sequence (*X*)WRWR(*X*), where *X* is K, R or H. The important role of large aromatic tryptophan residues in stabilizing the fibronectin fold by stitching together two β-strands is accompanied in D1 by a structural role for charged residues, especially arginines. Besides the discussed (*X*)WRWR(*X*) motif, we found a tight overlap of a continuous chain of residues R-L/V-R-A (residues Arg114-Leu115-Arg116-Ala117 in IFNγR2): the average r.m.s.d. between motifs from two receptors is 0.6 Å. The other important residues that are conserved in D1 of the available class 2 receptor structures are residues corresponding to Trp49, Ser124 and the Cys86/Cys96 pair forming a disulfide bond in IFNγR2. A unique feature of IFNγR2 D1 is a short helix (residues 78–85), which is present in no other discussed receptor structure.

Sequential and structural comparison of D2 revealed a considerable sequence variability, within which we identified the conservation of two proline residues, Pro142 and Pro143, and the structurally well conserved motif 175-YNVA*X*W-180, with r.m.s.d. values about 1 Å but low sequence similarity. Another characteristic structural feature of D2 is the formation of a disulfide bridge between Cys209 and Cys234. Higher values for the *B* factors in D2 indicate its higher flexibility compared with D1. A higher flexibility of D2 compared with D1 was also indicated in our previous studies of IFNγR1 (Mikulecký *et al.*, 2013; Černý *et al.*, 2015).

3.6. Structural alignment of domains D1 and D2 in IFNγR2 and in other class 2 receptors

We performed alignment of the IFNγR2 structure with the 11 remaining available structures of the class 2 receptor family in order to gauge their similarity and reveal their unique features. The alignment was measured by overlapping 34 residues in the N-terminal D1 and the same number of residues in the C-terminal D2; the r.m.s.d. values of the overlapped residues are listed in Fig. 3 and the overlapped residues are listed in Supplementary Table S2. The D1 domains are mutually more similar than the D2 domains, as highlighted by

PDB code	1fg9	5eh1	1j7v	3lqm	4doh	4doh	3dlq	3g9v	3og6	3se4	3se4	2puq
Chain ID	*C*	*A*	*R*	*A*	*E*	*B*	*R*	*A*	*B*	*A*	*C*	*T*
Receptor	IFNγR1	IFNγR2	IL10R1	IL10R2	IL20R1	IL20R2	IL22R	IL22BP	IL28R	IFNαR1	IFNαR2	TF
IFNγR1	*0.82*	0.87	0.73	0.73	0.80	0.78	0.90	0.71	0.82	1.35	1.12	0.85
IFNγR2	1.27	*0.79*	0.93	0.97	0.79	0.79	0.96	0.80	0.85	1.45	1.25	1.04
IL10R1	1.22	1.12	*1.07*	0.74	0.97	0.88	0.94	1.43	1.13	0.81	0.67	0.66
IL10R2	1.00	1.14	1.05	*0.89*	0.95	1.05	0.65	0.87	0.68	0.80	1.11	0.93
IL20R1	0.93	1.32	0.89	1.33	*1.59*	0.93	0.97	0.63	0.75	0.59	1.36	1.15
IL20R2	1.32	1.51	1.23	0.99	1.33	*0.91*	0.93	0.66	0.82	1.48	1.25	1.21
IL22R	1.20	1.49	1.09	0.86	1.04	1.16	*4.19*	1.20	1.19	0.87	0.98	1.03
IL22BP	1.32	1.48	0.94	1.20	1.18	1.30	1.17	*0.93*	1.02	1.16	1.11	0.96
IL28R	1.21	1.16	1.34	1.29	1.30	1.32	1.18	1.45	*0.91*	1.09	0.77	0.86
IFNαR1	1.16	0.96	1.31	1.33	1.33	0.98	1.75	1.26	1.59	*1.77*	0.94	1.20
IFNαR2	1.35	1.80	1.40	1.29	1.19	1.81	1.58	1.14	1.46	1.22	*1.43*	1.02
TF	1.38	1.16	1.23	1.26	1.44	1.04	1.16	1.40	1.43	1.16	1.81	*0.86*

Figure 3
Structural differences between the N-terminal (D1) and C-terminal (D2) domains of 12 class 2 cytokine receptors gauged by the r.m.s.d. values for backbone atoms of their 34 residues. R.m.s.d. values comparing D1 and D2 domains are shown above and below the diagonal, respectively. For instance, comparison between D1 of IL20R2 and IL22BP gives an r.m.s.d. of 0.66 Å; the r.m.s.d. between their D2 domains is 1.30 Å. R.m.s.d. values that are smaller and larger than the off-diagonal average r.m.s.d. value are highlighted in blue and red hues, respectively. The diagonal (in grey) shows the lowest r.m.s.d. values for 34 residues from D1 and D2 within each receptor structure; the r.m.s.d. between D1 and D2 of IL10R2 is 0.89 Å. References to the analyzed structures are as follows: IFNγR1, Thiel *et al.* (2000); IFNγR2, this work; IL10R1, Josephson *et al.* (2001); IL10R2, Yoon *et al.* (2010); IL20R1 and IL20R2, Logsdon *et al.* (2012); IL22R, Bleicher *et al.* (2008); IL22BP, de Moura *et al.* (2009); IL28R, Miknis *et al.* (2010); IFNαR1 and IFNαR2, Thomas *et al.* (2011); TF (human tissue factor), Larsen *et al.* (2007).

blue and red hues in Fig. 3; the average r.m.s.d. between two D1 domains is 0.95 Å and that between two D2 domains is 1.3 Å. A high similarity within D1 and D2, respectively, indicates that modulation of the specificity of receptors takes place in only a few variable regions, which are discussed below.

D1 domains bear two conflicting structural features: strict fold conservation reflected by high structural similarity of the selected residues, and at the same time the presence of two structurally highly variable loops corresponding to residues 70–73 and 97–107 in IFNγR2. The third variable loop was located in the D2 domain (residues 162–171 in IFNγR2); the loops are coloured red and yellow for D1 and green for D2 in Fig. 4. Given the fairly uniform core of both domains and variability concentrated in the three localized regions, we suggest that the binding specificity of the individual receptors is controlled by these variable regions. However, there is another factor that contributes to the receptor specificity, the different mutual orientation of D1 and D2 (Supplementary Table S3), which displaces these variable regions to different positions, thus providing a unique binding interface for each receptor.

