TECHNICKÁ UNIVERSITA V LIBERCI

HABILITAČNÍ PRÁCE

2005

Jiří Brynda

Určování struktury biologických makromolekul metodou rentgenové strukturní analýzy.

Jiří Brynda

Ústav fyzikální biologie

Jihočeská univerzita v Českých Budějovicích

a

Ústav molekulární genetiky

AKADEMIE VĚD ČESKÉ REPUBLIKY

HABILITAČNÍ PRÁCE
2005

Obsah

Řešení trojrozměrné struktury biologických makromolekul metodou rentgenové strukturní analýzy.

1. Úvod

Metoda rentgenové monokrystalové strukturní analýzy je často vysvětlována pomocí analogie s klasickou světelnou mikroskopií. Při určování struktury s pomocí mikroskopu využíváme interakce hmoty, preparátu s elektromagnetickým zářením v oboru viditelného světla (obor vlnových délek cca 350 – 800 nm). U metody rentgenové monokrystalové strukturní analýzy (v dalším textu budeme hovořit jen o **rentgenové krystalografii** nebo **rentgenové proteinové krystalografii**) je situace obdobná, strukturu molekuly tvořící krystal studujeme pomocí interakce s elektromagnetickým zářením, tentokráte v oboru rentgenového záření (obor vlnových délek cca 50 – 500 pm; 0.5 – 5 Å). Ze srovnání používaných vlnových délek vyplývá, jaké strukturní detaily je možné srovnávanými metodami studovat. Je známo, že maximální rozlišení jakékoli fyzikální metody využívající elektromagnetické záření je rovno polovině vlnové délky použitého záření, proto je světelná mikroskopie vhodná ke studiu například živých buněk a jejich struktur zatímco **rentgenová krystalografie** je vhodnou metodou pro studium chemických sloučenin, kde jednotlivé atomy spojené chemickými vazbami jsou ve vzájemných vzdálenostech v oboru jednotek Å. Dobře se také na této analogii vysvětluje důležitost matematických metod. U světelné mikroskopie máme zdroj záření, pozorovaný objekt, objektiv vytvářející zvětšený obraz a okulár pro sledování vzniklého obrazu. U **rentgenové krystalografie** je situace obdobná, máme zdroj RTG záření a studovaný monokrystal, který intereaguje se zářením. Zde však analogie končí, jelikož neexistují vhodné optické členy pro práci s rtg zářením, přebírá jejich úlohu matematické zpracování. Posledním fyzikálním jevem využitým v **rentgenové krystalografii** je detekce difraktovaného záření. Ve chvíli kdy je difrakční obrazec digitalizován, probíhá další zpracování a vyhodnocování jen výpočetně s využitím příslušných programů, tedy nástrojů numerické matematiky.

2. Fyzika základů difrakce

2.1. Difrakce jedním elektronem

Nedlouho po objevu RTG záření bylo předpovězeno a následně i ukázáno, že toto záření na krystalech difraktuje (rozptyluje se), tj. je odkláněno od původního směru do konkrétních směrů. Difrakce rtg paprsku na krystalu je výsledkem interakce elektrické složky elektromagnetického záření s elektrony v krystalu. Pokud monochromatické polarizované záření dopadne na elektron, elektron začne oscilovat ve směru elektrického vektoru dopadajícího záření.Tento oscilující elektron se stává sám zdrojem záření se stejnou frekvencí (elektron může taktéž vyzařovat na nižší frekvenci, kterou se však pro nedostatek prostoru nebudeme zabývat; Comptonuv neelastický rozptyl). Při elastickém rozptylu (Thomsonův jev) je:

1) Fázový rozdíl mezi dopadající a rozptýlenou vlnou π.

2) Amplituda elektrické rozptýlené části vlny ve vzdálenosti r (vzdálenost velká ve srovnání s vlnovou délkou záření) je

$$E_{el} = E_0 \frac{1}{r} \frac{e^2}{mc^2} \sin\varphi, \qquad (1) \ \textit{Thomsonův vztah}$$

kde E_0 je amplituda elektrického vektoru dopadajícího záření , e náboj elektronu, m jeho hmotnost, c rychlost světla a φ je úhel mezi směrem oscilace elektronu a rozptýleným zářením. Vyjádřeno v energii,

$$I_{el} = I_0 \frac{1}{r^2} \left(\frac{e^2}{mc^2} \right)^2 \sin^2\varphi, \qquad (2) \ \textit{Thomsonův vztah pro energii}$$

Bylo dokázáno (Klein&Nishima 1929 a také Heitler 1966), že rozptyl na jednom elektronu může být dobře popsán klasickým Thomsonovým rozptylem pokud pro užité záření platí $h\nu \ll mc^2$, takže vzhledem k rozsahu použitých vlnových délek (viz. úvod) je klasická aproximace vhodná. Zde je vhodné poznamenat že:

1) Intenzita rozptýlená volným elektronem je nezávislá na vlnové délce .

2) Thomsonův vztah platí též pro protony a atomová jádra, avšak vzhledem k jejich hmotnosti je jejich příspěvek zanedbatelný.

3) Vztah 2 platí pro polarizované záření a pro nepolarizováné záření musí být člen $\sin^2\varphi$ nahrazen vhodným polarizačním faktorem.

2.2. Difrakce na dvojici elektronů

Vhodný vztah pro popis difrakce systému lze odvodit klasicky výpočtem fázové diference mezi vlnami rozptýlenými oběma atomy. Na obrázku 1 je znázorněn systém dvou elektronů, počátek leží v pozici elektronu 1 a elektron 2 je v pozici $\vec{r}$. Elektrony rozptylují dopadající záření popsané vektorem $\vec{s}_0$ ve směru popsaném vektorem $\vec{s}$. Délka vektoru může být volena libovolně, ale je s výhodou zvolit délku $1/\lambda$. Oba elektrony rozptylují (difraktují) nezávisle na sobě a proto amplitudy rozptýlené elektronem jedna a dva jsou shodné, ale mají rozdíl ve fázi. Fázový rozdíl má původ v dráhovém rozdílu mezi paprsky prošlými elektronem jedna a dva. Dráhový rozdíl je $p + q = \lambda[\vec{r}.(\vec{s}_0 - \vec{s})]$ a výsledný fázový rozdíl je

$$-2\pi\lambda[\vec{r}.(\vec{s}_0 - \vec{s})]/\lambda = 2\pi\vec{r}.\vec{S} \tag{3}$$

kde $\vec{S} = \vec{s} - \vec{s}_0$.
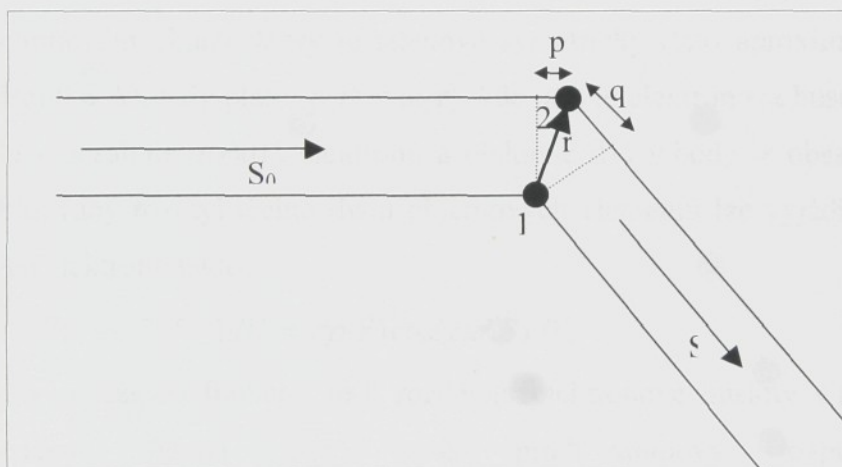
Obrázek 1. Černé body jsou elektrony. Počátek leží v elektronu 1 a elektron 2 leží v pozici $\vec{r}$. Elektrony jsou ozařovány rtg paprskem popsaným vektorem $\vec{s}_0$. Rozptýlené záření je pozorováno ve směru vektoru $\vec{s}$. Vzhledem k dráhovému rozdílu $p + q$, bude paprsek rozptýlený elektronem 2 pozadu ve fázi za paprskem rozptýleným elektronem 1.

Z obrázku 2 je zřejmé, že směr vektoru $\vec{S}$ je kolmý k imaginární reflexní rovině a že úhel dopadu záření a úhel ve kterém je záření rozptýleno jsou shodné. Dále lze ukázat že, délka $\vec{S}$ je dána vztahem

$$|\vec{S}| = 2\sin\theta/\lambda. \tag{4}$$
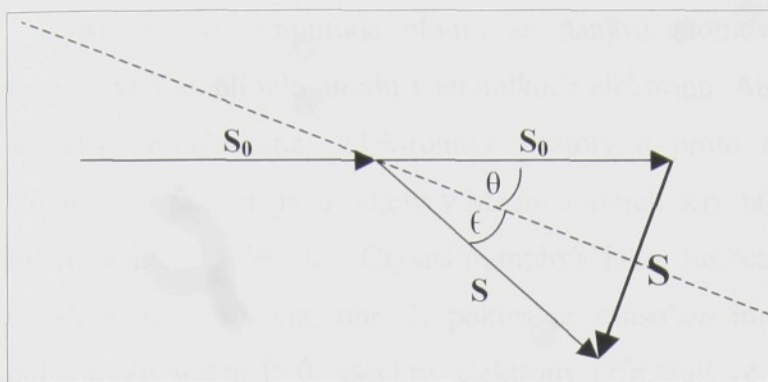
Obrázek 2. Difrakcí se dopadající vlna $\vec{s}_0$ mění ve vlnu $\vec{s}$. Velikost obou vektoru je rovna $1/\lambda$. Obrázek vysvětluje, jak může být difrakce popsána jako odraz na zrcadlové rovině (zrcadlová rovina čárkovaně).

2.3. Difrakce na atomu.

Pro volný elektron nebo elektrony volně vázané k atomu (tj. je-li vazebná energie a tlumení harmonického kmitání elektronu způsobené vazbou k atomu zanedbatelné vůči energii dopadajícího záření) lze amplitudu rozptylu atomem vyjádřit jako poměr rozptylu na atomu vůči rozptylu na jednom elektronu. Protože elektrony jsou okolo jádra rozmístěny v elektronovém oblaku který je středově symetrický (tato aproximace je použitelná až do rozlišení 0.6 Å), tedy platí: $\rho(r) = \rho(-r)$, kde $\rho(r)$ je elektronová hustota v bodě r. Objem dV_r v bodě r obsahuje $\rho(r)dV_r$ elektronů a obdobně dV_r v bodě $-r$ obsahuje $\rho(-r)dV_r$ elektronů. Kombinovaný rozptyl těchto dvou objemových elementů lze vyjádřit v jednotkách rozptylu volných elektronů takto:

$$\rho(\vec{r})(e^{2\pi i \vec{r}.\vec{S}} + e^{2\pi i(-\vec{r}).\vec{S}})dV_r = 2\rho(\vec{r})\cos(2\pi\vec{r}.\vec{S})dV_r \qquad (5),$$

jde tedy o reálnou funkci. Je-li rozdělení elektronové hustoty v atomu popsáno středově symetrickou funkcí $\rho(\vec{r})$, pak pro atomový rozptylový faktor platí:

$$f(\vec{S}) = \int_V \rho(\vec{r})\,e^{(2\pi r.S)}\,dV_r = \int_{V/2} 2\rho(\vec{r})\cos(2\pi\vec{r}.\vec{S})\,dV_r \qquad (6), \quad \text{atomový rozptylový faktor,}$$

kde $\vec{S} = (\vec{s} - \vec{s}_0)/\lambda$ je difrakční vektor zavedený na obr. 2 a integrace s proměnnou $\vec{r}$ probíhá přes celý prostor.
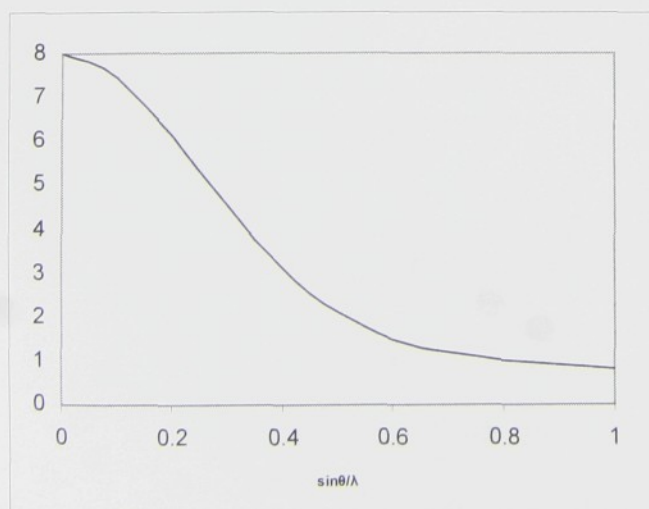
Obrázek 3. Hodnota atomového rozptylového faktoru f v závislosti na rozptylovém úhlu (např. pro atom kyslíku; osm elektronů).

Rozptylová amplituda atomu se nazývá atomový rozptylový faktor f. Vyjadřuje rozptylovou amplitudu atomu v jednotkách elektronu. Atomové f faktory jsou vypočteny pro sféricky průměrované elektronové hustoty a proto nejsou závislé na směru difrakce. Hodnoty f faktorů jsou udány v Mezinárodních Krystalografických tabulkách ITC (ITC – International Tables for Crystallography) jako funkce $\sin\theta/\lambda$. Hodnota f faktoru klesá s rostoucím $\sin\theta/\lambda$ viz. obr. 3, pokles je způsoben interferencí elektronů v elektronovém oblaku. Ve směru $\theta=0$, všechny elektrony difraktují ve fázi a atomový strukturní faktor je roven počtu elektronů atomu.

3. Rentgenová strukturní analýza

Rentgenová strukturní analýza je analýzou difrakčního obrazu vzniklého interakcí rentgenova záření s krystalem. Vše je založeno na faktu, že elektromagnetické záření je nabitými částicemi rozptylováno. Podle Thomsonova vztahu (viz výše) je intenzita rozptylu nepřímo úměrná hmotnosti částice, a proto se zajímáme o rozptyl na elektronech a ne na protonech. Krystal je periodicky se opakující prostředí obsahující periodicky se opakující elektronovou hustotu, a proto v něm dochází k difrakci elektromagnetického záření. Difrakce je pozorovatelná pro rentgenové záření, jehož vlnová délka je řádově srovnatelná s typickými mezirovinnými vzdálenostmi v krystalu. Kritérium pro pozorování difrakce je vyjádřeno Braggovým zákonem nebo v reciprokém prostoru pomocí Ewaldovy konstrukce, jak je to zachyceno na obrázku 4.

Braggův zákon   Ewaldova konstrukce

$d(s-s_0)=n\lambda$
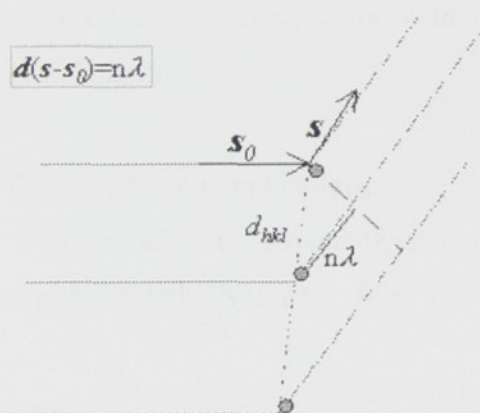
θ ....Bragguv úhel
2θ ...difrakcní úhel

s_0..jednotkový vektor ve smeru dopadu
s....jednotkový vektor ve smeru rozptylu
S...difrakcní vektor, $S=(s-s_0)/\lambda$

$2d_{hkl}\sin\theta=n\lambda$ hkl nesoudělná (Millerovy indexy), n lib. prirozené císlo
$|S|=|s-s_0|/\lambda=2\sin\theta/\lambda=1/d_{hkl}$ hkl soudelná (difrakcní indexy), n=1
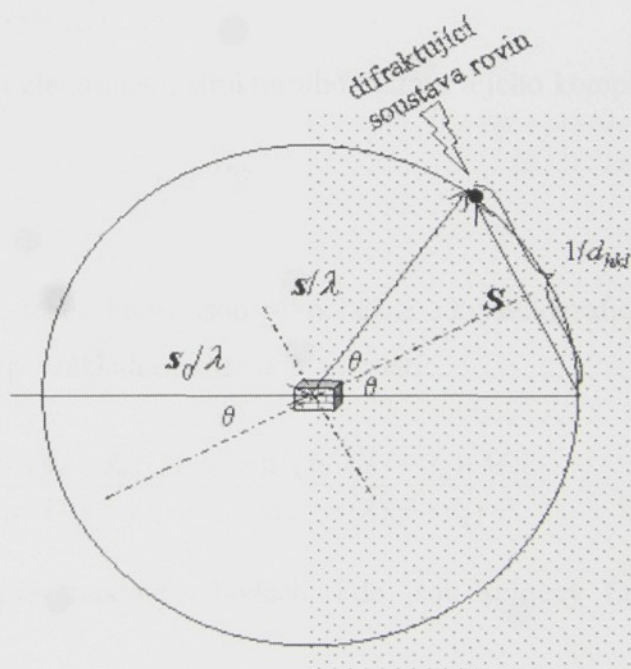
Obrázek 4. Braggův zákon, Ewaldova konstrukce.

Teorie rentgenové difrakce byla popsána v mnoha učebnicích (Valvoda, 1992; Giacovazzo, 2000; Clegg, 2002), zrovna tak jako specifika a praktické návody pro řešení struktur proteinů na základě rentgenové difrakce (McRee, 1993; Blundell a Johnson, 1976, Marek, 2002). Zde proto připomeňme jen některé klíčové vztahy kinematické teorie difrakce, na nichž je možnost detailní strukturní analýzy založena.

3.1. Rozptyl na libovolném souboru atomů (např. elementární buňce krystalu):

Amplituda záření rozptýleného ve směru $\vec{s}$ všemi atomy jednotkové buňky krystalu je úměrná strukturnímu faktoru

$$F(\vec{S}) = \sum_{n=1}^{N} f_n(\vec{S}) e^{\frac{2\pi i}{\lambda}(\vec{s}-\vec{s}_0)\cdot\vec{r}_n} \qquad (7) \quad strukturní \ faktor,$$

kde $\vec{r}_n$ je polohový vektor atomu v buňce.

Intenzita záření rozptýleného atomy v elementární buňce je potom rovna

$$I(\vec{S}) = I_e F(\vec{S}) F^*(\vec{S}). \qquad (8)$$

tedy součinu intenzity rozptýlené volným elektronem, strukturního faktoru a jeho komplexně sdružené hodnoty.

3.2. Difrakce na malém krystalu:

Pro krystal tvaru rovnoběžnostěnu, jehož hrany jsou rovnoběžné s krystalografickými osami a jehož délky jsou M_1a, M_2b, M_3c (při základní buňce a, b, c), platí:

$$I = I_e FF^* \frac{\sin^2(\pi/\lambda)(\vec{s}-\vec{s}_0)\cdot M_1\vec{a}}{\sin^2(\pi/\lambda)(\vec{s}-\vec{s}_0)\cdot\vec{a}} \cdot \frac{\sin^2(\pi/\lambda)(\vec{s}-\vec{s}_0)\cdot M_2\vec{b}}{\sin^2(\pi/\lambda)(\vec{s}-\vec{s}_0)\cdot\vec{b}} \cdot \frac{\sin^2(\pi/\lambda)(\vec{s}-\vec{s}_0)\cdot M_3\vec{c}}{\sin^2(\pi/\lambda)(\vec{s}-\vec{s}_0)\cdot\vec{c}} \qquad (9)$$

Intenzita difraktovaného záření I nabývá maxima v bodech, kde jsou splněny Laueho difrakční podmínky:

$$(\vec{s}-\vec{s}_0)\cdot\vec{a} = h\lambda$$
$$(\vec{s}-\vec{s}_0)\cdot\vec{b} = k\lambda \qquad (10) \ Laueho \ difrakční \ podmínky,$$
$$(\vec{s}-\vec{s}_0)\cdot\vec{c} = l\lambda$$

kde h, k, l jsou libovolná celá čísla. Čím je krystal větší, tím rychleji funkce I ubývá se vzdáleností od maxim, je tedy soustředěna do malých oblastí kolem maxim charakterizovaných příslušným h, k, l.

3.3. Vztah mezi strukturními faktory a elektronovou hustotou:

Nyní víme, že intenzity jsou nenulové jen v okolí maxim určených Laueho difrakčními podmínkami. Ve strukturních faktorech, jejichž absolutní hodnoty určíme z naměřených intenzit, můžeme proto také použít Laueho difrakční podmínky a víme tedy, že platí

$$\vec{S} \cdot \vec{r} = \frac{\vec{s} - \vec{s}_0}{\lambda} \cdot \vec{r} = \frac{\vec{s} - \vec{s}_0}{\lambda} \cdot (x, y, z) = hx + ky + lz = \vec{H} \cdot \vec{r},$$ (11)

tj. difrakční vektor lze vyjádřit jako vektor reciprokého prostoru $\vec{H} = (h, k, l)$, kde h, k, l jsou celá čísla a $\vec{r} = (x, y, z)$ jsou frakční souřadnice v elementární buňce. Mezi strukturními faktory $F(\vec{H})$ a elektronovou hustotou $\rho(\vec{r})$ pak platí vztahy:

$$F(\vec{H}) = \int_V \rho(\vec{r}) \, e^{(2\pi \vec{H} \cdot \vec{r})} \, dV \qquad (12) \qquad \text{\textit{strukturní faktor (F\textsubscript{o}, naměřený)}}$$

$$F(\vec{H}) = \sum_{n=1}^{N} f_n(\vec{H}) e^{2\pi \vec{H} \cdot r} \qquad (13) \qquad \text{\textit{strukturní faktor (F\textsubscript{c}, vypočtený)}}$$

$$\rho(x, y, z) = \sum_{h,k,l=-\infty}^{\infty} F_{(h,k,l)} e^{-2\pi i (hx + ky + lz)} \qquad (14) \qquad \text{\textit{elektronová hustota}}$$

kde V je objem elementární buňky. První výraz popisuje vlastní difrakční experiment, druhý pak výpočet strukturních faktorů na základě modelu a poslední rekonstrukci elektronové hustoty jako Fourierovu transformaci strukturních faktorů.

Ve skutečnosti se neměří maximální intenzita, ale tzv. integrální intenzita zahrnující jistou malou oblast kolem maxima. Při interpretaci naměřených intenzit je nutné brát v úvahu efekty, které zeslabují intenzitu difrakce. Vztah mezi intenzitami a strukturními faktory pak vypadá takto:

$$I(\vec{H}) = L \cdot P \cdot A \cdot E \cdot I_e \cdot F(\vec{H}) F^*(\vec{H}) \qquad (15)$$

kde L je Lorentzův faktor (korekce na různost délky průchodu reflexí Ewaldovou sférou, závisí na metodě měření), P je polarizační faktor ($P = (1 + \cos^2 \theta)/2$ pro ideálně nepolarizované záření), A je absorpční korekce (závisí na tvaru krystalu) a E je korekce na extinkci (efekt mnohonásobné difrakce na systému rovin a efekt zeslabení primárního paprsku díky difrakci na povrchových vrstvách krystalu).

Teplotní kmity atomu (a statické odchylky polohy atomu od jeho průměrné polohy v buňce průměrované přes soubor jednotlivých buněk v ozářené části krystalu) se projeví zeslabením difrakce podle rovnice (16), kde B_n je izotropní teplotní faktor n-tého atomu.

$$F(\vec{H}) = \sum_{n=1}^{N} f_n(\vec{H}) e^{2\pi \vec{H} \cdot r_n} e^{-B_n |\vec{H}|^2}$$ (16)

4. Měření difrakčních dat a jejich zpracování

Krystaly makromolekul jsou z hlediska měření difrakčních dat velice problematické. První problém vyplývá z velikosti elementární buňky, díky které jsou průměrné intenzity reflexí mnohem nižší a jejich počet vyšší. Za druhé, část objemu krystalu tvoří kanály vyplněné neuspořádaným rozpouštědlem, takže dále snižují intenzitu reflexí na vysokém rozlišení a ve většině případů omezují rozlišení na mnohem nižší než atomové. Za třetí, opět především díky obsahu rozpouštědla, jsou krystaly citlivé na radiační poškození. Nedávný technický pokrok v měřících metodách, jako např. používání synchrotronového záření, měření zmražených krystalů a vysoce účinné plošné detektory, usnadnil měření, ale to stále zůstává základem celé strukturní analýzy. Proto je nezbytné tomuto klíčovému kroku věnovat maximální pozornost.

Pro sběr dat z makromolekulárního monokrystalu při konkrétní vlnové délce potřebujete difraktometr složený z následujících součástí:

(1) zdroj RTG záření
(2) optickou soustavu pro fokusaci RTG záření na vzorek
(3) monochromátor pro selekci jedné vlnové délky záření
(4) kolimátor pro úpravu rozměru RTG záření
(5) „závěrku" (shutter) pro kontrolu délky expozice
(6) goniostat spojený s držákem krystalu
(7) účinný plošný detektor

a ještě:

(8) software pro měření, ukládání a zobrazování exponovaných dat
(9) software pro zpracování dat, tj. získání intenzit jednotlivých Braggovských difrakcí a jejich chyb

4.1. Zdroj RTG záření

Ve strukturní analýze makromolekul je praktiky vždy používána Cu anoda s fixní vlnovou délkou 1.542 Å (pokud není zdrojem synchrotronové záření). Rotační anoda má proti klasické rentgence výhodu vyšší intenzity RTG záření. Systémy jsou vybaveny grafitovým monochromátorem, nebo fokusačními zrcadly, nebo vícevrstvou optikou (Osmic), které zajišťují fokusaci a monochromatizaci RTG záření. Velice důležitá je volba apretury kolimátoru, ta by měla odpovídat velikosti krystalu, avšak pro velké krystaly s velkou elementární buňkou je lépe požít kolimator s menší apreturou něž velikost krystalu pro lepší prostorové rozlišení reflexí.

4.2. Goniostat a detektor

Difrakční podmínka je pro konkrétní reflexi splněna pokud odpovídající bod reciproké mříže protíná plochu Ewaldovy sféry. Pokud je ozařován stacionární krystal RTG zářením, pak pouze některé reflexe splňují difrakční podmínku. Proto pokud chceme změřit vetší množství dat (pokud možno všechny měřitelné, tj. kompletní soubor dat), musíme změnit buď velikost Ewaldovy sféry nebo orientaci krystalu. První možnost, s využitím polychromatického, „bílého" RTG záření je základ Laueho metody. Pokud používáme monochromatické záření vybrané vlnové délky, pak musíme geometrii krystalu během expozice měnit tak, aby byly difrakční podmínky splněny pro více reflexí. Proteinové difraktometry prakticky vždy požívají rotaci krystalu a plošný detektor.
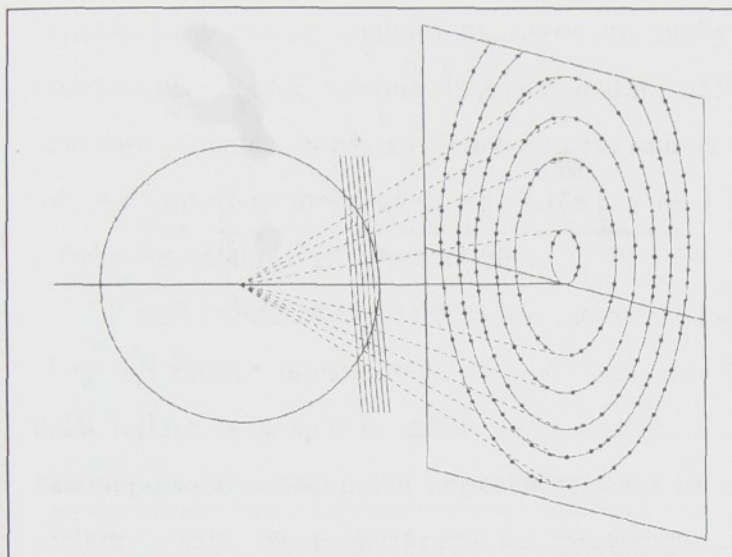
Prvním plošnými detektory byly fotografické filmy automaticky snímané optickými denzitometry. Takto digitalizované difrakční obrazce jsou dále zpracovávány. Hlavní nevýhody použití fotografických filmů byla pracnost, časová náročnost, vysoké pozadí a malý dynamický rozsah. Postupně byly vyvinuty další plošné detektory: drátové proporcionální čítače, „Imaging plates" a konečně CCD detektory (charge-coupled devices). Prakticky všechny současné plošné detektory jsou dnes používány v kombinaci s goniostatem, který zajišťuje rotaci krystalu okolo jedné osy během expozice. Zbylé přístroje vyžívají kappa geometrie (ω, κ, φ) kolíbky goniometru pro nastavení krystalu do další počáteční polohy, ale během měření taktéž rotují krystal kolem jedné osy.

4.3. Základy rotační metody

Při použití Ewaldovy konstrukce (obr. 4), která je geometrickou interpretací Braggova zákona, je monochromatické záření representováno povrchem koule o poloměru $1/\lambda$ a krystal reciprokou mřížkou. Reciproká mřížka sestává z bodů ležících na konci vektorů kolmých k difrakčním rovinám a jejich délka je úměrná převrácené hodnotě mezirovinné vzdálenosti $1/d$. Při rotační metodě je krystal otáčen o daný úhel definovaný jako úhel φ.

Pro krystal v jedné orientaci bude pouze malá část reflexí splňovat difrakční podmínku. Počet reflexí bude velice malý pro malé molekuly (v některých orientacích i nulový). Krystaly makromolekul mají však větší elementární buňky (a, b, c v řádu 100 Å), pokud srovnáváme s vlnovou délkou použitého záření, což v geometrické interpretaci znamená, že reciproká mřížka je velice hustě zaplněna vzhledem k velikosti Ewaldovy sféry. Z toho vyplývá, že mnoho reflexí difraktuje současně do různých směrů, jelikož mnoho bodů reciproké mřížky

leží současně na povrchu Ewaldovy sféry. To také vysvětluje výhodu velkých plošných detektorů pro měření krystalů s velkou elementární buňkou.


Obr. 5 Rovina reflexí, která je přibližně kolmá k záření dá vzniknout elipse reflexí na detektoru.

Body reciproké mříže leží v rovinách, takže protínají Ewaldovu sféru v kružnicích. Odpovídající difraktované reflexe z centra Ewaldovy sféry tak tvoří kužely a jelikož při většině experimentů je detektor umístěn kolmo na směr primárního svazku záření, tvoří reflexe rovin elipsy (viz obr. 5). Pokud některá osa krystalu leží téměř paralelně s primárním zářením, elipsy vytvářejí soustředné kruhy s centrem ve středu detektoru. Všechny reflexe ležící na jedné kružnici (elipse) mají jeden index společný a pro necentrované buňky postupně vzrůstá o jednotku. Vzdálenosti mezi kruhy závisejí na převrácené hodnotě parametru rovnoběžného s primárním zářením.
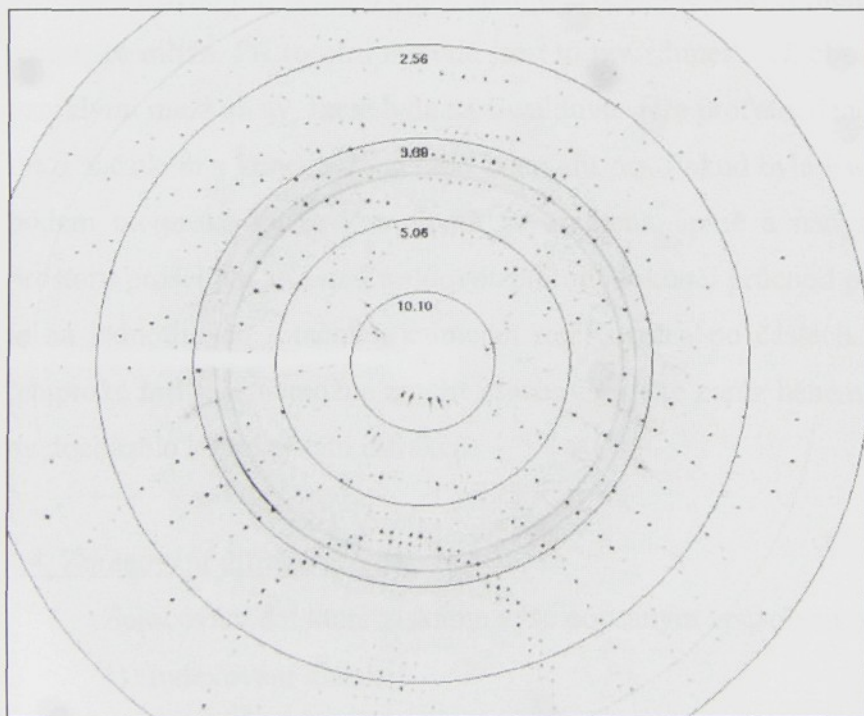
Obr. 6 Reálný difrakční obrazec tetragonální krystalové formy lysozymu. Pořízeno autorem na laboratorním difraktometru (zdroj záření s rotační měděnou anodou, FR591 fy Nonius Delft; záření upraveno zrcadlovou optikou XOS na totální odraz fy Nonius Delft; registrováno plošným detektorem MAR345 fy MarResearch).

Při Ewaldově konstrukci předpokládáme ideální zdroj záření s totálně paralelním zářením a ideální krystal se všemi elementárními buňkami stejně orientovanými, takže

výsledkem by byly bodové, nekonečně ostré Braggovské difrakce. Reálné experimentální podmínky se od tohoto ideálu liší ve třech aspektech. Za prvé, dopadající záření není nikdy dokonale paralelní, za druhé, reálné záření není striktně monochromatické a za třetí, reálné krystaly jsou tvořeny malými mosaikovými bloky, které nejsou vzájemně přesně stejně orientované. Pokud měříme krystaly makromolekul při laboratorní teplotě je rozpětí vzájemné orientace mosaikových bloků (mosaicita) většinou do 0.05°, ale pokud krystaly měříme zmražené může mosaicita vzrůst až na 1° i více. Všechny tyto efekty (rozšiřují difrakce na větší rozsah rotace krystalu.

Pokud máme změřit další reflexe, které v dané orientaci neleží na povrchu Ewaldovy sféry, při použití monochromatického záření, musíme krystalem pohhybovat tak, aby pro další reflexe byly splněny difrakční podmínky. Pokud je krystal otáčen okolo jedné osy během po sobě následujících expozicích, jedná se o rotační metodu. V praxi je rotační osa je skoro vždy osa rotace kolmá k dopadajícímu záření, tak aby byla zachována symetrie mezi oběma polovinami kompletního snímku. Toto je nejčastěji používaná metoda při měření difrakčních dat krystalů makromolekul. Pokud je krystal rotován během expozice, elipsy popsané pro fixní pozici krystalu mění svojí pozici. Výsledkem je, že snímek vzniklý během expozice obsahuje všechny difrakce mezi dvěma limitními pozicemi elips na začátku a konci daná rotace, které tvoří tzv. „půlměsíce" (lunes), viz obrázek 6.

U každého difrakčního experimentu získáváme více či méně deformovaný průmět části reciproké mříže. Při rotační metodě jsou to tzv. „lunes" jež obsahují difrakce mezi elipsami vzniklými mezi místy, která byla na Ewaldově sféře proťata danou rovinou reciproké mřížky mezi začátkem a koncem rotačního intervalu $\Delta\varphi$. Pokud byla Ewaldova sféra proťata daným bodem reciproké mříže kompletně, je změřena úplně a naopak, pokud bod reciprokého prostoru prošel jen z části Ewaldovou sférou, dokončí průchod při další rotaci a daná reflexe je na jednotlivých rotačních snímcích registrována po částech. Vzhledem k velké hustotě reciproké mříže je nemožné změřit všechny reflexe naráz během jedné velké rotace, jelikož by docházelo k překrývání difrakcí.

4.4. Zpracování difrakčních dat

Zpracování dat které získáme výše popsaným způsobem má několik fází:

1) indexování difrakcí
2) integrace intenzit difrakcí
3) škálování intenzit difrakcí

Při indexování se snažíme přiřadit indexy jednotlivých bodů reciproké mříže difrakčním maximům na všech postupných snímcích exponovaných v rámci měření krystalu v dané orientaci. Celkový postup je následující:

a) Nejprve vyhledáme difrakční maxima na jednom nebo více snímcích.

b) Určíme typ krystalové mříže a mřížkové parametry měřeného krystalu

c) a orientaci krystalu vzhledem k laboratornímu systému souřadnic.

d) Při postupném zpracování série difrakčních snímků zpřesňujeme, jak výše uvedené parametry popisující krystal, tak také další přístrojové parametry.

V průběhu integrace kvantifikujeme intenzitu difrakce jednoznačně popsané indexy h, k, l na daném snímku.   Metody kvantifikace intenzit se liší podle použitého software, nejpoužívanější metodou je profilová analýza, kdy program, na základě analýzy kvalitních difrakcí s nízkým poměrem signálu k šumu, muže zpětně lépe odhadnout tvar difrakce která je silně ovlivněna úrovní šumu. Program MOSFLM (Leslie, A.G.; 1999) například rozdělí plochu detektoru na několik částí, odděleně provede profilovou analýzu a ideální profily používá k fitování jednotlivých difrakcí i s odečtem pozadí. Velice obdobně postupuje program DENZO (Z. Otwinowski and W. Minor; 1997), který získává ideální profil v kruhové oblasti definované zadaným rádiem. Nejsofistokovaněji postupuje program XDS (Kabsch, W. 2001), který provádí profilovou analýzu ve třech rozměrech, tedy nejen plošně vzhledem k posizi x, y na detektoru, ale i na postupných snímcích, tedy i v rotačním úhlu φ.

Dalším úkolem na cestě k získání souborů intenzit a jejich σ je škálování. V průběhu škálovaní se snažíme korigovat případné fluktuace intenzity záření, vliv tvaru krystalu amnoho dalších vlivů, které mohli různě ovlivnit měření jednotlivých snímků, dále spojit intenzity difrakcí změřených na několika snímcích a zprůměrovat intenzity difrakcí které byly změřeny několikrát, například díky symetrii krystalu. Postupy se opět mírně různí v závislosti na použitém software a použitých korekčních parametrech, ale základem bývá minimalizace funkce:

$$\Psi = \sum_h \sum_l w_{hl} \left(I_{hl} - g_{hl} \langle I_h \rangle \right)^2 \qquad (17)$$

kde I_{hl} je l-té pozorování reflexe h, g_{hl} je asociovaný inverzní škálový parametr, $w_{hl} = 1/\sigma^2(I_{hl})$ a $\langle I_h \rangle$ je vážený průměr intenzity všech l měření reflexe h (Hamilton et al., 1965; Fox & Holmes, 1966). Intenzitu reflexe a odhad chyby pak získáme podle vztahů:

$$\langle I_h \rangle = \frac{\sum_l w_{hl} g_{hl} I_{hl}}{\sum_l w_{hl} g_{hl}^2} \quad , \tag{18}$$

$$\sigma(\langle I_h \rangle) = \frac{1}{\sum_l w_{hl} g_{hl}^2} \quad . \tag{19}$$

Kvalitu seškálovaných dat pak můžeme hodnotit pomocí různých statistických ukazatelů. Běžně užívaný parametr R_{merge} (někdy též R_{sym}) není zrovná nejlepším měřítkem kvality dat, jelikož pouze měří rozdíly mezi měřeními a nebere v úvahu zlepšení kvality průměrováním mnoha měření, R_{merge} faktor má samozřejmě tendenci vzrůstat s rostoucím počtem průměrovaných měření. Vylepšený, násobností měření vážené R faktory byly navrženy Diederichsem a Karplusem (1997), Weissem a Hilgenfeldem (1997) a Weissem (2001). Pokud n_h je počet měření jednotlivých reflexí **h**, pak

$$R_{merge} = R_{sym} = \sum_h \sum_l |I_{hl} - \langle I_h \rangle| \Bigg/ \sum_h \sum_l \langle I_h \rangle \tag{20}$$

je tradiční R_{merge},

$$R_{meas} = R_{r.i.m.} = \sum_h \left(\frac{n_h}{n_h - 1}\right) \sum_h \sum_l |I_{hl} - \langle I_h \rangle| \Bigg/ \sum_h \sum_l \langle I_h \rangle \tag{21}$$

je R faktor nezávislý na násobností měření (multiplicity-independent R factor) a

$$R_{p.i.m.} = \sum_h \left(\frac{1}{n_h - 1}\right) \sum_h \sum_l |I_{hl} - \langle I_h \rangle| \Bigg/ \sum_h \sum_l \langle I_h \rangle \tag{22}$$

je přesnost indikující R faktor (R faktor).