Several structural and sequential features of receptors of interferon-α and interferon-β (PDB entry 3se4; Thomas *et al.*, 2011), here labelled IFNαR1 and IFNαR2, distinguishes them from the other analyzed receptors. Specifically, IFNαR1 is composed of four instead of two domains; here, we analyzed D1 and D2. The D3–D4 pair cannot be analyzed as D4 is not resolved in the electron density. Further dissimilarities of IFNαR1 and IFNαR2 are found in the composition of their π–cation motifs: in IFNαR1 only four residues stack in D1 (Trp46, Arg76, Trp87 and Arg78) and three in D3 (Trp250, Arg279 and Trp291), while D1 of IFNαR2 does not have the motif at all and is replaced by the motif YVTV.

3.7. Similarity among receptors and consequences for evolution

As discovered previously (Yoon *et al.*, 2010), an aromatic tyrosine or phenylalanine residue situated in the cleft between D1 and D2 of gp130 (PDB entry 1bqu; Bravo *et al.*, 1998), γ_c (PDB entry 4gs7; Ring *et al.*, 2012) and IL10R2 (PDB entry 3lqm; Yoon *et al.*, 2010) serves as the key binding epitope of promiscuous class 1 and 2 receptors, implying the existence of a common ancestor. Superposition of these receptor structures shows that the orientation of Phe109 and three residues in gp130, γ_c and IL10R2 are quite different (Fig. 2*c*). No preferred rotamer of Phe109 overlaps the three former residues without significant rebuilding of the IFNγR2 backbone. The structural difference between IFNγR2 and the other receptors, especially IL10R2, is significant and suggests that there is not a common binding epitope for these receptors.

Significant sequence similarity between IFNγR2, IL10R2 and IL20R2 (20–25% sequence identity among different species; analysis not shown) points to their evolutionary relationship. If proven, it would be analogous to the evolutionary relationship between receptors 1 of the cytokines IFNγR1, IL10R1 and IL20R1 (Langer *et al.*, 2004). We may therefore hypothesize that these three cytokine systems have evolved from a common ancestral system: while interferon-γ evolved early in evolution and is known in fish species (Savan *et al.*, 2009), its receptor 2 emerged later in connection with the evolution of amphibians. The specific function of IFNγR2 therefore evolved from an older promiscuously functioning molecule. A likely candidate is IL10R2, because it is evolutionarily older and is known in primitive fishes, while IL20R2 emerges similarly to IFNγR2 in amphibians. The lack of a common binding epitope between IFNγR2 and the other class 2 receptors, notably IL10R2 (Fig. 2*c*), indirectly supports this hypothesis.

3.8. The sequence alignment of IFNγR2 from various species suggests its binding interface

In an attempt to identify the putative interface by which IFNγR2 forms a functional ternary complex with its binding partners interferon-γ and receptor 1, we aligned the IFNγR2 sequences from 90 species and used the *ConSurf* server to project the consensus onto the structure of IFNγR2 (Fig. 4*b*). The 32 conserved residues (purple in Fig. 4*b*) are mainly located in the inward arched part of the U-shaped receptor molecule. This part of the molecule contains the previously described stacking motif (Fig. 2*a*) and plays an important role in maintaining the overall structure. The 34 most variable residues are predominantly on the opposite side of the molecule (cyan in Fig. 4*b*). These regions of sequentially least conserved residues coincide with the location of the structurally variable loops derived by the superposition of receptor structures. We therefore conclude that the putative IFNγR2 interface for forming the active ternary complex with IFNγ and IFNγR1 is likely to be in the receptor 2 region with the most variable residues. This conclusion is supported by an analogous observation in the IFNγR1 system: the least sequentially conserved residues form the interface with the binding partner (Mikulecký *et al.*, 2013).

The composition of the ternary signalling complexes of dimeric cytokines discussed here, IFNγ and IL-10, is understood less than for monomeric examples such as IL-20, for which the ternary complex has a known crystal structure (Logsdon *et al.*, 2012). One of the reasons is the existence of two binding interfaces in the dimeric cytokines and the resulting different and more complex stoichiometry of the complexes; the crystal structure of an IFNγ–IFNγR1 complex with an unexpected 2:3 stoichiometry serves as an example (Thiel *et al.*, 2000). The topology and structure of the signalling ternary complex of IFNγ have been extensively studied and reviewed (Pestka *et al.*, 1997; Hoffmann *et al.*, 2015). Experiments in solution and on the cell surfaces indicated a 2:2:2 or 2:2:1 stoichiometry of the signalling IFNγ complex (Marsters *et al.*, 1995); cross-linking of different components of the IFNγ complex expressed in cloned cell lines have shown direct contact between IFNγR1 and IFNγR2 (Krause *et al.*, 2006) and also between IFNγ and IFNγR2 (Kotenko *et al.*, 1995). The newly determined structure of IFNγR2 may spur new experiments exploring the topology and three-dimensional structure of the signalizing ternary complex of IFNγ.

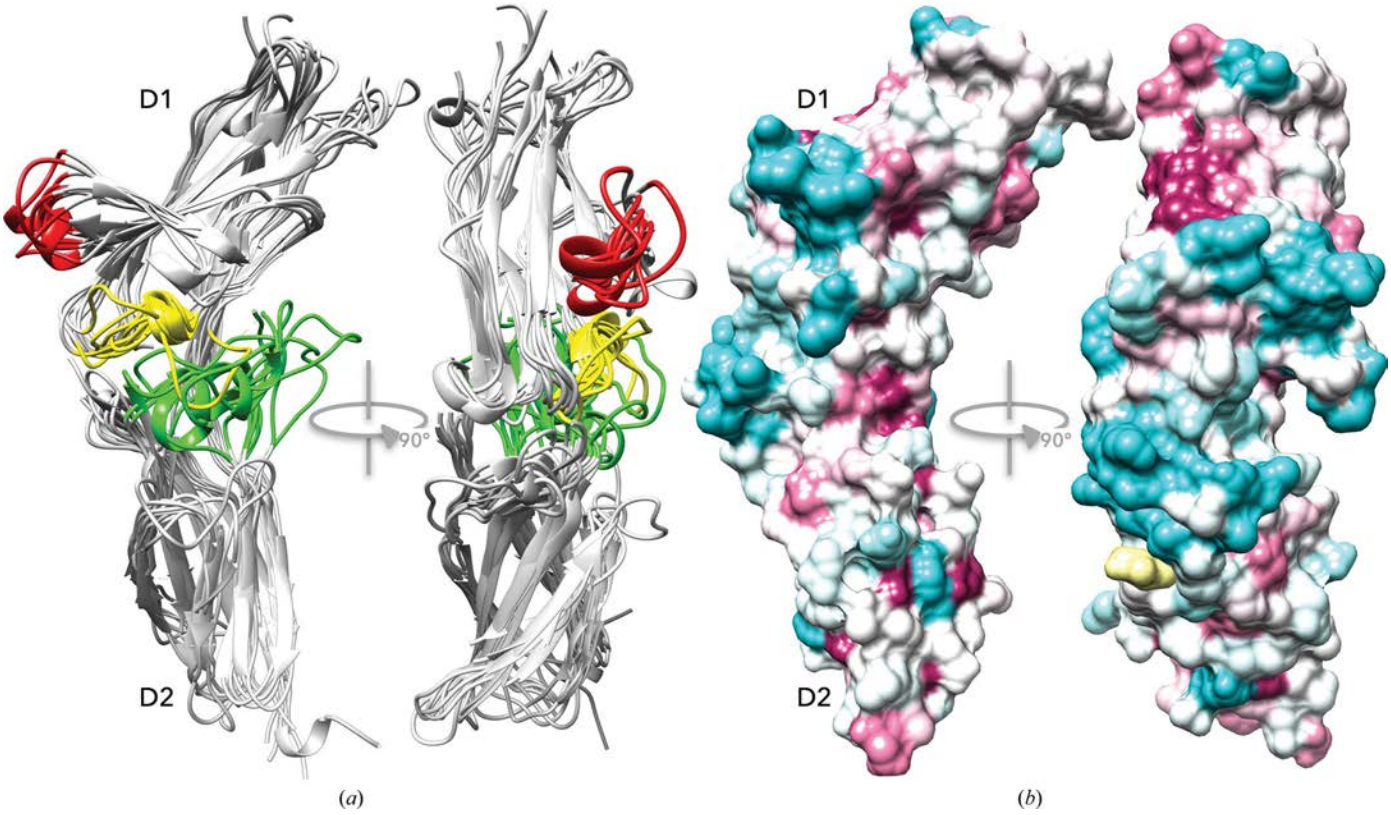


Figure 4
Structurally variable regions and a proposed interaction interface of the cytokine receptors. (*a*) Cartoon representation of the structural superposition of the D1 and D2 domains of the nine receptor structures. The backbone of the selected residues (listed in Supplementary Table S2) was superimposed independently for each domain and was then drawn on top of the IFNγR2 structure. The aligned cores of both domains overlap tightly (for r.m.s.d. values, see Fig. 3), but two regions in D1 (red and yellow) and one in D2 (green) are highly variable. (*b*) Sequences of IFNγR2 from 90 species with the sequentially most variable regions coloured cyan and conserved regions in purple drawn on the surface of the IFNγR2 structure by *ConSurf* (Glaser *et al.*, 2003). The IFNγR2 free cysteine Cys174 is highlighted in yellow. (*a*) and (*b*) show IFNγR2 in the same orientation.

4. Conclusions

A partially deglycosylated extracellular part of the interferon-γ receptor 2, IFNγR2, was crystallized and its structure was determined at 1.8 Å resolution and deposited in the PDB with accession code 5eh1. The electron-density map was interpreted for amino-acid residues 28–240, apart from two short loops. The IFNγR2 structure revealed the fold common to other cytokine receptors: two fibronectin type III domains connected by a short linker. IFNγR2 is a glycoprotein with five of the six potential N-linked glycosylation sites glycosylated, as confirmed by mass spectrometry and the crystal structure. Our analysis of glycosylation also uncovers the role of the oligosaccharide moieties at Asn110 and Asn137, which sandwich Trp131 and shield its hydrophobic aromatic ring from the solvent. Both potential disulfide bonds form but are not critical for the stability of IFNγR2, as it is also stable in a reducing environment. The fifth cysteine Cys174 is bound to the monomeric cysteine.

Structure and sequence alignments revealed some important features of the 12 class 2 receptors. Their N-terminal D1 domains are more mutually similar than their C-terminal D2 domains (Fig. 3). D1 carries a distinctive so far unrecognized structural feature: a π–cation motif of sequentially distant stacked residues (X)WRWR(X) (KWRWRH in IFNγR2; Fig. 2*a*). Analysis of the receptor structures revealed three structurally highly variable regions (Fig. 4*a*), which most likely bring about binding specificity for their interacting partners. This hypothesis is further supported by the alignment of IFNγR2 sequences from various species, which identified the highest sequence variability at positions coinciding with the structurally variable regions (Fig. 4*b*). An important structural feature distinguishing IFNγR2 from the related IL10R2, gp130 and γ_c receptors is the specific positioning of the aromatic recognition epitope in IFNγR2 (Fig. 2*c*).

We believe that the determination of the structure of the so-far missing component of the interferon-γ signalling complex will enable a deeper understanding of the functioning of this important immunity cascade.

5. Related literature

The following references are cited in the Supporting Information for this article: Cancino-Díaz *et al.* (2002), Černý *et al.* (2007), Furche *et al.* (2014), O'Boyle *et al.* (2011) and Rüger *et al.* (2015).

Acknowledgements

This study was supported by the Czech Science Foundation (GA CR) grant No. 16-20507S and by project BIOCEV

CZ.1.05/1.1.00/02.0109 from the ERDF. It was conducted at the Institute of Biotechnology CAS with help of the institutional grant RVO: 86652036. We thank HZB for the allocation of synchrotron-radiation beamtime.

References

Bach, E. A., Szabo, S. J., Dighe, A. S., Ashkenazi, A., Aguet, M., Murphy, K. M. & Schreiber, R. D. (1995). *Science*, **270**, 1215–1218.

Bazan, J. F. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 6934–6938.

Bleicher, L., de Moura, P. R., Watanabe, L., Colau, D., Dumoutier, L., Renauld, J.-C. & Polikarpov, I. (2008). *FEBS Lett.* **582**, 2985–2992.

Böhm, G., Muhr, R. & Jaenicke, R. (1992). *Protein Eng. Des. Sel.* **5**, 191–195.

Bravo, J., Staunton, D., Heath, J. K. & Jones, E. Y. (1998). *EMBO J.* **17**, 1665–1674.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. (2009). *BMC Bioinformatics*, **10**, 421.

Cancino-Díaz, J. C., Reyes-Maldonado, E., Bañuelos-Pánuco, C. A., Jiménez-Zamudio, L., García-Latorre, E., León-Dorantes, G., Blancas-González, F., Paredes-Cabrera, G. & Cancino-Díaz, M. E. (2002). *J. Investig. Dermatol.* **119**, 1114–1120.

Černý, J., Biedermannová, L., Mikulecký, P., Zahradnik, J., Charnavets, T., Šebo, P. & Schneider, B. (2015). *Biomed Res. Int.* **2015**, 716945.