Tyto faktory a především statistika I/σ_I v závislosti na rozlišení nám pomáhají rozhodnout jaký škálovací protokol, nebo jinak řečeno které korekce použít. A také nám pomáhají při rozhodování do jakého rozlišení jsou naměřená data použitelná. Většina krystalografů používá jako kriterium použitelnosti dat hodnotu $I/\sigma_I \geq 2$. Výsledkem snažení je tedy soubor Intenzit difrakcí a jejich σ_I. Tyto pak lze přepočíst na strukturní faktory a jejich σ_{SF}

5. Fázový problém

Elektronovou hustotu v obecném bodě elementární buňky lze vypočítat dle výše uvedeného vztahu (14). Ve výrazu se vyskytuje strukturní faktor $F_{(h,k,l)}$, jakožto komplexní číslo, jehož absolutní hodnota vyjadřuje amplitudu rozptýlené vlny a fáze vyjadřuje fázový posun vzhledem k vlně rozptýlené ve směru shodném s dopadajícím zářením. Vzhledem k tomu, že všechny difrakční experimenty průměrují v čase, nelze fázovou informací získat přímo z difrakčních dat. Z nich získáváme toliko absolutní hodnotu strukturního faktoru |F(h,k,l)|, která je úměrná odmocnině z intenzity difrakce I(h,k,l). Proto se v následujícím textu budeme zabývat způsoby určení fází strukturních faktorů, tak abychom mohli vypočítat elektronovou hustotu podle vztahu:

$$\rho(x,y,z) = \sum_{h,k,l=-\infty}^{\infty} \left| F_{(h,k,l)} \right| e^{-2\pi i (hx+ky+lz-\alpha_{(h,k,l)})} \tag{23},$$

kde $\alpha_{(h,k,l)}$ je fáze strukturního faktoru $F_{(h,k,l)}$.

V principu existují tři základní způsoby jak získat informace o fázích. V první skupině jsou metody založené na porovnávání Pattersonovy funkce pro naše naměřená data a pro podobný, homologní protein s již vyřešenou strukturou.

Druhou skupinu tvoří metody, které se někdy označují jako experimentální. Je to metoda využívající anomálního rozptylu a metoda izomorfní záměny.

Do třetí skupiny patří tzv. přímé metody stanovení fází založené, na statistických metodách. Tyto metody mají omezené použití v proteinové krystalografii vzhledem k velkému objemu dat a jejich nízkému rozlišení. Jsou však s úspěchem používány pro řešení substruktur, což je nezbytný postupný krok druhé skupiny.

Vzhledem k zaměření práce se omezíme na krátký nástin fungování první skupiny metod, jelikož tyto byly použity v další části práce. Pro tyto metody se vžil název „molekulové nahrazení" (Molecular replacement) a to proto, že pomocí známého modelu struktury podobné molekuly hledáme vhodnou polohu a vhodné umístění pro námi studovanou molekulu v elementární buňce krystalu, který jsme měřili.

Jak již bylo zmíněno, metoda molekulového nahrazení je založena na využití Pattersonovy funkce $P(u,v,w) = P(\vec{u})$. Ta je definována následujícím vztahem (viz. Patterson, 1934):

$$P(\vec{u}) = \frac{1}{V} \sum_{\vec{H}=-\infty}^{\infty} \left| F(\vec{H}) \right|^2 e^{2\pi i \vec{H}.\vec{u}} \cong \frac{1}{V} \sum_{\vec{H}=-\infty}^{\infty} \left| F(\vec{H}) \right|^2 \cos(2\pi \vec{H}.\vec{u})$$

$$\tag{24}$$

Je tedy zjevné, že Pattersonovu funkci lze získat bez znalosti fází, jelikož se jedná pouze o Furierovu sumaci s koeficienty $|F_{(h,k,l)}|^2$ namísto $F_{(h,k,l)} = |F_{(h,k,l)}|exp[2\pi i a_{(h,k,l)}]$. Periodicita a elementární buňka je stejná jak pro elektronovou hustotu tak pro Patersonovu funkci.

Pattersonova mapa tedy může být získána přímo po tom co byly intenzity reflexí změřeny a korigovány. Zbývá jen zodpovědět jakou informaci nám vlastně Pattersonova funkce přináší. Toto lze velice snadno nahlédnout z alternativního zápisu pro Pattersonovu funkci:

$$P(\vec{u}) = \int_V \rho(\vec{r}).\rho(\vec{r} + \vec{u})dV_r \qquad (25)$$

Funkce má tedy maxima v polohách vektoru $\vec{u}$, který se rovná některému meziatomovému vektoru. Pattersonova funkce je tak funkcí všech meziatomových vektorů. Jak již bylo uvedeno je to prostorová funkce definovaná na stejné elementární buňce jako měřený krystal a hodnoty maxim jsou úměrné součinu elektronových hustot v polohách $\vec{r}$ a $\vec{r} + \vec{u}$.

Metoda molekulového nahrazení je založena na porovnávání Pattersonovy funkce vypočtené z naměřených dat podle rovnice (24) a pro model molekuly podobného (homologního) proteinu ve zvolené orientaci a pozici spočtené podle identické rovnice, pro kterou jsme si připravili koeficienty $|F(\vec{H})|^2$ výpočtem podle rovnice (13 respektive 16). Prakticky tedy hledáme maximum funkce:

$$RT(\alpha, \beta, \gamma, r, s, t) = \int P_o(\vec{u}) \times P_c(\vec{u})d\vec{u} \qquad (26)$$

kde α, β, γ jsou tři rotační parametry; r, s, t translační parametry $P_o(\vec{u})$ je Pattersonova funkce vypočtená pro naměřená data a $P_c(\vec{u})$ je Pattersonova funkce vypočtená pro vyhledávací model v orientaci a pozici dané parametry $\alpha, \beta, \gamma, r, s, t$. V počátcích této metody, na konci šedesátých let minulého století, kdy náročnost numerického řešení narážela na hranice výpočetního výkonu soudobých počítačů byla úloha rozdělena na dvě postupné části. Zaprvé řešení rotační funkce:

$$Rot(\alpha, \beta, \gamma) = \int P_o(\vec{u}) \times P_c(\vec{u})d\vec{u} \qquad (27)$$

a následně, metodou pokusu a omylu, byla testována různá umístění jednotlivých řešení rotační funkce.

Pro charakterizaci shody strukturního modelu s naměřenými hodnotami intenzit reflexí se používá buď korelační koeficient:

$$C = \frac{\sum_{\vec{H}} \left(|F_{\vec{H}}^o|^2 - \overline{|F_{\vec{H}}^o|^2} \right) \times \left(|F_{\vec{H}}^c|^2 - \overline{|F_{\vec{H}}^c|^2} \right)}{\sqrt{\sum_{\vec{H}} \left(|F_{\vec{H}}^o|^2 - \overline{|F_{\vec{H}}^o|^2} \right)^2 \sum_{\vec{H}} \left(|F_{\vec{H}}^c|^2 - \overline{|F_{\vec{H}}^c|^2} \right)^2}} \qquad (28)$$

nebo častěji požívaný parametr shody modelu a difrakčních dat tzv. *R*-faktor (*R* z anglického reliability), definovaný vztahem:

$$R = \frac{\sum_{\vec{H}} || F_{\vec{H}}^o | - K | F_{\vec{H}}^c ||}{\sum_{\vec{H}} | F_{\vec{H}}^o |} \qquad \text{R-faktor}, \qquad (29)$$

kde $| F_{\vec{H}}^o |$ je naměřená velikost strukturního faktoru, $| F_{\vec{H}}^c |$ je hodnota vypočtená, příslušející danému 3-rozměrnému modelu struktury a K je škálovací faktor.

$$K = \frac{\sum_{\vec{H}} | F_{\vec{H}}^o |}{\sum_{\vec{H}} | F_{\vec{H}}^c |} \qquad (30)$$

Sčítá se přes všechny naměřené reflexe.

Konkrétní programy pro řešení problému fází metodou molekulového nahrazení v klasickém dvoukrokovém uspořádání jsou například **MolRep** (Vagin and Teplyakov, 1997) a **AmoRe** (Navaza 1994), součásti setu krystalografických programů **CCP4** (Collaborative Computational Project, number 4.) Vzhledem k nárůstu vypočetního výkonu počítačů v nedávné době a stále většímu tlaku na robustní výpočetní metody pro stále větší počet řešených struktur biologicky aktivních makromolekul jsou v současnosti ve stále větší míře využívány metody molekulového nahrazení, které šesti-dimenzionální prostor umístění molekul v elementární buňce prohledávají jednokrokově. Výhodou těchto metod je větší citlivost pro nalezení správného řešení, jelikož se netříští signál na příspěvek rotační a translační. Příkladem jsou programy **EPMR** (Kissinger 1999), který využívá evoluční mechanismus k redukci výpočetního času a **BRUTE** (Fujinaga & Read 1987), který opravdu systematicky prohledává N-rozměrný prostor se zvoleným krokem.

6. Budování modelu

V úvodu této kapitoly je na místě zopakovat základní princip monokrystalové rentgenová strukturní analýzy. Rtg záření difraktuje (je rozptylováno) na elektronech atomů molekul vytvářejících krystal. Elektrony se vyskytují v blízkosti atomových jader a také v oblastech kovalentních vazeb. Proto je difrakční experiment nejlépe vyjádřen rovnicí (12), tedy slovem vyjádřeno: k difrakci přispívají ta místa v elementární buňce, kde je elektronová hustota nenulová. Jakmile se nám podaří vyřešit problém fází, můžeme Fourierovou transformací podle rovnice (23) tuto elektronovou hustotu zrekonstruovat a například zobrazit jako síťově konturovanou plochu o dané úrovni elektronové hustoty, viz obrázek 7a a 7b. Toto je jedno z možných znázornění elektronové struktury, tak jak se jeví rtg záření, které na krystalu difraktuje. Avšak z hlediska chemické struktury, nebo lépe řečeno, prostorového
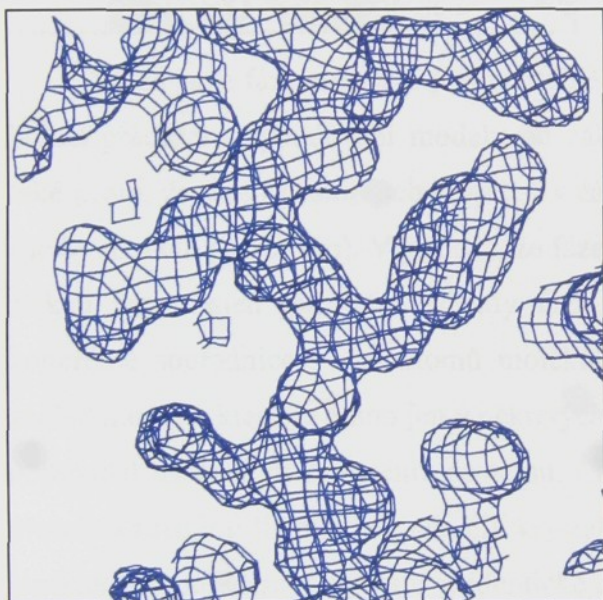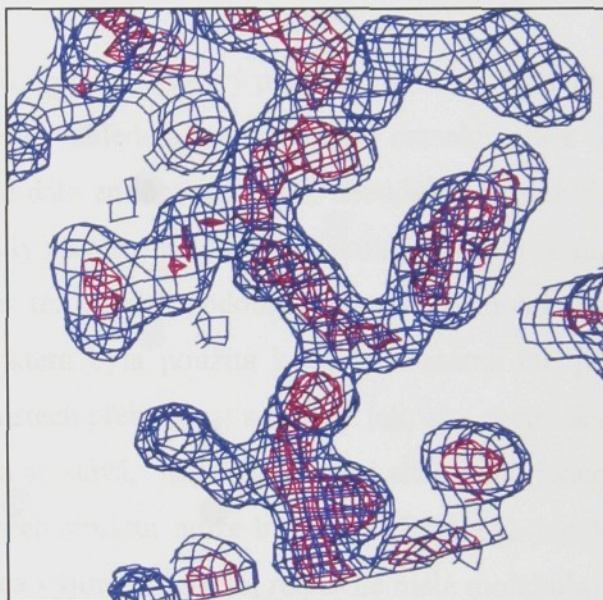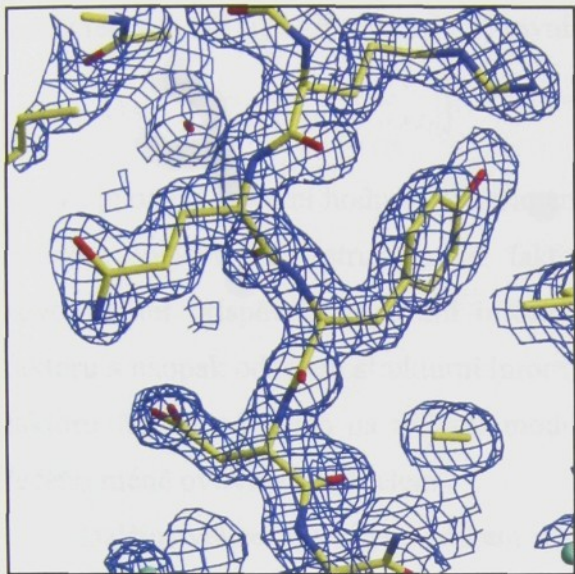

Obrázek 7a.   Obrázek 7b.

Výřez z mapy elektronové hustoty komplexu HIV-1 proteasy s inhibitorem. Na obrázku 5a. mapa elektronové hustoty konturována na úrovni 1σ, zde cca 0.8 elektronu na Å^3. Na obrázku 5b. mapa elektronové hustoty konturována na úrovni 1σ (modrá izoplocha cca 0.8 elektronu na Å^3) a 3σ (fialová izoplocha cca 2.4 elektronu na Å^3). Koeficienty pro výpočet map byly vypočteny programem Refmac5 (Murshudov, G. N. *et al* 1997), mapy byly vykresleny pomocí programu XtalView (McRee, D. E. 1999).

uspořádaní molekuly neposkytuje informaci téměř žádnou. Proto je jednou z hlavních úloh proteinové krystalografie výše znázorněné mapy interpretovat pomocí modelu struktury odpovídajícího proteinu. Přesněji řečeno, je třeba vysvětlit mapu elektronové hustoty na základě chemických znalostí, stereochemie, primární struktury (aminokyselinové sekvence) a struktury vázaných ligandů. V laboratorním slangu: "nafitovat strukturu do mapy elektronových hustot". Část modelu molekuly proteinu umístěná v mapě elektronové hustoty

je znázorněna na obrázku 8, jde o identickou mapu jako u obrázku 7a.

Obrázek 8. Výřez modelu HIV-1 proteasy s inhibitorem v mapě elektronové hustoty. Model molekuly HIV-1 proteasy je znázorněn jako tyčový model, bodové atomy jsou pospojovány ve směru kovalentních vazeb tyčkami, které nesou zbarvení podle druhu atomu ze kterého vycházejí (žlutá uhlík, modrá dusík a červená kyslík). Mapa elektronové hustoty konturována na úrovni 1σ, zde cca 0.8 elektronu na Å^3.

Pokud jsme fáze získali experimentálně (viz. kapitola Fázový problém) je nutno opravdu hodně práce pro vybudování modelu od základu (vzhledem k omezenému rozsahu práce a také proto, že tento postup nebyl využit v žádné dále zmiňované práci, nebude tento způsob stavby modelu popisován). V případě, že fáze byly získány metodou molekulového nahrazení, získali jsme nejen počáteční odhady fází, ale též model podobné proteinové molekuly, konkrétně souřadnice všech atomů molekuly, která byla použita k řešení a máme tak již základ modelu, který je nutno jen v některých místech přebudovat a upravit tak, aby chemicky odpovídal našemu studovanému proteinu. Často se stává, například pokud studujeme jeden protein s různými ligandy, že několik krystalových struktur může být izomorfních (t.j. jejich strukturní uspořádání v krystalu je identické až na výjimku ligandu, relativně malá molekula), v takovém případě je možné přímo použít model izomorfní struktury a jen mírně zrelaxovat jeho polohu v rámci elementární buňky (rigid body refinement, bude zmíněn v další kapitole).

Přes svou zdánlivou výhodu má tento postup svá negativa. Hlavním problémem je předpojatost fází a tedy i map elektronových hustot a výchozího modelu. Hlavní důvod předpojatosti je dominance fází při výpočtu mapy elektronových hustot (lze velice názorně dokumentovat na rozdílu reálné části strukturního faktoru při chybě amplitudy 25% oproti chybě fáze 25%). Platí tedy, že při použití fází získaných metodou molekulového nahrazení musíme vždy počítat s předpojatostí a ve vyhledávacím modelu pokud možno požívat jen tu část molekuly, která je pro obě molekuly shodná. Na tomto místě je též vhodné zmínit různé typy map, které se při budování a přestavbě modelu používá. Až dosud jsme hovořili o mapě elektronové hustoty počítané podle vztahu (23), takzvanou F_o mapu, kdy se pro výpočet

19

používá absolutní hodnota experimentálně získaného strukturního faktoru (F_o- observed) a jakkoli získaná fáze. Ve chvíli kdy je k dispozici model a fáze lze požívat kombinované koeficienty, například 2Fo-Fc, kdy se rovnice (23) „upraví" následujícím způsobem:

$$\rho(x,y,z) = \sum_{h,k,l=-\infty}^{\infty} \left[2\left| F_{(h,k,l)}^o \right| - \left| F_{(h,k,l)}^c \right| \right] e^{-2\pi i (hx+ky+lz-\alpha_{(h,k,l)})} \tag{31}$$

kde F_o je opět absolutní hodnota experimentálně získaného strukturního faktoru a F_c absolutní hodnota vypočteného strukturního faktoru. Smyslem této mapy je, nepřesně řečeno, zdvojnásobit příspěvek strukturní informace obsažený v amplitudě měřeného strukturního faktoru a naopak odečtení strukturní informace dodané modelem prostřednictvím strukturního faktoru F_c vypočteného na základě modelu, takto vzniklá mapa je méně předpojatá, jinak řečeno méně ovlivněná modelem.

Dalším velice důležitým druhem mapy je diferenční Fourierova mapa, nebo též F_oF_c mapa. Ta je počítána podle velice podobného vztahu jako výše zmíněná mapa, jen v hranaté závorce máme výraz $\left| F_{(h,k,l)}^o \right| - \left| F_{(h,k,l)}^c \right|$. Diferenční Fourierova mapa má zásadní význam při výstavbě a přebudovávání modelu. Její kladná maxima ukazují místa, kde v modelu ještě něco chybí a naopak místa se zápornou hodnotou ukazují oblasti, kde v modelu něco přebývá.

7. Upřesňování modelu.

Proces upřesňování modelu makromolekulární struktury může být definován jako postupné dolaďování parametrů modelu, konkrétně souboru souřadnic x_i, y_i, z_i, teplotního faktoru B_i a někdy též okupačního faktoru o_i každého i-tého atomu modelu tak, aby bylo dosaženo maximální možné shody s experimentálními daty. Význam kartézských souřadnic atomů x_i, y_i, z_i je zjevný. Okupační faktor popisuje s jakou četností se na pozici x_i, y_i, z_i atom vyskytuje a u stabilních struktur bez dynamických změn v polohách atomů by měl být roven 1. Význam okupačního faktoru lze dobře vysvětlit na příkladu atomu, který je kovalentně vázán, je však na konci řetězce, neinteraguje významně s okolím, má volnou torzi a tak může překmitávat mezi několika energetickými minimy a každá z těchto poloh je pak obsazena atomem s okupačním faktorem menším něž jedna, v součtu by pak okupační faktory jednoho atomu měly být rovny jedné.

Krystalografie biologických makromolekul není fundamentálně odlišná od krystalografie nízkomolekulárních sloučenin, ale rozhodně je složitější kvůli velikosti molekulárního systému. Pokud je struktura tvořena několika atomy, upřesnění lze provést iterativním výpočtem a do modelu není většinou nutné manuálně zasahovat (přímo přidávat nebo odebírat konkrétní atomy, případně měnit jejich polohu editací souřadnic). Typická makromolekula je však tvořena několika tisíci atomy a krystalizuje v elementárních buňkách o objemu milionů kubických ångströmů. Komplexnost této úlohy vyžaduje kombinování iterativních upřesňovacích (rafinačních) metod využívajících matematických principů a manuálních zásahů řešitele při budování, respektive přebudovávání, modelu ve 3D prostoru. Tento cyklus je znázorněn na obrázku 7. Výchozí pozicí v procesu upřesňování je model, tak jak byl vybudován do iniciální mapy elektronové hustoty, jak již bylo zmíněno v předešlé kapitole.

Nástrojem pro sledování úspěšnosti tohoto procesu je již dříve uvedený R-faktor (viz. rovnice 24). Pro upřesňování strukturního modelu se používá též vážený R_w-faktor (Hamilton, 1964):

$$R_w = \sqrt{\frac{\sum_{h,k,l} w_{(h,k,l)}\left(\left|F^o_{(h,k,l)}\right| - \left|KF^c_{(h,k,l)}\right|\right)^2}{\sum_{h,k,l} w_{(h,k,l)}\left|F^o_{(h,k,l)}\right|^2}} \qquad R_w\text{-vážený } R\text{-faktor (32)}$$

kde $w_{(h,k,l)}$ je váha příslušné reflexe (většinou je úměrná $\sigma^{-2}_{(h,k,l)}$). Dalším nástrojem využívaným při upřesňování je další faktor správnosti R_{free} (Brünger 1992). Tento faktor se

opět počítá naprosto identicky jako R-faktor jen s tím rozdílem, že je vypočten na souboru strukturních faktorů (SF), které se
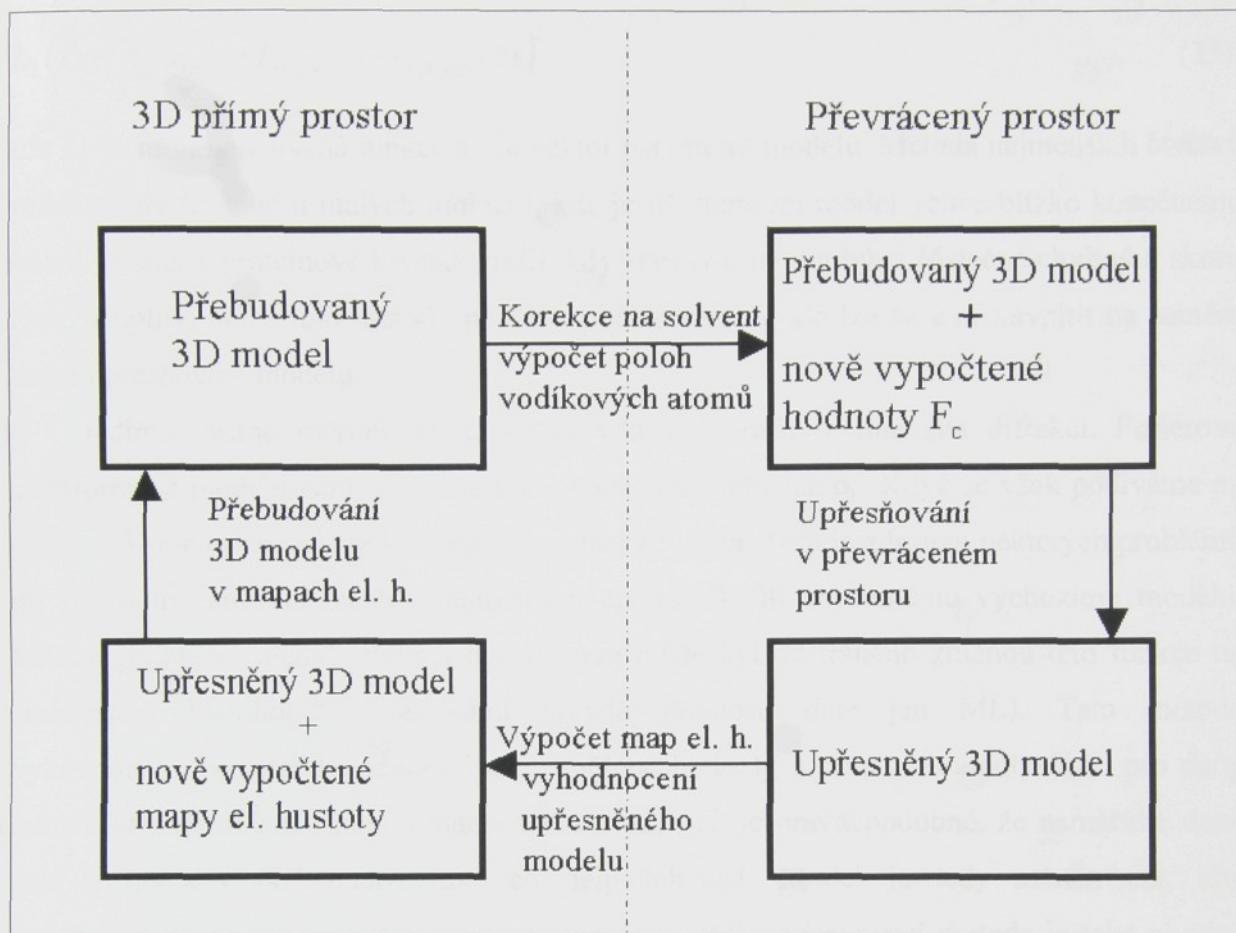
Obrázek 7: Blokové schéma procesu upřesňování modelu makromolekuly. Popisy v rámečcích označují stavy modelu v průběhu procesu a šipky označují jednotlivé kroky rafinačního cyklu.

zvolí náhodným výběrem a tyto SF se nepožívají k upřesňování. Smysl obětování časti dat je v tom, že můžeme kontrolovat průběh celého procesu srovnáním F_o s F_c, které nebyly nuceny konvergovat pomocí matematického aparátu. Pokud průběh faktoru R a R_{free} nekoreluje, je zřejmé, že jsme se dopustili nějaké zásadní chyby. Vzhledem k tomu, že je tento faktor vypočten na omezeném výběru dat (doporučuje se cca 5% ze souboru strukturních faktorů, pokud možno alespoň 1000 SF) je nutné počítat s větším kolísáním faktoru R_{free} a je proto lépe sledovat korelaci R a R_{free}, než konkrétní hodnoty.

7.1. Upřesňovací funkce

Pokud tedy chceme zlepšit shodu naměřených a vypočtených SF změnou parametrů modelu je potřeba mít nejprve nástroj míry shody. Klasicky u malých molekul a i u velkých

molekulárních systémů lze použít součet rozdílů, nejčastěji sumu druhých mocnim rozdílů, tedy :

$$L_2(\vec{x}) = \sum_{h,k,l} w_{(h,k,l)} \left(\left| F_{o(h,k,l)} \right| - \left| F_{c(h,k,l)}(\vec{x}) \right| \right)^2 \tag{33}$$

kde L_2 je minimalizovaná funkce a $\vec{x}$ je vektor parametrů modelu. Metoda nejmenších čtverců velice dobře funguje u malých molekul, kde je již startovní model velice blízko konečnému řešení. Avšak v proteinové krystalografii, kdy startovní model bývá částečně chybný a skoro vždy neúplný, nelze tuto metodu použít v počáteční fázi, ale lze se k ní navrátit na samém konci upřesňování modelu.

Jedinou přímo měřitelnou hodnotou v krystalografii je intenzita difrakcí. Furierova transformace modelu nám poskytne komplexní strukturní faktor. Když se však podíváme na rovnici (33) je jasné, že zde je informace o fázi ztracena. Toto je zdrojem některých problémů při požití upřesňovací metody nejmenších čtverců, kvůli nepřesnému výchozímu modelu. Mnoho problémů metody nejmenších čtverců může být odstraněno změnou této funkce na „maximum likelihood" (maximální pravděpodobnost, dále jen ML). Tato metoda vyhodnocuje, jak pravděpodobné jsou naměřené hodnoty a použité hodnoty fází pro daný model, přesněji řečeno: pokud máme tento model jak je pravděpodobné, že naměříme daná data a hodnoty fází budou také co nejpodobnější. Model je tedy měněn tak, aby pravděpodobnost daných měření byla co největší. ML upřesňovací metoda je také vhodná proto, že není tolik predeterminována případnými nepřesnostmi a chybějícími částmi modelu. Metoda je také velice vhodná pro kombinování informací o fázích z několika zdrojů, například pokud byl fázový problém vyřešen molekulovým nahrazením a navíc máme odhad fází, třeba i s nižším rozlišením z experimentu, lze tyto fázové informace zahrnout ve výpočtu pravděpodobnosti.

Takže pokud pravděpodobnost $P_a(\mathrm{F^o_i;F^c_i})$ je podmíněné pravděpodobnostní rozdělení strukturního faktoru $\mathrm{F^o_i}$ dané strukturním faktorem modelu $\mathrm{F^c_i}$, pak pravděpodobnost modelu je :

$$L = \prod_i P_a(F_i^o; F_i^c). \tag{34}$$

Což je většinou transformováno do lépe uchopitelné formy:

$$L = \sum_i \log P_a(F_i^o; F_i^c). \tag{35}$$

Jak již bylo uvedeno výše, pro proces upřesňování je třeba si vybrat jednak optimalizační funkci (nejmenší čtverce, ML nebo jinou) a numerickou metodu optimalizace. Dalším krokem v upřesňování je volba parametrů, kterými budeme model popisovat.

Závažným problémem v použití optimalizačních metod je poměr upřesňovaných parametrů ku počtu experimentálně získaných strukturních faktorů. V tabulce 1 je dobře vidět, že již při rozlišení o něco horším než 2.5Å se můžem s modelem se čtyřmi parametry pro atom dostat do situace, kdy máme více stupňů volnosti než dat. Problém nedostatku dat se v proteinové krystalografii řeší většinou přidáním strukturních dat a případně snížením počtu proměnných.

Rozlišení	Počet měření	Poměr měření a parameterů (x,y,z)	Poměr měření a parameterů (x,y,z, B)
3.0Å	3500	0.8	-
2.5Å	6800	1.6	1.2
1.9Å	13500	3.1	2.3
1.5Å	29800	6.8	5.1
1.2Å	58800	13.3	10.0
1.0Å	81300	18.5	13.8

Tabulka 1. Příklad závislosti počtu reflexí na rozlišení .

Jednou možností jsou optimalizační metody s pevnými vazebnými podmínkami (constrained) nebo lze uměle zvýšit počet pozorování. Pro optimalizační metodu s pevnými vazebnými podmínkami se též používá název „rigid-body refinement" (upřesňování pomocí tuhého tělesa). Klasickým příkladem je postup na začátku upřesňování, kdy například složitější strukturu s několika doménami rozdělíme na tyto podjednotky a pro tyto upřesňujeme tři translační a tři rotační parametry a tak zredukujeme počet proměnných na 6 x počet podjednotek. Tato metoda je možnou volbou v mnoha upřesňovacích programech jako například X-PLOR, CNS a Refmac. Výše uvedená možnost přidání dalších „chemických dat" do systému (nebo též měkké vazebné podmínky tzv. „restrained" optimalizační metody), využívá znalostí vazebných vzdáleností a dalších parametrů ze struktur malých fragmentů proteinů (malé peptidy, samostatné aminokysleiny a pod.). Například správné vazebné vzdálenosti pro identické typy atomů ve vazbě můžeme přidat jako pozorování (podobně jako změřené strukturní faktory) s průměrnou hodnotou délky vazby a její variancí. Další kritéria správné molekulové geometrie mohou být přidána do modelu podobným způsobem.

Například zakonzervování správné vazebné délky je dosaženo přidáním členu

$$\sum_{\substack{vaz. \\ atomy}} \left(\frac{1}{\sigma_{ij}^2} \right) \left(d_{ij} - d_{ij}^0 \right)^2 \tag{36}$$

kde d_{ij} je vazebná vzdálenost atomů i a j, d_{ij}^0 je ideální vazebná vzdálenost a $\frac{1}{\sigma_{ij}^2}$ je váha aplikovaná na vazbu (souvisí s variancí délky vazby ze známých struktur na vysokém rozlišení).

7.2. Optimalizační metody

Optimalizační metody pro nízkomolekulární sloučeniny jsou velice přímočaré, avšak struktury makromolekul jsou specielním problémem, kvůli jejich velikosti. Velký počet parametrů modelu extrémně zvyšuje prostor parametrů, který musí být prohledáván, aby bylo nalezeno vyhovující řešení a také vysoce zvyšuje nároky na objemy paměti nutné pro optimalizační procesy. Tato kombinace velkého počtu parametrů a také velkého počtu měření znamenají, že každý cyklus optimalizačního procesu je výpočetně náročný.

Optimalizační metody mohou být tříděny podle řádu derivace použité v algoritmu. Metody, které nepoužívají derivace a prohledávají vlastně náhodně prostor parametrů k nalezení řešení, jsou například: Monte Carlo metody a některé formy simulovaného žíhání. Metody prvního řádu, jenž počítají gradienty a tak známe směr změny parametrů, abychom se přiblížili žádané hodnotě zvolené optimalizační funkce. Metody druhého řádu počítají gradienty a zakřivení, tak známe nejen směr ale i vhodné velikosti změny parametrů tak jak postupujeme v průběhu upřesňování. Metody nultého řádu jsou všeobecně velice pomalé, jelikož multidimensionální prostor parametrů, který je prohledáván, je extrémně velký. Metody prvního řádu mohou být rychlé a kompaktní, ale nemohou rozlišit, zda nalezené řešení je lokální minimum nebo hledané správné řešení. Metody druhého řádu mohou odstranit problém uvíznutí v lokálním minimu v průběhu upřesňování, avšak výpočetní náročnost roste se třetí mocninou počtu parametrů a nároky na operační paměť jsou také mnohem vyšší. Až dosud jsou nejúspěšněji používány upřesňovací metody prvního řádu. Toto se zvolna mění s rostoucí dimenzací operačních pamětí a rychlostí procesorů (viz. například program SHELXH, součást balíku programu SHELX97).

7.3. Řešení upřesňovacích (rafinačních) rovnic.

Velice dobrým úvodem do problematiky řešení rafinačních rovnic je publikace: Mathematical techniques in crystallography and materials science (Prince, 1994). Nejběžnějším způsobem jak řešit výpočet parametrů, které minimalizují rovnici 33, je postup, kdy každý člen sumace z rovnice 33 přepíšeme jako samostatnou rovnici:

$$w_i[y_i - f_i(\vec{x})] = w_i \sum_{j=1}^{N} \left(\frac{\partial f_i(\vec{x})}{\partial x_j}\right)_{x^0} (x_j^0 - x_j) \tag{37}$$

kdy sumace probíhá v indexu j všech N parametrů modelu, y_i je naměřený strukturní faktor, $f_i(\mathbf{x})$ vypočtený *SF* a w_i je váha daného *SF*. Toto je jednoduchý rozvoj prvního řádu funkce $f_i(\mathbf{x})$ a vyjadřuje hypotézu, že vypočtená hodnota má být shodná s naměřenou. Systém rovnic může být řešen pro posun parametrů $(x_j^0 - x_j)$ za předpokladu, že máme nejméně tolik rovnic pro shodu *SF* jako parametrů (lze započíst i geometrické podmínky, viz. například vztah 36 pro délku vazby, který jsme přidali jako jedenu rovnici v soustavě rovnic 37).

V maticovém zápisu lze systém rovnic přepsat následujícím způsobem:

$$A\Delta = \vec{r} \tag{38}$$

kde **A** je M x N matice derivací, Δ je vektor změny parametrů a $\vec{r}$ vektor reziduí, viz. levá strana rovnice 37. Soustavu normálních rovnic lze vytvořit vynásobením rovnice 31 maticí $\mathbf{A}^T$. Výsledkem je systém se čtvercovou maticí N x N a výsledky v případě metody nejmenších čtverců, viz. níže:

$$A^T A\Delta = A^T \vec{r} \quad \text{nebo} \quad M\Delta = \vec{b}, \tag{39}$$

$$m_{ij} = \sum_{k=1}^{M} w_k \left(\frac{\partial f_k(\vec{x})}{\partial x_i}\right)\left(\frac{\partial f_k(\vec{x})}{\partial x_j}\right),$$

$$b_i = \sum_{k=1}^{M} w_k [y_k - f_k(\vec{x})]\left(\frac{\partial f_k(\vec{x})}{\partial x_i}\right).$$

Obdobně lze rozšířit rozvoj rovnice 33 o další člen Taylorova rozvoje v okolí $\vec{x}^0$:

$$\Phi(\vec{x} - \vec{x}^0) \approx \Phi(\vec{x}^0) + \left\langle \left(\frac{\partial \Phi}{\partial x_i}\right)_{x^0} (\vec{x} - \vec{x}^0)\right\rangle + \frac{1}{2}\left\langle (\vec{x} - \vec{x}^0)\left(\frac{\partial^2 \Phi}{\partial x_i \partial x_j}\right)_{x^0}(\vec{x} - \vec{x}^0)\right\rangle, \tag{40}$$

$$\left\langle \left(\frac{\partial \Phi}{\partial \vec{x}}\right)\right\rangle \approx \left\langle \left(\frac{\partial^2 \Phi}{\partial x_i \partial x_j}\right)_{x^0}(\vec{x} - \vec{x}^0)\right\rangle.$$

Tato aproximace druhého řádu předpokládá, že matice druhých derivací se nemění a lze jí počítat v $\vec{x}$ namísto v $\vec{x}^0$. Je namístě poznamenat, že výše uvedené řešení soustavy rovnic se používá běžně pro rafinaci struktur malých molekul a výjimečně pro makromolekulární struktury, pokud je k dispozici dostatek experimentálních dat (struktury na vysokém rozlišení při použití programu SHELXH, Sheldrick, G. M 1997).

7.4. Optimalizační programy

Jak již bylo výše uvedeno, nejčastěji jsou využívány metody prvního řádu. Nejběžnějším přístupem je řešit upřesňování struktury jako nelineární optimalizační úlohu od samého začátku. Tuto strategii používají například programy TNT (Tronrund D.E., 1992) a program X-PLOR (Brunger, Kurijan & Karplus,1987), odlišným způsobem.

TNT používá metodu podmíněného konjugovaného gradientu, kde podmiňovací funkce je parciální druhá derivace upřesňovací funkce podle daného parametru. Jinými slovy, krok v daném parametru je normalizován podle zakřivení upřesňovací funkce podle daného parametru a provede se krok v daném parametru podle konjugovaného gradientu. To způsobí, že krok parametru, který má strmé derivace je zmenšen a u měkých parametrů (jako jsou například teplotní B faktory) je krok zvětšen. To jednak zrychluje výpočet a také zvětšuje rádius konvergence metody.

Program X-PLOR (nyní CNS) používá simulované žíhání, které pomocí metody molekulové dynamiky (v pokročilých verzích CNS torzní dynamiky). Simulované žíhání je metoda, která mapuje chování upřesňovací funkce v konformačně dosažitelném prostoru parametrů (kartézských souřadnic jednotlivých atomů) v blízkosti původních pozic atomů. Tak může nalézt vhodné umístění části makromolekuly a překonat past lokálního minima funkce. K tomu navíc vyžívá Boltzmanovské pravděpodobnosti exp(-$\Delta E/kT$) pro různé polohy a tak rozhoduje o vhodnosti výběru nových parametrů. V nové poloze program sníží teplotní parametr, tak aby funkce jen neoscilovala, ale pokud možno konvergovala do globálního maxima. Program X-PLOR využívá difrakční data tak, že je vlastně uvažuje jako další člen v sumaci energií (optimalizační metody molekulové mechaniky/dynamiky sumují jednotlivé energetické příspěvky dané konformací molekuly a vzájemnými interakcemi atomů v rámci molekuly tak aby popsaly její pravděpodobné chování) a gradient celkové energie je použit jakožto gradient upřesňovací funkce, což vlastně řadí tento postup k metodám prvého řádu.

První široce používanou metodou v makromolekulárním upřesňování byl rafinační program PROLSQ (Konnert, 1976), který používal k řešení aproximaci druhého řádu, kde matice derivací byla velice zjednodušena. Parametry každého atomu byly řešeny jako

diagonální blok matice a mimodiagonální prvky byly použity jen tam, kde jsou atomy vázány kovalentní vazbou nebo jinak zjevně interagují. Tento částečný systém lineárních rovnic je pak řešen adaptací metody konjugovaného gardientu.

Nejkomplexnější rafinační program pro řešení makromolekulárních struktur je ten samý, jako nejpoužívanější program pro řešení pro struktur malých molekul – SHELX97 (Sheldrick, G. M 1997). Primární adaptací pro řešení makromolekulárních struktur bylo přidání konjugovaného gradientu, jakožto optimalizační metody pro případy, kdy plná matice metody nejmenších čtverců přesahuje možnosti paměti pro počet parametrů makromolekulárního modelu. S výhodou je tento program používán pro struktury na vysokém rozlišení ve finálních stadiích upřesňování. V závěrečných cyklech upřesňování lze použít metody nejmenších čtverců a tak získat velice dobré odhady σ jednotlivých parametrů. Pokud nám nestačí dimenzace polí na náš počet parametrů, lze rozdělit parametry do bloků, jejichž složení obměňujeme pro jednotlivé cykly.

Reference:

Blundell T. L. & Johnson, L. N. (1976) "Protein Crystallography" Academic Press Inc. London.

Bricogne, G. and Irwin, J. "Maximum-likelihood structure refinement: theory and implementation within BUSTER + TNT" in Macromolecular Refinement: Proceedings of the CCP4 Study Weekend January 1996, Dodson, E., Moore, M., Ralph, A. and Bailey, S. (eds.), CCLRC Daresbury Laboratory, pp. 85-92 (1996).

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

Brünger AT, Kuriyan J, Karplus M. 1987. Crystallographic R factor refinement by molecular dynamics. *Science* **235**:458–460.

Brünger AT, Leahy DJ, Hynes TR, Fox RO. 1991. 2.9 Å resolution structure of an anti-dinitrophenyl-spin label monoclonal antibody Fab fragment with bound hapten. *J Mol Biol* **221**:239–256.

Brünger, 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., Read, R. J., Rice, L. M., Simonson, T., andWarren, G. L. (1998). *Acta Cryst. D* **54**, 905-921.

Clegg, W. Blake, J. Gould R.O. and Main P.(2002) Crystal Structure Analysis, Oxford University Press.

COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.* **D50**, 760-763.

Diederichs, K. & Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography.*Nature Struct. Biol.* **4**, 269-275.

Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray proteinstructure refinement. *Acta Crystallogr* **A47**:392– 400.

Evans, P. R. (1993). Proceedings of the CCP4 Study Weekend. Data Collection and Processing, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114±122. Warrington: Daresbury Laboratory.

Fox, G. C. & Holmes, K. C. (1966). An alternative method of solving the layer scaling equations of Hamilton, Rollett and Sparks. *Acta Cryst.* **20**, 886-891.

Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517±521

Hamilton, W. C., Rollett, J. S. & Sparks, R. A. (1965). On the relative scaling of X-ray photographs. *Acta Cryst.* **18**, 129-130.

Hamilton, W. C. (1964). Statistics in Physical Science, Estimation Hypothesis Testing and Least Squares. New York: The Ronald Press.

Heitler W. (1966), Quantum Theory of Radiation, 3rd ed. Oxford University Press.

Hubbard, S.J. and Thornton, J.M. (1993). *'NACCESS', Computer Program.* Department of Biochemistry and Molecular Biology, University College London, UK.

Kabsch, W. (2001) Chapter **25.2.9.** *XDS* in *International Tables for Crystallography*, Volume F. Crystallography of Biological Macromolecules, Rossmann, M.G. and Arnold, E. (2001). Editors. Dordrecht: Kluwer Academic Publishers.

Kissinger, C. R.. Gehlhaar, D. K. & Fogel, D. B. (1999)"Rapid automated molecular replacement by evolutionary search", *Acta Crystallographica*, **D55**, 484-491.

Klein, O. Nishima, Y. (1929) Uber der Streuung von Strahlung durch freie Elektron nach der neuen relativistischen Quanten-dynamik von Dirac. Z. Physik **52** 853-868

Laskowski R A, MacArthur M W, Moss D S & Thornton J M (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283-291.

Leslie, A.G. Integration of macromolecular diffraction data. *Acta Cryst.* 1999, D55:1696-1702.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802±810.

Giacovazzo, C. Monaco, H.L. Viterbo, D. Scordari, F. Gilli, G. Zanotti, G. and Catti M.; Fundamentals of Crystallography; 2nd Edition; Oxford University Press.

McRee, D. E. (1999) *J Struct.Biol.* **125,** 156-165.

McRee, D. E. (1993) Practical Protein Crystallography. San Diego: Academic Press.

Marek J. & Trávníček Z. (2002) Monokrystalová rentgenová strukturní analýza, 1. vydání, Olomouc.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. "Refinement of macromolecular structures by the maximum-likelihood method", *Acta Cryst.* **D53**: 240-255 (1997).

Navaza, G. (1994) AMORE: an automated package for molecular replacement. *Acta Crystallogr.* **A50**, 157-163.

Otwinowski Z, MinorW. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**:307–326.

Pannu, N. S. and Read, R. J. "Improved structure refinement through maximum likelihood", *Acta Cryst.* **A52**: 659-668 (1996).

Patterson A. L. (1934) Fourier series method for the determination of the components of interatomic distances in crystals. *Phys. Rev.* **46**, 372-376

Prince E. Mathematical techniques in crystallography and materials science. Second edition. New York and Berlin: Springer-Verlag,1994.

Read, R. J. (1997). "Model phases: Probabilities and bias" *Methods Enzymol.* **277**, 110±128.

Sim, G. A. (1959). The distribution of phases for structures containing heavy atoms. *Acta Cryst.* **12**, 813±815.

Srinivasan, R. (1966). Weighting funtions for use in the early stages of structure analysis when part of the structure is known. *Acta Cryst.* **20**, 143±145.

Sheldrick, G. M.; Schneider, T. R. SHELXL: high-resolution refinement. (1997). *Methods Enzymol.* **277**, 319-343.

Sheldrick, G. M.; SHELX97 http://shelx.uni-ac.gwdg.de/SHELX/index.html

Trakhanov, S., Kreimer, D.I., Parkin, S., Ames, G.F., and Rupp, B. (1998). Cadmium-induced crystallization of proteins: II. Crystallization of the Salmonella typhimurium histidine-binding protein in complex with L-histidine, L-arginine, or L-lysine. *Protein Sci.* **7**, 600-604.

Tronrund D. E. Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Cryst.* (1992). A**48**, 912-916

Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022-1025.

Valvoda, V. Polcarová, M. Lukáč, P. (1992) Základy strukturní analýzy, Karolinum, Praha.

Vondrasek, J., van Buskirk, C.P., and Wlodawer, A. (1997). Database of three-dimensional structures of HIV proteinases.*Nat. Struct. Biol.* **4**, 8.

Weiss, M. S. & Hilgenfeld, R. (1997). On the use of the merging *R* factor as a quality indicator for X-ray data. *J. Appl. Cryst.* **30**, 203-205.

Weiss, M. S. (2001). Global indicators of X-ray data quality. *J. Appl. Cryst.* **34**, 130-135.

Wierenga, R. K., Kalk, K. H. & Hol, W. G. J. (1987). STRUCTURE DETERMINATION OF THE GLYCOSOMAL TRIOSEPHOSPHATE ISOMERASE FROM TRYPANOSOMA BRUCEI BRUCEI AT 2.4 ANGSTROMS RESOLUTION. *J.MOL.BIOL.***198,** 109

Wlodawer, A., and Vondrasek, J. (1998). Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Anu. Rev. Biophys. Biomol. Struct.* **27**, 249-284.

Wlodawer, A., Li, M., Gustchina, A., Dauter, Z., Uchida, K., Oyama, H., Goldfarb, N.E., Dunn, B.M., and Oda, K. (2001). Inhibitor complexes of the Pseudomonas serine-carboxyl proteinase. *Biochemistry* **40**, 15602-15611.

Dodatek A: Soubor publikací k habilitaci

Soubor obsahuje tyto následující publikace:

A1 <u>Brynda, J.</u> Řezáčová, P. Fábry, M. Hořejší, M. Stouračová, R. Sedléček, J. Souček, M. Hradílek, M. Lepšík, M. Konvalinka, J.: **A phenylnorstatine inhibitor binding to HIV1 protease: Geometry, protonation, and subsite-pocket interactions analyzed at atomic resolution**. (2004) *JOURNAL OF MEDICINAL CHEMISTRY* **47 (8)**: 2030-2036

A2 <u>Brynda J,</u> Rezacova P, Fabry M, Horejsi M, Stouracova R, Soucek M, Hradilek M, Konvalinka J, Sedlacek J.: **Inhibitor binding at the protein interface in crystals of a HIV-1 protease complex** (2004) *ACTA CRYSTALLOGRAPHICA SECTION D-BIOLOGICAL CRYSTALLOGRAPHY* **60**: 1943-1948

A3 Žáková, L. <u>Brynda, J.</u> Au-Alvarez, O. - Dodson, G. D. - Whittingham, J. L. Brzozowski, A. M. (2004) **Toward the insulin-IGF-I intermediate structures: Functional and structural properties of the [TyrB25NMePheB26] insulin mutant**. *Biochemistry*. **43**: 16293-16300. ISSN 0006-2960

A4 Ondráček, J. Weiss, M.S. <u>Brynda, J.</u> Fiala, J. Jursík, F. - Řezáčová, P. Jenner, L.B. Sedláček, J. (2005) **Regular arrangement of periodates bound to lysozyme**. *Acta Crystallographica Section D-Biological Crystallography*. **61, (9)**: 1181-1189. ISSN 0907-4449

A5 Cígler, P. Kožíšek, M. - Řezáčová, P. - <u>Brynda, J.</u> Otwinowski, Z. Pokorná, J. Plešek, J. Grüner, B. Dolečková-Marešová, L. Máša, M. Sedláček, J. - Bodem, J. - Kräusslich, H. G. - Král, V. - Konvalinka, J.: **From nonpeptide toward noncarbon protease inhibitors: Metallacarboranes as specific and potent inhibitors of HIV protease**. (2005), *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. **102, (43)**: 15394-15399. ISSN 1091-6490

A6 Kmoch, S. - <u>Brynda, J.</u> Bekefadu, A. - Bezouška, K. - Novák, P. Řezáčová, P. Ondrová, L. Filipec, M. Sedláček, J. Elleder, M. (2000), **Link between a novel human çD-crystallin allele and a unique cataract phenotype explained by protein crystallography**. *Human Molecular Genetics*. **9, (12)**: 1779-1786. ISSN 0964-6906

A7 Řezáčová, P. Lescar, J. - <u>Brynda, J.</u> - Fábry, M. Hořejší, M. Sedláček, J. Bentley, G. (2001), Structural basis of HIV-1 and HIV-2 protease inhibition by a monoclonal antibody. *Structure*. **9 (10)**: 887-895. ISSN 0969-2126

A8 Lescar, J. - <u>Brynda, J.</u> - Fábry, M. Hořejší, Řezáčová, P. M. Sedláček, J. Bentley, G. Structure of a single-chain Fv fragment of an antibody that inhibits the HIV-1 and HIV-2 proteases Acta Cryst. (2003). D59, 955±957

A9 Lescar, J., Stouracova R., Marie-Madeleine Riottot, M-M., Chitarra, V., <u>Brynda, J.,</u> Fabry, M., Magda Horejsi, M., Sedlacek, J., and Bentley, G.A., Three-dimensional Structure of an Fab-Peptide Complex: Structural Basis of HIV-1 Protease Inhibition by a Monoclonal Antibody. *J. Mol. Biol.* (1997) **267**, 1207±1222

A10 **Brynda, J.** (2003) Zpracování difrakčních dat krystalů biologických makromolekul získaných oscilační metodou. *Materials Structure in Chemistry, Biology, Physics and Technology*. Praha, ČR: Czech and Slovak Crystallographic Association,. s. 40-41. ISSN 1211-5894.

A Phenylnorstatine Inhibitor Binding to HIV-1 Protease: Geometry, Protonation, and Subsite–Pocket Interactions Analyzed at Atomic Resolution

Jiri Brynda,*,† Pavlina Rezacova,† Milan Fabry,† Magdalena Horejsi,† Renata Stouracova,† Juraj Sedlacek,†
Milan Soucek,‡ Martin Hradilek,‡ Martin Lepsik,‡ and Jan Konvalinka‡

Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo nám. 2,
16637 Prague 6, Czech Republic, and Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the
Czech Republic, Flemingovo nám. 2, 16610 Prague 6, Czech Republic

The X-ray structure of a complex of HIV-1 protease (PR) with a phenylnorstatine inhibitor
Z-Pns-Phe-Glu-Glu-NH$_2$ has been determined at 1.03 Å, the highest resolution so far reported
for any HIV PR complex. The inhibitor shows subnanomolar K_i values for both the wild-type
PR and the variant representing one of the most common mutations linked to resistance
development. The structure comprising the phenylnorstatine moiety of (2R,3S)-chirality displays
a unique pattern of hydrogen bonding to the two catalytic aspartate residues. This high
resolution makes it possible to assess the donor and acceptor relations of this hydrogen bonding
and to indicate a proton shared by the two catalytic residues. A structural mechanism for the
unimpaired inhibition of the protease Val82Ala mutant is also suggested, based on energy
calculations and analyses.

Introduction

A structure-aided design of HIV protease (PR) inhibitors has led to a class of drugs useful in clinical anti-HIV intervention (see ref 1 for review). Nevertheless, mutational development of HIV PR drug resistance presents a major medical complication. Thus, the present state of anti-AIDS therapies calls for the design of novel compounds that would overcome the problem of HIV PR resistance to current drugs. Inhibitors with unconventional binding modes (or those interacting with HIV PR domains other than the active site, e.g., the dimerization domain[2]) might open a way to a new generation of anti-HIV PR compounds capable of inhibiting drug-resistant enzyme species (e.g., refs 3–6).

Recently, a combinatorial chemistry approach[7] yielded a series of novel pluripotent HIV PR inhibitors having a picomolar range for their K_i values for the wild-type HIV PR as well as various degrees of insensitivity of their inhibitory potency to HIV PR variants with mutations in positions 48, 82, 84, and 90, often found in drug-resistant PR strains.[8] Detailed kinetic analysis of three chosen inhibitors has even shown somewhat better inhibition of a drug-resistant HIV PR mutant compared with the wild-type enzyme. In the present paper, the structure of wild-type HIV PR complexed with one of these inhibitors, Z-Pns-Phe-Glu-Glu-NH$_2$ (Z, benzyloxycarbonyl; Pns, phenylnorstatine, (2R,3S)-3-amino-2-hydroxy-4-phenylbutanoic acid, Figure 1a, termed hereafter KI2), is described. The phenylnorstatine group, an untypical inhibitor moiety, served the purpose of investigating the potential of replacement of the peptide bond with larger groups (the "main chain" between the aromatic groups occupying S1' and S1

pockets is longer by one carbonyl group compared to common hydroxyethylene or hydroxyethylamine isosteres). This compound inhibits the Val82Ala mutant of HIV PR with a K_i value 0.11 nM, while the wild-type HIV PR inhibition has K_i = 0.18 nM.[8]

Since the combinatorial chemistry approach and consecutive primary screening against drug-resistant HIV PR mutants are rather empirical in their nature, only direct structural studies may elucidate subtleties of the inhibitory potential of such developed lead compounds. This paper presents the wild-type HIV PR/phenylnorstatine inhibitor complex structure solved at 1.03 Å, which is the best resolution for HIV PR reported to date. With an excess of inhibitor, the protein/inhibitor crystals grew with an extraordinary diffraction quality, apparently because of the presence of a second inhibitor molecule bound in the crystal lattice as an outer ligand at the protein interface (Brynda et al., manuscript in preparation). Though the compound used for complex formation and crystal growth was a 2:1 mixture of (21R,11S)- and (21S,11S)-diastereoisomers, both protease-bound inhibitor molecules had a (21R,11S)-configuration.

The present paper is focused on two basic structural questions. First, what type of binding makes the (2R,3S)-phenylnorstatine group, an untypical inhibitor moiety, functional? Second, what structural elements contribute to the unimpaired inhibition of the drug-resistant Val82Ala HIV PR mutant? The first question should raised because the unusual (2R,3S)-chirality makes the present inhibitor structurally dissimilar to the common (2S,3S)-phenylnorstatine compounds with their well-described mode of binding of transition-state isosteres to the catalytic center.[9] Regarding the resistance mechanisms in the Val82Ala HIV PR mutant, comparison of wild-type and mutant complex structures was feasible

---

* To whom correspondence should be addressed. Telephone: +420 20183212. Fax: +420 224310955. E-mail: brynda@img.cas.cz.
† Institute of Molecular Genetics.
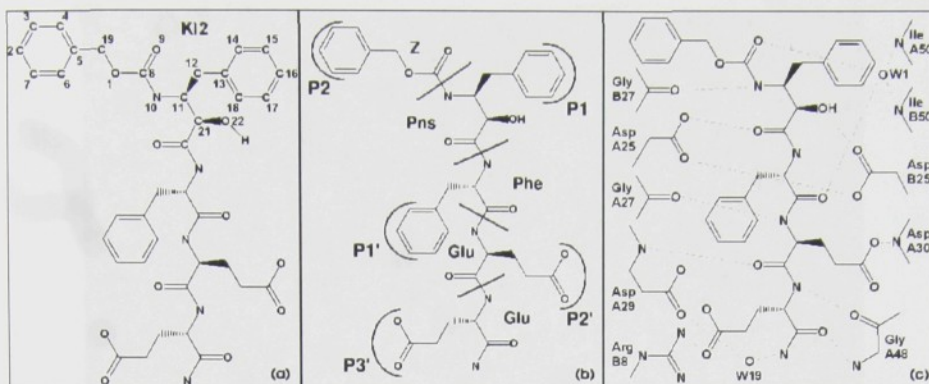‡ Institute of Organic Chemistry and Biochemistry.

Figure 1. Binding of the inhibitor in the active site. (a) Chemical structure of the inhibitor KI2, where numbering of atoms corresponds to the deposited PDB file (1NH0). (b) Positioning of the inhibitor side chains in the subsites of the active site. The segmentation (bold lines cutting the backbone) was used for energy calculations (see the text). (c) System of hydrogen bonds between inhibitor and the protease active site.

so far for two compounds of different types: Ritonavir itself[10] and a C_2 symmetry based diol inhibitor.[11]

While our solved structure directly shows, in considerable detail, the distinctive pattern of the hydrogen bonding of the present inhibitor to the enzyme catalytic aspartic acid residues, the question of resistance-overcoming mechanisms has been approached using molecular modeling methods and energy calculations.

Results

Inhibitor Binding Mode. The overall scheme of the inhibitor binding is shown in Figure 1b. The inhibitor bound to the active site displays the usual extended conformation, as directed by the occupation of the binding pockets. The S1 pocket consisting of residues Pro A81, Val A82, Ile A84, Gly B27, Gly B49, and Ile B50 contains the aromatic part of the phenylnorstatine group of the inhibitor (Pns group). The S1' pocket consisting of Gly A27, Gly A49, Arg B8, Pro B81, Val B82, and Ile B84 is occupied by the phenylalanine side chain of the inhibitor. The S2 pocket consisting of Asp B29, Asp B30, Val B32, Ile B47, and Gly B48 contains the aromatic part of the terminal benzyloxycarbonyl group (Z group), while the S2' pocket consisting of Ala A28, Asp A29, Asp A30, Ile A47, Gly A48, and Ile B50 is occupied by the glutamic acid side chain of the inhibitor. The poor quality of the omit map found for the Z group of the inhibitor in the S2 pocket is best explained by an alternative conformation of the benzyloxy substituent. The S3' pocket consisting of Ile A47, Gly A48, and Arg B8 contains the terminal glutamic acid group of the inhibitor.

The hydrogen-bonding interactions between the inhibitor and the protein are extensive. The schematic diagram showing the hydrogen bonds is depicted in Figure 1c. While most of the observed hydrogen bonds are common to many pseudopeptide HIV PR inhibitors, the central moiety of the present inhibitor KI2 maintains a unique type of hydrogen bonding. The common hydrogen bonds include interaction of the inhibitor main chain atoms, i.e., O9 oxygen of the Z group with both N Ile A50 and N Ile B50 via one water molecule W1; N10 of the Pns group with O Gly B27; the carbonyl oxygen O Phe I2 with both N Ile A50 and N Ile B50 via water molecule W1; the nitrogen N Glu I3 to O Gly A27; the carbonyl oxygen O Glu I3 to N Asp A29; the nitrogen N

Glu I4 to O Gly A48; the carbonyl oxygen O Glu I4 to N Gly A48; and the terminal amide nitrogen to OD2 Asp A29 via a water molecule W19. Also, the carboxylates of both glutamic acid residues of the inhibitor maintain hydrogen bonding (OE1 Glu I3 to OD2 Asp A30, and OE2 Glu I4 to NH2 Arg B8) in their respective binding pockets.

The phenylnorstatine moiety of the inhibitor has two hydrogen-bonding oxygen atoms, i.e., O22 I1 of the central hydroxyl group and the oxygen atom O I1 of the adjacent carbonyl group. The oxygen atom of the carbonyl group (O I1) forms a hydrogen bond with Asp A25, and the oxygen atom of the hydroxyl group (O22 I1) forms a hydrogen bond with Asp B25 (Figure 2). Furthermore, the nitrogen atom of the phenylalanine residue next to the backbone carbonyl group of the inhibitor is remarkably close to the hydrogen-bonding oxygen of Asp B25; the distance of 3.38 Å suggests additional partial hydrogen bonding. The hydrogen atom between OD2 of Asp A25 and OI1 is not directly visible at the actual resolution. Nevertheless, the length of the CG−OD2 bond in the Asp A25 residue is 1.28 Å, significantly more than the parametrized length[12] 1.249-(19) Å indicating protonation of the oxygen atom forming the hydrogen bond with the inhibitor carbonyl group (cf. refs 13 and 14 for the assessment method). In contrast, the bond length CG−OD2 in Asp B25 residue is 1.24 Å, which indicates deprotonation; the hydrogen-bonding hydrogen atom comes from the inhibitor O22 hydroxyl group. The CG−OD1 in the Asp A25 and the Asp B25 residues are 1.24 and 1.25 Å, respectively, indicating that both OD1 oxygen atoms form hydrogen bonds with a shared hydrogen atom (Figure 2).

The observed intermolecular hydrogen bonding does not represent the conventional transition-state analogues where the hydroxyl group of a nonscissile junction is positioned between the two catalytic aspartate carboxyl groups within hydrogen-bonding distance.[1] Complex structures have been determined for two inhibitors that have an (S,S)-allophenylnorstatine central part, where the hydrogen-bonding hydroxyl group also points between the aspartate carboxyl groups but is positioned markedly out of their plane.[15] The present structure further differs from these. While each oxygen atom of the inhibitor central part maintains separate hydrogen bonding with only one of the carboxyl groups,
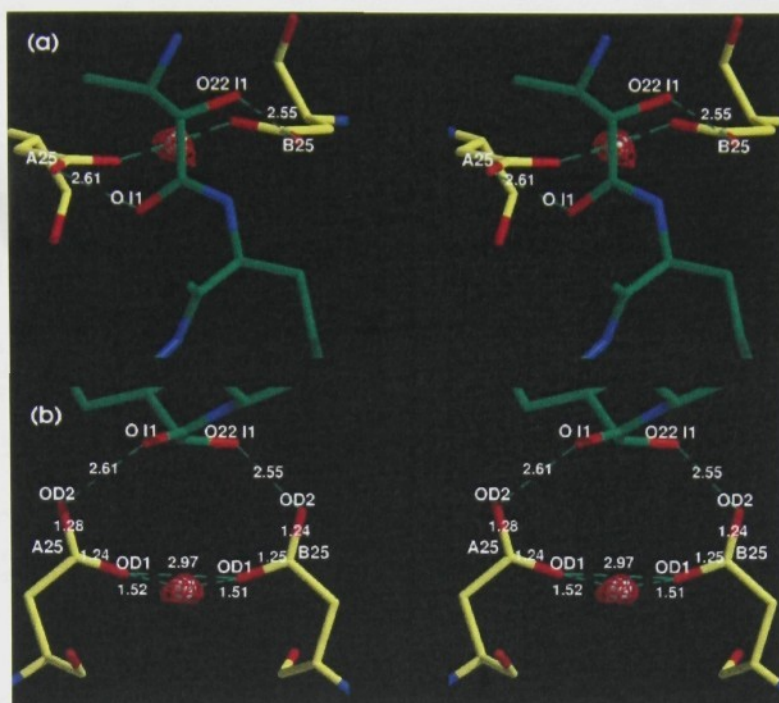
Figure 2. Schematic view of hydrogen bonding of the central part of the inhibitor and the catalytic aspartates. (a) Stereoview along the symmetry axis of the HIV PR dimer. (b) Stereoview perpendicular to the symmetry axis of the HIV PR dimer. Inhibitor carbon atoms are color-coded green, and the protein carbon atoms are color-coded yellow. The hydrogen atom shared by the Asp A25 and Asp B25 residues is shown (white) in the local maximum of the mFo − DFc electron density map (contoured at 2.6σ, red).

the C21−O22 bond has a unique orientation, roughly in the plane of the adjacent carbonyl C=O bond, the torsion angle O−C−C21−O22 being −156° (Figure 2).

To summarize, the present inhibitor KI2 bound in the HIV PR active site displays interactions of its main chain and side chains that are similar to most HIV PR complexes with peptidomimetic inhibitors except for the phenylnorstatine moiety of the KI2 inhibitor that displays unusual positioning relative to the catalytic aspartic acid residues with which it maintains a unique type of hydrogen bonding.

Structural Basis for Unimpaired Inhibition of Val82Ala Mutant. Background information shows that amino acid substitutions in position 82 of HIV PR are frequently associated with drug resistance.[16] Specifically for the Val82Ala mutation, the structural change from the wild-type side chain corresponds to the net loss of two methyl groups from each valine 82 residue, A and B, thus leaving more room in the S1′ and S1 binding pockets for ligand binding. Decreased van der Waals interactions in these pockets can then be considered as the cause of an 8-fold-reduced sensitivity to the drug Ritonavir.[17] Remarkably, KI2 shows an opposite trend: inhibition of the Val82Ala mutant HIV PR is somewhat better than that of the wild-type enzyme.[8] In the present complex structure, an unusual angle of the aromatic "side chain" in the phenylnorstatine moiety (C11−C12−C13 = 105.6°) appears suggestively as an element involved in the difference of the inhibitor binding to the wild-type and mutant HIV PR. A tendency to adopt a normal angle value (which is 113.8° in phenylalanine[12]) could then be expected to improve van der Waals interactions with the mutant enzyme in the larger S1 binding pocket.
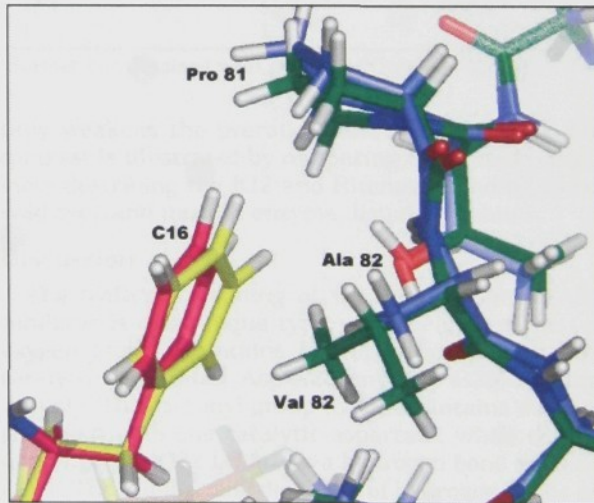

Figure 3. Details of interactions at the P1 subsite. The wild-type protein structure is green, and the modeled Val82Ala mutant is blue. The inhibitor is magenta for the wild-type structure and is yellow for the mutant model.

A more informative view of a structural basis for the unimpaired inhibition of the Val82Ala mutant HIV PR was obtained by a comparison of the present wild-type complex structure with its in silico modeled mutant-enzyme counterpart (see Experimental Procedures). From a structural point of view, the aromatic "side chain" of the phenylnorstatine moiety in the mutant complex adopts a new position in the S1 binding pocket, as depicted in Figure 3. Quantitative data were obtained by molecular-mechanics-based calculations of interaction energies of individual inhibitor segments defined
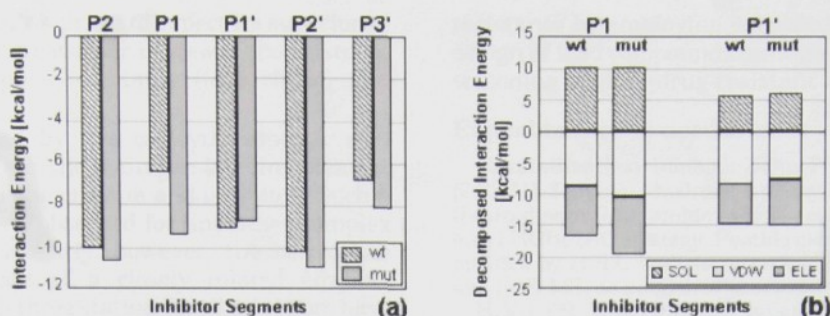
Figure 4. Comparisons of total and decomposed interaction energies calculated for individual subsites. The segmentation of the inhibitor structure is shown in Figure 1b: W, wild-type; MUT, Val82Ala mutant; SOL, solvation; ELE, electrostatic, including hydrogen bonding; VDW, van der Waals.

Table 1. Comparison of Parameters Describing the Inhibitor Binding

	1HXW wt + RIT	1N49 D25N/V82A + RIT	1NH0 wt + KI2	model V82A + KI2
	Entire Complex			
inhibitor buried surface area^a (Å²)	872.2	812.6	930.6	936.5
inhibitor buried surface area^a (%)	91.5	83.0	94.9	94.8
gap volume^b (Å3)	204.2	532.5	164.9	193.5
no. H bonds^c	9	6	13	14
K_i values^d (nM)	0.015 ± 0.003	0.12 ± 0.02 5	0.18 ± 0.02	0.11 ± 0.01
	S1 Subsite			
inhibitor buried surface area^a (Å²)	199.3	179.9	243.1	244.4
inhibitor buried surface area^a (%)	98.3	84.4	96.6	96.0
gap volume^b (Å³)	72.2	179.2	62.5	94.2
no. H bonds^c	4	3	4	5
	S1′ Subsite			
inhibitor buried surface area^a (Å²)	124.7	139.7	163.0	168.0
inhibitor buried surface area^a (%)	95.1	98.5	99.9	99.9
gap volume^b (Å³)	55.8	45.1	50.7	57.9
no. H bonds^c	1	1	1	1

^{a} Calculated with program NACCESS.[32] ^{b} Calculated with program Surfnet.[34] ^{c} Obtained with program CONTACT (CCP4).[35] ^{d} As determined in ref 8.

as shown in Figure 1b. The interaction energies of individual inhibitor segments and their respective binding pockets are plotted in the histogram in Figure 4a. While the interaction in S1 is markedly increased with the mutation, the interaction in S1′ is only slightly decreased. A decomposition of the interaction energies into van der Waals, electrostatic (including hydrogen bonding), and solvation energy terms (Figure 4b) shows that the gain in the S1 binding pocket comprises an increase of van der Waals plus electrostatic interactions while the loss in S1′ is contributed by solvation and van der Waals contacts.

The actual loss of the interaction energy in S1′ clearly would not suffice to compensate for the gain found in the S1 binding pocket. The summarized interactions in S1 plus S1′ binding pockets are fully consistent with a better fit of the KI2 structure with the mutant enzyme active site compared to the wild-type. Not surprisingly, the mutation appears almost neutral in regard to the interaction energies calculated for S2, S2′, and S3′ binding pockets (Figure 4a). Such a structural mechanism sharply contrasts with that involved in the Ritonavir resistance, as is apparent from the complex structures solved by others.[10,18] Here, the mutant S1′ shows good occupation by the P1′ aromatic group of Ritonavir but interaction in the S1 is substantially weakened. This loss also involves a shift of the entire Ritonavir molecule in the direction from S1 toward S1′. The shift interferes also with the proper binding of Ritonavir in other binding pockets and thus incremen-

tally weakens the overall Ritonavir interaction. The contrast is illustrated by comparing calculated parameters describing the KI2 and Ritonavir binding to the wild-type and mutant enzyme, listed in Table 1.

Discussion

The hydrogen bonding of the central part of this inhibitor is of a unique type: neither one of the two oxygen atoms maintains hydrogen bonds with both catalytic aspartates, Asp A25 and Asp B25, simultaneously. The carbonyl group (O I1) maintains a hydrogen bond with one catalytic aspartate, while the hydroxyl group (O22 I1) forms a hydrogen bond with the other. The number and distances of hydrogen bonds are comparable to other, more frequently used isosteres, rendering thus the phenylnorstatine group, an untypical inhibitor moiety, a full functionality.

The observed geometry of the inhibitor phenylnorstatine moiety, which is "out of plane" of the catalytic aspartic acid residues, is determined by the (R)-configuration on the C21 carbon atom and by the mostly normal occupation of the substrate binding pockets. Surprising in this context is the exclusive presence of the (21R,11S)-diastereoisomer in the active site (as clearly evidenced by the omit map) despite the fact that the (21S,11S)-diastereoisomer was also present in an amount sufficient for complex formation before and during the crystal growth. Thus, the empirical finding is that the apparently more tightly bound (R,S)-diastereoisomer is "extracted" into the enzyme active site,

while it seems difficult a priori to expect its superiority above the (*S*,*S*)-diastereoisomer in view of the existence of tightly bound allophenylnorstatine ((*S*,*S*)-chiral) compounds.[15]

The proton shared by two catalytic aspartic acid residues completes a cyclic hydrogen-bonding element in the interaction of the enzyme and inhibitor. Such a network has not been observed for any other complex of HIV PR. Quite recently, however, the active site protons in complexes of a closely related enzyme, endothiapepsin, with three statine-based inhibitors have been either identified directly in the electron densities or assessed from the carboxyl bond lengths.[14] Distinct protonation states of Asp A25 and Asp B25 in HIV-1 PR/drug complexes have been assessed previously on the basis of NMR evidence.[19] Also, the hydrogen-bonding network we describe here resembles to some extent the cyclic network assessed recently with the crystal structure of an oligosaccharide hydrate.[20] By use of knowledge of the hydrogen donor/acceptor relations (vide supra), the cyclic hydrogen-bonding element of HIV PR may be classified as a homodromic hexagon.

Though not originally "tailored" to inhibit the Ritonavir-resistant Val82Ala HIV PR mutant, the solved complex wild-type HIV PR structure and in silico calculations show that KI2 has specific characteristics to serve such a purpose. Owing to the relatively long "main chain" of the critical central part of KI2, the occupation of the wild-type S1 pocket requires a deformation, as though the wild-type pocket were too "tight" for the respective aromatic "side chain". All calculated parameters show that the strain is relieved with the Val82Ala mutation. Such findings can be valued as a special case of the contribution to the original ideas on the structural mechanisms of the HIV PR drug resistance.[17]

Conclusions

This paper presents the structure of the HIV-1 protease in complex with (2*R*,3*S*)-phenylnorstatine compound Z-Pns-Phe-Glu-Glu-NH₂ (KI2), determined at the highest resolution so far reported for any HIV protease complex (1.03 Å). The observed details of structural elements found excellently match the earlier findings on the inhibitor properties.[8] First, the mostly normal occupation of the HIV PR binding pockets by the inhibitor "side chains" and the apparently stable hydrogen bonding of the central phenylnorstatine moiety to the catalytic aspartate residues correspond to the subnanomolar inhibition constant. Second, the solved wild-type complex structure also gives clues to explain unimpaired inhibition of the Ritonavir resistance conferring mutant. While the solved structure reveals a strained conformation of the aromatic "side chain" of phenylnorstatine moiety in the wild-type S1 binding pocket, this deformation is not present in our model of the inhibitor in the "looser" Val82Ala-mutated S1 binding pocket. Naturally, such a structural mechanism is absent with Ritonavir and other inhibitors.[11,21] Because of a shorter main chain, Ritonavir cannot form a proper binding in the Val82Ala mutant S1 and S1′ binding pockets simultaneously.

Taken together, these results further validate the particular approach to overcome the HIV PR drug resistance by employing extended central parts in the design of lead compounds combined with their primary screening against drug-resistant HIV PR mutants.

Experimental Procedures

Crystallization. Inhibitor Z-Pns-Phe-Glu-Glu-NH₂, (Pns = (2*RS*,3*S*)-3-amino-2-hydroxy-4-phenylbutanoic acid) was synthesized on a Rink amide MBHA resin using the Fmoc/*tert*-butyl-HOBt/DIC strategy. Peptide cleaved from the resin was purified by HPLC and characterized by amino acid analysis and FAB MS, as published previously.[8]

HIV-1 PR (wild-type, Bru isolate) used for crystallization was prepared as described previously.[22] The complex was prepared by admixing the HIV-1 PR with a 6-fold molar excess of the inhibitor (from a 10 mM stock solution in dimethyl sulfoxide). For crystallization experiments, the complex was transferred to a buffer solution containing 10 mM sodium acetate, pH 5.6, 0.05% 2-mercaptoethanol, and 1 mM ethylenediaminetetraacetic acid (EDTA), and the mixture was concentrated by means of ultrafiltration using Centricon-10 (Millipore). Crystallization trials were performed by the hanging-drop technique, and the best crystals were obtained under the following conditions: 1 μL of the reservoir solution (50 mM MES, pH 6.5, 2.4 M ammonium sulfate) was mixed with 2 μL of the solution containing the HIV-1 PR/inhibitor complex at a concentration of 2.2 mg/mL and was allowed to equilibrate over 1 mL of the reservoir solution at 18 °C. After several days, crystals with the dimensions 0.2 mm × 0.2 mm × 0.3 mm appeared.

X-ray Data Collection. Crystals were soaked in the reservoir solution with 20% (v/v) glycerol and transferred into liquid nitrogen. Diffraction data were collected using synchrotron radiation, wavelength of 0.93 Å (ESRF Grenoble, ID14-2), and ADSC Q4 CCD based detector at 100 K (Oxford Cryosystem). Two crystals were measured: crystal 1 diffracted up to 1.03 Å, data were collected in two passes, a high-resolution pass comprising data from 10.0 to 1.03 Å, and a low-resolution pass comprising data from 55.0 to 1.9 Å. Data were measured with 1° oscillation, and the exposure time was 8 and 1 s, respectively. Data for the second crystal were measured with 0.2° oscillation, and the exposure time was 2 and 1 s, respectively. All data sets were integrated by XDS[23] and scaled by XSCALE.[24] The small slices technique gave better R_{meas} values. However, the best data set was obtained by scaling all data sets together. The data collection statistics are shown in Table 2.

Structure Determination, Refinement, and Analysis. The HIV-1 PR/inhibitor structure was solved by molecular replacement using the EPMR program.[25] HIV-1 PR from a complex structure (PDB code: 1VIK[26]) was used as a search model. Structure factors in the 15−4 Å resolution range were used for both the rotational and translational searches. The solution yielded a correlation coefficient of 0.553 and an *R* factor of 45.3%. After the initial rigid-body refinement, using the CNS program[27] at 50−1.6 Å resolution (monomers allowed to move independently), the *R* factor and R_{free} values fell to 42.7% and 43.6%, respectively. Further refinement of the model containing only HIV-1 PR was carried out with CNS using the simulated annealing protocol.[27] The R_{free} factor[28] was used to monitor the progress of the refinement by omitting 2.5% of the data (2164 reflections), randomly chosen. Building of the inhibitor was performed using XtalView.[29]

At this stage, a model of the inhibitor molecule was built in the enzyme active site. After repeated cycles of the refinement, the model of an inhibitor bound at the protein interface was built by fitting it into difference electron density maps. The completed model was then refined by conjugate gradient least-squares methods against structure factors using SHELXL-97.[30] Binning of the *R* value by resolution showed that the bulk solvent correction, as implemented in SHELXL-97, did not describe the low-resolution data well. Therefore, low-resolution cutoff at 6 Å was applied.

Twenty amino acids in multiple conformations for protein and "inhibitor side chains" were introduced for residues

Table 2. Crystallographic Statistics

parameter	
Data Collection Statistics	
space group	*P*212121
unit cell	
a (Å)	28.85
b (Å)	66.52
c (Å)	93.00
diffraction limits (Å)	54.1−1.03
no. of measured reflections	872 620
no. of unique reflections	88 784
average $I/\sigma(I)$ [a]	10.5 (1.85)
R_{sym} (%) [a]	7.8 (37.9)
completeness (%) [a]	99.0 (93.3)
Wilson *B* factor (Å²)	7.9
Refinement Statistics	
R factor (%)	13.0
R-free factor (%)	16.5
non-hydrogen atoms	1937
non-hydrogen atoms in inhibitors	106
heterogeneous atoms	15
water molecules	233
reflections used in refinement	86 020
reflections in test set	2205
rmsd from ideal bond distances (Å)	0.016
rmsd from ideal valence angle values (deg)	0.036

[a] Values in parentheses correspond to the last resolution shell (1.06−1.03 Å).

exhibiting alternative conformations in 1.5σ 2mFo − DFc and 3σ mFo − DFc electron density maps. During further refinement, anisotropic displacement parameters for protein and active site inhibitor atoms were introduced and the refinement using SHELXL-97 was continued. Next, anisotropic displacement parameters of atoms of outer ligand and solvent molecules were added to the model.

Hydrogen atoms of amino acids, placed according to stereochemistry, further reduced the R_{work} and R_{free} values. At the end of the refinement, a round of blocked full-matrix least-squares refinement was performed using all reflections (including those previously used for R_{free}) to obtain a proper estimate of standard deviations of all refined parameters.

The stereochemical quality of the model was controlled with PROCHECK.[31] Atomic coordinates have been deposited into the Protein Data Bank, PDB code 1NH0. The refinement statistics is shown in Table 2.

Buried surface areas were calculated using the program NACCESS,[32] an implementation of the Lee and Richards algorithm.[33] The default values (probe radius, 1.4 Å; z-slices, 0.05 Å; van der Waals radii) were used in the calculations. The volume of the gaps between the inhibitor and the protease was calculated using the program SURFNET[34] (maximal and minimal radii of 4.0 and 1.0 Å, respectively). Hydrogen bonds (≤3.2 Å) and van der Waals contacts (≤4.0 Å) were analyzed by the program CONTACT (CCP4).[35]

Molecular Modeling. The simulations and energy analyses were carried out in AMBER 7.0.[36] Structure manipulation was performed with INSIGHT II 2000.[37] Charges for the inhibitor atoms were obtained from quantum mechanical calculations using Gaussian 98, revision A6.[39]

Hydrogen atoms were added to the crystal structure of the wild-type HIV protease/inhibitor complex. The protonation of the active site aspartic acid residues was done according to the findings of the X-ray analysis (this work). The Val82Ala mutant was created by replacing the two valine side chains in the wild-type protease with alanine. The model included the crystallographically determined water molecules.