Černý, J., Jurečka, P., Hobza, P. & Valdés, H. (2007). *J. Phys. Chem. A*, **111**, 1146–1154.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Cook, J. R., Emanuel, S. L., Donnelly, R. J., Soh, J., Mariano, T. M., Schwartz, B., Rhee, S. & Pestka, S. (1994). *J. Biol. Chem.* **269**, 7013–7018.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fountoulakis, M., Juranville, J.-F., Stüber, D., Weibel, E. K. & Garotta, G. (1990). *J. Biol. Chem.* **265**, 13268–13275.

Furche, F., Ahlrichs, R., Hättig, C., Klopper, W., Sierka, M. & Weigend, F. (2014). *Wiley Interdiscipl. Rev. Comput. Mol. Sci.* **4**, 91–100.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *Bioinformatics*, **19**, 163–164.

Harlos, K., Martin, D. M., O'Brien, D. P., Jones, E. Y., Stuart, D. I., Polikarpov, I., Miller, A., Tuddenham, E. G. & Boys, C. W. (1994). *Nature (London)*, **370**, 662–666.

Hemmi, S., Böhni, R., Stark, G., Di Marco, F. & Aguet, M. (1994). *Cell*, **76**, 803–810.

Hoffmann, H. H., Schneider, W. M. & Rice, C. M. (2015). *Trends Immunol.* **36**, 124–138.

Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.

Josephson, K., Logsdon, N. J. & Walter, M. R. (2001). *Immunity*, **15**, 35–46.

Jung, V., Rashidbaigi, A., Jones, C., Tischfield, J. A., Shows, T. B. & Pestka, S. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 4151–4155.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kotenko, S. V., Izotova, L. S., Pollack, B. P., Mariano, T. M., Donnelly, R. J., Muthukumaran, G., Cook, J. R., Garotta, G., Silvennoinen, O., Ihle, J. N. & Pestka, S. (1995). *J. Biol. Chem.* **270**, 20915–20921.

Krause, C. D., Lavnikova, N., Xie, J., Mei, E., Mirochnitchenko, O. V., Jia, Y., Hochstrasser, R. M. & Pestka, S. (2006). *Cell Res.* **16**, 55–69.

Landar, A., Curry, B., Parker, M. H., DiGiacomo, R., Indelicato, S. R., Nagabhushan, T. L., Rizzi, G. & Walter, M. R. (2000). *J. Mol. Biol.* **299**, 169–179.

Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.

Langer, J. A., Cutrone, E. C. & Kotenko, S. (2004). *Cytokine Growth Factor Rev.* **15**, 33–48.

Larsen, K. S., Østergaard, H., Bjelke, J. R., Olsen, O. H., Rasmussen, H. B., Christensen, L., Kragelund, B. B. & Stennicke, H. R. (2007). *Biochem. J.* **405**, 429–438.

Lin, F.-C. & Young, H. A. (2013). *Adv. Biosci. Biotechnol.* **04**, 6–13.

Logsdon, N. J., Deshpande, A., Harris, B. D., Rajashankar, K. R. & Walter, M. R. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 12704–12709.

Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.

Marsters, S. A., Pennica, D., Bach, E., Schreiber, R. D. & Ashkenazi, A. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 5401–5405.

McElroy, C. A., Dohm, J. A. & Walsh, S. T. R. (2009). *Structure*, **17**, 54–65.

Miknis, Z., Margracheva, E., Li, W., Zdanov, A., Kotenko, S. W. & Wlodawer, A. (2010). *J. Mol. Biol.* **404**, 650–664.

Mikulecký, P., Černý, J., Biedermannová, L., Petroková, H., Kuchař, M., Vondrášek, J., Malý, P., Šebo, P. & Schneider, B. (2013). *Biomed Res. Int.* **2013**, 752514.

Moncada-Vélez, M. *et al.* (2013). *Blood*, **122**, 2390–2401.

Moura, P. R. de, Watanabe, L., Bleicher, L., Colau, D., Dumoutier, L., Lemaire, M. M., Renauld, J.-C. & Polikarpov, I. (2009). *FEBS Lett.* **583**, 1072–1077.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T. & Hutchison, G. R. (2011). *J. Cheminform.* **3**, 33.

Okonechnikov, K., Golosova, O. & Fursov, M. (2012). *Bioinformatics*, **28**, 1166–1167.

Pestka, S., Kotenko, S. V., Muthukumaran, G., Izotova, L. S., Cook, J. R. & Garotta, G. (1997). *Cytokine Growth Factor Rev.* **8**, 189–206.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Plíhal, O., Sklenář, J., Kmoníčková, J., Man, P., Pompach, P., Havlíček, V., Křen, V. & Bezouška, K. (2004). *Biochem. Soc. Trans.* **32**, 764–765.

Pompach, P., Man, P., Kavan, D., Hofbauerová, K., Kumar, V., Bezouška, K., Havlíček, V. & Novák, P. (2009). *J. Mass Spectrom.* **44**, 1571–1578.

Ring, A. M., Lin, J.-X., Feng, D., Mitra, S., Rickert, M., Bowman, G. R., Pande, V. S., Li, P., Moraga, I., Spolski, R., Özkan, E., Leonard, W. J. & Garcia, K. C. (2012). *Nature Immunol.* **13**, 1187–1195.

Rüger, R., van Lenthe, E., Lu, Y., Frenzel, J., Heine, T. & Visscher, L. (2015). *J. Chem. Theory Comput.* **11**, 157–167.

Savan, R., Ravichandran, S., Collins, J. R., Sakai, M. & Young, H. A. (2009). *Cytokine Growth Factor Rev.* **20**, 115–124.

Schilling, B., Row, R. H., Gibson, B. W., Guo, X. & Young, M. M. (2003). *J. Am. Soc. Mass Spectrom.* **14**, 834–850.

Schoenborn, J. R. & Wilson, C. B. (2007). *Adv. Immunol.* **96**, 41–101.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol. Syst. Biol.* **7**, 539.

Thiel, D. J., le Du, M.-H., Walter, R. L., D'Arcy, A., Chène, C., Fountoulakis, M., Garotta, G., Winkler, F. K. & Ealick, S. E. (2000). *Structure*, **8**, 927–936.

Thomas, C., Moraga, I., Levin, D., Krutzik, P. O., Podoplelova, Y., Trejo, A., Lee, C., Yarden, G., Vleck, S. E., Glenn, J. S., Nolan, G. P., Piehler, J., Schreiber, G. & Garcia, K. C. (2011). *Cell*, **146**, 621–632.