In the simulations, the biomolecular force field ff99[39] was used for the protease and general AMBER force field (gaff)[36] parameters were used for the inhibitor. Partial charges for the inhibitor atoms were determined by fitting electrostatic potential calculated at the HF/6-31G* level. In the wild-type complex, we optimized hydrogen atom positions only. In the mutant, we first relaxed the hydrogen atoms followed by

10 000 cycles of minimization in the neighborhood of the mutated residue (residues 23, 80−84 from both protease monomers) plus all atoms of the inhibitor.

Using the resulting structures, we calculated protease−inhibitor interaction energies with an MM-GBSA approach (molecular mechanics generalized Born surface area).[40] Energy terms were separated into van der Waals, electrostatic, and solvation terms. For the purposes of decomposition, the inhibitor structure was split into amino-acid-like segments (from P2 to P3′ in Figure 1b).

References

(1) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 Protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249−284.
(2) Lescar, J.; Brynda, J.; Rezacova, P.; Stouracova, R.; Riottot, M.-M.; Chitarra, V.; Fabry, M.; Horejsi, M.; Sedlacek, J.; Bentley, G. A. Inhibition of the HIV-1 and HIV-2 proteases by a monoclonal antibody. *Protein Sci.* **1999**, *8*, 2686−2696.
(3) Weber, J.; Mesters, J. R.; Lepsik, M.; Prejdova, J.; Svec, M.; Sponarova, J.; Mlcochova, P.; Skalicka, K.; Strisovsky, K.; Uhlikova, T.; Soucek, M.; Machala, L.; Stankova, M.; Vondrasek, J.; Klimkait, T.; Kraeusslich, H. G.; Hilgenfeld, R.; Konvalinka, J. Unusual binding mode of an HIV-1 protease inhibitor explains its potency against multi-drug-resistant virus strains. *J. Mol. Biol.* **2002**, *324*, 739−754.
(4) Dohnalek, J.; Hasek, J.; Duskova, J.; Petrokova, H.; Hradilek, M.; Soucek, M.; Konvalinka, J.; Brynda, J.; Sedlacek, J.; Fabry, M. Hydroxyethylamine isostere of an HIV-1 protease inhibitor prefers its amine to the hydroxy group in binding to catalytic aspartates. A synchrotron study of HIV-1 protease in complex with a peptidomimetic inhibitor. *J. Med. Chem.* **2002**, *45*, 1432−1438.
(5) Yoshimura, K.; Kato, R.; Kavlick, M. F.; Nguyen, A.; Maroun, V.; Maeda, K.; Hussain, K. A.; Ghosh, A. K.; Gulnik, S. V.; Erickson, J. W.; Mitsuya, H. A potent human immunodeficiency virus type 1 protease inhibitor, UIC-94003 (TMC-126), and selection of a novel (A28S) mutation in the protease active site. *J. Virol.* **2002**, *76*, 1349−1358.
(6) Yoshimura, K.; Kato, R.; Yusa, K.; Kavlick, M. F.; Maroun, V.; Nguyen, A.; Mimoto, T.; Ueno, T.; Shintani, M.; Falloon, J.; Masur, H.; Hayashi, H.; Erickson, J.; Mitsuya, H. JE-2147: A dipeptide protease inhibitor (PI) that potently inhibits multi-PI-resistant HIV-1. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8675−8680.
(7) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* **1991**, *354*, 84−86.
(8) Rinnova, M.; Hradilek, M.; Barinka, C.; Weber, J.; Soucek, M.; Vondrasek, J.; Klimkait, T.; Konvalinka, J. A picomolar inhibitor of resistant strains of human immunodeficiency virus protease identified by a combinatorial approach. *Arch. Biochem. Biophys.* **2000**, *382*, 22−30.
(9) Kiso, Y.; Matsumoto, H.; Mizumoto, S.; Kimura, T.; Fujiwara, Y.; Akaji, K. Small dipeptide-based HIV protease inhibitors containing the hydroxymethylcarbonyl isostere as an ideal transition-state mimic. *Biopolymers* **1999**, *51*, 59−68.
(10) Prabu-Jeyabalan, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy. *J. Virol.* **2003**, *77*, 1306−1315.
(11) Baldwin, E. T.; Bhat, T. N.; Liu, B.; Pattabiraman, N.; Erickson, J. W. Structural basis of drug resistance for the V82A mutant of HIV-1 proteinase. *Nat. Struct. Biol.* **1995**, *2*, 244−249.
(12) Engh, R. A.; Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **1991**, *47*, 392−400.
(13) Wlodawer, A.; Li, M.; Gustchina, A.; Dauter, Z.; Uchida, K.; Oyama, H.; Goldfarb, N. E.; Dunn, B. M.; Oda, K. Inhibitor complexes of the *Pseudomonas* serine-carboxyl proteinase. *Biochemistry* **2001**, *40*, 15602−15611.

(14) Coates, L.; Erskine, P. T.; Crump, M. P.; Wood, S. P.; Cooper, J. B. Five atomic resolution structures of endothiapepsin inhibitor complexes: implications for the aspartic proteinase mechanism. *J. Mol. Biol.* **2002**, *318*, 1405−1415.

(15) Reiling, K. K.; Endres, N. F.; Dauber, D. S.; Craik, C. S.; Stroud, R. M. Anisotropic Dynamics of the Je-2147-HIV protease complex: drug resistance and thermodynamic binding mode examined in a 1.09 Å structure. *Biochemistry* **2002**, *41*, 4582−4594.

(16) D'Aquila, R. T.; Schapiro, J. M.; Brun-Vezinet, F.; Clotet, B.; Conway, B.; Demeter, L. M.; Grant, R. M.; Johnson, V. A.; Kuritzkes, D. R.; Loveday, C.; Shafer, R. W.; Richman, D. D. Drug resistance mutations in HIV-1. *Top. HIV Med.* **2002**, *10*, 11−19.

(17) Erickson, J. W.; Burt, S. K. Structural mechanisms of HIV drug resistance. *Annu. Rev. Pharmacol. Toxicol.* **1996**, *36*, 545−571.

(18) Kempf, D. J.; Marsh, K. C.; Denissen, J. F.; McDonald, E.; Vasavanonda, S.; Flentge, C. A.; Green, B. E.; Fino, L.; Park, C. H.; Kong, X. P.; et al. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2484−2489.

(19) Wang, Y. X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry* **1996**, *35*, 9945−9950.

(20) Imamura, K.; Nimz, O.; Jacob, J.; Myles, D.; Mason, S. A.; Kitamura, S.; Aree, T.; Saenger, W. Hydrogen-bond network in cyclodecaamylose hydrate at 20 K; neutron diffraction study of novel structural motifs band-flip and kink in alpha-(1→4)-D-glucoside oligosaccharides. *Acta Crystallogr. B* **2001**, *57*, 833−841.

(21) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* **1995**, *3*, 581−590.

(22) Sedlacek, J.; Fabry, M.; Horejsi, M.; Brynda, J.; Luftig, R. B.; Majer, P. A rapid screening method for biological activity of human immunodeficiency virus proteinase inhibitors by using a recombinant DNA-derived bacterial system. *Anal. Biochem.* **1993**, *215*, 306−309.

(23) Kabsch, W. XDS. In *International Tables for Crystallography*, 1st ed.; Rossmann, M. G., Arnold, E., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; Vol. F, pp 730−734.

(24) Kabsch, W. Integration, scaling, space-group assignment and post refinement. In *International Tables for Crystallography*, 1st ed.; Rossmann, M. G., Arnold, E., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; Vol. F, pp 218−224.

(25) Kissinger, C. R.; Gehlhaar, D. K.; Fogel, D. B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* **1999**, *55*, 484−491.

(26) Lange-Savage, G.; Berchtold, H.; Liesum, A.; Budt, K. H.; Peyman, A.; Knolle, J.; Sedlacek, J.; Fabry, M.; Hilgenfeld, R. Structure of HOE/BAY 793 complexed to human immunodeficiency virus (HIV-1) protease in two different crystal forms—structure/function relationship and influence of crystal packing. *Eur. J. Biochem.* **1997**, *248*, 313−322.

(27) 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.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography and NMR system: new software suite for macromolecular structure determination. *Acta Crystallogr. D* **1998**, *54*, 905−921.

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

(29) McRee, D. E. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **1999**, *125*, 156−165.

(30) Sheldrick, G. M.; Schneider, T. R. SHELXL: high-resolution refinement. *Methods Enzymol.* **1997**, *277*, 319−343.

(31) Laskowski, R. A.; McArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.

(32) Hubbard, S. J.; Thornton, J. M. *NACCESS* (Computer Program); Department of Biochemistry and Molecular Biology, University College London: London, 1993.

(33) Lee, B.; Richards, F. M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.

(34) Laskowski, R. A. SURFNET: A program for visualizing molecular surfaces, cavities and intermolecular interactions. *J. Mol. Graphics* **1995**, *13*, 323−330.

(35) Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **1994**, *50*, 760−763.

(36) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 7*; University of California, San Francisco, 2002.

(37) *InsightII 2000*; Molecular Simulations Inc.: San Diego, CA, 2000.

(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.

(39) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(40) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.

JM031105Q

Inhibitor binding at the protein interface in crystals of a HIV-1 protease complex

Jiří Brynda,[a]* Pavlína Řezáčová,[a] Milan Fábry,[a] Magdalena Hořejší,[a] Renata Štouračová,[a] Milan Souček,[b] Martin Hradílek,[b] Jan Konvalinka[b] and Juraj Sedláček[a]

[a]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 16637 Praha 6, Czech Republic, and [b]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 16610 Praha 6, Czech Republic

Correspondence e-mail: brynda@img.cas.cz

Depending on the excess of ligand used for complex formation, the HIV-1 protease complexed with a novel phenylnorstatine inhibitor forms crystals of either hexagonal ($P6_1$) or orthorhombic ($P2_12_12_1$) symmetry. The orthorhombic form shows an unusual complexity of crystal packing: in addition to one inhibitor molecule that is bound to the enzyme active site, the second inhibitor molecule is bound as an outer ligand at the protein interface. Binding of the outer ligand apparently increases the crystal-quality parameters so that the diffraction data allow solution of the structure of the complex at 1.03 Å, the best resolution reported to date. The outer ligand interacts with all four surrounding HIV-1 protease molecules and has a bent conformation owing to its accommodation in the intermolecular space. The parameters of the solved structures of the orthorhombic and hexagonal forms are compared.

1. Introduction

The retroviral protease of HIV (HIV PR) is an enzyme that has been extensively studied by protein crystallography methods (reviewed, for example, in Wlodawer & Vondrasek, 1998). Structure-based drug design then yielded eight drugs that have now been approved for clinical anti-AIDS use (for a review, see Wlodawer & Vondrasek, 1998). The search for novel compounds that would overcome known drug resistance (Erickson & Burt, 1996) has included a combinatorial approach in which large sets of potential inhibitors are screened against resistant HIV PR species as primary targets (Houghten et al., 1991). Using combinatorial peptide libraries, we have already identified several compounds that have good (subnanomolar) inhibitory capacity towards representative drug-resistant HIV PR forms (Rinnova et al., 2000). The phenylnorstatine group, an atypical inhibitor moiety, served the purpose of investigating the potential of replacement of the peptide bond with larger groups. In the previous paper (Brynda et al., 2004) and in the present paper, the structure of wild-type HIV-1 PR complexed with one of these inhibitors, Z-Pns-Phe-Glu-Glu-NH_2 [Z, benzyloxycarbonyl; Pns, phenylnorstatine, ($2RS,3S$)-3-amino-2-hydroxy-4-phenylbutanoic acid] is described. Using a sixfold molar excess of ligand for pre-forming the HIV-1 PR complex, the protein–inhibitor crystals grew to an extraordinary diffraction quality, allowing thus the structure to be determined at 1.03 Å resolution; to our knowledge, this is the highest resolution of any HIV PR–inhibitor complex reported to date (PDB as of June 2004).

electronic reprint

Table 1
X-ray data-collection and processing statistics.

Values in parentheses correspond to the last resolution shell.

	$P2_12_12_1$	$P6_1$
Space group	$P2_12_12_1$	$P6_1$
Unit-cell parameters (Å)	$a = 28.85$, $b = 66.52$, $c = 93.10$	$a = 61.37$, $b = 61.37$, $c = 80.52$
Diffraction limits (Å)	54.1–1.03 (1.06–1.03)	26.54–2.20 (2.33–2.20)
No. measured diffraction maxima	872620	54226
No. unique reflections	88784	8742
Average $I/\sigma(I)$	10.5 (1.85)	10.4 (2.56)
R_{sym} (%)	7.8 (37.9)	11.3 (58.7)
Completeness (%)	99.0 (93.3)	99.3 (97.1)
Wilson B factor (Å^2)	7.9	46.9

Table 2
Refinement statistics.

	$P2_12_12_1$	$P6_1$
Space group	$P2_12_12_1$	$P6_1$
R factor (%)	13.0	20.9
R_{free} factor (%)	16.5	26.3
Non-H atoms in model	1937	1669
Water molecules	233	17
Reflections used in refinement	86020	8326
Reflections in test set	2205	413
Rm.s.d. from ideal bond distances (Å)	0.016	0.17
R.m.s.d. from ideal valence-angle values (°)	0.036 Å†	2.94
Program used	$SHELX$97	$REFMAC$ 5.1.24

† Distance between two atoms that are both bonded to the same atom.

2. Experimental

2.1. Crystallization

The inhibitor Z-Pns-Phe-Glu-Glu-NH$_2$ belongs to a series of inhibitors that replace the scissile bond at the cleavage site of the substrate with a phenylnorstatine group. The inhibitor was synthesized on Rink amide MBHA resin using the Fmoc/*t*-butyl-HOBt/DIC strategy, as published previously (Rinnova *et al.*, 2000). HIV-1 PR (wild type, Bru isolate) used for complex formation and crystallization was obtained by recombinant expression as described previously (Sedlacek *et al.*, 1993).

In trials using the hanging-drop technique, hexagonal as well as orthorhombic crystals were obtained with buffer containing 50 m*M* MES pH 6.5, 2.4 *M* ammonium sulfate as the precipitating agent. The procedure comprised pre-formation of the complex using a sixfold molar excess of inhibitor in the case of orthorhombic crystals and a threefold to fourfold molar excess of the inhibitor in the case of hexagonal crystals. The complex solution containing 10 m*M* sodium acetate pH 5.6, 0.05%(*v*/*v*) 2-mercaptoethanol, 1 m*M* ethylenediamine-tetraacetic acid (EDTA) was concentrated to 2.2 mg ml^{-1} protease in a Centricon-10 (Millipore) cell. Each hanging drop consisted of 2 µl concentrated complex solution and 1 µl reservoir solution. Hexagonal crystals appeared overnight, whereas orthorhombic crystals appeared after several days of equilibration at 291 K. Crystals were mounted in a nylon loop, soaked in cryoprotectant solution [reservoir solution with 20%(*v*/*v*) glycerol] for a few seconds, bathed in paraffin oil and finally transferred into liquid nitrogen and stored frozen.

2.2. X-ray data collection, structure determination and analysis

Data collection from orthorhombic crystals and determination and refinement of the orthorhombic structure are described in detail in Brynda *et al.* (2004). Coordinates and structure factors for this structure have been deposited with PDB code 1nh0.

The best hexagonal crystal, with dimensions 0.08 × 0.08 × 0.5 mm, was used for measurements. Diffraction data were collected on beamline ID14-2 at the European Synchrotron Radiation Facility, Grenoble at 0.93 Å wavelength, using an ADSC Q4 CCD-based detector at 100 K (Oxford Cryosys-

tems). The data were integrated using *XDS* (Kabsch, 2001*a*) and scaled using *XSCALE* (Kabsch, 2001*b*), *i.e.* by the same procedure as described for the orthorhombic crystals (Brynda *et al.*, 2004). Table 1 summarizes the data-collection statistics.

Since our hexagonal crystal appeared to be isomorphous with all other $P6_1$ crystals of HIV-1 protease complexes, structure determination was performed by the rigid-body refinement protocol using 1vij as the initial model and then by the restrained refinement protocol, both using *REFMAC* v.5.1.24 (Collaborative Computational Project, Number 4, 1994). At this point, the inhibitor was built in the active site according to the difference electron-density map ($m|F_o| - D|F_c|$ coefficients) and its known position in the orthorhombic structure using *XtalView* (McRee, 1999). After a few cycles of restrained refinement, water molecules were added and the model was again refined using *REFMAC*. Coordinates and structure factors for this hexagonal structure have been deposited as PDB code 1u8g. The final refinement statistics are summarized in Table 2.

Analysis of crystal contacts was performed using *CONTACT* (Collaborative Computational Project, Number 4, 1994). The parameters for contact counts were as follows: hydrogen-bond distance limits were 2.25–3.25 Å (non-H–non-H atom) and the van der Waals bond distance limit was 3.9 Å (non-H–non-H atom). The molecular models used comprised only protein and inhibitor molecules; water, ions and other small molecules were ignored. In the cases of dual conformations, the atoms having partial occupancy were counted with their respective occupancy factors.

Density of crystal packing is expressed as the volume of the unit cell divided by the number of liganded HIV PR dimers in it; this (reciprocal) parameter is termed 'per HIV PR molecule volume'. The solvent contents were calculated with *MATTHEWS_COEF* program (Collaborative Computational Project, Number 4, 1994).

3. Results

3.1. Resolution limit and model quality

The three-dimensional structure of HIV-1 protease complexed with a recently characterized inhibitor (Rinnova *et al.*, 2000), Z-Pns-Phe-Glu-Glu-NH$_2$ (see §1 for abbreviations and Fig. 1 for the formula) was determined (see also Brynda *et al.*, 2004). Orthorhombic crystals grown from a co-crystal-

lization mixture containing a sixfold molar excess of the inhibitor (see §2) diffracted to nearly 1 Å resolution. For the chosen resolution of 1.03 Å, reflections in the last shell had an average $I/\sigma(I)$ ratio of higher than 1.85. After molecular replacement and rigid-body refinement, electron density for the inhibitor molecule in the active site was clearly recognizable. On the basis of additional continuous positive electron density in the solvent region, the second molecule of the inhibitor bound to the outer protein surface was built into the model (Fig. 2). Although the compound used for complex formation and crystal growth had mixed R and S chirality at the C21 C atom (the only racemic atom), both bound inhibitor molecules were of R chirality at C21.

Remarkably, co-crystallization from a mixture having the same composition except for a lower (threefold) molar excess of the inhibitor led to the growth of hexagonal crystals (space group $P6_1$) that diffracted to 2.2 Å resolution (PDB code 1u8g). Table 2 summarizes the data-collection, processing and structure-refinement statistics. In brief, a single inhibitor molecule was found in the active site with a mixed orientation, as consistently observed with all other hexagonal crystals (e.g. Dohnálek et al., 2001). The inhibitor structure is identical to that of the active-site inhibitor molecule in the $P2_12_12_1$ form: the r.m.s.d.s for the two orientations in the $P6_1$ form are 0.101 and 0.109 Å, respectively, calculated against all 53 non-H atoms of the prevalent conformation of the active-site inhibitor of the $P2_12_12_1$ form.

The final map obtained for the orthorhombic form showed very clear electron densities for all side chains except for those of two HIV PR surface amino-acid residues, LysA14 and ArgA41. The refined model contains 198 amino acids, two inhibitor molecules, 230 water molecules, two sulfate anions, a complete mercaptoethanol molecule bound to residue CysA67 and another mercaptoethanol molecule represented only by its S atom at a distinctive covalent-binding distance to atom SG of CysB67, similar to that of other PDB entries (e.g. 1dif, 1hih). The model of the protease has 94.9% of the non-glycine and non-proline residues in the most favoured region of the Ramachandran plot and the remainder in the additionally allowed region. Remarkably, the model accuracy of the atom coordinates and their statistical deviations permits an assessment of the protonation/deprotonation of carboxylic O atoms of the enzyme catalytic aspartates on the basis of C—O distances (Wlodawer et al., 2001), as detailed elsewhere (Brynda et al., 2004).

3.2. Crystal packing: contacts of the inhibitor with outer protein surface

The complex crystallized in space group $P2_12_12_1$ and the asymmetric part of the unit cell contained one molecule of HIV PR dimer and two molecules of the inhibitor, one in the active site and one as an outer ligand (Fig. 3). The arrangement of the protein molecules in the crystal lattice is distinctly different from that commonly found in other $P2_12_12_1$ orthorhombic crystals of HIV PR complexes (March et al., 1996; Prabu-Jeyabalan et al., 2000; Mahalingam et al., 2001, 2002;

King et al., 2002). As shown in Fig. 3 (see also the stereoview in Fig. 4b), the orientation of the protein molecules is basically such that a region to the 'left' of the four-termini region of a HIV PR molecule points towards one arm of the flap structure of one neighbouring HIV PR molecule, while the region to the 'right' of the four-termini region points towards one arm of the flap structure of another neighbouring HIV PR molecule. Each HIV PR has one inhibitor molecule in its active site and the additional inhibitor molecules consistently occupy the sites where the flap and the four-termini regions of HIV PR molecules are close to each other in the crystal lattice.

Each molecule of the inhibitor bound to the outer protein surfaces makes intermolecular contacts with the protein chains of four HIV PR molecules. The outer ligand inhibitor molecule has a bent conformation, in contrast to the extended conformation of the inhibitor molecule bound to the active site; a more detailed comparison is given in the following section. The outer ligand interacts via two hydrogen bonds with two amino-acid residues in the terminal region of one HIV PR molecule (J1 O9 to TrpA6 NE1 and GluJ4 OE1 to ThrA4 OG1) and via one hydrogen bond to a symmetry-related HIV PR molecule (GluJ4 NXT to CysB67 O).

The outer ligand makes one hydrogen bond to the N-terminal region of one HIV PR molecule (J1 O9 to TrpA6 NE1, 2.91 Å). The outer ligand displays a number of
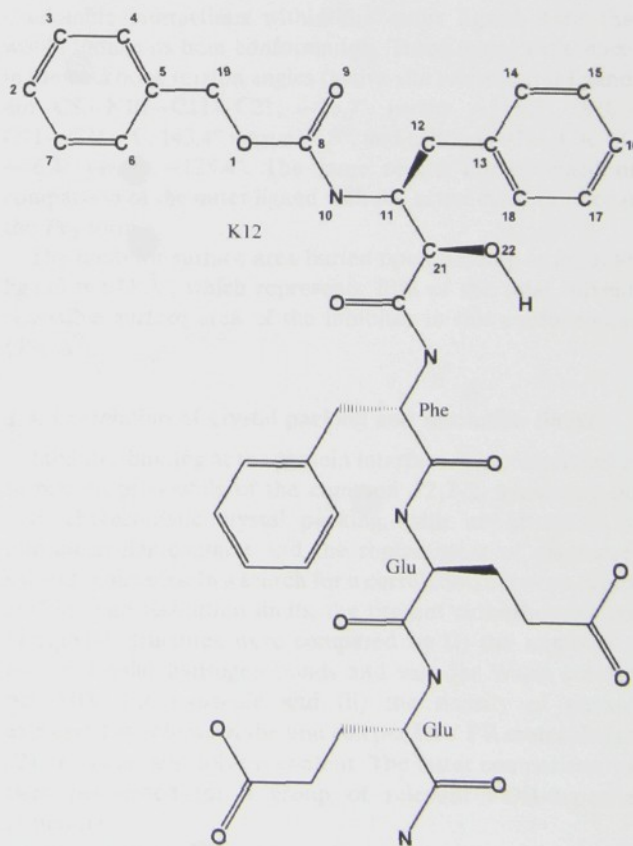

Figure 1
Chemical structure of the inhibitor; the numbering of atoms and labelling of residues correspond to the deposited PDB file.
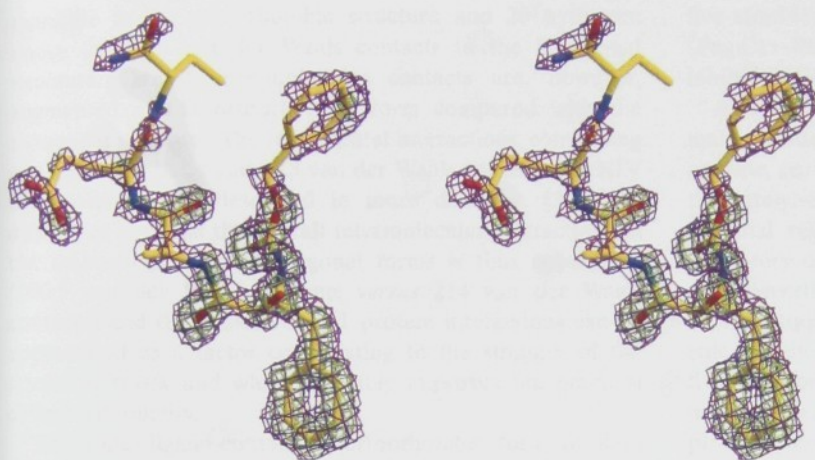
Figure 2
Inhibitor bound at the protein interface shown in an omit map (blue, 3σ level; green, 6σ level).

Figure 3
Crystal packing. One unit cell is shown; the orientation of its axes is marked at the bottom right. Protein molecules are represented by ribbon models coloured red for the molecule in the asymmetric part of the unit cell and coloured green, blue, and yellow for the first, second and third symmetry-related molecules, respectively. Stick models, coloured orange for the molecules bound in the active site and magenta for the molecules contacting the outer protein surface, represent the inhibitor molecules.

van der Waals interactions with all four surrounding HIV PR molecules: 35 van der Waals interactions with the dimer in the asymmetric unit (red in Fig. 4b), seven van der Waals interactions with a dimer in the neighbouring unit cell (yellow molecule in Fig. 4b), 14.5 van der Waals interactions with the dimer in the first symmetry-related position (green molecule in Fig. 4b) and 30 van der Waals interactions with the dimer in the second symmetry-related position (blue molecule in Fig. 4b).

When counted per HIV PR dimeric molecule, interactions with outer ligands comprise one hydrogen bond and 35 + 7 + 14.5 + 30 (i.e. a total of 86.5) van der Waals contacts.

The polypeptide chains of HIV PR in the binding regions of the outer ligand remain undistorted as revealed by direct comparison with the present hexagonal form (r.m.s.d. is 0.18 Å for the main-chain atoms) or with other HIV PR structures, e.g. PDB code 1hxw (Kempf et al., 1995).

3.3. Inhibitor conformation: inside and outside the active site

The exclusive presence of the R stereo-isomer in the complex (despite the presence of sufficient amounts of the S stereoisomer before and during crystal growth) is clearly evidenced by the omit map for the active-site molecule (Brynda et al., 2004) as well as for the outer ligand, as shown in Fig. 2. The inhibitor molecule bound in the active site displays interactions of its main chain and side chains that are similar in most HIV PR complexes with peptidomimetic inhibitors, except for the phenylnorstatine moiety of the present inhibitor, which maintains a unique type of hydrogen bonding (Brynda et al., 2004). The inhibitor bound to the active site displays the usual extended conformation, corresponding to its accommodation in the enzyme active site. When this conformation is superimposed on the conformation of the same compound bound as the outer ligand (Fig. 4a) many differences become evident. The outer ligand has a bent conformation corresponding to accommodation in the intermolecular space (Fig. 4b) and to the above-described interaction with the protein-surface residues. There are, however, no evident hydrogen-bonding or Coulombic interactions within the outer ligand itself that would induce its bent conformation. Three major differences in the backbone torsion angles (active site versus outer ligand) are C8—N10—C11—C21, −75.7° versus −116.2°, N10—C11—C21—C, 143.4° versus 74.8°, and φ Phe C—N—CA—C, −66.4° versus −125.4°. The same results are obtained on comparison of the outer ligand with the active-site inhibitor of the P6$_1$ form.

The inhibitor surface area buried upon binding as an outer ligand is 674 Å^2, which represents 70% of the total solvent-accessible surface area of the inhibitor in this conformation (954 Å^2).

3.4. Correlation of crystal packing and resolution limits

Inhibitor binding at the protein interface is accompanied by formation of crystals of the common $P2_12_12_1$ symmetry but with characteristic crystal packing, with an alteration in intermolecular contacts and the replacement of disordered solvent molecules. In a search for a correlation between crystal packing and resolution limits, the present orthorhombic and hexagonal structures were compared by (i) the numbers of intermolecular hydrogen bonds and van der Waals contacts per HIV PR molecule and (ii) the density of packing expressed as volume of the unit cell per HIV PR molecule (see §2) or as per cent solvent content. The latter comparison was then performed for a group of relevant PDB-deposited structures.

The present orthorhombic and hexagonal structures display similar extents of direct protein–protein interaction: 24 hydrogen bonds and 214 van der Waals contacts per HIV PR

molecule in the orthorhombic structure and 20 hydrogen bonds and 206 van der Waals contacts in the hexagonal structure. Overall intermolecular contacts are, however, augmented in the orthorhombic form compared with the hexagonal structure. The incremental interactions, comprising one hydrogen bond and 86.5 van der Waals contacts per HIV PR molecule, are described in more detail in §3.2. The difference between the overall intermolecular interactions in the orthorhombic and hexagonal forms is thus substantial (300.5 van der Waals contacts *versus* 214 van der Waals contacts) and the protein–ligand–protein interactions can be appreciated as a factor contributing to the strength of the crystal network and which probably improves the practical diffraction quality.

The outer-ligand-containing orthorhombic form displays very dense crystal packing: its per HIV PR molecule volume, 44 815.5 Å^3, is the fourth smallest value among the 168 PDB-deposited structures of HIV-1 PR complexes and its solvent content, 36.1%, is the lowest. However, the importance of the dense crystal packing cannot simply be inferred since the hexagonal form is packed with a similarly high density (43 771.9 Å^3, the smallest value of the 168 structures, and 36.7%, the second lowest solvent content). Also, the overall ranking of the 168 structures according to their density of crystal packing (data not shown) gives no obvious clues. Contribution of the density of packing to the crystal quality becomes apparent with a correlation performed separately for structures of $P2_12_12_1$ symmetry (55 PDB entries, when not including another three of >59 000 Å^3 per HIV PR molecule volume). The plot of the calculated 'per HIV PR molecule volumes' and the stated resolution limits (Fig. 5) displays a direct proportionality trend for this group, although with a mediocre correlation coefficient of 0.498. In qualitative terms, this correlation can probably be understood as the crystal quality being improved with denser crystal packing since the protein molecules are less exposed to solvent and their side chains are more restricted in movement.

4. Discussion

Several aspects of the binding of a novel compound Z-Pns-Phe-Glu-Glu-NH$_2$ to HIV protease have already been examined at the level of an atomic resolution complex structure (Brynda *et al.*, 2004). The central part of this compound includes the pseudodipeptide phenylnorstatine-phenyla-lanine (Pns-Phe), which has an extended

five-atom-long backbone between two key aromatic groups (Pns C11–Phe C^α) and resembles successful Pns-Pro-based inhibitors (Reiling *et al.*, 2002).

A second molecule of inhibitor, bound as an outer ligand, makes contacts with functionally important regions of the enzyme, *i.e.* with the flap structure that moves in each turn of the catalytic cycle and with the Trp6 residue in the amino-terminal region of the polypeptide chain involved in the obligatory dimerization of the enzyme subunits (Figs. 3 and 4b). Nevertheless, at least two arguments exist against the consideration of an inhibitory role of the outer ligand mole-cule. Firstly, assembly in solution of a complex consisting of five components (four protein molecules and one ligand molecule, as in the crystal) seems unlikely. Secondly, the protein regions in contact with the ligand are of undistorted conformations and this contrasts with known mechanisms of
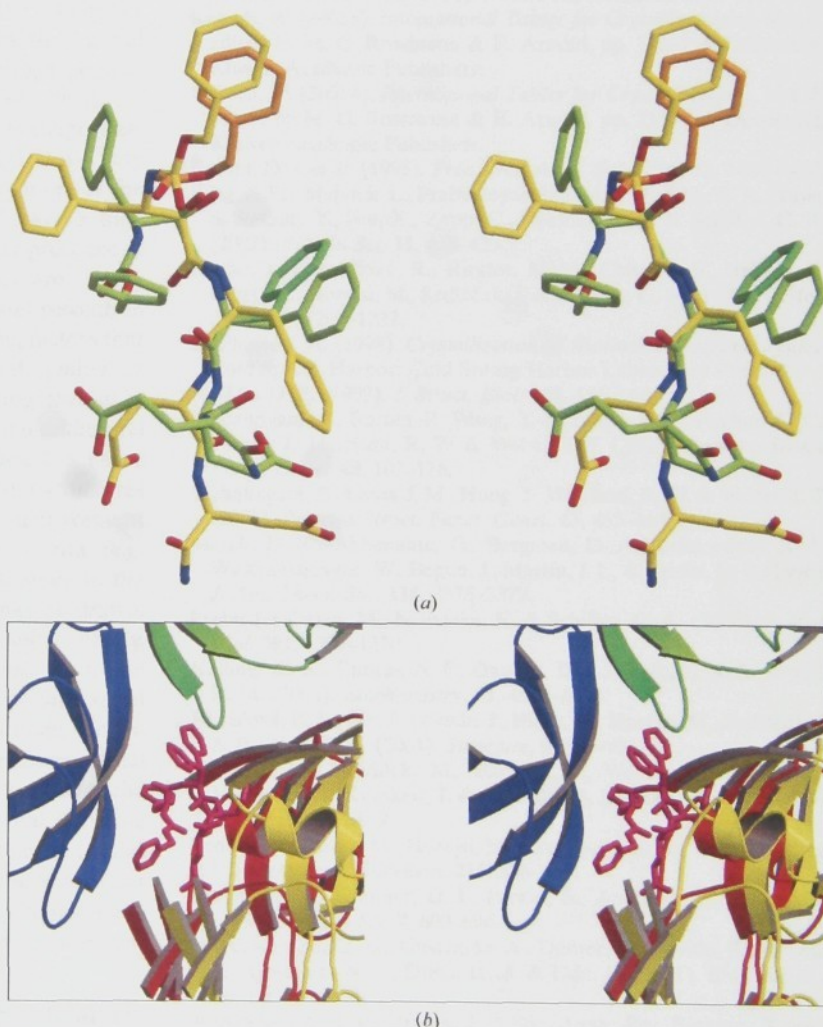

(a)

(b)

Figure 4
Details and comparisons of the ligand structures. (a) Stereoview comparison of the conformations of the active-site inhibitor (yellow) and the outer ligand (green). (b) Stereoview of inhibitor (magenta) maintaining contacts with four protein molecules: red colour corresponds to the protein molecule in the asymmetric part of unit cell and yellow shows that in the neighbouring cell; green corresponds to the third symmetric position and blue to the fourth symmetric position.
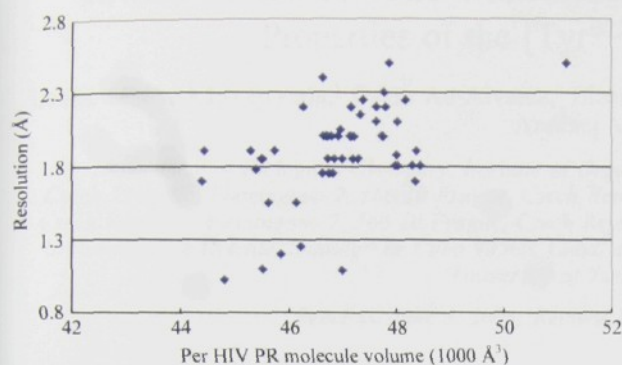
Figure 5
Plot of the calculated 'per HIV PR molecule volumes' and the stated resolution limits for 55 PDB entries. For details, see text.

HIV PR inhibition outside its active side: the inhibitory antibodies with epitopes in the flap or in the amino-terminal HIV PR region do substantially distort the epitope peptide conformations (Lescar *et al.*, 1997; Rezacova *et al.*, 2001).

The quality of a protein crystal is defined by its shape, size, mosaicity and resolution limit (McPherson, 1998). Very generally, resolution limit is determined by the order of the crystal structure: a highly ordered crystal will have a high resolution limit. While negative influences, such as presence of chemical or conformational heterogeneities, are easily understandable, positive factors improving crystal resolution limit often lack rigorous explanation. In literature, factors that facilitate initial crystal growth, improve crystal quality or expand resolution limit are frequently discussed together, as in the case of metal ions, additives or detergents. An additional inhibitor molecule bound at the protein interface has never been found in the large variety of HIV PR crystal structures solved before and described cases of the improvement achieved by co-crystallization with a specific ligand (*e.g.* Trakhanov *et al.*, 1998) do not bear sufficient analogy to the present case. We believe that two conclusions may be drawn, fully based on the above experimental results. Firstly, improvement of the resolution evidently arises from the characteristic crystal packing in view of the substantial difference between the diffraction of the orthorhombic crystals containing the outer ligand and that of the conventional hexagonal ones. Second, all quantitative parameters match the criteria for supplementary ordering of the crystal with the interface ligand binding: the movement of protein molecules relative to each other may become restricted owing to stronger networking and the movement of side chains may be restricted in larger areas of the protein molecules owing to replacement of disordered solvent.

References

Brynda, J., Rezáčová, P., Fábry, M., Horejsi, M., Štouračová, R., Sedláček, J., Souček, M., Hradílek, M., Lepsik, M. & Konvalinka, J. (2004). *J. Med. Chem.* **47**, 2030–2036.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dohnálek, J., Hašek, J., Dušková, J., Petroková, H., Hradílek, M., Souček, M., Konvalinka, J., Brynda, J., Sedláček, J. & Fábry, M. (2001). *Acta Cryst.* D**57**, 472–476.

Erickson, J. W. & Burt, S. K. (1996). *Annu. Rev. Pharmacol. Toxicol.* **36**, 545–571.

Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T. & Cuervo, J. H. (1991). *Nature (London)*, **354**, 84–86.

Kabsch, W. (2001a). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 730–734. Dordrecht: Kluwer Academic Publishers.

Kabsch, W. (2001b). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 218–224. Dordrecht: Kluwer Academic Publishers.

Kempf, D. J. *et al.* (1995). *Proc. Natl Acad. Sci. USA*, **92**, 2484–2488.

King, N. M., Melnick, L., Prabu-Jeyabalan, M., Nalivaika, E. A., Yang, S. S., Gao, Y., Nie, X., Zepp, C., Heefner, D. L. & Schiffer, C. A. (2002). *Protein Sci.* **11**, 418–429.

Lescar, J., Štouračová, R., Riottot, M.-M., Chitarra, V., Brynda, J., Fábry, M., Horejsi, M., Sedláček, J. & Bentley, G. A. (1997). *J. Mol. Biol.* **267**, 1207–1222.

McPherson, A. (1998). *Crystallization of Biological Macromolecules*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.

Mahalingam, B., Boross, P., Wang, Y.-F., Louis, J. M., Fischer, C. C., Tozser, J., Harrison, R. W. & Weber, I. T. (2002). *Proteins Struct. Funct. Genet.* **48**, 107–116.

Mahalingam, B., Louis, J. M., Hung, J., Harrison, R. W. & Weber, I. T. (2001). *Proteins Struct. Funct. Genet.* **43**, 455–464.

March, D. R., Abbenante, G., Bergman, D. A., Brinkworth, R. I., Wickramasinghe, W., Begun, J., Martin, J. L. & Fairlie, D. P. (1996). *J. Am. Chem. Soc.* **118**, 3375–3379.

Prabu-Jeyabalan, M., Nalivaika, E. & Schiffer, C. A. (2000). *J. Mol. Biol.* **301**, 1207–1220.

Reiling, K. K., Endres, N. F., Dauber, D. S., Craik, C. S. & Stroud, R. M. (2002). *Biochemistry*, **41**, 4582–4594.

Řezáčová, P., Lescar, J., Brynda, J., Fábry, M., Horejsi, M., Sedláček, J. & Bentley, G. A. (2001). *Structure*, **9**, 887–95.

Rinnova, M., Hradílek, M., Barinka, C., Weber, J., Souček, M., Vondrasek, J., Klimkait, T. & Konvalinka, J. (2000). *Arch. Biochem. Biophys.* **382**, 22–30.

Sedláček, J., Fábry, M., Horejsi, M., Brynda, J., Luftig, R. B. & Majer, P. (1993). *Anal. Biochem.* **215**, 306–309.

Trakhanov, S., Kreimer, D. I., Parkin, S., Ames, G. F. & Rupp, B. (1998). *Protein Sci.* **7**, 600–604.

Wlodawer, A., Li, M., Gustchina, A., Dauter, Z., Uchida, K., Oyama, H., Goldfarb, N. E., Dunn, B. M. & Oda, K. (2001). *Biochemistry*, **40**, 15602–15611.

Wlodawer, A. & Vondrasek, J. (1998). *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284.