Walter, M. R., Windsor, W. T., Nagabhushan, T. L., Lundell, D. J., Lunn, C. A., Zauodny, P. J. & Narula, S. K. (1995). *Nature (London)*, **376**, 230–235.

Yoon, S.-I., Jones, B. C., Logsdon, N. J., Harris, B. D., Deshpande, A., Radaeva, S., Halloran, B. A., Gao, B. & Walter, M. R. (2010). *Structure*, **18**, 638–648.

STRUCTURAL BIOLOGY

Volume 72 (2016)

Supporting information for article:

Crystal structure of the human interferon gamma receptor 2 reveals the structural basis for receptor specificity

Pavel Mikulecký, Jirí Zahradník, Petr Kolenko, Jirí Cerný, Tatsiana Charnavets, Lucie Kolárová, Iva Necasová, Phuong Ngoc Pham and Bohdan Schneider

S1. Methods

S1.1. Cloning, expression, and purification of endoglycosidases. Endoglycosidases Endo H (residues 42-313; UniProt P04067) and PNGase F (residues 8-299; UniProt Q9XBM8; (Cancino-Diaz et al., 2002)) were produced in *Escherichia coli* BL21 (λDE3) and purified before deglycosylation trials. Briefly, genes were obtained as a GeneArt Strings DNA Fragments and cloned into previously modified vector pET26b(+) bearing StrepTag instead of 6x HisTag using *Nde*I and *Xho*I restriction enzymes. Cells were grown in 1 L of LB medium at 30 °C until the OD_{600} = 0.6 was reached, then the temperature was decreased to 16 °C and protein expression was induced by addition of 1 mM IPTG, and cells were incubated for another 20 hours. Two-step purification of enzymes consisted of gravity flow StrepTactin Sepharose followed by size exclusion chromatography on a HiLoad 16/600 Superdex 75 pg (GE Healthcare).

S1.2. Thermal shift assay (TSA). Melting temperature curves was also obtained from fluorescence-based Thermal shift assay using fluoroprobe. Experiment was performed in "CFX96 Touch Real-Time PCR Detection System" (Bio-Rad) using FRET Scan Mode. The concentration of fluorescent SYPRO Orange dye (Sigma Aldrich) was 8-fold dilution from 5,000-fold stock and protein concentration was 0.05 mg/mL in final volume of 25 µL. As a reference was used only "HN Buffer" without protein. Thermal denaturation of proteins was performed in capped "Low Tube Strips, CLR" (Bio-Rad) and possible air bubbles in samples were removed by centrifugation immediately before the assay. The samples were heated from 20 to 95 °C with stepwise increment of 0.5 °C/min and a 30 s hold step for every point, followed by the fluorescence reading. Data (after subtraction of reference sample) were normalized and used for first derivative calculation to estimate the melting temperature.

Table S1. Primers for mutagenesis. The IFNgR2 variants bearing single mutations: N110Q, N137Q, and N231Q were produced using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies). Primer pair (forward primer – Fwd and reverse – Rev) are marked by mutation code.

Primer name	Sequence (mutant position highlighted)
FwdN137Q:	5'-GTTTCAACACTATCGGCAAGTGACTGTCGGGCCTC-3'
RevN137Q:	5'-GAGGCCCGACAGTCACTTGCCGATAGTGTTGAAAC-3'
FwdN110Q:	5'-CTTCCCAATGGATTTCCAGGTCACTCTACGCCTTCG-3'
RevN110Q:	5'-CGAAGGCGTAGAGTGACCTGGAAATCCATTGGGAAG-3'
FwdN231Q:	5'-GAGTCGGGCATTTAAGCCAGATATCTTGCTACGAAACAATG-3'
RevN231Q:	5'-CATTGTTTCGTAGCAAGATATCTGGCTTAAATGCCCGACTC-3'

Table S3. The mutual orientation of the D1 and D2 domains of the cytokine receptors measured as torsion angles. The torsion angles are defined by CA atoms of four residues highly conserved among all receptors.

PDB and chain IDs	Receptor name	Torsion	Residues defining torsion between D1 and D2			
			Residue 1	Residue 2	Residue 3	Residue 4
5eh1_A	IFNgR2	-38º	A92	R89	Y150	V152
3lqm_A	IL10R2	-49º	A91	R88	Y154	V156
1fg9_C	IFNgR1	-30º	A87	R84	Y155	V157
1j7v_R	IL10R1	-14º	A81	R78	Y146	I148
4doh_B	IL20R2	-82º	A115	R112	F165	V167
4doh_E	IL20R1	-27º	A114	R111	Y181	V183
3dlq_R	IL22R	-21º	A98	R95	Y165	L167
3g9v_A	IL22BP	-40º	A105	R102	Y171	V173
3og6_B	IL28R	-17º	T82	R79	Y138	V140

Table S2. Amino acid residues (numbered as in the PDB structures) used to calculate root mean square deviations (rmsd, [Å]) for the D1 and D2 domains of the twelve analyzed receptor molecules listed in Figure 3. Note that the number of superimposed residues is the same in all domains, 34.

			Calculation of rmsd of 34 residues of the N-terminal D1 domain																																	
PDB ID	chain ID	receptor	residue number as numbered in the PDB file																																	
1fg9	C	IFNgR1	28	29	30	31	41	42	43	44	45	46	47	56	57	58	59	60	65	66	67	68	69	81	82	83	84	85	86	87	88	94	96	97	98	99
5eh1	A	IFNgR2	21	22	23	24	37	38	39	40	41	42	43	49	50	51	52	61	66	67	68	69	70	86	87	88	89	90	91	92	93	99	101	102	103	104
1j7v	R	IL10R1	21	22	23	24	35	36	37	38	39	40	41	48	49	50	51	52	57	58	59	60	61	75	76	77	78	79	80	81	82	88	90	91	92	93
3lqm	A	IL10R2	37	38	39	40	51	52	53	54	55	56	57	62	63	64	65	66	71	72	73	74	75	85	86	87	88	89	90	91	92	98	100	101	102	103
4doh	B	IL20R2	51	52	53	54	65	66	67	68	69	70	71	83	84	85	86	89	94	95	96	97	98	109	110	111	112	113	114	115	116	122	124	125	126	127
4doh	E	IL20R1	53	54	55	56	68	69	70	71	72	73	74	81	82	83	84	87	92	93	94	95	96	108	109	110	111	112	113	114	115	121	123	124	125	126
3dlq	R	IL22R	38	39	40	41	52	53	54	55	56	57	58	65	66	67	68	71	76	77	78	79	80	92	93	94	95	96	97	98	99	106	107	108	109	110
3g9v	A	IL22BP	44	45	46	47	59	60	61	62	63	64	65	72	73	74	75	78	83	84	85	86	87	99	100	101	102	103	104	105	106	112	114	115	116	117
3og6	B	IL28R	20	21	22	23	34	35	36	37	38	39	40	48	49	50	51	54	59	60	61	62	63	76	77	78	79	80	81	82	83	89	91	92	93	94
3se4	A	IFNabR1	17	18	19	20	33	34	35	36	37	38	39	46	47	48	49	52	57	58	59	60	61	73	74	75	76	77	78	79	80	86	87	88	89	90
3se4	C	IFNabR2	24	25	26	27	38	39	40	41	42	43	44	52	53	54	55	58	63	64	65	66	67	79	80	81	82	83	84	85	86	93	94	95	96	97
2puq	T	Tissue_F	22	23	24	25	33	34	35	36	37	38	39	45	46	47	48	49	54	55	56	57	58	71	72	73	74	75	76	77	78	93	94	95	96	97