Toward the Insulin–IGF-I Intermediate Structures: Functional and Structural Properties of the [Tyr^{B25}NMePhe^{B26}] Insulin Mutant[†]

Lenka Žáková,[‡] Jiří Brynda,[§] Oskar Au-Alvarez,[‖] Eleanor J. Dodson,[⊥] Guy G. Dodson,[⊥] Jean L. Whittingham,[⊥] and Andrzej M. Brzozowski*,[⊥]

Department of Biological Chemistry, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo 2, 166 10 Prague, Czech Republic, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo 2, 166 10 Prague, Czech Republic, Departamento de Quimica, Facultia De Cienicias Naturales, Universidad de Oriente, Santiago de Cuba 90500, Cuba, and York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, U.K.

Received June 3, 2004; Revised Manuscript Received September 13, 2004

ABSTRACT: The origins of differentiation of insulin from insulin-like growth factor I (IGF-I) are still unknown. To address the problem of a structural and biological switch from the mostly metabolic hormonal activity of insulin to the predominant growth factor activities of IGF-I, an insulin analogue with IGF-I-like structural features has been synthesized. Insulin residues Phe^{B25} and Tyr^{B26} have been swapped with the IGF-I-like Tyr^{24} and Phe^{25} sequence with a simultaneous methylation of the peptide nitrogen of residue Phe^{B26}. These modifications were expected to introduce a substantial kink in the main chain, as observed at residue Phe^{25} in the IGF-I crystal structure. These alterations should provide insight into the structural origins of insulin–IGF-I structural and functional divergence. The [Tyr^{B25}NMePhe^{B26}] mutant has been characterized, and its crystal structure has been determined. Surprisingly, all of these changes are well accommodated within an insulin R6 hexamer. Only one molecule of each dimer in the hexamer responds to the structural alterations, the other remaining very similar to wild-type insulin. All alterations, modest in their scale, cumulate in the C-terminal part of the B-chain (residues B23–B30), which moves toward the core of the insulin molecule and is associated with a significant shift of the A1 helix toward the C-terminus of the B-chain. These changes do not produce the expected bend of the main chain, but the fold of the mutant does reflect some structural characteristics of IGF-1, and in addition establishes the CO^{A19}–NH^{B25} hydrogen bond, which is normally characteristic of T-state insulin.

Despite a wealth of structural information about human insulin, the exact nature of the binding of this hormone to its cognate insulin receptor remains unknown. Some complementary data about the structure–function relationship of insulin come from studies on human insulin-like growth factor I (hIGF-I).[1] hIGF-I, as opposed to two-chain insulin, is a single polypeptide 70-amino acid hormone with close structural and biological connections to insulin, reflected in its ability to bind to both the hIGF-I receptor and, with lower affinity, the insulin receptor. Insulin consists of two disulfide-linked chains (residues A1–A21 and B1–B30), where the B-chain is homologous with the amino-terminal 29-residue region of hIGF-I. A 12-amino acid linking sequence ("C-loop") of hIGF-I joins its B-domain to a region corre-sponding to the A-chain of insulin. The terminus of the C-end of hIGF-I contains eight additional amino acids named the D-domain (Figure 1).

The comparative studies of insulin, IGF-I, and their analogues are complicated further by the ability of insulin to form at least two, different hexamer (T6 and R6) conformers. The so-called T (tense) state occurs when the A and B chains form a more compact molecule with the first six N-terminal residues (B1–B6) adopting an extended chain fold. In the R (relaxed) conformer, all these residues are in a fully α-helical conformation extending the already existing B9–B19 α-helix (*1*). However, several other intermediate T → R hexamers and insulin forms (e.g., T3R3 and T3R3^f) have also been observed (*2*)

Recent crystal structure determinations of hIGF-I (*3*, *4*) revealed a structure that might represent one of the active

* To whom correspondence should be addressed. E-mail: marek@ysbl.york.ac.uk. Telephone: +44-(0)1904-328265. Fax: +44-(0)1904-328266.

[‡] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

[§] Institute of Molecular Genetics, Academy of Sciences of the Czech Republic.

[‖] Universidad de Oriente.

[⊥] University of York.

[1] Abbreviations: hIGF-I, human insulin-like growth factor I; TPCK-trypsin, trypsin treated with L-1-tosylamine-2-phenylethyl chloromethyl ketone; DOI, desoctapeptide$^{B23-B30}$ insulin; Tris, TRIZMA buffer; Fmoc, 9-fluorenylmethoxycarbonyl; HBTU, 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; DIPEA, diisopropylethyl-amine; PyBroP, bromo tris(pyrrolidinophosphonium) hexafluorophos-phate; TFE, trifluoroethanol; DCM, dichloromethane; TFA, trifluoro-acetic acid; TIS, triisopropylsilane; DMF, *N*,*N*-dimethylformamide; GFYNMeFTPK(Pac)T, Gly-Phe-Tyr-NMePhe-Thr-Pro-Lys(Pac)-Thr; Pac, phenylacetyl group; RP-HPLC, reverse phase high-performance liquid chromatography; SEM, standard error of the mean.
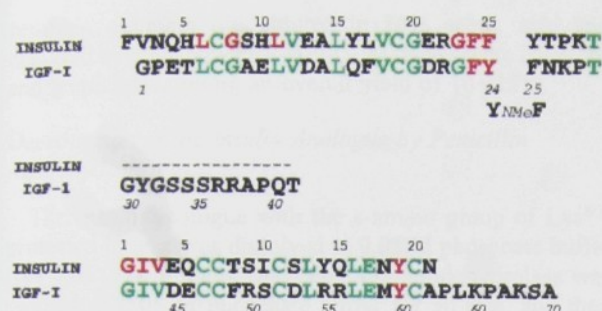
```
        1    5         10        15        20        25
INSULIN FVNQHLCGSHLVEALYLVCGERGFF      YTPKT
IGF-I   GPETLCGAELVDALQFVCGDRGFY       FNKPT
        1                              24   25
                                       Y NMe F
```

```
INSULIN -------------------
IGF-1   GYGSSSRRAPQT
        30     35    40
```

```
        1    5         10        15        20
INSULIN GIVEQCCTSICSLYQLENYCN
IGF-I   GIVDECCFRSCDLRRLEMYCAPLKPAKSA
        45        50        55        60      60      70
```

FIGURE 1: Sequence and chain organization of the [Tyr^{B25}NMe-Phe^{B26}] mutant, insulin, and hIGF-I. The sequence changes in [Tyr^{B25}NMePhe^{B26}] are given in the bottom part of the top row (B-chain) [otherwise, the mutant sequence has been omitted in the middle (C-chain) and bottom (A-chain) rows as it is identical to that of insulin]. The top numbering in each row corresponds to insulin and the bottom one to the IGF-I sequence. Identical residues in insulin and IGF-I are colored green, and residues that are important for receptor binding are colored red.

conformations of insulin adopted upon receptor binding. This interaction is known to involve the N-terminal part of the A-chain of insulin uncovered by the displacement of the B-chain C-terminus. However, the molecular bases of the structural switch that determine biological specificities of both hormones are still unknown. The putative "insulin-active" conformation described in hIGF-I is promoted by the C-loop which introduces a sharp bend at residue Phe^{25}, exposing residues 42−44 (equivalent to residues A1−A3 in insulin). This observation is in agreement with other studies, which have established the link between exposure of residues A1−A3 and insulin activity (5−7). On the other hand, there is other evidence (8−11) which demonstrates that not only the rigid, covalent bond between the N-terminus of the A-chain and the C-terminus of the B-chain but also an excessive flexibility of C-terminal residues of the B-chain leads to a low binding affinity of such insulin analogues. The three evolutionarily conserved aromatic residues, Phe^{B24}, Phe^{B25}, and Tyr^{B26} of insulin and Phe^{23}, Tyr^{24}, and Phe^{25} of hIGF-I, are among the most important residues for the interaction with insulin or IGF-I receptors (12−14). The alterations of these amino acids (especially those diminishing the hydrophobic character of these side chains) drastically affect the affinity of the hormones for their receptors (15).

Both the insulin and IGF-I receptors belong to the tyrosine kinase family of receptors with sequence homologies varying from 41 to 84% between different domains and subunits (16, 17). Despite this high degree of similarity and analogous functional organizations of these receptors, insulin and IGF-I bind to the noncognate receptor with low affinity (18). The insulin−IGF-I mutant studies indicated also that their level of binding to the noncognate receptor is proportional to the degree of similarity of a particular analogue to the physiological ligand (14, 19).

These observations pushed us toward a more detailed analysis of the possible "IGF-like" structural transitions within the insulin molecule. In this study, we have tried to probe the structural effects of swapping insulin residues Phe^{B25} and Tyr^{B26} with the IGF-I-like Tyr^{24} and Phe^{25} residues, respectively, coupled with the methylation of the main chain nitrogen atom of residue B26 (Figure 1). This additional modification of the main chain was introduced with the aim of forcing the terminal part of the B-chain of

insulin to move away into an IGF-like conformation, mimicking the kink in IGF-I at residue Phe^{B25}. This new insulin mutant will here be termed [Tyr^{B25}NMePhe^{B26}]. The corresponding chain numberings in insulin (or the [Tyr^{B25}NMePhe^{B26}] mutant) and h-IGF-I are given in brackets and braces, respectively.

MATERIALS AND METHODS

Synthesis of the [Tyr^{B25}NMePhe^{B26}] Mutant

The preparation of [Tyr^{B25}NMePhe^{B26}] involved several steps. First, zinc-free porcine insulin was treated with TPCK-trypsin dissolved in 0.1 M CaCl$_2$ to give desoctapeptide$^{B23−B30}$ insulin (DOI). Second, DOI was coupled with the Tyr^{B25}NMePhe^{B26} octapeptide C-terminal analogue of the human insulin B chain, and finally, the protecting groups were removed.

Preparation of Desoctapeptide$^{B23−B30}$ Insulin (DOI)

To a solution of Zn-free porcine insulin (20) dissolved in 0.05 M Tris-HCl buffer (pH 9.0−9.2) was added TPCK-trypsin in 0.1 M CaCl$_2$ to a final enzyme:substrate ratio of 1:50. This solution was left for 20 h at room temperature, after which the pH of the solution was adjusted to 5.4 using 1 M HCl. The solution was then centrifuged at 2000g for 20 min, and the resulting supernatant was lyophilized. The lyophilized desoctapeptide insulin was subsequently dissolved in 10% acetic acid and purified by gel filtration using a Sephadex G50 matrix. This procedure yielded 75% pure desoctapeptide$^{B23−B30}$ insulin.

Preparation of the Gly-Phe-Tyr-NMePhe-Thr-Pro-Lys(Pac)-Thr-OH Octapeptide

The octapeptide analogue of the C-terminus of the human insulin B-chain was synthesized on a 2-chlorotrityl chloride resin. This process employed the Fmoc protection strategy for α-amino groups of individual amino acids, with *tert*-butyl group protection for the side chain of the tyrosine and threonine residues and phenylacetyl (Pac) group protection for the lysine residue. Coupling of the amino acids was carried out in a HBTU/DIPEA environment, with the exception of that of the *N*-methylphenylalanine, which was carried out in a PyBroP/DIPEA solution. The resulting peptide was cleaved from the resin using an acetic acid/TFE/DCM mixture (1:1:8 by volume), and the protecting groups of the sides were cleaved off using a TFA/TIS/H$_2$O/DCM solution (50:2:2:46 by volume), with the exception of phenylacetyl, which was removed later. The final yield was 85%.

Enzymatic Semisynthesis of the Insulin Analogue

The synthesis of the analogue was performed according to the method of Inouye (21, 22). The amino component [G-F-Y-NMeF-T-P-K(Pac)-T, 150 mM] and 30 mM DOI were dissolved in a solution containing 55% (v/v) DMF, 10 mM CaCl$_2$, and TPCK-trypsin (enzyme:substrate molar ratio of 1:50), to a total volume of 200 μL. The pH value was adjusted to pH 6.9−7.0 using *N*-methylmorpholine. The resulting mixture was then incubated for 4−7 h at room temperature. The reaction, monitored by analytical RP-HPLC, was stopped by the addition of cold acetone. The

resulting sediment was diluted in 10% acetic acid and analyzed by preparative RP-HPLC. All peaks were collected and lyophilized, giving an overall yield of 10−15%.

Deprotection of the Insulin Analogue by Penicillin Amidohydrolase

The insulin analogue with the ϵ-amino group of Lys[B29] protected by Pac was dissolved in 0.05 M phosphate buffer (pH 7.6) (*12*). Immobilized penicillin amidohydrolase was swelled in 0.05 M phosphate buffer for 30 min, and then the analogue was added. The reaction was allowed to run for 2.5 h, and then two peaks arising from RP-HPLC analysis were collected and lyophilized, giving a final of [Tyr[B25]NMePhe[B26]] yield of 50−60%.

Biological Assays

The assay described by Zorad *et al.* (*24*) was used to determine the relative binding affinities of insulin and [Tyr[B25]NMePhe[B26]] for the insulin receptor of rat adipose tissue plasma membranes. The membranes were isolated from epidymal fat of male rats (Wistar, 210−250 g); 50 μg of plasma membrane was incubated in a 5 mL test tube with [^{125}I]insulin (2 × 10^{-10} M ∼ 70 000 cpm). Then various concentrations of insulin or the insulin analogue (ranging from 10^{-13} to 10^{-6} M) in a buffer consisting of 2 mM *N*-ethylmaleimide, 13.2 mM $CaCl_2$, 0.1% (w/v) bovine serum albumin, and Tris-HCl (pH 7.6) were added to a total volume of 250 μL. Each solution was incubated for 21 h at 4 °C, and then the reaction was terminated by the addition of ice-cold 120 mM NaCl followed by a quick filtration on a Brandell (Gaithersburg, MD) cell harvester. The bound radioactivity was determined by γ-counting. Total binding (in the absence of unlabeled insulin) was ∼10% of the total radioactivity. Nonspecific binding (in the presence of 10^{-5} M insulin) amounted to less than 15% of the total binding.

Analysis of the Binding Data

The relative receptor binding potencies of the insulin analogue were determined by measuring the concentration of the analogue causing half-maximal inhibition (IC_{50}) of labeled insulin binding and the concentration of standard human insulin causing the equivalent inhibition within a single experiment. Competitive binding curves were plotted using Graph-Pad (San Diego, CA) Prizm 3, comparing the best fit for single-binding site models. Relative receptor binding potency is defined as (IC_{50} value of human insulin)/(IC_{50} value of analogue) × 100.

Crystallization

[Tyr[B25]NMePhe[B26]] was crystallized by the hanging drop vapor diffusion method. Several attempts to crystallize the mutant using new protocols, different from pre-established crystallization methods for insulin, were unsuccessful. Eventually, crystals were obtained from drops consisting of 1 μL of protein solution (10 mg/mL analogue in 0.02 M HCl) and 1 μL of the reservoir solution [0.1 M trisodium citrate, 0.02% (w/v) zinc acetate, 6% (w/v) Tris-HCl buffer (pH 8.2)] suspended over 1 mL reservoirs. Different concentrations of phenol 0.05, 0.10, and 0.15% (w/v) were added to the reservoir solution prior to mixing with the protein. Crystals

Table 1: X-ray Data and Refinement Statistics	
Data Processing Statistics	
data collection site	in-house
wavelength (Å)	1.54
space group	$P2_1$
unit cell dimensions	$a = 59.90$ Å, $b = 62.12$ Å, $c = 47.80$ Å, $\beta = 110.6°$
diffraction limits (Å)	25.56−2.08 (2.19−2.08)[g]
no. of observations	96148
no. of unique reflections	17496
completeness of the data (%)	88.1 (89.2)[g]
R_{merge}[a]	3.4 (12.0)[g]
$I/\sigma\langle I\rangle$	5.7 (1.8)[g]
Refinement Statistics	
resolution range (Å)	25.57−2.08
no. of reflections used (R_{free} set)	17491 (956)
R_{cryst} (%)/R_{free}[b]	18.9/25.6
no. of protein atoms/ water molecules/metal ions	2475/83/2
rmsd for bonds (Å)/angles (deg)[c]	0.018/1.7
rmsd for main chain ΔB (Å)[d]	1.16
mean *B*-factor (Å²)[e]	43.3/56.3/27.1/59.1/
% A, B, L/a, b, l, p/XX[f]	94.4/4.8/0.8

[a] $R_{merge} = 100\Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$. [b] $R_{crys} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$. R_{free} is like R_{cryst} but calculated over 5.2% of data that were excluded from the refinement process. [c] Root-mean-square deviations in bond lengths and angles from Engh and Huber ideal values. [d] Root-mean-square deviations between $B_{factors}$ for bonded main chain atoms. [e] Mean temperature factor for the whole molecule, main chain, side chain, metal, and water atoms, respectively. [f] Percentage of residues located in the most favored, additional allowed, and disallowed regions of the Ramachandran plot as determined by PROCHECK (*33*). [g] Highest-resolution shell statistics given in parentheses.

grew in all wells within 1 week. They belong to the $P2_1$ space group with one hexamer in the asymmetric unit (Matthews coefficient of 2.38 Da Å^{-3}, solvent content of 48%).

X-ray Crystallography

Data Collection. As the crystals were found to be extremely sensitive to changes in the crystallization environment, flash-cooling experiments were unsuccessful. Therefore, a 2.08 Å resolution data set was collected at room temperature from a crystal mounted in a quartz capillary tube on in-house X-ray equipment (mar345 Image Plate Detector, Nonius FR 591 generator) using Cu Kα_1 radiation. All data were integrated and reduced using MOSFLM (*25*). X-ray data collection and processing statistics are summarized in Table 1.

Structure Determination and Refinement. The insulin coordinates of Protein Data Bank entries 1EVR and 1EV6 (*26*) with the water molecules removed were used as models in a preliminary search in MOLREP (*27*). Slightly better results were obtained with 1EVR, which was subsequently used in final search in AMoRe (*28*) (used here due to its more powerful rigid body refinement tool) to determine the structure by molecular replacement. Final restrained and rigid body refinements were performed with REFMAC5 (*29*) employing maximum likelihood procedures. NCS restraints were not imposed during any stages of the refinement. Water molecules were located using ARP/warp (*30*). Successive rounds of refinement using all data between 25.57 and 2.08 Å, in conjunction with manual rebuilding, gave a final model with an R_{cryst} of 18.9% and an R_{free} of 24.9%. The final model comprises 2409 protein atoms, two metal atoms, and 83 water

Table 2: Values of IC_{50}[a] and Receptor Binding Affinities[b] of Insulin and the [Tyr^{B25}NMePhe^{B26}] Insulin Mutant

ligand	$IC_{50} \pm$ SEM (nM)	n	potency (%)
human insulin	1.83 ± 1.18	10	100
[Tyr^{B25}NMePhe^{B26}] insulin	7.82 ± 1.95	4	23

[a] IC_{50} values represent concentrations of insulin or the insulin analogue causing half-maximal inhibition of binding of [^{125}I]insulin to the insulin receptor. Each value represents the mean $\pm$ SEM of multiple determinations; the number of separate determinations is shown in parentheses. [b] Relative receptor binding affinity defined as (IC_{50} of human insulin)/(IC_{50} of analogue) $\times$ 100. See Experimental Procedures for details.
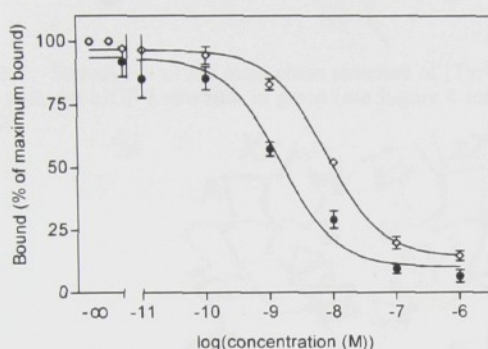

FIGURE 2: Inhibition of binding of [^{125}I]insulin to rat adipose tissue plasma membranes by human insulin (●) and [Tyr^{B25}NMePhe^{B26}] insulin (◇). Values are means $\pm$ SEM. See Experimental Procedures for details. Quantitative information is provided in Table 2.

molecules. All model building was carried out using QUANTA (QUANTA98, Accelrys Inc., San Diego, CA). Programs of the CCP4 suite (*31*) were used during the structure solution and refinement. A summary of the refinement statistics is given in Table 1. Superpositions of [Tyr^{B25}NMePhe^{B26}] and insulin and hIGF-I were carried out using QUANTA with the insulin [B9–B19] helix ({B8–B19} of IGF-I) as the main template. Figures 3 and 5 were produced using MOLSCRIPT (*32*), and Figures 4, 6, and 7 were created using QUANTA. The atomic coordinates of the final model have been deposited in the Protein Data Bank as entry 1W8P.

RESULTS AND DISCUSSION

Biological Activity of the [Tyr^{B25}NMePhe^{B26}] Mutant. The insulin receptor binding affinity of [Tyr^{B25}NMePhe^{B26}] insulin is only 25% of that of human insulin (Table 2 and Figure 2). Previous studies of the replacement of Phe^{B25} with tyrosine have already shown a small decrease in the binding potency (*12*). The substitution of the hydrogen atom of the amide nitrogen with an *N*-methyl group in [Tyr^{B25}NMePhe^{B26}] insulin eliminates its ability to form one of the β-strand hydrogen bonds required for dimer formation. However, the retention of full receptor affinity by despentapeptide$^{B26-B30}$ insulin-B25-amide (*34, 35*) suggests that this hydrogen bond is not crucial for interaction with the insulin receptor (*8*), although the potential involvement of the [B25] amide group in this insulin bond in forming a hydrogen bond with [A19] cannot be ruled out.

The C-terminus of the B-chain of insulin is extremely sensitive to any conformational alteration of the backbone. Studies on insulin with [B24–B25/B25–B26] peptide bonds replaced with an ester bond (*36, 37*) or a methylene group

(*38*) showed extremely low receptor binding affinity, which could indicate a very high flexibility of this part of the molecule, which may cause its inability to be stabilized in an "active" conformation. The roles of [B25] and [B26] amides are likely to be substantially different as methylation of NH[B26] in destetrapeptide$^{B27-B30}$ insulin-B26-amide (DTIA) does not significantly affect its ability to bind to a IR or its biological activity, while [NMeB25]DTIA variants retain no more than 2.3% of the potency of human insulin in *in vitro* studies (*11*). On the contrary, the N-methylation of the [B26] peptide in DTIA associated with [B26]Tyr $\rightarrow$ His substitution gives rise to a potency of 5250%, yielding the most potent insulin ever obtained (*11*). Hence, the 25% potency of [Tyr^{B25}NMePhe^{B26}] is more likely to result from the converging and mutual effects of both the [Phe^{25} $\rightarrow$ Tyr] swap and the NMe[B26] methylation than from a particular single modification.

The discussion of these differing results will follow here a more detailed description of the [Tyr^{B25}NMePhe^{B26}] structure and the conformational changes found in the crystals of this mutant insulin.

Insulin, [Tyr^{B25}NMePhe^{B26}], and IGF-I Structural Relationships. In the discussion below, we analyze the structural relationships of the molecules (here termed *xz* forming one of the dimers of [Tyr^{B25}NMePhe^{B26}] with (a) its structural equivalents (*XZ*) from the monoclinic human insulin {for simplicity, here termed Wild Type, WT [PDB entry 1ZNJ (*39*)]} and (b) the single molecule of hIGF-I [PDB entry 1GZR (*4*)]). This particular (1ZNJ) insulin structure was chosen for a more detailed comparison because of its high degree of similarity in crystal packing with [Tyr^{B25}NMePhe^{B26}] ($a = 59.90$ Å, $b = 62.12$ Å, $c = 47.80$ Å, and $\beta = 110.6°$ for [Tyr^{B25}NMePhe^{B26}] vs $a = 61.23$ Å, $b = 61.65$ Å, $c = 48.05$ Å, and $\beta = 110.5°$ for insulin), to minimize the bias that could result from different interhexamer contacts. In addition, the 1ZNJ insulin model was also probed in the molecular replacement searches, giving solutions very comparable to the other models. The comparisons result from the alignments of these structures using part of their B helix, [Ser9]{Ala8}[Val18]{Val17}, as a template (Figure 3).

The structure of [Tyr^{B25}NMePhe^{B26}] in this crystal is similar to that of the R6 insulin with three dimers assembled into a hexameric 32 organization. There are two zinc ions on the 3-fold axis, each tetrahedrally coordinated to three histidines and one water molecule. The formation of the R6 hexamer in [Tyr^{B25}NMePhe^{B26}] was likely to be due to the presence of phenol in the crystallization medium. Six phenols in the structure form hydrogen bonds, one to the carbonyl oxygen of [A6] and the other to the peptide nitrogen of [A11].

Inspection of [Tyr^{B25}NMePhe^{B26}] and WT hexamers (Figure 4) showed that the structural differences are not evenly propagated throughout these assemblies. However, the differences have local character and concern mostly the subtle changes of the conformation of the C-terminal part of the B-chain (the overall rmsd between the dimers of the mutant was less than 0.1 Å). All main new structural features such as CO^{B25}–NH^{A19} hydrogen bond were fully preserved in all dimers (e.g., the length of this interaction varied from 2.79 Å, through 2.95 Å, to 3.1 Å). The *xz* dimer (equivalent of AB and CD chains in the PDB file) that exhibited the
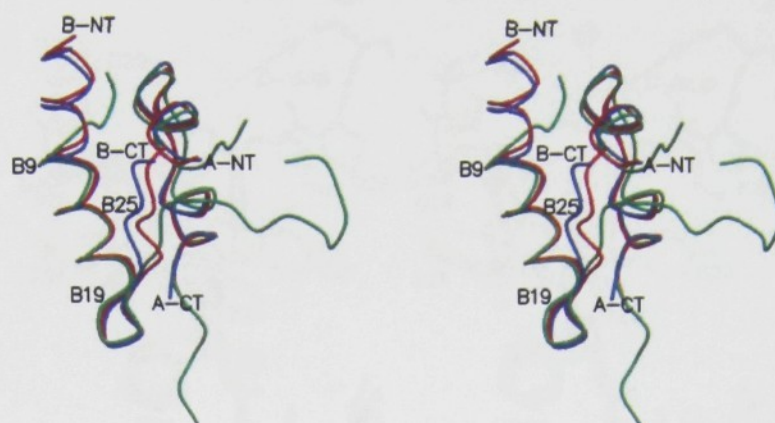
FIGURE 3: Stereoview of the main chain structure of [Tyr[B25]NMePhe[B26]] molecules [blue (*x*) and red (*z*)] superimposed on the [B9−B18] helix, with the hIGF-I structure in green (see Figure 4 for the definition of *x* and *z*).
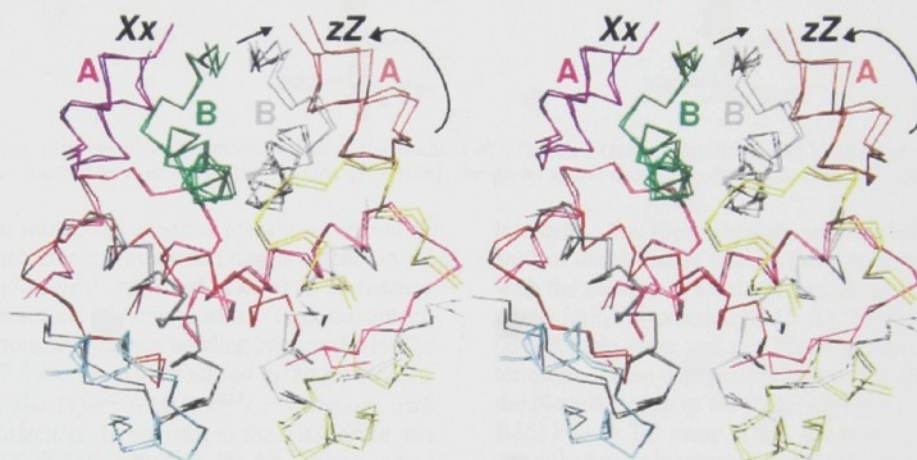

FIGURE 4: Comparison of the [Tyr[B25]NMePhe[B26]] hexamer with the wild-type monoclinic insulin structure [a certain nonstandard projection (3-fold hexamer axis not perpendicular to the plane of the figure) has been chosen to visualize differences in hexamer packing]. The molecules and dimers that are discussed in the text in more detail are marked as *xz* for [Tyr[B25]NMePhe[B26]] and *XZ* for the wild-type insulin; A and B refer to A- and B-chains, respectively, of insulin. The arrows show the direction of movements within B- and A-chains (divergent stereo).

highest degree of conformational changes was used for further comparisons with its *XZ* equivalent from the WT crystal.

As both molecules of the *XZ* dimer of the WT structure are virtually identical (average rmsd on Cα atoms of ∼0.13 Å) only one comparison, with the *X* molecule, has been carried out.

Comparison of the [Tyr[B25]NMePhe[B26]] x Monomer with the X Molecule. The overall structures of the *x* and *X* molecules (Figure 4) are very similar, with the rmsd for Cα atoms being 0.31 Å. The conformations of practically all of the side chains of the N-terminus and the α-helix of the B-chain are very alike, with the rings of [B1]Phe, [B16]Phe, and [B24]Phe being virtually fully superimposable. Surprisingly perhaps, the methylation of the nitrogen atom from the peptide bond of Phe[B26] does not produce any perturbation of the main chain in the *x* molecule.

The position of the new [B25]Tyr side chain is well-defined in the electron density map (Figure 5). It is situated between two alternative conformations of [B25]Phe found in the *X* molecule. The phenyl rings of the new [B26]Phe and WT [B26]Tyr occupy similar positions, wedged between [B12]Val and [B28]Pro. However, [B26]Phe of [Tyr[B25]NMePhe[B26]] has moved into the protein core by ca.
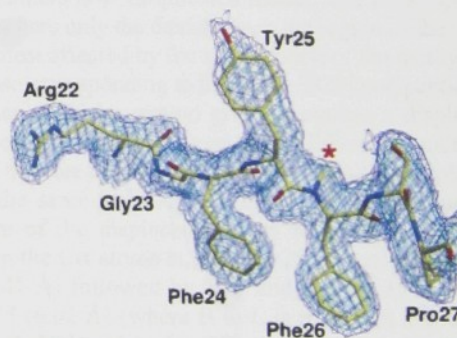

FIGURE 5: Electron density [1σ level (blue) and 0.5σ (purple)] of the [B22−B27] region. The position of methyl group present in the [B25]−[B26] peptide bond is marked with the red asterisk; residues are shown in stick representation with nitrogen, oxygen, and carbon atoms colored blue, red, and yellow, respectively.

1 Å (Figure 6). As the phenyl ring of [B26]Tyr sits in a fully hydrophobic environment in the WT molecule and is not (cannot be) engaged in any hydrogen bonds, this small displacement probably reflects more favorable surroundings for the new benzene side chain of [B26]Phe. The largest differences between the C-termini of the B-chain in *x* and *X* molecules (1.74 Å) are at the [B28]Pro Cα atoms, but the overall conformations of the B-chains here are very alike.
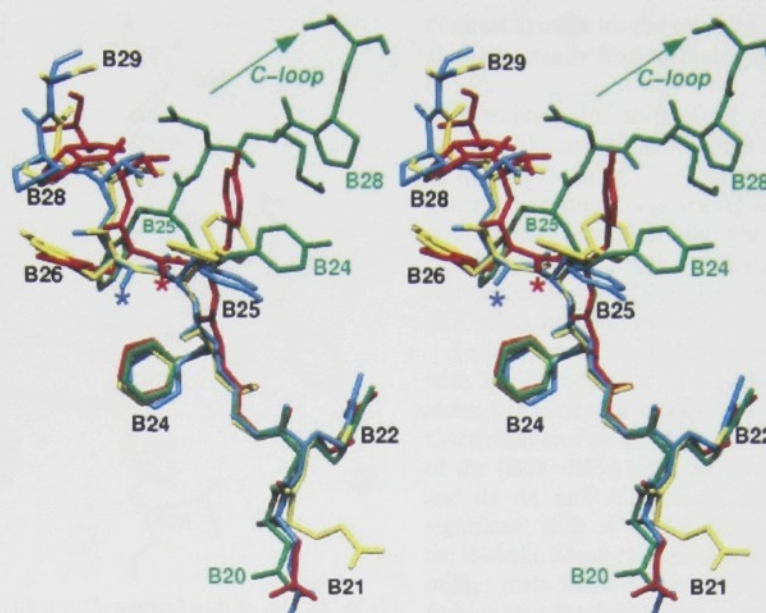
FIGURE 6: Stereoview (divergent) of C-terminal parts of the B-chain of *x* (blue), *z* (red), *X* (yellow), and hIGF-I (green) molecules. Blue and red asterisks are located next to the methyl groups of N[B26]; the green arrow shows the direction of the C-loop in hIGF-I.

The A1 helices in molecules *x* and *X* are almost identical. The A2 helix in molecule *x* shows small displacement toward the A1 helix of molecule *X*, reflected in the 0.63 Å distance between the Cα positions of A19 tyrosines. This movement, resulting also in stronger hydrogen bonding between [A19]Tyr and [A5]Gln (2.27 Å), is more pronounced in the *z* molecule.

Comparison of the [Tyr^{B25}NMePhe^{B26}] z Monomer with the x, X, and Z Molecules. In contrast to the *x* molecule, the *z* molecule of [Tyr^{B25}NMePhe^{B26}] insulin shows some interesting structural changes. First, the [B9−B18] α-helix overlaps less with its *x, X,* and *Z* counterparts (e.g., N-termini [B1]Phe residues are displaced by ca. 2 Å). The changes in the C-terminal part of the B-chain are already initiated at the beginning of its β-strand, and the position of the Cα atom of [B23]Gly deviates by 0.67 Å from that in its WT site. The [B24]Phe Cα atoms are also displaced, by 0.67 Å, while their phenyl rings occupy the same position. The maximum displacement of the main chain from its WT conformation in the *z* molecule occurs at the site of [B25]Tyr; its Cα atom is 1.87 Å from the WT position. As Figure 6 shows, the phenyl ring of [B25]Tyr is located between the alternative positions of this side chain found in molecules *X* and *Z* but opposite in direction to that seen in the *x* molecule.

Because of the large repositioning of its Cα atom, the phenolic ring of [B25]Tyr acquires a conformation almost perpendicular to the plane containing all of the [B25] aromatic rings in other insulin molecules (Figure 6).

The perturbation of the remaining part of the *z* molecule main B-chain continues further toward its C-terminus (by ca. 1 Å), but the side chains of [B26]Phe, [B27]Thr, and [B28]Pro maintain conformations similar to their *x, X,* and *Z* conformations. The methyl group is approximately 0.9 Å from the main chain of [B26]Phe. It is interesting that the overall changes of the C-terminal part of the B-chain and the general route of its displacement are not toward the "outer" side of the insulin molecule but are pushing the [B23−B29] polypeptide more toward the core of the insulin molecule and its A1 and A2 helices (see Figure 4). It is also interesting that these changes are coupled with a significant shift of the A1 helix toward the C-terminus of the B-chain, with the maximum displacement along the axis of the helix at the [A3]Val position (1.15 Å). This rearrangement puts [B29]Lys in close van der Waals contact with the [A1]Gly terminus. These changes are opposite to those resulting from the N-methylation of the [B26] nitrogen atom and the [B25−B26] Phe ↔ Tyr swap. Thus, the overall structural changes not only do not increase the mobility of the B-chain but also bring about an overall contraction and increased compactness of the insulin molecule.

Comparison of the Dimers and Hexamer of [Tyr^{B25}NMe-Phe^{B26}] and Human Insulin. The comparison of the dimers and hexamers is a complicated matter, and for simplicity, it involves here only the deviations in the region of the structure that is most affected by the introduction of the methyl group, i.e., those corresponding to the [B21−B29] polypeptides. The introduction of the methyl group provokes a displacement of these peptides away from the WT insulin position and toward the core of the molecule. However, this displacement is not the same in all of the dimers of the hexamer. As a measure of the displacement, we found that the distance between the Cα atoms at position 25 is greater for L25 and J25 (6.41 Å) followed by B25 and D25 (6.35 Å) and F25 and H25 (6.02 Å) (where B to L is the chain numbering in the final PDB file). In WT insulin, the corresponding distances are 4.38, 4.16, and 4.28 Å, respectively. The increased lengths of the new distances provoke the breaking of the intermolecular β-sheet in our structure (Figure 7). Only one of the four CO−NH hydrogen bonds of this β-sheet is almost preserved in the *z* molecule [CO−N, 3.43 Å (2.99 Å in WT)] at the [B26^z−B24^x] contact. The three remaining main chain−main chain hydrogen bonds are broken: [B26^z−B24^x], 4.09 Å (2.63 Å in the WT); [B24(CO)^z−B26^x], 4.42 Å (2.57 Å); and [B24(NH)^z−B26^x], 4.46 Å (3.00 Å).

Comparison of the [Tyr^{B25}NMePhe^{B26}] x and z Monomers with Human IGF-I. It was assumed that the methylation of the [B25] N peptide bond atom and the swap of [B25]Tyr
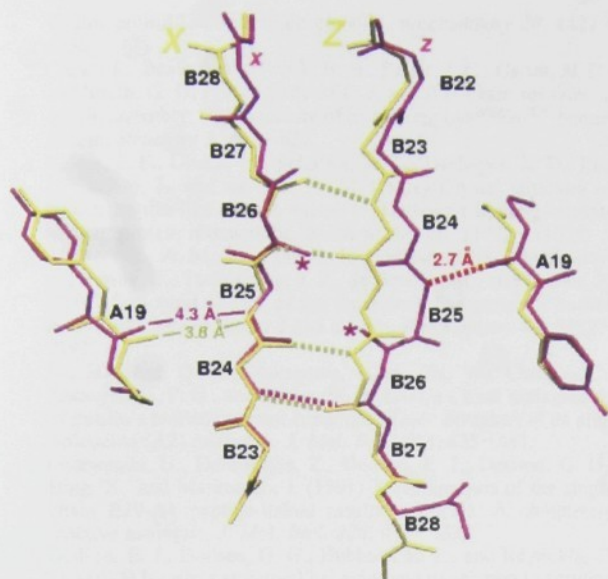
FIGURE 7: Effect of methylation (CH$_3$ marked with an asterisk) of peptide nitrogen atom of residue [B26] on the dimer/monomer hydrogen bond network: wild-type insulin (R-state) (XZ molecules) in yellow and [Tyr^{B25}NMePhe^{B26}] (xz molecules) in magenta. Distances in angstroms indicate the separations between the CO group of [A19] and the NH group of [B25]; the new [A19]−[B25] hydrogen bond in [Tyr^{B25}NMePhe^{B26}] (2.7 Å), typical for the T-state insulin and hIGF-I, is colored red.

and [B26]Phe for the typical IGF sequence of Phe and Tyr might produce a bend or distortion of the main chain at the [B26]{B25} Cα position, as seen in the IGF-I structure associated with its unique C-loop. Although this conformational change does not occur in [Tyr^{B25}NMePhe^{B26}], the position of the [B26]Phe Cα atom in the z molecule is nonetheless halfway between insulin-like and IGF-I-like conformations (Figure 6). One of the consequences of this motion is the introduction of a new hydrogen bond between the CO group of [A19] (one of the key residues for receptor binding) and the NH group of [B25] (2.79 Å), which also occurs in the hIGF-I (2.99 Å) structure but not in R-state molecule x (4.48 Å) or in molecules X and Z (3.81 Å) of insulin (see Figure 7) (the CO^{A19}−NH^{B25} distances in two other dimers of the mutant are 2.95 and 3.1 Å, respectively). This hydrogen bond (length usually in the range of 2.79−3.39 Å) is one of the characteristic structural features typical for T-state insulins and is adopted in [Tyr^{B25}NMePhe^{B26}] R-like hexamers by half of the protein molecules. In hIGF-I, this hydrogen bond is linked with other strong interactions between the C-terminus of the A-chain and the terminal fragment of the B-chain, and contributes significantly to the stabilization of the carboxyl part of the A2 helix, the conformation of the B-chain, and the rigidity of the C-neck region. It is tempting to see the emergence and strengthening (shortening) of the Phe^{B25}−Tyr^{A19} hydrogen bond as one of the potential sources of lower receptor affinity of the [Tyr^{B25}NMePhe^{B26}] mutant due to interference of this interaction with the conformational rearrangement of this region that is required for an effective receptor binding.

Although the B-chain in the z molecule does not bend in an IGF-I-like fashion, it seems that its distortion, due to the methylation of the [B26] N atom, leads to an even more stable conformation of this part of the molecule. This is in contrast to other insulin mutants, where modifications of the B-chain usually lead to higher mobility and local disorder (7−9).

In addition, the shift of the A1 helix in the z molecule parallels the overall shift of the A1 and A2 α-helices seen in hIGF-I. Although the key role of the IGF-I C-loop in generating compactness and rigidity of the region is clear, it is interesting that hIGF-like Phe ↔ Tyr swap at positions [B25] and [B26] and the methylation of the N atom of the residues at position [B26] have resulted in a more compact (IGF-like) structure in the insulin molecule.

To summarize, the [B25]Phe ↔ [B26]Tyr swap, coupled with the introduction of a methyl group on the [B26]N atom, has produced modest but significant changes in the C-terminal part of the B-chain. They result in a movement of the [B23−B29] polypeptide toward the core of insulin and its A1 and A2 helices, which is associated with a significant shift of the A1 helix toward the C-terminus of the B-chain. These changes do not produce the expected bend of the main chain at the [B26] Cα atom in hIGF-I-like fashion, but the fold of the mutant undergoes restructuring in that direction, resulting in an overall contraction of the molecule and the break of the main chain hydrogen bonds within the dimer. These results also support the observation that methylation of the [B26] peptide bond is not crucial for the insulin function and that the chemical character of the side chain [e.g., [B26]Tyr → His (11)] but not of the main chain (after [B26]) is more important for insulin potency.

The combination of chemical and sequence modifications in the [Tyr^{B25}NMePhe^{B26}] mutant illustrates the plasticity of the insulin molecule. In a general sense, it is interesting that the steric pressure introduced into the hexamer by the N−CH$_3$ bond at position [B26] has been accommodated by a similar adjustment within each of the 3-fold related dimers, and the 3-fold symmetry is thus well preserved. However, the adjustment does not reflect the local 2-fold symmetry; it moves only one main chain segment away from the β-sheet structure. With this, the NH^{B25}−CO^{A19} hydrogen bond, a feature of the T-state (the free dimer and two-Zn hexamer) insulin and hIGF-I, is restored. Similar heterogeneous responses of insulin aggregates to nonstandard modifications described here have also been observed in allo-Ile[A2] (40), where chiral perturbation was accommodated differently in T- and R-state insulin cores.

Our results indicate again the sophisticated complexity of the insulin molecule, which despite its very limited size, is capable of the accommodation of significant challenges coming from the chemical modifications of primary and main chain structures despite restraints of the hexamer assembly. The more unambiguous insight into the insulin−IGF-I structural relationship still requires further research, including the addition of more IGF-like sequence (e.g., [B28]Pro ↔ [B29]Lys swap) to the [Tyr^{B25}NMePhe^{B26}] mutant. This may provide some evidence about whether the [B28] ↔ [B29] residue swap, which has been shown to cause the monomerization of insulin (41), will be capable of promoting the hIGF-like conformational distortion of the B-chain of the insulin polypeptide.

REFERENCES

1. Kaarsholm, N. C., Ko, H. C., and Dunn, M. F. (1989) Comparison of solution structural flexibility and zinc-binding domains for

insulin, proinsulin, and miniproinsulin, *Biochemistry 28*, 4427–4435.

2. Ciszak, E., Beals, J. M., Frank, B. H., Baker, J. C., Carter, N. D., and Smith, G. D. (1995) Role of C-terminal B-chain residues in insulin assembly: The structure of hexameric Lys^{B28}Pro^{B29}-human insulin, *Structure 3*, 615–622.

3. Vajdos, F. F., Ultsch, M., Schaffer, M. L., Deshayes, K. D., Liu, J., Skelton, J., and de Vos, A. M. (2001) Crystal structure of human insulin-like growth factor-1: Detergent binding inhibits binding protein interactions, *Biochemistry 40*, 11022–11029.

4. Brzozowski, A. M., Dodson, E. J., Dodson, G. G., Murshudov, G., Verma, C., Turkenburg, J. P., de Bree, F. M., and Dauter, Z. (2002) Structural origins of the functional divergence of human insulin-like growth factor-I and insulin, *Biochemistry 41*, 9389–9397.

5. Xu, B., Hua, Q. X., Nakagawa, S. H., Jia, W., Chu, Y. C., Katsoyannis, P. G., and Weiss, M. A. (2002) Chiral mutagenesis of insulin's hidden receptor-binding surface: Structure of an allo-isoleucine (A2) analogue, *J. Mol. Biol. 316*, 435–441.

6. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Bing, X., and Markussen, J. (1991) X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue, *J. Mol. Biol. 220*, 425–433.

7. Dodson, E. J., Dodson, G. G., Hubbard, R. E., and Reynolds, C. D. (1983) Insulin's structural behavior and its relation to activity, *Biopolymers 22*, 281–291.

8. Leyer, S., Gattner, H. G., Leithauser, M., Brandenburg, D., Wollmer, A., and Hocker, H. (1995) The role of the C-terminus of the insulin B-chain in modulating structural and functional properties of the hormone, *Int. J. Pept. Protein Res. 46*, 397–407.

9. Kurapkat, G., Siedentop, M., Gattner, H. G., Hagelstein, M., Brandenburg, D., Grotzinger, J., and Wollmer, A. (1999) The solution structure of a superpotent B-chain shortened single-replacement insulin analogue, *Protein Sci. 8*, 499–508.

10. Lenz, V., Gattner, H. G., Sievert, D., Wollmer, A., Engels, M., and Hocker, H. (1991) Semisynthetic des-(B27–B30)-insulins with modified B26-tyrosine, *Biol. Chem. Hoppe-Seyler 372*, 495–504.

11. Žáková, L., Barth, T., Jiráček, J., Barthová, J., and Zórad, S. (2004) Shortened insulins analogues: Marked changes in biological activity resulting from replacement of TyrB26 and N-methylation of peptide bonds in the C-terminus of the B-chain, *Biochemistry 43*, 2323–2331.

12. Mirmira, R. G., Nakagawa, S. H., and Tager, H. S. (1991) Importance of the character and configuration of residues of B24, B25, and B26 in insulin receptor interactions, *J. Biol. Chem. 266*, 1428–1436.

13. Nakagawa, S. H., and Tager, H. S. (1987) Role of the COOH-terminal B-chain domain in insulin-receptor interactions. Identification of perturbations involving the insulin main chain, *J. Biol. Chem. 262*, 12054–12058.

14. Cascieri, M. A., Chicchi, G. G., Applebaum, J., Hayes, N. S., Green, B. G., and Bayne, M. L. (1988) Mutants of Human Insulin-Like Growth Factor-I with Reduced Affinity for the Type-1 Insulin-Like Growth-Factor Receptor, *Biochemistry 27*, 3229–3233.

15. Tager, H. S. (1990) Mutant human insulins and insulin structure–function relationships, in *Handbook of Experimental Pharmacology* (Cuatrecasas, P., and Jacobs, S., Eds.) pp 41–64, Springer-Verlag, Berlin.

16. De Meyts, P., and Whittaker, J. (2002) Structural biology of insulin and IGF1 receptors: Implications for drug design, *Nat. Rev. Drug Discovery 1*, 769–783.

17. Adams, T. E., Epa, V. C., Garrett, T. P., and Ward, C. W. (2000) Structure and function of the type 1 insulin-like growth factor receptor, *Cell. Mol. Life Sci. 57*, 1050–1093.

18. Soos, M., Field, C. E., and Siddle, K. (1993) Purified hybrid insulin/insulin-like growth factor-I receptors bind insulin-like growth factor-I, but not insulin, with high affinity, *Biochem. J. 290*, 419–426.

19. Bayne, M. L., Applebaum, J., Chicchi, G. G., Hayes, N. S., Green, B. G., and Cascieri, M. A. (1988) Structural Analogs of Human Insulin-Like Growth Factor-I with Reduced Affinity for Serum

Binding-Proteins and the Type-2 Insulin-Like Growth-Factor Receptor, *J. Biol. Chem. 263*, 6233–6239.

20. Young, J. D., and Carpenter, F. H. (1961) Isolation and characterization of products formed by the action of trypsin on insulin, *J. Biol. Chem. 236*, 743–748.

21. Inouye, K., Watanabe, K., Morihara, K., Tochino, Y., Kanaya, T., Sakibara, S., and Emura, J. (1979) Enzyme-Assisted Semisynthesis of Human Insulin, *J. Am. Chem. Soc. 101*, 751–752.

22. Kubiak, T., and Cowburn, D. (1986) Enzymatic semisynthesis of porcine despentapeptide (B26–30) insulin using unprotected desoctapeptide (B23–30) insulin as a substrate. Model studies, *Int. J. Pept. Protein Res. 27*, 514–521.

23. Svoboda, I., Brandenburg, D., Barth, T., Gattner, H. G., Jiracek, J., Velek, J., Blaha, I., Ubik, K., Kasicka, V., and Pospisek, J. (1994) Semisynthetic insulin analogues modified in positions B24, B25 and B29, *Biol. Chem. Hoppe-Seyler 375*, 373–378.

24. Zorad, S., Golda, V., Fickova, M., Macho, L., Pinterova, L., and Jurcovicova, J. (2002) Terguride treatment attenuated prolactin release and enhanced insulin receptor affinity and GLUT 4 content in obese spontaneously hypertensive female, but not male rats, *Ann. N.Y. Acad. Sci. 967*, 490–499.

25. Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data, *CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, Number 26.

26. Smith, G. D., Ciszak, E., Magrum, L. A., Pangborn, W. A., and Blessing, R. H. (2000) R6 Hexameric Insulin Complexed with m-Cresol or Resorcinol, *Acta Crystallogr. D56*, 1541–1548.

27. Vagin, A., and Teplyakov, A. (1997) MOLREP: An automated program for molecular replacement, *J. Appl. Crystallogr. 30*, 1022–1025.

28. Navaza, J. (1994) AMoRe: An automated package for molecular replacement, *Acta Crystallogr. A50*, 157–163.

29. Murshudow, G. N., Vagin, G. N., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum likelihood method, *Acta Crystallogr. D57*, 122–133.

30. Lamzin, V. S., and Wilson, K. S. (1993) Automated refinement of protein models, *Acta Crystallogr. D49*, 129–149.

31. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760–763.

32. Kraulis, P. J. (1991) MOLSCRIPT: A Program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946–950.

33. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *PROCHECK*: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283–291.

34. Cosmatos, A., Frederigos, N., and Katsoyannis, P. G. (1979) Chemical synthesis of [des(tetrapeptide B27–B30), Tyr(NH2)-26-B] and [des(pentapeptide B26–B30) human (porcine) insulin, *J. Chem. Soc., Perkin Trans. 1 11*, 1311–1317.

35. Fischer, W. H., Saunders, D., Brandenburg, D., Wollmer, A., and Zahn, H. (1985) A shortened insulin with full in vitro potency, *Biol. Chem. Hoppe-Seyler 366*, 521–525.

36. Wollmer, A., Gilge, G., Brandenburg, D., and Gattner, H. G. (1994) An insulin with the native sequence but virtually no activity, *Biol. Chem. Hoppe-Seyler 375*, 219–222.

37. Kurapkat, G., De Wolf, E., Grotzinger, J., and Wollmer, A. (1997) Inactive conformation of insulin despite its wild-type sequence, *Protein Sci. 6*, 580–587.

38. Nakagawa, S. H., Johansen, N. L., Madsen, K., Schwartz, T. W., and Tager, H. S. (1993) Implications of replacing peptide bonds in the COOH-terminal B-chain domain of insulin by the ψ (CH2–NH) linker, *Int. J. Pept. Protein Res. 42*, 578–584.

39. Smith, G. D., and Dodson, G. G. (1992) The structure of rhombohedral R6 insulin hexamer that binds phenol, *Biopolymers 32*, 441–445.

40. Wan, Z.-l., Xu, B., Chu, Y.-C., Katsoyannis, P. G., and Weiss, M. (2003) Crystal structure of allo-IleA2-Insulin, an inactive chiral analogue: Implications for the mechanism of receptor binding, *Biochemistry 42*, 12770–12783.

41. Howey, D. C., Bowsher, R. R., Brunelle, R. L., and Woodworth, J. R. (1994) [Lys(B28), Pro(B29)]-human insulin. A rapidly absorbed analogue of human insulin, *Diabetes 43*, 396–402.

Regular arrangement of periodates bound to lysozyme

Jan Ondráček,[a]* Manfred S. Weiss,[b] Jiří Brynda,[c] Jaroslav Fiala,[d] František Jursík,[e] Pavlína Řezáčová,[c] Lasse B. Jenner[f] and Juraj Sedláček[c]

[a]Department of Recombinant Expression and Structural Biology, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo n. 2, CZ-16637 Praha 6, Czech Republic, [b]EMBL Hamburg Outstation, c/o DESY, Notkestrasse 58, D-22603 Hamburg, Germany, [c]Department of Recombinant Expression and Structural Biology, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo n. 2, CZ 166 37 Praha 6, Czech Republic, [d]New Technologies Research Centre of the West Bohemian University, Univerzitní 22, CZ-30614 Plzeň, Czech Republic, [e]Department of Inorganic Chemistry, Prague Institute of Chemical Technology, Technická 5, CZ-16628 Praha 6, Czech Republic, and [f]Department of Molecular and Structural Biology, University of Aarhus, Gustav Wiedsvej 10C, DK-8000, Aarhus, Denmark

Correspondence e-mail: ondracek@img.cas.cz

The structure of tetragonal hen egg-white lysozyme soaked in a periodate solution has been determined to a resolution of 1.8 Å. Four high-occupancy periodate positions have been identified on the basis of the anomalous signal of the I atoms. The four periodates exhibit a regular rectangular arrangement on the surface of the lysozyme molecule. No similar regular arrangement was found either in lysozyme crystals soaked in other heavy-atom anions or in other structures from the Protein Data Bank. Depending on their position on the surface of the protein, the periodate ions deviate to a varying extent from ideal octahedral geometry.

1. Introduction

Many complexes of proteins with inorganic oxoanions can be found in the Protein Data Bank (PDB). As of January 10, 2005 there are complexes of proteins with sulfates (2960 entries; PDB chemical component ID SO4), nitrates (70 entries; ID NO3), phosphates (804 entries; ID PO4), carbonates (81 entries; ID CO3) *etc.* Only 63 PDB entries containing heavy-atom oxoanions have been found (Table 1): 31 for vanadate (IDs VO3, V4O, V7O, VO4), 21 for tungstate (ID WO4), eight for molybdate (ID MOO), two for perrhenate (ID REO) and one for selenate (ID SE4); no entries were found for bromate, tellurate or periodate. In the Heavy Atom Databank, only sodium perrhenate and tungstates are mentioned.

The chemistry of protein–heavy atom complexes in which the heavy atoms comprise inorganic oxoanions has as yet been little studied. Some of the abovementioned heavy-atom anions have a simple and regular shape: perrhenate and selenate are tetrahedra, bromate is a deformed tetrahedron and tellurate and periodate are octahedra. In addition, vanadate, tungstate and molybdate can polymerize. The structure of the resulting polymeric anion depends strongly on both the local pH and the surface contour of the protein. In the protein crystal, polymeric ions can be formed that may differ from those present in solution. Thus, new inorganic compounds may be obtained inside protein crystals (Ondráček & Mesters, 2002). Similarly, the protein surface may stabilize intermediates and catalyse chemical reactions.

Anomalously scattering heavy atoms can also be exploited for experimental phase determination. Recently, the quick soaking of protein crystals with a solution containing 0.25–1.0 M Br^- or I^- was proposed as a new and universal method for protein phase determination (Dauter & Dauter, 1999, 2001; Dauter *et al.*, 2000, 2001). Bromine has an absorption edge at 0.92 Å and is therefore suitable for MAD analysis (Hendrickson & Ogata, 1997). With $\Delta f'' = 1.28 \, e^-$ at $\lambda = 1.54$ Å, it can be used successfully for structure solution by

doi:10.1107/S090744490501543X **1**

Files: d/av5036/av5036.3d d/av5036/av5036.sgml **AV5036** FA IU-0510/43(6)6 511/20(31)5   D:FA:2005:61:7:0:0–0

research papers

Table 1
Protein structures containing selected heavy-atom oxoanions as found in the PDB as of January 10, 2005.

Oxoanion	PDB code	No. sites	Resolution (Å)
Selenate SeO_4^{2-}	1fga	2	2.2
Perrhenate ReO_4^{2-}	1hnu	2	2.1
	1k4j	7	2.5
Molybdate MoO_4^{2-}	1h9m	8	1.6
	1h9s	2	1.8
	1amf	1	1.7
	1coi	3	2.4
	1gun	8	1.8
	1guo	8	2.5
	1h9j	2	1.8
	1o7l	5	2.8
Tungstate WO_4^{2-}	3kbp	4	3.0
	1e3p	1	2.5
	1dko	1	2.4
	1gug	16	1.6
	1h9r	2	1.9
	1lgp	1	2.0
	1fez	2	3.0
	1fr3	16	1.5
	1h9k	4	1.8
	2ush	4	2.2
	1atg	1	1.2
	1bys	1	2.0
	1ckj	5	2.5
	1cws	1	2.0
	1j9k	2	2.1
	1wod	1	1.8
	1ytw	1	2.4
	2hnq	1	2.8
	1mu7	2	2.0
	1ohd	2	1.9
	1u7p	2	1.9
Vanadate $V_4O_{12}^{4-}$ (V4O)	1l7v	2	3.2
	1uiz	1	1.9
Vanadate $V_4O_{12}^{4-}$ (VO3)	1e59	1	1.3
Vanadate $V_7O_{19}^{3-}$ (V7O)	1dk1	1	2.9
	1rxs	20	2.8
Vanadate VO_4^{3-} (VO4)	1bo6	2	2.1
	1vnc	1	2.1
	3rnt	1	1.8
	1dfl	2	4.2
	1idq	1	2.0
	1idu	1	2.2
	1rpt	1	3.0
	1qi9	2	2.1
	1j9l	2	1.9
	1vnf	1	2.4
	1vne	1	2.2
	1vng	1	2.2
	1vom	1	1.9
	1vni	1	2.2
	1vnh	1	2.1
	1lkx	4	3.0
	1mu9	2	2.1
	1nop	2	2.3
	1ovi	1	2.5
	1rff	2	1.7
	1rfi	2	2.2
	1rg1	2	2.1
	1uzi	3	1.89
	1rg2	2	2.1
	1rgt	2	2.0
	1rgu	2	2.2
	1rho	2	2.3

Table 2
Data-collection and processing parameters.

Values in parentheses correspond to the last resolution shell.

Unit-cell parameters	
a (Å)	77.03
c (Å)	37.12
Wavelength	1.54178
Crystal-to-detector distance (mm)	100
Resolution (Å)	99.0–1.83 (1.88–1.83)
No. images	2×118
Exposure time per image (s)	900
Φ rotation per image (°)	1
Completeness (%)	99.3 (99.3)
No. unique reflections	10310
No. unique reflections, $F > 4\sigma(F)$	9837
R_{merge} (%)	7.3 (15.6)
R_{anom} (%)	6.0 (8.3)
$I/\sigma(I)$	6.6 (4.4)
Redundancy	15.9 (15.7)

by the use of I_3^- (Evans & Bricogne, 2002). The study has been expanded to include lanthanides (Nagem et al., 2003) and uranyl compounds (Djinovic Carugo et al., 2005) and may also be applicable to heavy-atom oxoanions.

This study is a contribution to establishing a relation between the size, shape and chemical properties of the various heavy-atom oxoanions and their binding modes to the proteins. In addition, it may be possible to exploit the anomalous signal for the determination of new protein structures. The structure and binding mode of periodate anion is the first presented result.

2. Materials and methods

2.1. Crystal soaking

Tetragonal crystals of lysozyme (lysozyme chloride from egg white, dried; Roanal, Budapest, Hungary) of size 0.1–0.2 mm grown by the hanging-drop method from 50 mM sodium acetate pH 4.5, 1–1.2 M NaCl were used for this study. Lysozyme crystals were soaked at room temperature for 5 min in solutions made up from one part of aqueous saturated NaIO$_4$ and 11, five and two parts of 50 mM acetate pH 4.5 and then for 15 min in a solution made up from two parts aqueous saturated NaIO$_4$ and one part 50% PEG 2000. Before flash-cooling, crystals were briefly immersed in paraffin oil.

2.2. Data collection and refinement

Diffraction data were collected on a Cu rotating-anode generator (FR591, Nonius, The Netherlands) operated at 50 kV and 80 mA equipped with an cold nitrogen-stream generator (Oxford Cryosystems, UK) at 150 K and a MAR imaging-plate detector (MAR345, X-ray Research, Germany). The inverse-beam technique was used for data collection. The raw data were integrated using *DENZO* (Otwinowski & Minor, 1997) and scaled using *SCALA* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) using the SCALA-Det scaling protocol as described in Mueller-Dieckmann et al. (2004). Further data reduction and refinement of the structure were carried out with other programs

SIRAS (Dauter & Dauter, 1999). Iodine, with $\Delta f'' = 6.84$ e$^-$ at $\lambda = 1.54$ Å, has been used successfully for structure solution by SAD (**Chen et al., 1991**). The use of I$^-$ was extended recently

Table 3
HipHop refinement parameters.

The values given are for reflections with $F > 4\sigma(F)$.

Refinement	Input	1	2	3	4	5	6	7	8	9	10	Mean (e.s.d.)	Average structure
R, full data (%)	17.35	15.63	15.79	15.79	15.57	15.78	15.79	15.73	15.68	15.77	15.67	15.72 (7)	15.31
R (90% set) (%)	16.89	15.38	15.41	15.48	15.28	15.52	15.43	15.45	15.42	15.54	15.36	15.43 (7)	15.03
R_{free} (10% set) (%)	24.4	21.8	23.2	22.4	21.9	22.1	22.6	22.2	22.1	22.1	21.9	22.2 (4)	21.0
wR^2, full data	43.53	39.08	39.49	39.53	39.00	39.47	39.56	39.57	39.30	39.40	39.25	39.37 (19)	—
S, full data	2.98	2.653	2.676	2.679	2.646	2.671	2.677	2.677	2.661	2.671	2.659	2.668 (11)	—
Waters	155	185	176	176	185	173	172	170	175	178	178	177 (5)	—
$R_{free} - R$ (%)	7.55	6.4	7.8	6.9	6.6	6.6	7.2	6.8	6.7	6.6	6.5	6.8 (4)	6.0

Table 4
Bond lengths (Å) in periodate I_A and possible hydrogen bonds (Å).

The bond lengths in the periodate I_A correspond to the structure $[I(HO)_4O_2]^-$.

Refinement	1	2	3	4	5	6	7	8	9	10	Mean
I—O1	1.40	1.58	1.58	1.57	1.65	1.52	1.62	1.52	1.48	1.52	1.54 (7)
I—O2	1.96	2.01	2.04	1.94	1.85	1.96	2.07	1.97	2.01	1.89	1.97 (6)
I—O3	1.91	1.89	1.87	1.74	1.76	1.88	1.89	1.77	1.84	1.80	1.84 (6)
I—O4	1.78	1.87	1.84	1.77	1.83	1.79	1.75	1.80	1.83	1.86	1.81 (4)
I—O5	1.66	1.67	1.65	1.60	1.51	1.74	1.70	1.68	1.74	1.65	1.66 (6)
I—O6	1.80	1.88	1.88	1.83	1.86	1.80	1.88	1.85	1.86	1.82	1.85 (3)
O1···Asn93 OD1	3.39	3.26	3.28	3.48	3.41	3.31	3.09	3.17	3.31	3.19	3.29 (11)
O1···Asn93 ND2	3.60	3.37	3.32	3.48	3.19	3.48	3.35	3.28	3.45	3.34	3.39 (11)
O2···His15 ND1	2.87	2.68	2.61	2.72	2.59	2.76	2.56	2.74	2.78	2.91	2.72 (11)
O2···Arg128 NH2^i	2.70	2.87	3.01	2.90	3.29	2.79	2.69	2.50	2.52	2.52	2.78 (24)
O3···His15 ND1	3.27	3.27	2.98	3.11	3.31	2.63	2.63	2.83	2.65	2.68	2.94 (27)
O4···Lys96 NZ	3.45	3.45	3.39	3.47	3.48	3.47	3.36	3.42	3.47	3.50	3.45 (4)
O4···Arg128 NH2^i	2.52	2.46	2.58	2.56	2.58	2.63	2.57	2.55	2.59	2.52	2.56 (4)
O4···His15 O	3.66	3.57	3.56	3.64	3.55	3.79	3.73	3.59	3.68	3.61	3.64 (7)
O6···His15 O	3.40	3.36	3.32	3.22	3.27	3.30	3.38	3.35	3.34	3.34	3.33 (5)
O6···His15 ND1	3.27	3.27	3.33	3.41	3.29	3.39	3.39	3.35	3.33	3.32	3.34 (5)
O6···Asn93 ND2	2.83	2.75	2.81	2.86	2.90	2.88	2.80	2.81	2.81	2.78	2.82 (4)
O6···Lys96 NZ	3.60	3.61	3.46	3.46	3.54	3.38	3.24	3.44	3.43	3.57	3.46 (11)
O1···water 1173	—	—	3.30	—	2.71	—	—	3.30	—	—	3.10 (28)
O1···water 1179	—	—	—	—	—	—	—	—	—	3.37	—
O3···water 1179	—	—	—	—	—	—	—	—	—	3.21	—

† Symmetry code: (i) $-y, -x, \frac{1}{2} - z$.

from the *CCP*4 suite. Unit-cell parameters, data-collection details and statistics are given in Table 2.

The PDB entry 1hel (Wilson *et al.*, 1992) without solvent molecules was used as the starting model for structure determination. For preliminary refinement, the conjugate-gradient algorithm (CGLS) was used as implemented in the program *SHELX*97 (Sheldrick & Schneider, 1997). After determination of the whole structure of lysozyme and the positions of periodates and nearly all waters, the full-matrix (least-squares) refinement was also used. The parameters of this refinement are given in Table 3 in the 'input' column.

Although the reliability parameters for the preliminary refinement appear to be satisfactory for the resolution, it was impossible to determine the individual iodine–oxygen distances in the periodate structures. The structure was therefore re-refined using the *HipHop* refinement method (Ondráček, 2005; Ondráček *et al.*, 2005; http://www.img.cas.cz/hiphop). A paper with further details of *HipHop* is currently in preparation (Ondráček *et al.*, 2005).

Briefly, in the 'Hip' step 150 maxima in the difference Fourier synthesis were filled with water molecules provided they fulfilled the geometrical criteria for water molecules. The

'Hop' steps were run five times and waters not fulfilling the ball shape or criteria of density greater than 0.09, 0.13, 0.18, 0.23 and 0.25 e⁻ Å⁻³ were gradually removed. Every Hip and every Hop step was followed by 1000 cycles of CGLS in *SHELXH*.

During the preliminary *HipHop* cycles the structure was inspected and some side chains were reoriented manually. In this way the static disorder in Ile55, Asn77 and Lys97 and the different side-chain orientations of Leu83 and Val109 were found. After stabilization of the model, the final ten cycles were run and statistics of *HipHop* refinement were calculated.

Table 3 shows the final refinement results of the final ten *HipHop* cycles and their mean values. In the column 'Average structure' crystallographic (vector structure factors) averages of all ten structures are given. In the multiconformer PDB entry 1hc0 the single structures are labelled by different MODEL numbers.

2.3. Structure comparisons

Three-dimensional structure superpositions were carried out using the program *LSQKAB* (Collaborative Computational Project, Number 4, 1994). Since the *HipHop* refinement protocol yields an ensemble of ten models describing the refined structure, the values for the superpositions given are average values over all ten structures.

3. Results

3.1. The overall structure of lysozyme

The global refinement parameters (Table 3) are comparable in quality with the parameters for other published structures of native hen egg-white tetragonal lysozyme (Table 9). The three-dimensional structure of lysozyme is very similar to other published hen egg-white lysozyme structures. A comparison of the ten models generated by *HipHop* refine-

research papers

Table 5
Bond lengths (Å) in periodate I_B and possible hydrogen bonds (Å).

The bond lengths in the periodate I_B correspond to the structure [I(HO)$_5$O].

Refinement	1	2	3	4	5	6	7	8	9	10	Mean
I—O1	1.59	1.61	1.56	1.60	1.64	1.63	1.66	1.72	1.77	1.82	1.66 (8)
I—O2	1.81	1.89	1.88	1.85	1.88	1.84	1.83	1.84	1.87	1.84	1.85 (2)
I—O3	1.92	1.98	2.01	1.96	1.89	1.97	1.90	1.95	1.92	1.96	1.95 (4)
I—O4	1.99	1.95	2.00	1.98	1.90	2.07	2.02	1.98	1.99	2.02	1.99 (4)
I—O5	1.97	2.02	1.99	2.01	2.02	1.97	2.06	2.07	2.04	2.06	2.02 (3)
I—O6	1.88	2.05	2.10	1.96	1.91	1.84	1.90	1.99	1.93	1.98	1.95 (4)
O1···Glu7 OE2[i]	2.80	2.75	2.72	2.75	2.76	2.75	2.65	2.58	2.59	2.48	2.68 (10)
O1···Arg14 NH2	2.49	2.49	2.49	2.49	2.49	2.50	2.49	2.49	2.50	2.50	2.49 (1)
O2···Ile88 N	2.83	2.84	2.86	2.91	2.90	2.91	2.92	2.98	2.94	3.00	2.91 (5)
O2···Ser86 O	3.39	3.32	3.35	3.36	3.32	3.33	3.33	3.30	3.36	3.33	3.34 (3)
O3···His5 NE2	3.70	3.65	3.59	3.72	3.75	3.74	3.74	3.71	3.73	3.62	3.70 (5)
O4···Glu7 OE2[i]	3.54	3.44	3.38	3.46	3.54	3.58	3.51	3.48	3.49	3.42	3.48 (6)
O4···Asp87 OD1	3.69	3.69	3.64	3.65	3.69	3.64	3.59	3.64	3.61	3.61	3.65 (3)
O4···I_F O2	2.41	2.38	2.53	2.49	2.56	2.37	2.46	2.51	2.44	2.49	2.46 (6)
O5···Arg14 NH2	3.20	3.29	3.32	3.22	3.27	3.24	3.35	3.37	3.31	3.26	3.28 (5)
O5···His15 NE2	2.79	2.84	2.90	2.83	2.78	2.82	2.87	2.79	2.77	2.70	2.81 (5)
O5···Asp87 OD1	2.96	2.78	2.69	2.80	2.83	2.93	2.87	2.89	2.90	2.90	2.86 (8)
O5···Ile88 N	3.48	3.47	3.49	3.54	3.50	3.57	3.43	3.51	3.61	3.60	3.52 (6)
O1···water 1080	2.75	2.84	2.85	2.84	2.92	2.93	2.99	3.02	3.02	3.16	2.93 (11)
O1···water 1064[†]	3.45	3.54	3.49	3.51	3.49	3.47	3.51	3.52	3.51	3.51	3.50 (2)
O2···water 1181	2.90	2.78	2.79	2.73	2.66	2.73	2.78	2.75	2.73	2.72	2.76 (6)
O3···water 1080	2.35	2.33	2.32	2.30	2.53	2.49	2.60	2.53	2.58	2.61	2.46 (12)
O6···water 1064	2.83	2.47	2.53	2.63	2.82	2.86	2.70	2.63	2.71	2.65	2.68 (12)
O6···water 1064[†]	3.09	3.29	3.29	3.19	3.16	3.09	3.17	3.11	3.26	3.17	3.18 (7)
O6···water 1181	2.67	2.59	2.51	2.54	2.56	2.56	2.60	2.78	2.86	2.74	2.64 (11)

† Symmetry code: (i) $-y, -x, \frac{1}{2} - z$.

Table 6
Bond lengths (Å) in periodate_C and possible hydrogen bonds (Å).

The bond lengths in the periodate correspond to the structure [I(HO)$_3$O$_3$]$^{2-}$.

Refinement	1	2	3	4	5	6	7	8	9	10	Mean
I—O1	1.68	1.60	1.64	1.64	1.64	1.61	1.59	1.56	1.65	1.56	1.62 (4)
I—O2	1.56	1.40	1.45	1.46	1.49	1.51	1.48	1.38	1.45	1.49	1.47 (5)
I—O3	2.03	2.03	2.04	2.06	2.04	2.07	2.02	2.06	2.07	2.07	2.05 (2)
I—O4	1.92	1.87	1.97	1.85	1.97	1.93	1.83	1.98	1.92	1.95	1.92 (5)
I—O5	1.66	1.71	1.64	1.58	1.55	1.41	1.71	1.67	1.46	1.67	1.61 (10)
I—O6	2.07	2.13	2.02	2.08	2.00	2.11	2.19	2.08	2.05	2.08	2.08 (5)
O1···Arg68 NH1[ii]	3.17	3.17	3.14	3.23	3.17	3.12	3.16	3.16	3.10	3.14	3.16 (3)
O3···Glu35 O	2.65	2.66	2.72	2.66	2.68	2.62	2.58	2.57	2.56	2.56	2.63 (5)
O3···Asn44 ND2	2.76	2.74	2.76	2.70	2.71	2.72	2.79	2.75	2.75	2.71	2.74 (3)
O5···Asn44 OD1	3.30	3.24	3.36	3.45	3.50	3.54	3.30	3.31	3.51	3.39	3.39 (10)
O5···Asn44 ND2	2.88	2.87	2.89	2.89	3.04	2.94	2.87	2.85	2.91	2.98	2.91 (6)
O5···Arg45 NH2[ii]	3.00	2.96	2.99	3.00	3.06	3.19	2.89	2.99	3.11	2.99	3.02 (8)
O5···Arg68 NH1[ii]	3.61	3.51	3.43	3.63	3.56	3.69	3.58	3.50	3.54	3.57	3.56 (7)
O6···Ser36 O	3.53	3.37	3.37	3.32	3.39	3.26	3.13	3.21	3.27	3.22	3.31 (11)
O6···Asn37 N	3.45	3.33	3.40	3.37	3.38	3.31	3.24	3.34	3.34	3.34	3.35 (5)
O1···water 1024	2.66	2.81	2.65	2.63	2.68	2.77	2.78	2.82	2.71	2.94	2.75 (9)
O2···water 1039	3.18	3.30	3.17	3.23	3.37	3.29	3.55	3.40	3.40	3.34	3.32 (11)
O2···water 1094	3.57	3.56	3.49	3.57	3.30	3.29	3.45	3.69	3.35	3.40	3.47 (13)
O3···water 1094	3.94	3.93	3.81	3.90	3.64	3.54	3.55	3.94	3.60	3.55	3.74 (17)
O5···water 1067	2.60	2.67	2.71	2.68	2.63	2.73	2.50	2.59	2.80	2.54	2.65 (9)
O6···water 1039	2.75	2.83	2.85	2.98	3.00	3.12	3.33	3.19	3.10	3.11	3.03 (17)
O6···water 1024	3.69	3.59	3.48	3.49	3.54	3.50	3.41	3.49	3.41	3.48	3.51 (8)

† Symmetry code: (ii) $y, x, 1 - z$.

ment yielded r.m.s. deviations in the range of the expected coordinate error of the structure (Table 9).

The static disorder of Arg14 and the dynamic disorder of Trp62 will be discussed in detail in the paper describing the basis of *HipHop* refinement (Ondráček *et al.*, 2005). Other cases of disorder were observed for Lys97, Asn103, Arg21 and Arg125. These and the side-chain disorder observed for Ile55,

Asn77 and Lys97 were not present in PDBentry 1hel, which was used as the starting model for refinement. Furthermore, PDB entry 1hel significantly differs from our refined model in the side-chain orientations of Lys13, Arg21, Asn59, Arg61, Trp62, Ser86, Lys97, Asn103 and Arg125.

In PDB entry 1dpx, the disorder observed for Arg14 (atoms NE, CZ, NH1 and NH2) differs from that found in our final ten models (only the CG and CD atoms are disordered; the remaining atom positions are again similar). The disorder for Lys1 NZ observed in 1dpx was not found in our models. The disorder of Asn77 determined in 1dpx is practically identical to that found in the first steps of *HipHop* refinement.

No disordered residues were observed in 1jis and 1lz8. In 1jis, the orientation of Arg14 is identical to that found in one set of our models, the orientations of Ile55, Asn77 and Lys97 are similar and the orientations of Arg21, Asn103 and Arg125 are different. In 1lz8, the orientations of Ile55 and Asn77 are identical with that found in one set of our models, the orientations of Arg21, Lys97 and Arg125 are similar and the orientations of Arg14 and Asn103 are again different.

More interesting is the comparison with the high-resolution structure 1iee (resolution 0.94 Å). The disorder observed for Ile55, Asn77 and Lys97 found and refined during *HipHop* refinement in our model is almost identical to that found in 1iee. The side-chain orientations for Asn59, Ser86 and Lys97 in our models are identical to one of the multiple side-chain conformations in 1iee, while the side-chain orientations for Lys13, Arg14, Arg21, Arg61, Trp62, Asn103, Arg125 and Arg128 differ between the two structures.

With respect to the water molecules, a total of 291 water molecules were found in different positions in the ten models. The average number of waters per model is 177 (5) (Table 3). The distribution of water molecules is as follows: 112 water molecules occurred in all ten models, 18 waters in nine, ten in eight, 11 in seven, nine in six, 12 in five, 12 in four, 17 in three, 26 in two

Table 7
Bond lengths (Å) in periodate_D and possible hydrogen bonds (Å).

The bond lengths in the periodate correspond to the structure $[I(HO)_6]^-$.

Refinement	1	2	3	4	5	6	7	8	9	10	Mean
I—O1	1.84	1.80	1.79	1.77	1.77	1.78	1.82	1.77	1.78	1.70	1.78 (4)
I—O2	1.99	2.01	1.94	1.83	1.96	1.88	1.96	1.84	1.93	1.82	1.92 (7)
I—O3	1.93	1.89	1.86	1.91	1.82	1.83	1.83	1.75	1.82	1.84	1.85 (5)
I—O4	2.01	2.03	2.04	2.06	2.09	2.08	2.07	2.00	2.04	2.02	2.04 (3)
I—O5	1.80	1.88	1.85	1.86	1.92	1.77	1.95	1.90	1.91	1.91	1.88 (5)
I—O6	1.74	1.91	1.89	1.84	1.91	1.85	1.89	1.95	1.96	1.86	1.88 (6)
O1···Glu35 OE1	2.59	2.62	2.68	2.68	2.65	2.65	2.62	2.67	2.69	2.76	2.66 (4)
O1···Asp52 OD2	3.14	3.15	3.12	3.17	3.19	3.19	3.19	3.19	3.14	3.13	3.16 (3)
O2···Ala107 O	3.41	3.34	3.39	3.44	3.34	3.39	3.35	3.42	3.31	3.50	3.39 (5)
O2···Val109 N	2.64	2.64	2.69	2.79	2.74	2.75	2.71	2.75	2.69	2.74	2.71 (5)
O3···Glu35 OE1	3.11	3.14	3.09	3.14	3.05	3.10	3.10	3.05	3.07	3.12	3.10 (3)
O3···Glu35 OE2	2.30	2.30	2.30	2.30	2.29	2.30	2.29	2.30	2.30	2.32	2.30 (1)
O3···Ala109 N	2.90	2.84	2.89	2.79	3.00	2.91	2.96	3.06	2.99	2.95	2.93 (8)
O3···Ala110 N	2.89	2.92	2.98	2.91	2.98	2.98	2.97	3.07	3.01	2.98	2.97 (5)
O4···Asp52 OD2	2.32	2.30	2.30	2.29	2.29	2.30	2.30	2.30	2.29	2.30	2.30 (1)
O5···Glu35 OE1	3.30	3.36	3.41	3.34	3.42	3.32	3.41	3.37	3.40	3.45	3.38 (5)
O5···Glu35 OE2	3.11	3.10	3.19	3.13	3.20	3.12	3.18	3.17	3.07	3.14	3.14 (4)
O5···Gln57 O	3.20	3.10	3.10	3.07	3.05	3.19	3.03	3.03	3.05	3.05	3.09 (6)
O5···Asp52 OD2	3.63	3.61	3.53	3.53	3.55	3.54	3.52	3.51	3.50	3.50	3.54 (4)
O1···water 1007	2.99	3.08	3.02	3.01	3.07	3.01	3.00	3.10	3.10	3.14	3.05 (5)
O1···water 1146	3.16	—	—	—	—	—	—	—	—	—	3.16
O1···water 1168	—	—	—	—	—	3.01	3.08	3.13	3.20	3.21	3.13 (7)
O2···water 1169	—	2.51	2.63	2.66	2.49	—	—	—	—	—	2.57 (7)
O2···water 1184	—	—	—	—	—	2.63	2.61	2.64	2.61	2.63	2.62 (1)
O3···water 1146	3.41	—	—	—	—	—	—	—	—	—	3.41
O4···water 1059	2.29	2.30	2.29	2.29	2.29	2.29	2.29	2.29	2.29	2.29	2.29 (1)
O4···water 1181	—	—	—	—	—	2.25	—	—	—	—	2.25
O4···water 1184	—	—	—	—	—	3.06	3.12	3.11	3.06	3.18	3.11 (4)
O4···water 1169	—	2.99	3.10	3.12	3.00	—	—	—	—	—	3.05 (6)
O4···water 1169	—	—	—	—	—	3.27	3.22	3.21	3.33	3.29	3.26 (4)
O6···water 1146	3.33	—	—	—	—	—	—	—	—	—	3.33
O6···water 1168	—	2.89	2.67	2.78	—	—	—	—	—	—	2.78 (9)
O6···water 1169	—	—	—	—	—	2.89	2.94	2.89	3.09	2.98	2.96 (7)
O6···water 1168	—	—	—	—	—	3.37	3.33	3.33	3.40	3.21	3.33 (6)

and 64 in only one. The stable water-molecule content and the limited water capacity is in agreement with our other results.

3.2. The structure of the periodate ions

Nine possible periodate positions, four maxima at the sulfur positions of Cys6, Cys76, Cys80 and Cys115 and one for a chloride corresponding to the chloride 205 in 1lz8 were found on the basis of the anomalous diffraction (Fig. 1). Four periodates exhibited higher occupancy than 0.5. These were termed periodates I_A, I_B, I_C and I_D. After *HipHop* refinement and on the basis of the static evaluation it was possible to study their structures in detail.