			Calculation of rmsd of 34 residues of the C-terminal D2 domain																																	
PDB ID	chain ID	receptor	residue number as numbered in the PDB file																																	
1fg9	C	IFNgR1	113	114	115	116	117	118	124	125	126	127	128	129	152	154	155	156	157	158	159	184	185	186	187	197	198	199	200	201	202	203	204	213	214	216
5eh1	A	IFNgR2	118	120	121	122	123	124	130	131	132	133	134	135	148	149	150	151	152	153	154	171	172	173	174	184	185	186	187	188	189	190	191	203	204	206
1j7v	R	IL10R1	106	108	109	110	111	112	118	119	120	121	122	123	144	145	146	147	148	149	150	168	169	170	171	181	182	183	184	185	186	187	188	197	198	200
3lqm	A	IL10R2	116	117	118	119	120	121	127	128	129	130	131	132	152	153	154	155	156	157	158	175	176	177	178	188	189	190	191	192	193	194	195	204	205	207
4doh	B	IL20R2	141	142	143	144	145	146	152	153	154	155	156	157	163	164	165	166	167	168	169	189	190	191	192	202	203	204	205	206	207	208	209	218	219	221
4doh	E	IL20R1	140	141	142	143	144	145	151	152	153	154	155	156	179	180	181	182	183	184	185	202	203	204	205	215	216	217	218	219	220	221	222	232	233	234
3dlq	R	IL22R	124	125	126	127	128	129	135	136	137	138	139	140	163	164	165	166	167	168	169	185	186	187	188	198	199	200	201	202	203	204	205	214	215	216
3g9v	A	IL22BP	131	132	133	134	135	136	142	143	144	145	146	147	169	170	171	172	173	174	175	192	193	194	195	206	207	208	209	210	211	212	213	222	223	225
3og6	B	IL28R	108	109	110	111	112	113	119	120	121	122	123	124	136	137	138	139	140	141	142	161	162	163	164	175	176	177	178	179	180	181	182	192	193	195
3se4	A	IFNabR1	104	105	106	107	108	109	115	116	117	118	119	120	136	137	138	139	140	141	142	159	160	161	162	172	173	174	175	176	177	178	179	188	189	191
3se4	C	IFNabR2	110	111	112	113	114	115	121	122	123	124	125	126	139	140	141	142	143	144	145	167	168	169	170	180	181	182	183	184	185	186	187	195	196	198
2puq	T	Tissue_F	111	112	113	115	116	117	123	124	125	126	127	128	151	152	153	154	155	156	157	174	175	176	177	186	187	188	189	190	191	192	193	204	205	207

Calculation of rmsd of 34 residues providing the best overlap between D1 and D2 within each receptor molecule

PDB ID	chain ID	recepto	residue number as numbered in the PDB file																																	
1fg9	C	IFNgR1N	15	18	19	20	24	26	27	30	40	42	43	44	45	46	47	48	64	66	67	68	80	81	82	84	85	86	87	88	89	91	92	93	102	103
1fg9	C	IFNgR1C	113	115	116	117	121	123	124	127	152	155	156	157	158	159	160	161	172	183	184	185	195	196	197	199	200	201	202	203	204	208	209	210	219	220
5eh1	A	IFNgR2N	8	9	10	11	12	13	14	15	21	22	23	24	38	39	40	41	42	45	64	65	66	67	68	69	81	87	88	89	90	91	93	97	101	102
5eh1	A	IFNgR2C	141	142	143	144	145	149	151	152	153	155	168	169	170	171	172	173	174	175	177	178	179	180	181	182	183	184	185	195	196	198	199	200	201	202
1j7v	R	IL10R1N	7	8	9	10	11	13	20	21	24	25	32	39	40	41	45	46	47	50	57	58	59	60	61	62	67	74	76	77	78	79	81	82	89	90
1j7v	R	IL10R1C	105	106	107	108	109	111	118	119	122	123	142	149	150	151	156	157	158	161	166	167	168	169	170	171	174	179	181	182	183	184	186	187	196	197
3lqm	A	IL10R2N	27	29	37	48	50	51	52	53	54	56	60	63	64	71	72	73	74	75	76	86	87	88	89	90	91	92	93	95	96	103	104	105	106	107
3lqm	A	IL10R2C	118	120	128	149	152	153	154	155	156	158	164	167	168	173	174	175	176	177	178	188	189	190	191	192	193	194	195	199	200	208	209	210	211	212
4doh	B	IL20R1N	63	66	67	68	69	70	71	72	73	74	75	77	79	88	89	102	104	106	107	108	109	110	111	112	113	114	115	119	121	122	123	124	125	126
4doh	B	IL20R1C	171	174	176	178	179	180	181	182	183	184	185	186	191	193	194	195	196	197	198	210	213	214	215	216	217	218	219	220	221	222	224	226	234	235
4doh	E	IL20R2N	38	52	64	65	66	67	68	69	70	71	72	98	99	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	130	131	132
4doh	E	IL20R2C	141	154	163	164	165	166	167	168	169	170	171	191	192	200	201	202	203	204	205	206	207	208	209	210	213	214	215	216	217	218	219	223	224	225
3dlq	R	IL22RN	20	22	25	26	27	28	29	30	39	41	43	44	45	46	47	48	51	52	53	54	55	56	57	58	59	62	90	98	100	104	105	106	107	112
3dlq	R	IL22RC	121	122	123	124	125	126	134	164	165	166	167	168	169	170	171	198	200	202	203	204	210	211	212	213	214	215	216	217	218	219	220	221	222	224
3g9v	A	IL22BPN	30	31	34	35	36	43	44	45	46	48	49	50	58	59	60	61	62	63	64	75	81	82	83	100	101	102	103	104	105	111	112	113	114	115
3g9v	A	IL22BPC	130	131	133	134	135	142	143	144	145	147	148	149	169	170	171	172	173	174	175	186	189	190	191	206	207	208	209	210	211	219	220	221	222	223
3og6	B	IL28RN	7	17	19	20	21	24	31	32	34	35	36	37	38	39	40	46	58	62	63	64	65	74	76	79	80	81	82	83	87	88	89	90	91	92
3og6	B	IL28RC	108	116	119	120	121	124	134	135	137	138	139	140	141	142	143	147	157	162	163	164	165	172	174	177	178	179	180	181	188	189	190	191	192	193
3se4	A	IFNabR1N	11	13	18	19	20	33	34	35	36	37	38	39	40	41	47	48	54	58	59	60	71	72	73	74	75	76	77	78	79	80	87	88	93	94
3se4	A	IFNabR1C	107	109	117	118	119	137	138	139	140	141	142	143	144	145	151	152	155	158	159	160	169	170	171	172	173	174	175	176	177	178	188	189	194	195
3se4	C	IFNabR2N	39	40	41	42	43	50	52	53	54	59	66	67	68	69	70	72	73	74	77	78	79	80	81	82	83	84	85	94	95	96	97	98	99	100
3se4	C	IFNabR2C	141	142	143	144	145	149	151	152	153	155	168	169	170	171	172	173	174	175	177	178	179	180	181	182	183	184	185	195	196	198	199	200	201	202
2puq	T	TissueFN	6	13	14	22	23	24	25	26	35	36	39	40	41	44	45	46	47	54	55	56	57	58	59	70	71	72	73	74	75	76	93	94	99	100
2puq	T	TissueFC	108	115	116	124	125	126	127	128	154	155	158	159	160	164	165	166	167	172	173	174	175	176	177	184	185	186	187	188	189	190	203	204	209	210