The bond lengths in periodate I_A (occupancy 0.6) and the possible hydrogen bonds to protein atoms and water molecules are shown in Fig. 2(*a*) and the distances are given in Table 4. The values show that two of the iodine–oxygen bonds seem to be significantly shorter then the other four. Thus, this structure corresponds to the formula $[I(OH)_4O_2]^-$, with two short I—O bonds and four longer I—OH bonds. One of the negative charges of the periodate is in the neighbourhood of the positively charged Arg128 of a symmetry-related molecule described by the symmetry operator $-y, -x, 1/2 - z$.

The bond lengths of periodate I_B (occupancy 0.7) and possible hydrogen bonds are summarized in Table 5 and shown in Fig. 2(*b*). Only one bond length in this periodate differs statistically from the others. Thus, this periodate corresponds to the formula $[I(OH)_5O]$. This formula is also supported by its neighbourhood: the two negatively charged side chains Glu7 (position $-y, -x, 1/2 - z$) and Asp87 and one positively charged side chain Arg14 are close to I_B. The interaction of Arg14 with I_B may explain the static disorder observed for Arg14. The side chain of Arg14 exhibits two different orientations with relative frequencies of 6:4 in the ten final models.

The geometry and interactions of periodate I_C (occupancy 0.5) are given in Table 6 and Fig. 2(*c*). The bond lengths of this periodate correspond to the formula $[I(OH)_3O_3]^{2-}$. The two negative charges of this periodate are stabilized by Arg45 and Arg68 (both in position $y, z, 1 - z$).

The structure of periodate I_D (occupancy 0.6) is described in Table 7 and Fig. 2(*d*). The bond lengths correspond to the formula $[I(OH)_6]^+$. The
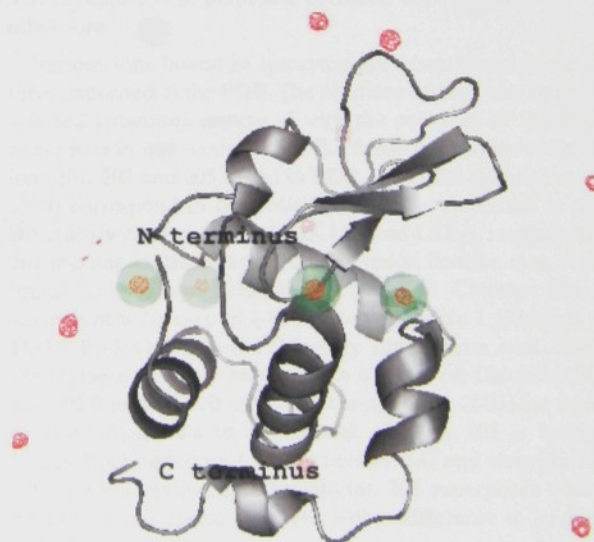

Figure 1
Anomalous difference electron-density map (contoured at 3 standard deviations above the mean value) and regular geometric arrangement of periodates (green spheres). The additional maxima are from symmetry-related molecules. The figure was produced using the program *PyMol* (DeLano, 2002).

research papers

Refinement	1	2	3	4	5	6	7	8	9	10	Mean
Distances ($\mathring{A}$)											
I_A$\cdots$I_B	10.00	10.01	10.03	10.01	10.01	9.99	9.99	9.96	9.95	9.95	9.99 (3)
I_B$\cdots$I_C	20.50	20.50	20.50	20.48	20.50	20.48	20.49	20.50	20.49	20.50	20.49 (1)
I_C$\cdots$I_D	10.41	10.41	10.40	10.39	10.39	10.39	10.41	10.42	10.42	10.40	10.40 (1)
I_D$\cdots$I_A	22.94	22.91	22.91	22.91	22.81	22.92	22.92	22.92	22.90	22.91	22.91 (3)
Angles (°)											
I_A$\cdots$I_B$\cdots$I_C	107.94	107.78	107.73	107.78	107.63	107.85	107.90	107.94	107.90	107.81	107.83 (9)
I_B$\cdots$I_C$\cdots$I_D	86.37	86.34	86.40	86.47	86.36	86.43	86.32	86.32	86.33	86.37	86.37 (5)
I_C$\cdots$I_D$\cdots$I_A	91.43	91.50	91.50	91.43	91.54	91.40	91.47	91.37	91.31	91.35	91.43 (7)
I_D$\cdots$I_A$\cdots$I_B	74.24	74.36	74.35	74.30	74.44	74.29	74.27	74.35	74.44	74.45	74.35 (7)
Distances from plane ($\mathring{A}$)											
I_A	0.061	0.071	0.057	0.077	0.071	0.072	0.074	0.080	0.074	0.070	0.071 (7)
I_B	0.069	0.079	0.064	0.086	0.079	0.081	0.083	0.090	0.082	0.078	0.079 (7)
I_C	0.064	0.073	0.059	0.080	0.073	0.075	0.077	0.083	0.076	0.072	0.073 (7)
I_D	0.056	0.065	0.052	0.071	0.065	0.070	0.068	0.073	0.067	0.064	0.065 (6)
R.m.s. deviation of fitted atoms	0.063	0.072	0.058	0.079	0.072	0.073	0.076	0.082	0.075	0.071	0.072 (7)

positive charge of the periodate is stabilized by the presence of Glu35 and Asp52 in its neighbourhood.

The I—O bond lengths given in Tables 4–7 are in good agreement with the published mean value of 1.803 $\mathring{A}$ for the distance $I^{7+}-O^{2-}$ (Wilson & Prince, 1999).

3.3. Exploitation of anomalous signal for structure solution

Our data measured were more then good enough for the determination of periodate positions, the verification of these positions on the basis of the anomalous difference map and the determination of the periodate oxygen positions. However, attempts to use the anomalous signal for phase determination were unsuccessful. No easily interpretable map was obtained by the SAD method using the programs *MLPHARE* and *DM* (Collaborative Computational Project, Number 4, 1994). However, it is planned to reperform the experiment more carefully and to enhance the anomalous signal by carrying out data collection at longer wavelengths (Mueller-Dieckmann *et al.*, 2004, 2005). This will be the subject of a separate study.

3.4. Geometric arrangement of the four periodates with high occupancy

Selected distances, angles and the plane through the I atoms of the periodates I_A, I_B, I_C and I_D are described in Table 8. The I atoms of I_A, I_B, I_C and I_D form a deformed rectangle *ABCD* (Fig. 1). The deformation occurs because of the deviation of I_A of about 3 $\mathring{A}$ in the direction *DA*. The angles *BCD* and *CDA* may be considered to be regular, the sides *AB* and *CD* as equivalent and the side *BC* as double this value.

The regular arrangement of the periodate ions is in good agreement with the unit-cell parameters of several minerals and may help to explain the phenomenon of epitaxy of lysozyme on the minerals lepidolite and fluorite (McPherson & Shlichta, 1988). The only commonly accepted necessary condition for epitaxy of one compound on another is the presence of the same *d* value in both compounds. In the case of proteins, the *d* value of one compound can be changed by

the distance of possible binding points, represented by periodate positions. The average value of 10.23 $\mathring{A}$ obtained from *AB*, *BD* and *CD* used for this evaluation is in an excellent agreement with the unit-cell parameters $c = 10.13$ $\mathring{A}$ in lepidolite $C2/m$ (Bernard & Rost, 1992), $c = 20.20$ $\mathring{A}$ in lepidolite $C2/c$ or $a = 5.20$ $\mathring{A}$ in lepidolite $P3_112$. The agreement is evident for fluorite $P3_121$ with $a = 4.913$ $\mathring{A}$ or for fluorite $P4_12_12$ with $a = 4.97$ $\mathring{A}$.

The epitaxy of proteins on minerals has been described (McPherson & Shlichta, 1988) for proteins such as concanavalin B, catalase and canavalin, but a similar analysis for these proteins is so far lacking.

3.5. Comparison of periodate positions with positions of other ions

Various ions bound to lysozyme are often found in structures deposited in the PDB. The positions of chlorides in three selected structures compared with the positions of the periodate ions in our model yields the following results: chloride ions 201, 202 and 205 found in PDB entry 1lz8 (Dauter *et al.*, 1999) correspond to the positions of water molecules in our structure, with deviations of 1.05, 1.22 and 1.57 $\mathring{A}$, respectively. No maxima in the anomalous difference Fourier map were found for chlorines 203, 206, 207 and 208. Chlorine 204 is situated at a distance of 2.45 $\mathring{A}$ from periodate I_B bound to His15. Both chlorides in PDB entry 1dpx (Weiss *et al.*, 2000) are replaced by water molecules in our crystal. Chlorides 201 and 202 found in PDB entry 1iee (Sauter et al., 2001) are again replaced by waters in our crystal, chloride 203 is located 1.13 $\mathring{A}$ from periodate I_H (occupancy 0.4) and chloride 204 1.41 $\mathring{A}$ from periodate I_B. Chloride 205 superposes with a chloride in our crystal structure, with a difference in position of 1.67 $\mathring{A}$.

The ion positions in two other lysozyme structures determined under xenon and krypton pressure (PDB entries 1c10 and 1qtk; Prangé *et al.*, 1998) were also compared with our model (Table 9). In 1c10, the chlorides 160, 161 and 164 correspond to water molecules in our model, chloride 162

Table 9
Published structures of native hen egg-white tetragonal lysozyme and average values of r.m.s. deviations from our models (Å).

PDB code	Structure	Resolution (Å)	R (%)	R_{free} (%)	R.m.s.d. (Å)	Reference
1dpx	177 waters, 2 chlorides	1.65	18.7	24.6	0.223 (5)	Weiss et al. (2000)
1iee	233 waters, 1 sodium, 5 chlorides	0.94	12.3	15.1	0.271 (6)	Sauter et al. (2001)
1jis	141 waters	1.90	19.0	23.4	0.313 (3)	Datta et al. (2001)
1lz8	224 waters, 1 sodium, 8 chlorides	1.53	22	31	0.343 (3)	Dauter et al. (1999)
1hel	319 waters	1.70	15.2	Not given	0.279 (4)	Wilson et al. (1992)
1c10	68 waters, 1 sodium, 6 chlorides and 3 xenons	2.03	17.2	22	–	Prangé et al. (1998)
1qtk	77 waters, 1 sodium, 3 chlorides and 1 krypton	2.03	17.3	21.6	–	Prangé et al. (1998)
1gwd	103 waters, 1 sodium, 7 chlorides and 5 iodides	1.77	16.8	21.1	–	Evans & Bricogne (2002)

approximately corresponds to the position of the low-occupancy periodate I_H (distance 2.09 Å), chloride 163 roughly superimposes with periodate I_B (distance 2.69 Å) and

chloride 165 is 1.54 Å from the position of the chloride in our model. In 1qtk, chlorides 160 and 161 correspond to water molecules in our model and the position of chloride 162 corresponds to the position of the low-occupancy periodate I_H (distance 2.01 Å). As might be expected, all xenon and krypton positions are far from periodate positions.

The periodate positions were also compared with ions found in the tri-iodide derivative of tetragonal lysozyme (Evans & Bricogne, 2002; Table 9). Tri-iodide is formed by iodides A1130-A1147-A1130, with the I—I distance being 2.72 Å (A1147 is placed on a special position). Tri-iodide, iodides A1131 and A1133 and chlorides A1137, A1139, A1142 and A1143 are distant from any periodate positions in our model. The iodide A1132 position approximately corresponds to the position of periodate I_E (occupancy 0.21); here, the side chain of the nearby Asp18 exhibits a different orientation. The chlorides A1138 and A1140 practically superimpose with periodate I_H and periodate I_B, with differences in position of 2.01 and 2.41 Å, respectively. The chloride A1141 corresponds to the chloride ion in our model, with a distance difference of 1.39 Å.

The structure of lysozyme in complex with 2-methyl-2,4-pentanediol and 2-amino-2-hydroxymethyl-propane-1,3-diol (Weiss et al., 2000) and in complex with tri-N-acetyl-chitotriose (Cheetham et al., 1992) as well as the structural effects of other monovalent anions such as para-toluenesulfonate, NO_3^-, SCN^- and I^- on polymorphic lysozyme crystals (Vaney et al., 2001) have also been studied. No similarities between the binding of these ions and periodates in tetragonal HEW lysozyme were found.

The characteristics of periodate-binding positions are determined by its octahedral shape, size and charge flexibility. Periodate can change its charge and I—O bond lengths and thus adjust itself to the lysozyme surface.
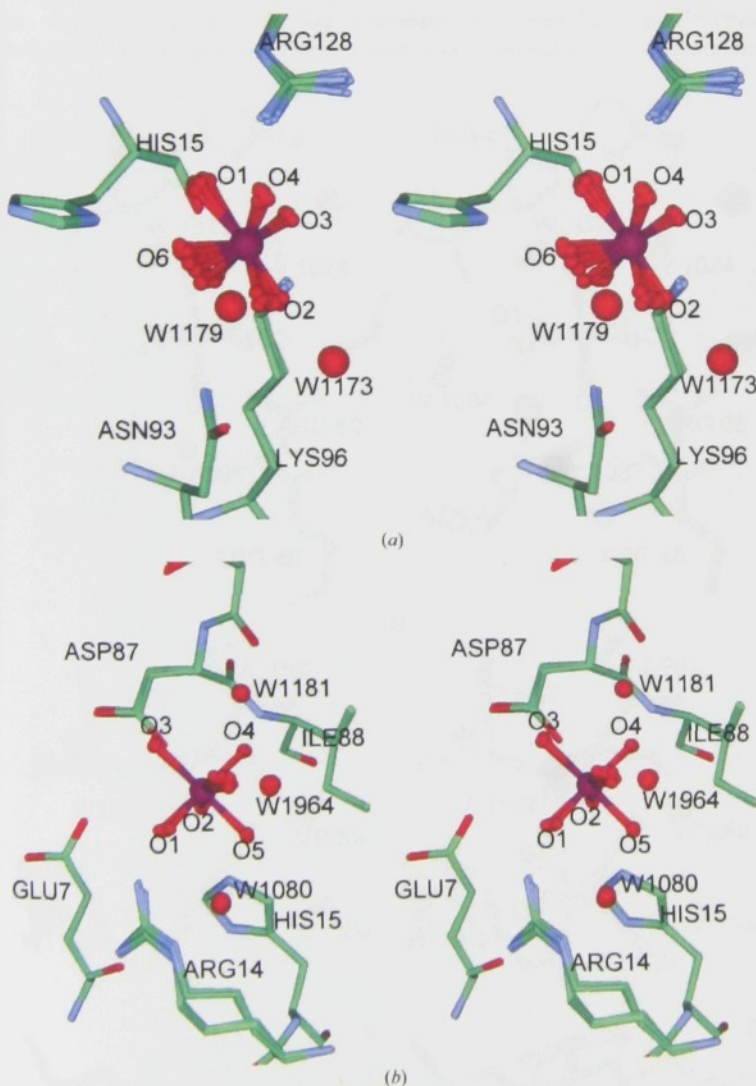
Figure 2
Stereoscopic views of the four periodate ions bound to lysozyme exhibiting ocupancy >0.5. The figure was prepared using the programme *Viewer Lite* 4.2. (*a*) Binding position of periodate I_A. (*b*) Binding position of periodate I_B. (*c*) Binding position of periodate I_C. (*d*) Binding position of periodate I_D.

Surprisingly, periodate oxidizes neither cysteines or methionines nor any other parts of the lysozyme molecule. No similar structural flexibility of inorganic oxoanions such as bromate, selenate, tellurate *etc.* was found within lysozyme crystals (Ondráček, unpublished results).

4. Conclusions

The soaking of tetragonal lysozyme crystals in periodate solutions led to the binding of periodate in four highly occupied positions. Sstandard refinement of the structure did not yield the positions of periodate O atoms. Thus, the structures of the periodate ions could not be determined exactly. A standard single model is not capable of describing every significant maximum in the difference Fourier map. However, the oxygen positions in all four periodate ions could be determined using the novel *HipHop* refinement procedure.

The structures of the four bound periodates differ and depend to some extent on the binding position of the periodate on the surface of the lysozyme molecule. The periodate anion is able to accept or release H^+ and balance a partial charge within lysozyme crystal. Thus, the structure of the periodate depends on the local pH in the binding position. Three of four periodates are bound to two neighbouring symmetry-related lysozyme molecules.

Four periodates form a regular rectangular arrangement on the surface of lysozyme. A similar arrangement was not found after soaking tetragonal lysozyme in other heavy-atom oxoanion solutions such as tellurate, bromate, selenate, vanadate, tungstate or perrhenate or in the structure of other complexes of heavy-atom oxoanions with proteins listed in Table 1. The reason can be found in the octahedral symmetry of periodate and its structural flexibility as presented in this article. There is a relationship between the dimensions of this rectangle and the unit-cell parameters of some inorganic minerals that might enable epitaxy of lysozyme on these minerals.

(c)

(d)

References

Bernard, J. H. & Rost, R. (1992). *Encyklopedický Přehled Minerálů*. Prague: Academia CSAV.
Cheetham, J. C., Artymiuk, P. J. & Phillips, D. C. (1992). *J. Mol. Biol.* **224**, 613–628.
Please provide reference – Chen *et al.* (1991).
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Datta, S., Biswal, B. K. & Vijayan, M. (2001). *Acta Cryst.* D**57**, 1614–1620.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.
Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
Dauter, Z. & Dauter, M. (2001). *Structure*, **9**, 21–26.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Dauter, Z., Li, M. & Wlodawer, A. (2001). *Acta Cryst.* D**57**, 239–249.
Djinović Carugo, K., Helliwell, J. R., Stuhrmann, H. & Weiss, M. S. (2005). In the press.
Evans, G. & Bricogne, G. (2002). *Acta Cryst.* D**58**, 976–991.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
McPherson, A. & Shlichta, P. (1988). *Science*, **239**, 385–387.
Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2005). In the press.
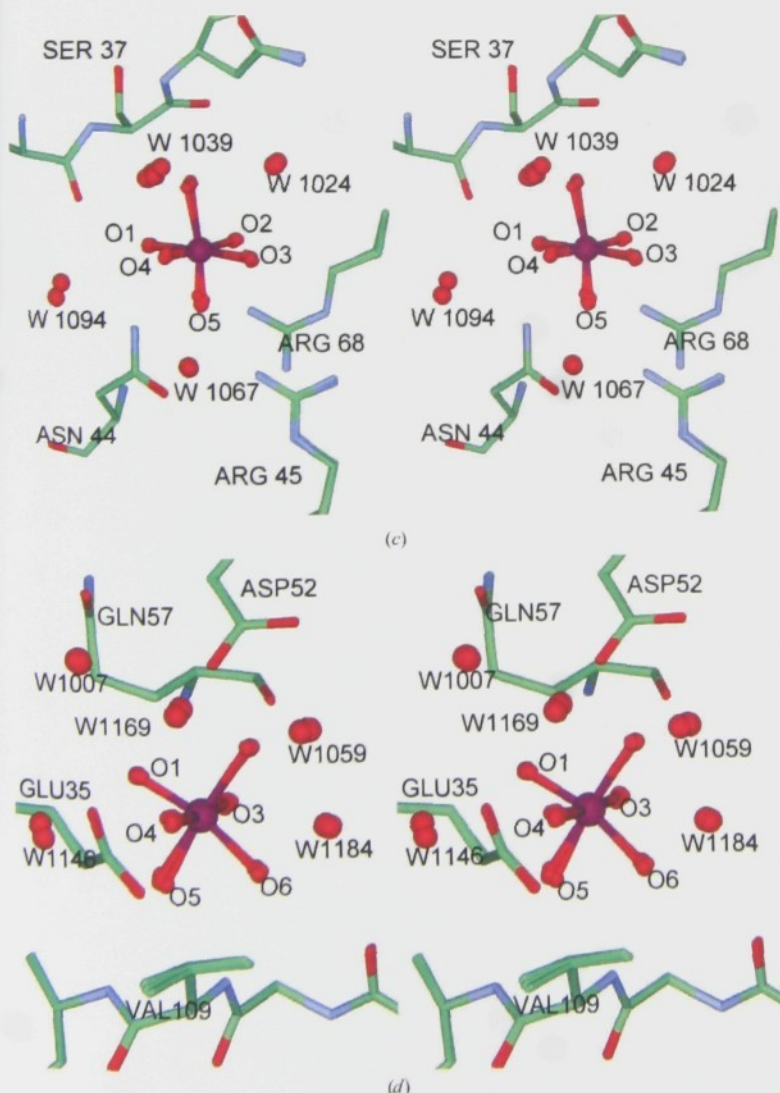
Mueller-Dieckmann, C., Polentarutti, M., Djinovic Carugo, K., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2004). *Acta Cryst.* D**60**, 28–38.

Nagem, R. A., Polikarpov, I. & Dauter, Z. (2003). *Methods Enzymol.* **374**, 120–137.

Ondráček, J. & Mesters, J. (2002). Hamburger Synchrotronstrahlungslabor HASYLAB Am Deutschen Elektronen-Synchrotron DESY, Annual Report.

Not cited in text – Ondráček, J. & Mesters, J. (2004). Hamburger Synchrotronstrahlungslabor HASYLAB Am Deutschen Elektronen-Synchrotron DESY, Annual Report.

Ondráček, J. (2005). *HHRR: Program For Multi-Model HipHop REFMAC5 Protein Structure Refinement.* Institute of Molecular Genetics, Prague, Czech Republic.

Ondráček, J., Weiss, M. S. & Mesters, J. R. (2005). In preparation.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Prangé, T., Schiltz, M., Pernot, L., Colloc'h, N., Longhi, S., Bourguet, W. & Fourme, R. (1998). *Proteins*, **30**, 61–73.

Sauter, C., Otalora, F., Gavira, J.-A., Vidal, O., Giegé, R. & Garcia-Ruiz, J.-M. (2001). *Acta Cryst.* D**57**, 1119–1126.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Vaney, M. C., Broutin, I., Retailleau, P., Douangamath, A., Lafont, S., Hamiaux, C., Prangé, T., Ducruix, A. & Riès-Kautt, M. (2001). *Acta Cryst.* D**57**, 929–940.

Weiss, M. S., Palm, G. J. & Hilgenfeld, R. (2000). *Acta Cryst.* D**56**, 952–958.

Wilson, A. J. C. & Prince, E. (1999). Editors. *International Tables for Crystallography*, Vol. C, p. 777. Dordrecht: Kluwer Academic Publishers.

Wilson, K. P., Malcolm, B. A. & Matthews, B. W. (1992). *J. Biol. Chem.* **267**, 10842–10849.

From nonpeptide toward noncarbon protease inhibitors: Metallacarboranes as specific and potent inhibitors of HIV protease

Petr Cigler*[†‡], Milan Kožíšek*[‡], Pavlína Řezáčová[‡§¶], Jiří Brynda[§], Zbyszek Otwinowski[¶], Jana Pokorná*, Jaromír Plešek[∥], Bohumír Grüner[∥], Lucie Dolečková-Marešová*, Martin Máša*, Juraj Sedláček[§], Jochen Bodem**, Hans-Georg Kräusslich**, Vladimír Král[†,††], and Jan Konvalinka*[‡‡§§]

Institutes of *Organic Chemistry and Biochemistry and [§]Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic; [†]Department of Analytical Chemistry, Institute of Chemical Technology, Technicka 5, 166 28 Prague 6, Czech Republic; [¶]Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-8816; [∥]Institute of Inorganic Chemistry, Academy of Sciences of the Czech Republic, Area of Research Institutes 1001, 250 68 Husinec-Rez near Prague, Czech Republic; **Department of Virology, University of Heidelberg, Im Neuenheimer Feld 324, D-69120 Heidelberg, Germany; and [‡‡]Department of Biochemistry, Faculty of Natural Science, Charles University, Hlavova 2030, Prague 2, 128 43 Czech Republic

HIV protease (PR) represents a prime target for rational drug design, and protease inhibitors (PI) are powerful antiviral drugs. Most of the current PIs are pseudopeptide compounds with limited bioavailability and stability, and their use is compromised by high costs, side effects, and development of resistant strains. In our search for novel PI structures, we have identified a group of inorganic compounds, icosahedral metallacarboranes, as candidates for a novel class of nonpeptidic PIs. Here, we report the potent, specific, and selective competitive inhibition of HIV PR by substituted metallacarboranes. The most active compound, sodium hydrogen butylimino bis-8,8-[5-(3-oxa-pentoxy)-3-cobalt bis(1,2-dicarbollide)]di-ate, exhibited a K_i value of 2.2 nM and a submicromolar EC_{50} in antiviral tests, showed no toxicity in tissue culture, weakly inhibited human cathepsin D and pepsin, and was inactive against trypsin, papain, and amylase. The structure of the parent cobalt bis(1,2-dicarbollide) in complex with HIV PR was determined at 2.15 Å resolution by protein crystallography and represents the first carborane–protein complex structure determined. It shows the following mode of PR inhibition: two molecules of the parent compound bind to the hydrophobic pockets in the flap-proximal region of the S3 and S3′ subsites of PR. We suggest, therefore, that these compounds block flap closure in addition to filling the corresponding binding pockets as conventional PIs. This type of binding and inhibition, chemical and biological stability, low toxicity, and the possibility to introduce various modifications make boron clusters attractive pharmacophores for potent and specific enzyme inhibition.

rational drug design | aspartic proteases | carboranes | x-ray structure analysis | virostatics

HIV protease (PR) is responsible for cleaving viral polyprotein precursors into mature, functional viral enzymes and structural proteins. This process, called viral maturation, is required for the progeny virion to become replication competent and infectious. Chemical inhibition or inactivation by mutation of PR blocks the infectivity of the virus (1). PR has thus become a prime target for therapeutic intervention in AIDS. Academic and industrial research has led to the rapid development of eight effective inhibitors that are currently in clinical use, with several others still in the pipeline (for review, see refs. 2 and 3).

Despite the considerable success of rational drug design, the need for effective PR inhibitors (PIs) is still urgent. Most of the current PIs are pseudopeptide compounds with limited bioavailability and stability. Moreover, their clinical use is compromised by high production cost, various side effects, and rapid development of resistant viral strains (4). Therefore, there is a continuing need for

the design of new PIs with an emphasis on broad specificity against PI-resistant HIV mutants (5, 6).

Molecular modeling and/or random testing of compound libraries revealed several PR inhibitors with unexpected structures. Most of the first-generation PIs were pseudopeptides. Some recent compounds involve nonpeptidic structures, such as cyclic ureas, sulfonamides, etc. (2). However, even inorganic compounds, Nb-containing polyoxometalates, specifically inhibit HIV PR with submicromolar EC_{50} values in tissue cultures (7). In this case, the inhibitors were shown to be noncompetitive, and a model suggested binding to the cationic pocket on the outer surface of the flaps. There is also an evidence for compounds with unexpected chemistry capable to target the active site of the enzyme. The HIV PR-binding cleft was shown to accommodate C_{60} fullerenes, hydrophobic and electrophilic spheric compounds, and some fullerene derivatives are indeed weak inhibitors of HIV PR (8–11).

We searched for other types of unconventional chemical structures that would fit into the PR-binding cleft, would be biologically stable, and would enable facile chemical modification. When screening a number of structural motifs, we identified a group of inorganic compounds, icosahedral boranes, carboranes, and namely 12-vertex metal bis(dicarbollides), as promising frameworks for a novel class of nonpeptide PIs. These boron/carbon clusters are polyhedra based on a three-dimensional skeleton with triangular facets.

Boron-containing polyhedral compounds have been intensively studied because of their use in boron neutron capture therapy (12) and in radioimaging (13, 14). From the structural point of view, the variety of known structural types of boranes, heteroboranes, and metallaboranes represents an interesting counterpart to organic compounds, especially to aromates. With the icosahedral cage being only slightly larger than the phenyl ring rotation envelope, carboranes were used as stable hydrophobic pharmacophores (e.g., refs. 15–17). There is little information on the use of carboranes as

---

Table 1. Structures and activities of metallacarborane inhibitors

Compound no.	Structure	Molecular weight of anion	K_i, nM	EC_{50}, μM
1		323.74	66 ± 30	6
2		339.74	$6,800 \pm 500$	20
3		427.85	$2,500 \pm 400$	6
4		624.09	20 ± 5	13
5		837.71	4.9 ± 2.1	3
6		893.82	2.2 ± 1.2	0.25

All compounds were prepared as their sodium salts. Color coding: orange, BH groups; black, CH groups; blue, Co atom. For details see *Materials and Methods*.

enzyme inhibitors. Few examples reported in the literature involve benzolactams bearing dicarba-*closo*-dodecaborane (16, 18) or carborane substitution of the phenyl ring in the phenyl-phthalimido-imid, yielding a tumor necrosis factor-α modulator with the activity comparable with the parental compound (19). Porphyrins substituted with dicarba-*closo*-dodecaboranes were found as inhibitors of HIV PR, with IC_{50} in the submicromolar range (20).

Our main attention has been focused on ionic metal bis(dicarbollides) that consist of two dicarbollide subclusters sandwiching the central metal atom. In metal bis(dicarbollides), the equal 11-vertex dicarbollide subclusters are connected by a *commo* metal vertex, forming two 12-vertex metal dicarbollide subclusters. These *closo* 26-electron compounds with "peanut-like 12-vertex geometry" were described as early as 1965 by Hawthorne *et al.* (21) and thus form basic stones in metallacarborane chemistry.

Among other transition metal metallacarboranes (22), cobalt bis(1,2-dicarbollide) ion (23) shows certain unique features: synthetic availability, wide possibilities of *exo*-skeletal modifications, high stability, charge delocalization, low nucleophilicity, strong acidity of conjugated acids, and high hydrophobicity. These properties are reflected in unique solution properties and ion-pairing behavior of this ion, which in turn led to its known applications in extraction chemistry (24, 25) and in the development of lowest coordinating anions (26) and compounds for radioimaging (14). However, the metal bis(dicarbollides) have never been considered as biologically active compounds or pharmacophores.

In this article, we report the potent, specific, and selective inhibition of HIV PR by parental and substituted metallacarbo-

ranes, namely cobalt bis(1,2-dicarbollides). We provide evidence for the mechanism of action of these compounds, show their antiviral activity in tissue cultures, analyze their binding toward the enzyme by x-ray crystallography, and show the potential of this class of compounds to become a novel pharmacophore for enzyme inhibition.

Materials and Methods

Chemical Syntheses. Compound **1** (Table 1) was converted to

Table 2. Data collection and refinement statistics for crystallographic structure determination

Space group	C2
Unit cell	$a = 85.3$ Å, $b = 67.2$ Å, $c = 42.5$ Å, $\beta = 95.0°$
Data collection resolution, Å	52.7–2.14
Completeness, %	99.2 (99.2)*
Average $I/\sigma(I)$	9.3 (3.0)*
R_{merge},[†] %	5.6 (22.8)*
Refinement resolution, Å	30–2.15
R factor[‡]/R_{free} factor,[§] %	17.59/23.25
No. of atoms (protein/water/others)	1,535/202/46
rms deviation bond length/angles, Å/°	0.009/1.92

*Values in parentheses refer to the highest resolution shell 2.19–2.14.
[†]$R_{merge} = \Sigma_{hkl} | I - \langle I \rangle | \Sigma_{hkl} I$.
[‡]$R = \Sigma | F_{calc} - F_{obs} | / \Sigma | F_{obs} |$.
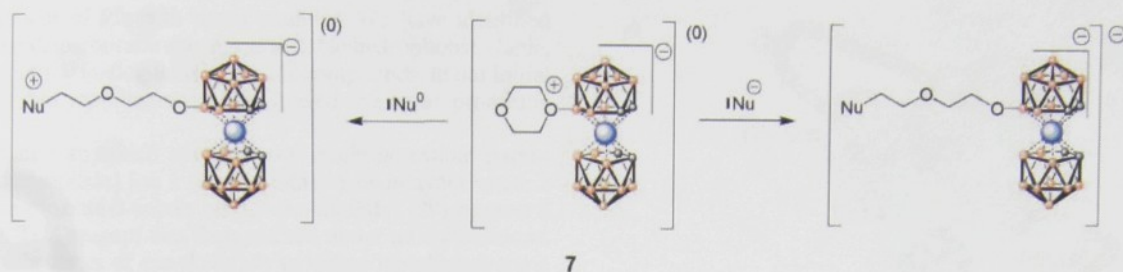[§]The R_{free} as defined in ref. 45 was calculated for 5% of reflections.

7

sodium salt from the commercially available cesium salt (Katchem, Rez u Prahy, Czech Republic) by using the extraction procedure described in ref. 27. Compound **2** was obtained by direct hydroxylation of **1** by using warm diluted sulfuric acid according to procedures in ref. 28. The starting dioxanate intermediate was prepared as described in ref. 29. The synthesis of compounds **3–6** is described in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site.

Enzymes. The expression, refolding, and purification of HIV-1 PR, HIV-2 PR, and the HIV-1 PR variant (Q7K, L33I, L63I) bearing three mutations that minimize the autoproteolytic cleavage (30) were performed as described (31). PRs from mouse intracisternal A particles (MIA14 PR) (32), human cathepsin D, and pepsin were prepared as described (33, 34). Porcine α-amylase, bovine trypsin, and papain were purchased from Sigma.

Inhibition Assays. *Inhibition of HIV PRs.* The IC_{50} and K_i values were determined by spectrophotometric assay with the chromogenic substrate KARVNleNphEANle-NH₂ as described (31). The inhibition constants were estimated by using a competitive inhibition equation according to ref. 35. The mechanisms of inhibition were derived from initial reaction rates versus concentrations of substrate in the presence of various concentrations of inhibitor by using a Lineweaver–Burk plot.

Inhibition of MIA14 protease. A spectrophotometric assay was used to determine inhibition characteristics by using chromogenic substrate DSAYNphVVS as described (32).

Inhibition of human cathepsin, pepsin, trypsin, papain, and α-amylase. For experimental details, see *Supporting Materials and Methods*.

Testing of Antiviral Activities in Tissue Cultures. Antiviral activity was analyzed by using PM-1 cells infected with HIV-1 strain NL4-3 modified from a published procedure (36). PM-1 cells were infected

by coculture and washed 4 h after infection, and compounds **1–6** or the solvent DMSO, respectively, was added after the wash. Newly produced virus was harvested at 48 h postinfection and cleared by brief centrifugation, and infectious titer was determined on TZM cells, which express β-galactosidase from a Tat-responsive promoter. Viral titers and standard deviations are derived from three independent experiments.

Crystallization, Data Collection, and Structure Solution. The complex for crystallization was prepared by mixing HIV-1 PR (Q7K, L33I, L63I) with 3.7-fold molar excess of compound **1** dissolved in DMSO and concentrated by ultrafiltration to final concentration of 7.5 mg/ml. The crystals were grown by hanging drop vapor diffusion technique at 19°C by using 0.1 M Tris·HCl (pH 8.5) and 2.0 M ammonium dihydrogen phosphate as the precipitating solution. Diffraction data were collected at 100 K by using synchrotron radiation of wavelength 0.8 Å [X13 beamline, Deutsches Elektronen-Synchrotron (DESY) Hamburg, Germany] and were processed by using the HKL 2000 software package (37). The HIV PR structure was solved by molecular replacement by using protein coordinates from Protein Data Bank (PDB) structure 1NH0 (38). The structure solution and refinement were performed by using the CCP4 program suite.

Crystal parameters and data collection statistics are summarized in Table 2. Atomic coordinates and structure factors have been deposited to PDB: code 1ZTZ. The details of structure determination are found in Table 4, which is published as supporting information on the PNAS web site.

Results and Discussion

Inhibitor Design and Synthesis. Most of the HIV PIs currently used in clinics are pseudopeptide or peptide mimetics based on a limited number of structural building blocks. Our intention was to identify novel core structures and thus expand chemical space available for
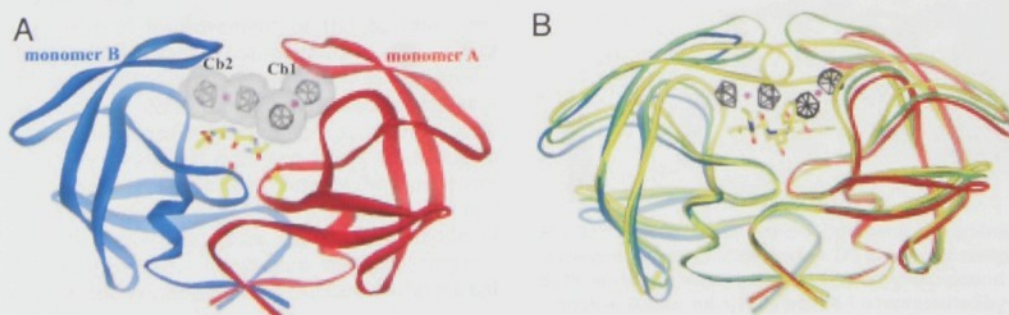
the development of PIs with novel qualities. We have identified 12-vertex metallacarborane clusters as suitable hydrophobic, stable, nontoxic structural analogues of aromatic compounds. In our initial tests, substituted metallacarboranes showed the most promising results.

Our synthetic approach started from easily accessible parent cobalt bis(dicarbollide) ion **1** to yield either 8-hydroxyderivative **2** (28) or the 8-dioxane-3-cobalt bis(1,2-dicarbollide) (29) reagent **7** (see Fig. 1). This reagent was then reacted under mild conditions, giving rise to a series of *exo*-skeletally modified metallacarborane cluster anions **3–6** for testing. For synthesis of compounds **3–6**, ring cleavage reaction of **7** with *O*- and *N*-nucleophiles was used (see Fig. 1), thus producing ionic species if the ring opener is an anion, and betain-type zwitterions if the base is not charged.

The ring opening procedure of **7** zwitterion has already become a widely applicable method for attachment of the cobalt bis(dicarbollide) moiety to various organic substances (39–41). However, compounds **5** and **6** represent examples of zwitterionic-anionic structures containing two cobalt bis(1,2-dicarbollide) subunits bonded via flexible organic spacer chain.

Inhibition Constants and Antiviral Activities. Compounds **1–6** were tested as potential inhibitors of HIV PR *in vitro* and in tissue cultures. The corresponding inhibition constants (K_i values) and antiviral activities (EC_{50} values) are summarized in Table 1. All compounds exhibit classical competitive binding (data shown in *Supporting Materials and Methods* and Fig. 4, which is published as supporting information on the PNAS web site). This kinetic analysis suggests that tested cobalt bis(dicarbollide) competes with the peptide substrate and, therefore, binds to the active cleft of the enzyme. This suggestion has been confirmed by x-ray analysis of the complex of HIV PR with compound **1** (see below). Parent compound **1** shows tight inhibition *in vitro* and micromolar antiviral potency. Derivatization of compound **1** by hydroxyl and 2-(2-hydroxyethoxy) ethoxy groups yielded compounds **2** and **3**, exhibiting much weaker activity *in vitro* and comparable antiviral activities in tissue cultures. Simple visual inspection of the size of compounds **1–3** (Table 1) in comparison with the volume of the closed form of the HIV PR active cleft led to the notion that these compounds would not have sufficient contacts with the corresponding substrate-binding clefts. The solvent accessible area of compound **1** is more then two times lower when compared with a representative conventional pseudopeptide PI, lopinavir (LPV). However, the x-ray structure analysis solved this apparent contradiction, showing that two inhibitor units are needed for the efficient binding to the PR active cleft (see Figs. 2 and 3). Because the relative molecular weight of compound **1** is one of the lowest ever reported to inhibit HIV PR, it provides enough room for further improvement by means of structure-activity analyses, and therefore it was selected as the lead compound of our series of metallacarborane inhibitors of HIV PR.

Approximately a 100-fold improvement of the K_i value was achieved by enlarging the side chain of compound **3** in position 8 of the cage by the addition of a 1,2-diphenyl-2-hydroxy-ethoxy group, yielding compound **4** with 20 nM K_i. The binding of the compound was further improved four times by designing a symmetric compound **5** and, even further, by alkylating the secondary amino group with a butyl moiety (compound **6**), which represents the most active inhibitor of the compound series, with a K_i value of 2.2 nM and submicromolar antiviral activity in tissue culture. The mode of binding of this compound and its interaction with the enzyme-binding pockets could be inferred from the structure of the parent compound **1**.