Figure S1. SDS-PAGE analysis of the purified glycosylated and deglycosylated IFNgR2. The 12.5% Tris-Glycine gel was stained using Coomassie Brilliant Blue. Lane M – molecular-weight markers with depicted mass in kDa; Lane R2g – glycosylated IFNgR2; Lane R2d – IFNgR2 deglycosylated by Endo H endoglycosidase.

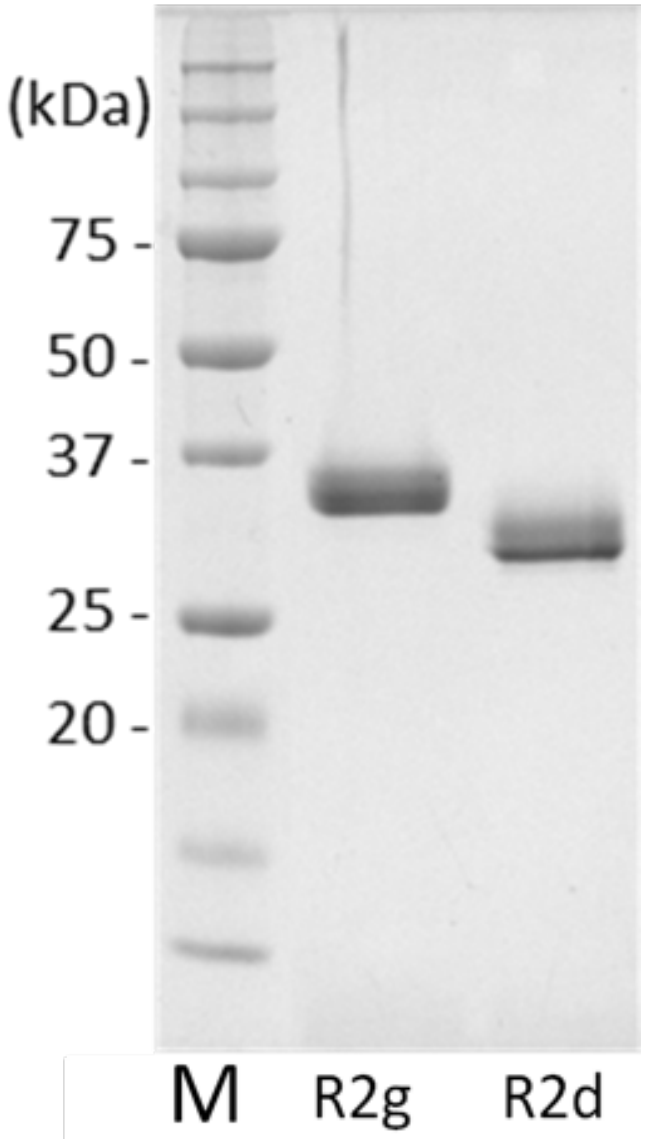


Figure S2. Normalized melting curves measured by thermal shift assay (TSA). Data showed that melting temperature of deglycosylated IFNgR2 (48 °C) is slightly lower than of glycosylated variant (50.5 °C). The addition of 5 mM TCEP to break the disulfide bonds lower the melting temperature just by one degree of Celsius of both deglycosylated and glycosylated IFNgR2 (47 and 49.5 °C, respectively).

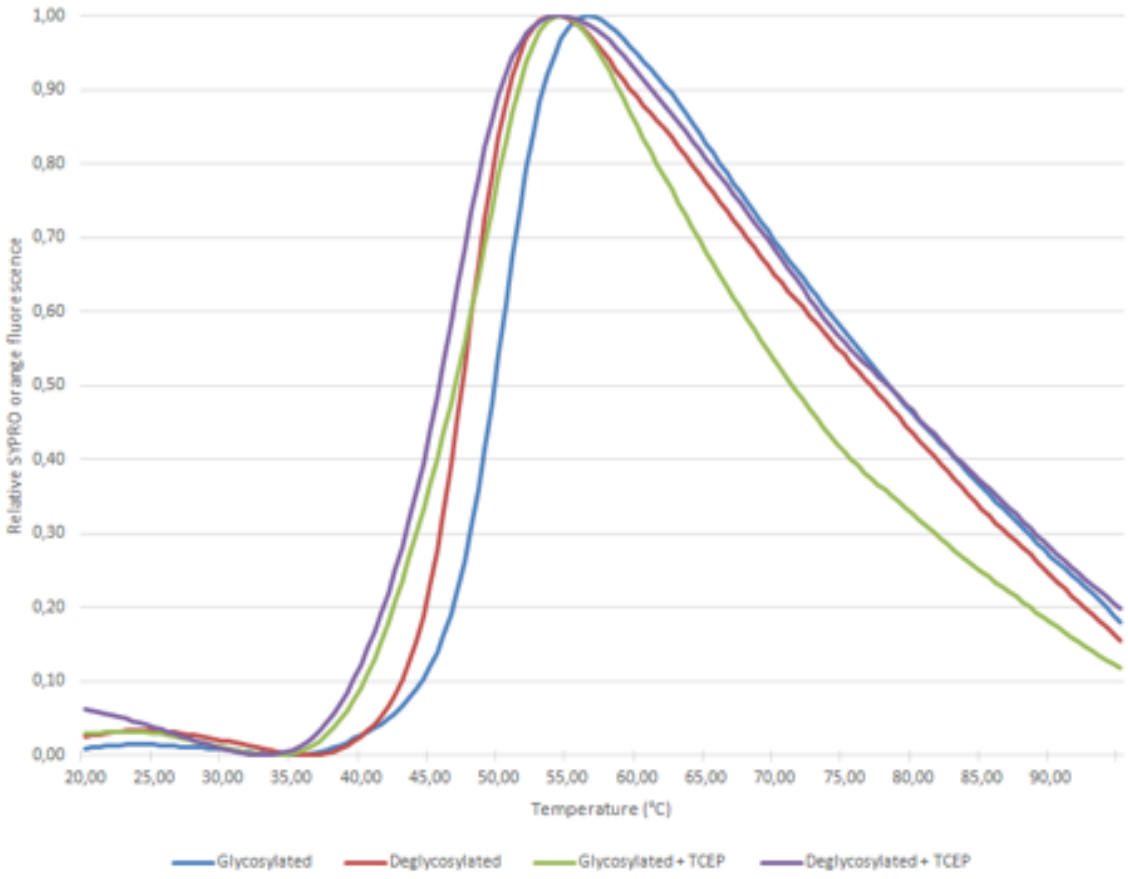
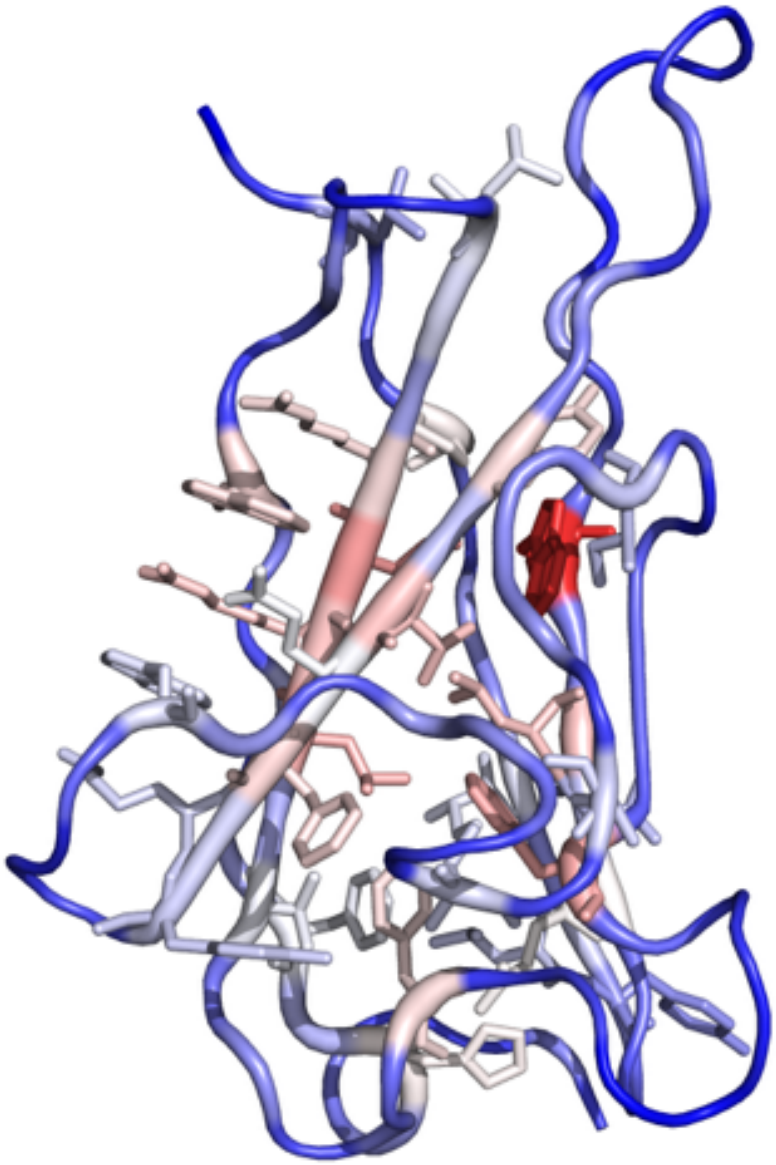


Figure S3. Residues responsible for stability of the IFNgR2 N-terminal (D1) domain. The color and thickness of the cartoon representation show relative interaction energy per residue ranging from low (blue) to high stabilizing values (red). The interaction energy within D1 was calculated at the DFT-D level (RI-TPSS/TZVP augmented with empirical dispersion term (Černý et al., 2007)) with the solvation effects described using COSMO in program TurboMole (Furche et al., 2014). All pairs of D1 residues within 4 Å were extracted from the 5eh1 structure using VMD; it was 113 residues forming 580 pairs. Hydrogen atoms were added using OpenBabel (O'Boyle et al., 2011) and their positions were optimized at the DFTB-D level of theory as implemented in dftb+ program (Rüger et al., 2015). Pairwise interaction energy was calculated at the DFT-D level and contributions involving each residue were summed.



Bibliography

Cancino-Diaz, J. C., Reyes-Maldonado, E., Banuelos-Panuco, C. A., Jimenez-Zamudio, L., Garcia-Latorre, E., Leon-Dorantes, G., Blancas-Gonzalez, F., Paredes-Cabrera, G. & Cancino-Diaz, M. E. (2002). *J. Investig. Dermatol.* **119**, 1114-1120.

Černý, J., Jurečka, P., Hobza, P. & Valdés, H. (2007). *The Journal of Physical Chemistry A* **111**, 1146-1154.

Furche, F., Ahlrichs, R., Hättig, C., Klopper, W., Sierka, M. & Weigend, F. (2014). *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 91-100.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T. & Hutchison, G. R. (2011). *Journal of Cheminformatics* **3**, 33-33.

Rüger, R., van Lenthe, E., Lu, Y., Frenzel, J., Heine, T. & Visscher, L. (2015). *Journal of Chemical Theory and Computation* **11**, 157-167.