It is striking that compound **6** showed an EC_{50} for inhibition of HIV-1 in tissue culture of 250 nM, which was ≈10-fold better than that observed for the structurally very similar compound **5**. In contrast, the K_i value for PR inhibition *in vitro* exhibited only a 2-fold difference. This result indicates that subtle differences in
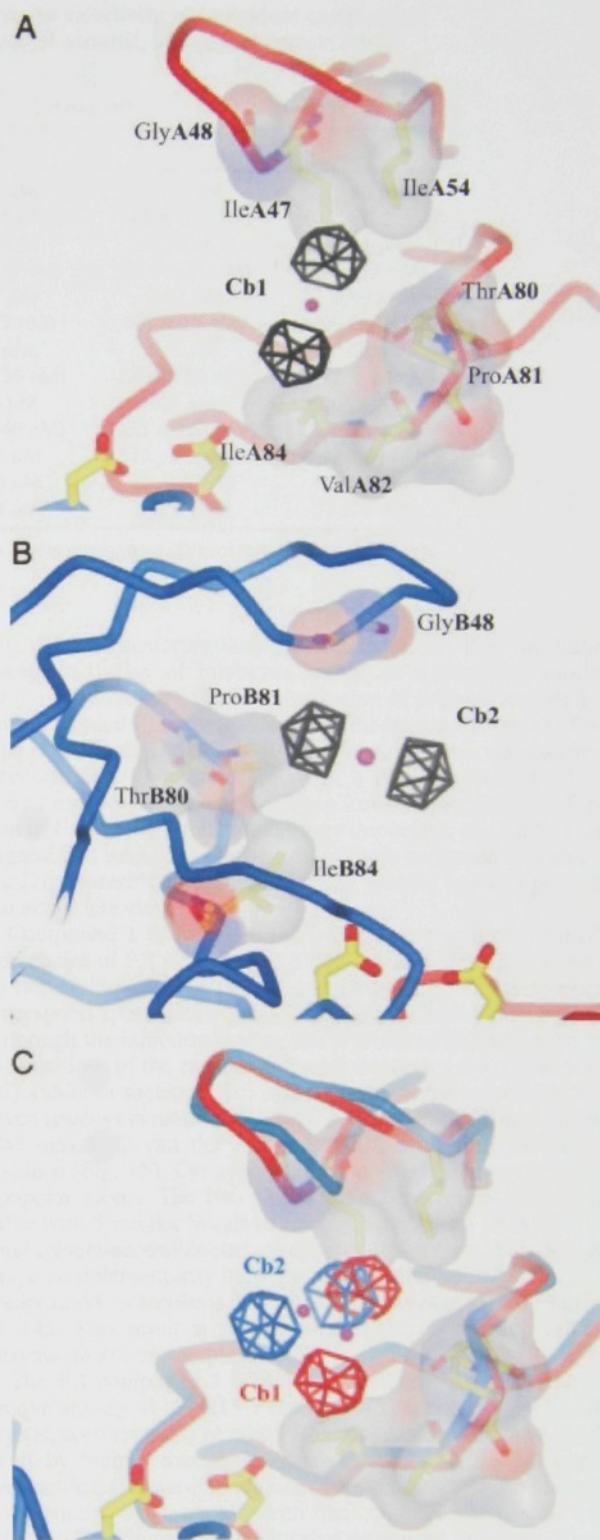

Fig. 3. Interactions of compound **1** with the amino acid residues in the corresponding PR-binding pocket. (A) Binding of compound **1** molecule Cb1 by PR monomer A (red tube). (B) Binding of compound **1** molecule Cb2 by PR monomer B (blue tube). Compound **1** is represented by a stick model in gray, with cobalt shown as a magenta sphere. PR residues in contact with compound **1** are represented by stick models, and their solvent-accessible surfaces are colored by atom charge (blue, positive; red, negative). (C) Superposition of the two compound **1**-binding modes. The color scheme and representation for PR is the same as in A and B, and atoms in compound **1** are colored with the color of the interacting PR chain.

Table 3. The IC$_{50}$ values that demonstrate specificity and selectivity of individual compounds as analyzed with other retroviral PRs, representatives of aspartic, serine and cystein PRs, and amylase

Enzyme	Compound			
	1	4	5	6
WT HIV-1 PR	1.4 μM	0.13 μM	0.14 μM	0.10 μM
	(66 $\pm$ 30 nM)	(20 $\pm$ 5 nM)	(4.9 $\pm$ 2.1 nM)	(2.2 $\pm$ 1.2 nM)
WT HIV-2 PR	1.5 μM	0.76 μM	0.35 μM	0.31 μM
	(220 $\pm$ 34 nM)	(140 $\pm$ 8 nM)	(110 $\pm$ 17 nM)	(39 $\pm$ 1 nM)
MIA14 PR	1.0 μM	0.21 μM	0.63 μM	0.59 μM
	(85 $\pm$ 17 nM)	(22 $\pm$ 7 nM)	(60 $\pm$ 22 nM)	(85 $\pm$ 4 nM)
Human cathepsin D	2.1 μM	1.3 μM	1.9 μM	0.50 μM
	(1,100 $\pm$ 100 nM)	(670 $\pm$ 30 nM)	(960 $\pm$ 30 nM)	(250 $\pm$ 30 nM)
Pepsin	1.5 μM	0.86 μM	1.3 μM	0.73 μM
	(760 $\pm$ 90 nM)	(430 $\pm$ 40 nM)	(630 $\pm$ 160 nM)	(360 $\pm$ 50 nM)
Trypsin	$\gg$50 μM	$\gg$50 μM	10 μM (ND)	$\gg$50 μM
Papain	$\gg$50 μM	$\gg$50 μM	46 μM (ND)	$\gg$50 μM
Amylase	$\gg$50 μM	$\gg$50 μM	3 μM (ND)	17 μM (ND)

The experimental error in the IC$_{50}$ determination is <10% of the given value. The inhibition constants K_i are shown in parentheses when applicable.

structure may lead to significant alterations in potency and suggests that further derivatization of this new group of PIs may significantly enhance their potential as antiretroviral drugs.

Analysis of the polyprotein processing by Western blotting shows a processing defect in the virus grown in the presence of active compounds (data not shown). No significant toxicity of tested compounds in tissue cultures was observed in the concentration range up to 50 μM.

Specificity and Selectivity Testing. The selectivities of the lead compound **1** and the more potent compounds **4–6** were tested on a panel of seven enzymes, including PR from the highly homologous HIV-2 virus, PR from more distantly related retrovirus MIA 14, prototype human aspartic PRs cathepsin D and pepsin, serine PR trypsin, cystein PR papain, and amylase as a representative of nonproteolytic enzymes with an anionic active-site cleft. The results are summarized in Table 3 in terms of IC$_{50}$ values; the corresponding K_i values are shown in parentheses when appropriate.

All tested compounds inhibit homologous HIV-2 PR and MIA PR, although less tightly, suggesting that they might be active against mutated resistant PR species selected under the pressure of clinically used PIs in HIV-positive patients. The activity of tested compounds toward cathepsin D and pepsin is two orders of a magnitude lower in terms of K_i when compared with HIV-1 PR. The tested compounds do not significantly inhibit any other enzyme analyzed.

Crystal Structure of PR–Compound 1 Complex. Structure of HIV PR–compound **1** complex was determined at 2.15 Å resolution with R factor of 17.6% and R_{free} of 23.6%. The final model comprises a PR dimer (chains A and B) with two molecules of compound **1** bound in the active site (labeled Cb1 and Cb2 in Fig. 2A). Because compound **1** is highly symmetrical in shape, it is not, however, possible to distinguish unambiguously the positions of carbons and borons in electron density maps at 2.15 Å resolution.

With the two molecules of compound **1** bound, the overall conformation of the PR is similar to the open conformation typical for free PR. Most of the structures of substrate-based active site inhibitor complexes exhibit flaps closed over the active site. However, the PR complexed with **1** can be superimposed with the unliganded PR structure [PDB ID code 1HHP (42)] with an rms deviation in α-carbon positions of 0.99 Å (Fig. 2B). Flaps are obviously held in the open conformation by binding of the inhibitor molecules to the flap-proximal, "upper" part of the active site cleft

(Fig. 2). The structure thus provides evidence that the hitherto unexplored class of inhibitors shows an unexpected mode of inhibition. So far, the open conformation of flap was reported only for unliganded PR, whereas all inhibitor-bound structures of wild-type HIV PR show closed flap conformation. Here, we describe the structure of a complex of PR with a relatively potent inhibitor bound uniquely in the enzyme open form. Because the conformation of flap in PR is functionally very important, this finding might suggest that inhibition mechanism of these compounds is blocking the flap closure rather than filling the specific bonding pockets in the active site cleft.

Compound **1** is bound in the hydrophobic pockets formed by side-chains of PR residues Pro-81, Ile-84, and Val-82 and covered by flap residues Ile-47, Gly-48, and Ile-54 (Fig. 3). These pockets correspond approximately to S3 and S3′ substrate-binding subsites. Although the inhibitor-binding site is identical in both monomers, the positions of the two compound **1** molecules are different (Fig. 3C). Inhibitor molecule Cb1 makes 21 van der Waals contacts with seven residues in monomer A (Fig. 3A) whereas inhibitor molecule Cb2 makes 12 van der Waals contacts with four monomer B residues (Fig. 3B). On average, 84% contacts are made with PR nonpolar atoms. The two Cb1 and Cb2 molecules contact each other with 3 van der Waals contacts. Compound **1** loses 89% of its total solvent-accessible surface upon complex formation, and good shape complementarity between both S3 and S3′ PR subsites and compound **1** molecules is illustrated by an average gap volume index of 0.45. This result is slightly lower than an average value for enzyme–inhibitor complexes (43).

The PR compound **1** complex crystallized in C2 crystal form, unique among all 169 HIV PR structures deposited in PDB. In the crystal, symmetrically related PR complexes are oriented head-to-head by their active sites. As a result of this crystal packing, symmetrical flaps are in contact with each other and molecules of compound **1** are in contact with their symmetry mates, as well as with flaps belonging to neighboring symmetry molecule. Thus, in addition to the above described contacts, interactions of compound **1** with PR based on crystal contacts can be observed with residues Gly-48, Gly-49, Ile-50, and Phe-53. However, formation of these interactions in solution seems unlikely because assembly of a complex consisting of two PR dimers and four inhibitor molecules is highly improbable.

Five of the PR residues that are in contact with compound **1** are often mutated in drug-resistant PR variants (Ile-47, Ile-48, Ile-54, Val-82, and Ile-84). The question whether the presence of these

mutations affects compound **1** binding needs to be answered by further biochemical and structural studies. Nevertheless, compound **1** binds by two different modes to two identical binding pockets formed by monomers A and B (Fig. 3C). Thus, we can expect that compound **1** is able to adapt its position so that it could bind into an appropriate pocket altered by mutations.

In addition to the two inhibitor molecules, a continuous electron density map at the bottom of the PR active site allowed modeling of the tetrapeptide Ala-Gly-Ala-Ala, which represents a product of PR autoproteolytic cleavage, often observed during cocrystallization of PR with weak active-site inhibitors. The product of PR degradation is then found in the active site instead of inhibitor (ref. 46; PDB ID code 1SP5). In the present structure, however, the peptide and compound **1** occupy different sites of the active cleft, and, therefore, they can bind simultaneously. Nevertheless, high temperature factors of the peptide main chain atoms point to high mobility and/or its probable lower occupancy. The lack of electron density for side chains further suggests that several peptides originating from three possible cleavage sites (44) are present in the structure. Therefore, we presume that the presence of peptide in the active site is not required for compound **1**-specific binding.

In conclusion, we have shown that boron clusters represent convenient building blocks that can create important interactions with hydrophobic patches of the HIV PR-binding site. X-ray structure analysis of the metallaborane–PR complex brings up compelling evidence about the inhibitor-binding site and type of interaction. *Exo*-skeletal substitution of the parent metallaborane cluster tecton introduces additional noncovalent interactions leading to the dramatic improvement in inhibition efficacy and selectivity. The combination of the hydrophobic interactions of the scaffold with substitutions allowing for specific H-bonding and coulombic interactions might further increase the potency of this class of nonpeptide PR inhibitors based on an inorganic framework. Chemical and biological stability, low toxicity, and the possibility to introduce heteroatoms into the cage or polar group modifications to the side chains make boron clusters very attractive pharmacophores.

1. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M. & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4686–4690.
2. Prejdova, J., Soucek, M. & Konvalinka, J. (2004) *Curr. Drug Targets: Infect. Disord.* **4**, 137–152.
3. Clavel, F. & Hance, A. J. (2004) *N. Engl. J. Med.* **350**, 1023–1035.
4. Coffin, J. M. (1995) *Science* **267**, 483–489.
5. Surleraux, D. L. N. G., De Kock, H. A., Verschueren, W. G., Pille, G. M. E., Maes, L. J. R., Peeters, A., Vendeville, S., De Meyer, S., Azijn, H., Pauwels, R., *et al.* (2005) *J. Med. Chem.* **48**, 1965–1973.
6. Surleraux, D. L. N. G., Tahri, A., Verschueren, W. G., Pille, G. M. E., De Kock, H. A., Jonckers, T. H. M., Peeters, A., De Meyer, S., Azijn, H., Pauwels, R., *et al.* (2005) *J. Med. Chem.* **48**, 1813–1822.
7. Judd, D. A., Nettles, J. H., Nevins, N., Snyder, J. P., Liotta, D. C., Tang, J., Ermolieff, J., Schinazi, R. F. & Hill, C. L. (2001) *J. Am. Chem. Soc.* **123**, 886–897.
8. Bosi, S., Da Ros, T., Spalluto, G. & Prato, M. (2003) *Eur. J. Med. Chem.* **38**, 913–923.
9. Friedman, S. H., DeCamp, D. L., Sijbesma, R. P., Srdanov, G., Wudl, F. & Kenyon, G. L. (1993) *J. Am. Chem. Soc.* **115**, 6506–6509.
10. Sijbesma, R., Srdanov, G., Wudl, F., Castoro, J. A., Wilkins, C., Friedman, S. H., DeCamp, D. L. & Kenyon, G. L. (1993) *J. Am. Chem. Soc.* **115**, 6510–6512.
11. Zheng, Z., Juodawlkis, A. S., Wirtz, S. S., Schinazi, R. F., Zeng, H., Bellavia, C., Wudl, F. & Hill, C. L. (1998) in *Fullerenes: Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials*, eds. Kadish, K. M. & Ruoff, R. S. (Electrochem. Soc., San Diego, CA), Vol. 6, No. PV98-8.
12. Hawthorne, M. F. (1993) *Angew. Chem.* **105**, 997–1033.
13. Soloway, A. H., Tjarks, W., Barnum, B. A., Rong, F. G., Barth, R. F., Codogni, I. M. & Wilson, J. G. (1998) *Chem. Rev. (Washington, DC)* **98**, 1515–1562.
14. Hawthorne, M. F. & Maderna, A. (1999) *Chem. Rev. (Washington, DC)* **99**, 3421–3434.
15. Endo, Y., Ohta, K., Yoshimi, T. & Yamaguchi, K. (2004) *Phosphorus Sulfur Silicon Relat. Elem.* **179**, 799–802.
16. Valliant, J. F., Guenther, K. J., King, A. S., Morel, P., Schaffer, P., Sogbein, O. O. & Stephenson, K. A. (2002) *Coord. Chem. Rev.* **232**, 173–230.
17. Fujii, S., Hashimoto, Y., Suzuki, T., Ohta, S. & Endo, Y. (2005) *Bioorg. Med. Chem. Lett.* **15**, 227–230.
18. Endo, Y., Yoshimi, T., Kimura, K. & Itai, A. (1999) *Bioorg. Med. Chem. Lett.* **9**, 2561–2564.
19. Tsuji, M., Koiso, Y., Takahashi, H., Hashimoto, Y. & Endo, Y. (2000) *Biol. Pharm. Bull.* **23**, 513–516.
20. DeCamp, D. L., Babe, L. M., Salto, R., Lucich, J. L., Koo, M. S., Kahl, S. B. & Craik, C. S. (1992) *J. Med. Chem.* **35**, 3426–3428.
21. Hawthorne, M. F., Young, D. C. & Wegner, P. A. (1965) *J. Am. Chem. Soc.* **87**, 1818–1819.
22. Saxena, A. K. & Hosmane, N. S. (1993) *Chem. Rev. (Washington, DC)* **93**, 1081–1124.
23. Sivaev, I. B. & Bregadze, V. I. (1999) *Collect. Czech. Chem. Commun.* **64**, 783–805.
24. Plesek, J. (1992) *Chem. Rev. (Washington, DC)* **92**, 269–278.
25. Rais, J. & Gruner, B. (2005) in *Solvent Extraction*, eds. Marcus, I. & SenGupta, A. K. (Dekker, New York), pp. 243–334.
26. Reed, C. A. (1998) *Acc. Chem. Res.* **31**, 133–139.
27. Plešek, J., Base, K., Mares, F., Hanousek, F., Stibr, B. & Hermanek, S. (1984) *Collect. Czech. Chem. Commun.* **49**, 2776–2789.
28. Plešek, J., Gruner, B., Baca, J., Fusek, J. & Cisarova, I. (2002) *J. Organomet. Chem.* **649**, 181–190.
29. Plešek, J., Hermanek, S., Franken, A., Cisarova, I. & Nachtigal, C. (1997) *Collect. Czech. Chem. Commun.* **62**, 47–56.
30. Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarcich, J. L., Howe, W. J. & Tomich, C. S. (1994) *Biochemistry* **33**, 9405–9413.
31. Weber, J., Mesters, J. R., Lepsik, M., Prejdova, J., Svec, M., Sponarova, J., Mlcochova, P., Skalicka, K., Strisovsky, K., Uhlikova, T., *et al.* (2002) *J. Mol. Biol.* **324**, 739–754.
32. Strisovsky, K., Smrz, D., Fehrmann, F., Kraeusslich, H. G. & Konvalinka, J. (2002) *Arch. Biochem. Biophys.* **398**, 261–268.
33. Mares, M., Meloun, B., Pavlik, M., Kostka, V. & Baudys, M. (1989) *FEBS Lett.* **251**, 94–98.
34. Kucerova, Z., Pohl, J. & Korbova, L. *J. Chromatogr.* **376**, 409–412.
35. Williams, J. W. & Morrison, J. F. *Methods Enzymol.* **63**, 437–467.
36. Benyoucef, S., Hober, D., Shen, L., Ajana, F., De Groote, D., Bocket-Mouton, L., Gerard, Y., Lion, G., Vilain, V. & Wattre, P. (1997) *Microbiol. Immunol.* **41**, 939–946.
37. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307–326.
38. Brynda, J., Řezáčová, P., Fabry, M., Horejsi, M., Stouracova, R., Soucek, M., Hradilek, M., Konvalinka, J. & Sedláček, J. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1943–1948.
39. Gruner, B., Plešek, J., Baca, J., Cisarova, I., Dozol, J. F., Rouquette, H., Vinas, C., Selucky, P. & Rais, J. (2002) *New J. Chem.* **26**, 1519–1527.
40. Sivaev, I. B., Starikova, Z. A., Sjoberg, S. & Bregadze, V. I. (2002) *J. Organomet. Chem.* **649**, 1–8.
41. Olejniczak, A. B., Plešek, J., Kriz, O. & Lesnikowski, Z. J. (2003) *Angew. Chem. Int. Ed.* **42**, 5740–5743.
42. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H. & Poljak, R. J. (1991) *Biochimie* **73**, 1391–1396.
43. Jones, S. & Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13–20.
44. Rose, R. B., Craik, C. S. & Stroud, R. M. (1998) *Biochemistry* **37**, 2607–2621.
45. Bruenger, A. T. (1992) *Nature* **355**, 472–475.
46. Buchtelova, E., Hasek, J., Dohnalek, J., Petvokova, H., Duskova, J., Skalova, T., Brynda, J., Sedlacek, J., Hradilek, J. Konvalinka, J., *et al.* (2001) *Mater. Struct.* **8**, 31–32.

Link between a novel human γD-crystallin allele and a unique cataract phenotype explained by protein crystallography

Stanislav Kmoch, Jiří Brynda[1], Befekadu Asfaw, Karel Bezouška[2], Petr Novák[3], Pavlína Řezáčová[1], Lenka Ondrová, Martin Filipec[4], Juraj Sedláček[1] and Milan Elleder[+]

Institute of Inherited Metabolic Diseases, Division B, Building D, Ke Karlovu 2, 12808 Prague 2, Czech Republic, [1]Department of Gene Manipulation, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic, [2]Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic, [3]Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic and [4]Ophthalmologic Clinic, Charles University, 1st Faculty of Medicine and University Hospital, Prague, Czech Republic

We describe a 5-year-old boy with a unique congenital cataract caused by deposition of numerous birefringent, pleiochroic and macroscopically prismatic crystals. Crystal analysis with subsequent automatic Edman degradation and matrix-associated laser desorption ionization time-of-flight mass spectrometry have identified the crystal-forming protein as γD-crystallin (CRYGD) lacking the N-terminal methionine. Sequencing of the *CRYGD* gene has shown a heterozygous C→A transversion in position 109 of the inferred cDNA (36R→S transversion of the processed, N-terminal methionine-lacking CRYGD). The lens protein crystals were X-ray diffracting, and our crystal structure solution at 2.25 Å suggests that mutant R36S CRYGD has an unaltered protein fold. In contrast, the observed crystal packing is possible only with the mutant protein molecules that lack the bulky Arg36 side chain. This is the first described case of human cataract caused by crystallization of a protein in the lens. It involves the third known mutation in the *CRYGD* gene but offers, for the first time, a causative explanation of the phenotype.

INTRODUCTION

Cataract (lens opacification) is an etiologically and genetically heterogeneous process, the common denominator of which is loss of lens mass physical homogeneity leading to abrupt changes of the refractive index resulting in light scattering and loss of transparency.

Cataractogenesis is associated with a long list of metabolic disorders either inherited or acquired, leading to breakdown of the lens supramolecular architecture mainly through loss of osmoregulation or oxidative damage (1,2). As the major proteins maintaining the lens transparency are crystallines (representing up to 90% of the soluble lens proteins) it is

reasonable to consider them as a frequent final target of the pathologic events leading to their precipitation and formation of high molecular weight aggregates. Besides that, a number of structural alterations at the microscopic level have been described encompassing vacuoles, globular and membranous formations together with a number of alterations defined solely at the macroscopic level without microscopic correlation (3).

Recently it was shown that cataractogenesis can be primarily induced by mutations in the family of genes coding for crystallines, which in the mutant form are prone to aggregation. The disorders are congenital and are transmitted as an autosomal dominant trait. In laboratory animals the process was associated with mutations in the β and γ crystallin genes (4–7). In humans, cataracts of various appearances were found to be associated with mutations in the *CRYBB2* (8), *CRYBA1* (9) and *CRYAA* (10) genes. Reactivation of the *CRYGE* pseudogene mutated in the promotor region and overexpressing the improperly folded incomplete single motif protein was found in a family with the hereditary Coppock-like cataract (11), which was later determined to be due to mutation in the *CRYGC* (12). Finally, cataracts with either dust-like opacities or with aculeiform phenotype were determined to be caused by mutation in the *CRYGD* gene (13,14). In some of the above-mentioned cases, the problem of causative explanation of the cataract phenotype seems to depend much, if not entirely, on understanding the propensity of the respective mutant crystallin to form deposits. A need for understanding the protein structural basis that underlies formation of protein aggregates has already been recognized and attempts have been made to accumulate relevant information. Up to now, the most advanced results in this direction were obtained by molecular modeling of the R14C CRYGD structure (13). These results confirm a substantial role for protein surface residues but, however plausible in general, do not unequivocally solve the problem as to how the altered primary structure may project itself into macroscopic properties of the mutated protein. We believe that our present study further establishes the structural basis of pathogenicity of mutant crystallines. In this report we describe a novel allele of

[+]To whom correspondence should be addressed. Tel: +420 2 2491 8283; Fax: +420 2 2491 9392; Email: melleder@beba.cesnet.cz

the *CRYGD* gene that confers R36S mutation and is associated with unique deposition of crystals of the mutant gene product in the eye lens. As detailed below, X-ray diffraction data were obtained with isolated crystals of R36S mutated CRYGD and the solved crystal structure has directly shown that the mutation may fundamentally promote the crystal packing.

RESULTS

Microscopic analysis and histochemistry of the aspirated lens material

The lens extract smears showed crystals slightly stained by the Giemsa–May–Gruewald technique, resistant to Sudan dyes, to 0.1 N HCl, to 0.1 N NaOH and to chloroform–methanol extraction. There was no detectable acid phosphatase activity either around the crystals or in the rest of the smeared lens fibers. When suspended in phosphate-buffered saline (PBS) the crystals were macroscopically prismatic, pleiochroic and birefringent (Fig. 1b and c).

Ultrastructure

Many of the lens fiber cells displayed ultrastructure which could be classified as normal. There was a uniformly fine granular appearance of the cytoplasm and varied outline of the cell borders which were either even and almost linear, or highly undulating (ball-and-socket arrangement). Some were present as invaginations, which were either small (vermiform) or large (globular) up to 1 μM in diameter, from the neighboring cells. The invaginations, especially those of highly dense cyto-plasms, could be distinguished from residential cell inclusions by the presence of gap junctions which were frequently seen as focal around 0.5 μM long plaque-like structures marking the cell membranes. Similar structures with highly irregular outlines were seen without apparent gap junction structures. There were marked differences in density of individual fiber cells ranging from very light to very dense. There were foci of massive globular cytoplasmic transformation with cytoplasmic globules 0.1–0.6 μM in diameter. The surface of the globules was covered by either a single or multilayered periodic membrane system of uncertain origin. The crystals were seen in a small number of cells of the sample processed for electron microscopy. They were either deposited free in the cytoplasm (Fig. 2a) or surrounded by concentrically arranged membranes (Fig. 2b and c) in foci rich in globular transformation.

Characterization of crystals

Resistance of the crystals to the above-mentioned procedures suggested their protein nature. To determine their protein identity, isolated crystals were analyzed on a 10–20% gradient SDS–PAGE gel. The gel was stained with Coomassie blue. The resulting predominant protein (20 kDa band) (Fig. 3b) was either transferred on polyvinylidene difluoride (PVDF) membrane and sequenced using automatic Edman degradation or cut from the gel, trypsin digested and subjected to matrix-associated laser desorption ionization time-of-flight (MALDI–TOF) mass spectrometry analysis. Both the 17 amino acid N-terminal sequence (Fig. 3b) and the 40% sequence coverage (for further increase, see below and Fig. 3d) showed the protein
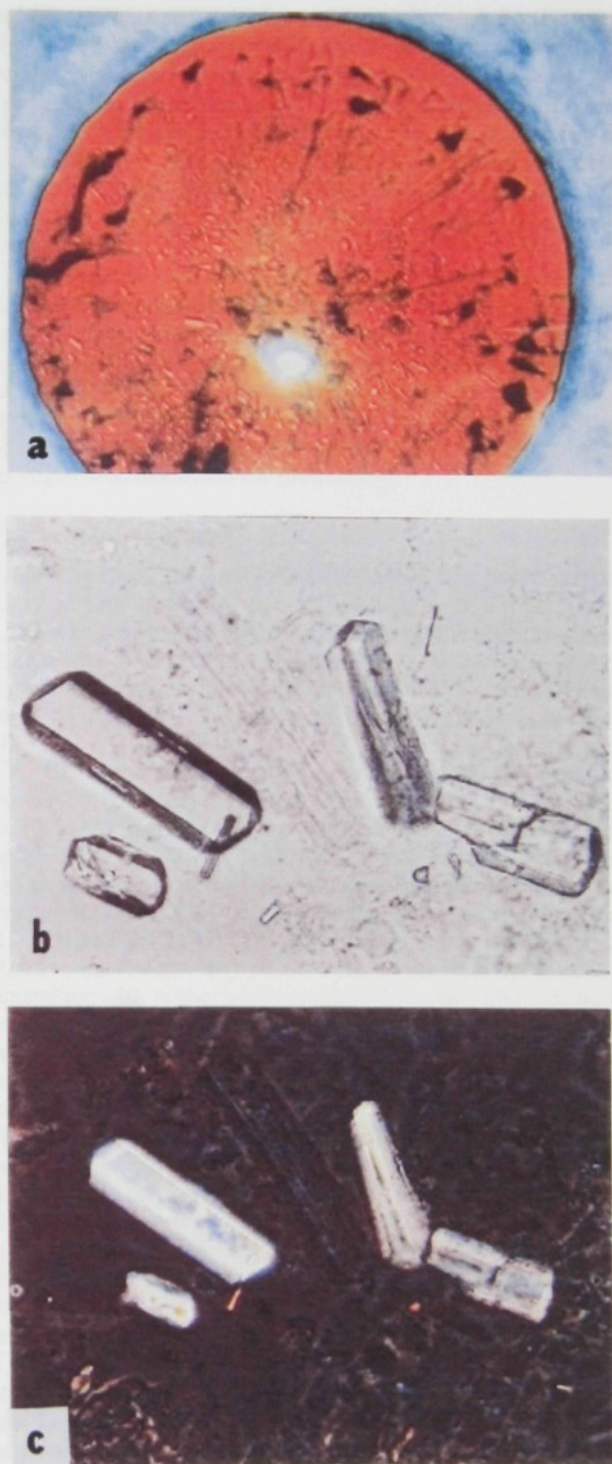

Figure 1. (a) Patient's lens before surgery with numerous transparent longitudinal crystals reaching 1.0 mm in length, with a higher density in the fetal and infantile nucleus and a lower density in the lens cortex. A fine diffuse granular opacity was observed in the nucleus. In the lens cortex, grayish lines were present extending from the center towards the periphery as well as larger, irregular opacities concentrated mainly in the periphery. No pathology was observed by biomicroscopy in the conjunctiva, cornea, anterior chamber, iris or retina. (b) Isolated crystals suspended in PBS. Unstained (phase contrast), ×375. (c) The same as in (b), in polarized light, ×375.

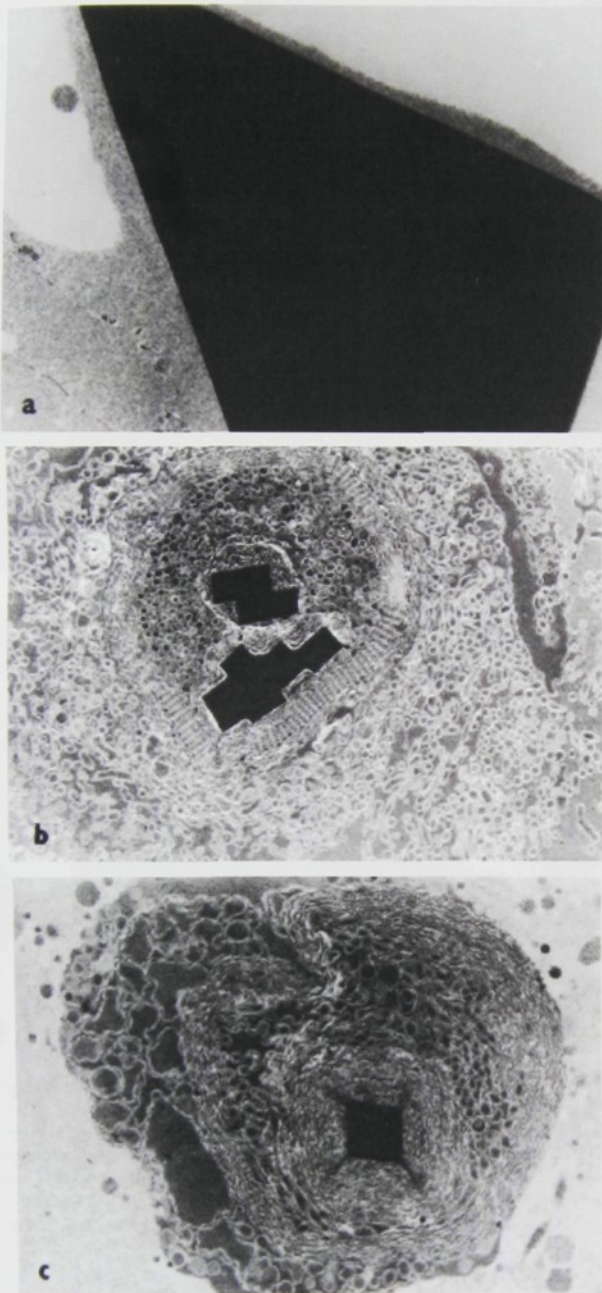**Figure 2.** Electron micrograph of the removed lens tissue. (**a**) Large crystal deposited in the cytosol of the lens fibre cell, ×6000. (**b** and **c**) Foci of extensive lens cell degeneration with globular cytoplasmic transformation. The globules are 0.1–0.6 μM in diameter, with surfaces often covered by concentric membranes forming layers of various thickness. Bulky membranous coverings (of uncertain origin) surrounding also the crystal deposits, ×6000.

to be human γD-crystallin (CRYGD) (NCBI Protein Database accession no. I77413).

DNA sequencing and mutation analysis

Sequence analysis of the propositus *CRYGD* gene showed a heterozygous transversion in the coding region of exon 2 [109C→A in the corresponding cDNA (Fig. 3a)]. This mutation

predicts an arginine to serine change at position 36 (R36S) of the processed, initiation-methionine-lacking protein. The predicted R36S replacement in the processed CRYGD was confirmed at the protein level by mass spectrometry of tryptic peptides of the crystal-forming protein (Fig. 3d) showing the expected mass difference in the fragment spanning amino acids 32–58 as well as perfect CRYGD identity in the remaining coverage of ~75% of the polypeptide length (see above).

No mutation was identified either in the mother or in the healthy sibling of the propositus and in the 100 independent control alleles, using PCR–RFLP analysis (Fig. 3c). No DNA was available for analysis from the father, who died accidentally at age 33 years. He suffered from easily correctable myopia and there were no signs of an overt cataract on ophthalmologic examination of his eye fundus carried out by the ophthalmologist who examined the proband.

Protein crystallography

A striking feature of the R36S γD-crystallin-composed crystals obtained from the lens deposits was their ability to diffract X-rays so that they provided fully measurable sets of reflections. From the viewpoint of protein crystallography, the crystals were of suboptimal quality (mosaicity: 0.9°, *B* value from the Wilson plot: 55 Å^2), but allowed us to solve the crystal structure (Fig. 4). We solved the structure of human R36S-mutated CRYGD by molecular replacement using the known structure of bovine CRYGD (15) as a search model. The unit cell dimensions at 100K are: a = 54.24 Å, b = 81.98 Å and c = 105.54 Å. The space group is P2$_1$2$_1$2$_1$ with two molecules of CRYGD per asymmetric unit (termed A and B). In the refinement of the crystal structure, 22 930 reflections were used (working set 21 753, test set 1177). The structure is presently refined to R/Rfree 26.4%/28.6% (by slow-cooling protocol of simulated annealing with non-crystallographic symmetry applied; no σ cut-off). A total of 342 residues plus 25 solvent molecules were refined (total 2823 atoms). The model lacks the C-terminal amino acids 172 and 173, where the electron density is poor. The stereochemistry of the final model shows that 259 residues (85.8% of the total 302 non-end, non-glycine and non-proline residues) have most favored main-chain dihedral angles, and 43 residues (14.2%) have additional allowed angles. Omit density maps clearly indicate the presence of serine side chains in both molecules at amino acid position 36, and make it obvious that the bulky arginine side chains cannot be accommodated there in any rotamer position. Although the protein fold of the mutated human CRYGD (Fig. 4d) is almost identical to that of bovine CRYGD, the respective molecular contacts (Fig. 4a, b and c) differ substantially. In crystals of bovine CRYGD (15) the side chains of arginine 36 do not point towards neighboring protein molecules. The structure determined for R36S human CRYGD shows that the Ser36 side chain of molecule A interacts with Asn24 of molecule B and the Ser36 side chain of molecule B points to Asn24 of the symmetry-related molecule A′ (SerOγ to AsnNδ distances being 3.0 and 2.7 Å, respectively) (Fig. 4b and c).

It is believed that many structural features determine the relative orientation of molecules in crystal lattices, the distribution of the surface charges being prominent among them. The absence of the Arg36 charge (Fig. 4e) could be seen as the
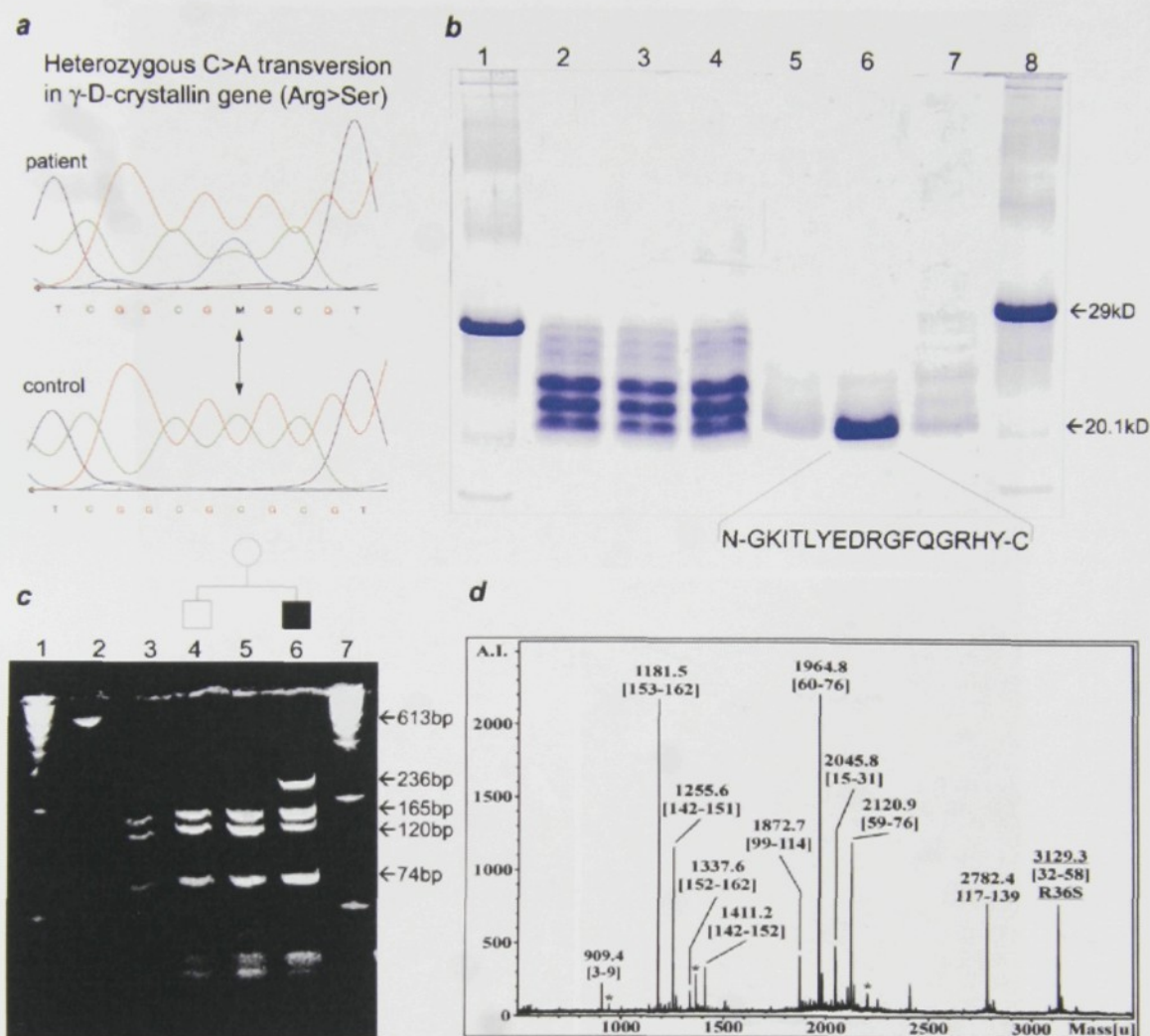
Figure 3. (a) Genomic sequence of CRYGD exon 2. Heterozygous C→A transversion in proband is shown above the normal sequence. (b) The extracted lens proteins (lanes 2–4) and dissolved isolated crystals (lane 6) analyzed on 10–20% gradient SDS–PAGE gel and Coomassie blue R stained. Lanes 5 and 7 represent less effective isolation procedures (gradient centrifugation fractions); lanes 1 and 8, molecular weight markers. (c) PCR–RFLP analysis of exons 1 and 2 genomic fragments showing mutation-induced loss of *Bsh*1236I restriction site in the proband (lane 6) but not in the mother (lane 5) and healthy sibling (lane 4). Lanes 1 and 7, DNA size markers; lane 2, non-cleaved PCR product; lane 3, control sample. (d) γD-crystallin crystal tryptic map showing abnormal MH+3129 kDa and 3145 kDa masses resulting from the absence of the Arg36 cleavage site matching the R36S amino acid exchange. Mass of non-mutant tryptic peptide 36–58 is apparently absent.

factor promoting the resultant orientation of the molecules and, in turn, the crystal formation (and pathogenesis). At least, with all caveats of the complicated theory of mechanisms and kinetics of crystal formation, the missing charge can be seen as the feature that decreases the solubility of the R36S-mutated protein. Protein crystallography, however, gives us the most solid and clear-cut clue to the pathogenic crystal formation: the crystals cannot form with wild-type protein (either as the major or as an admixed component) because of steric hindrances imposed by the bulky Arg36 side chains. It is tempting to extend this line of reasoning to explain at least qualitatively the dominance of the mutant allele: the molecules of the protein product of the normal allele cannot, because of steric hindrance, enter (and perturb) the crystals that form with the pathogenic product.

DISCUSSION

Our case belongs to an extremely rare group of cataracts associated with crystal deposition. Exhaustive survey of the literary data showed, however, that the crystals described so far were composed of cholesterol, amino acids or calcium salts (16). The lens crystals strongly resembling those of our case were described in a case of crystalline cataract associated with uncombable hair (17). However, the crystals were suggested to be composed of cystine due to their high sulphur content. To the best of our knowledge, our case represents the first cataract caused by deposition of defined crystallized protein, proved to be CRYGD. *CRYGD* gene analysis identified a single transversion in heterozygous state (cDNA 109C→A) predicting R37S substitution (of unprocessed protein) at the protein level. The mutation was absent in both the patient's mother and
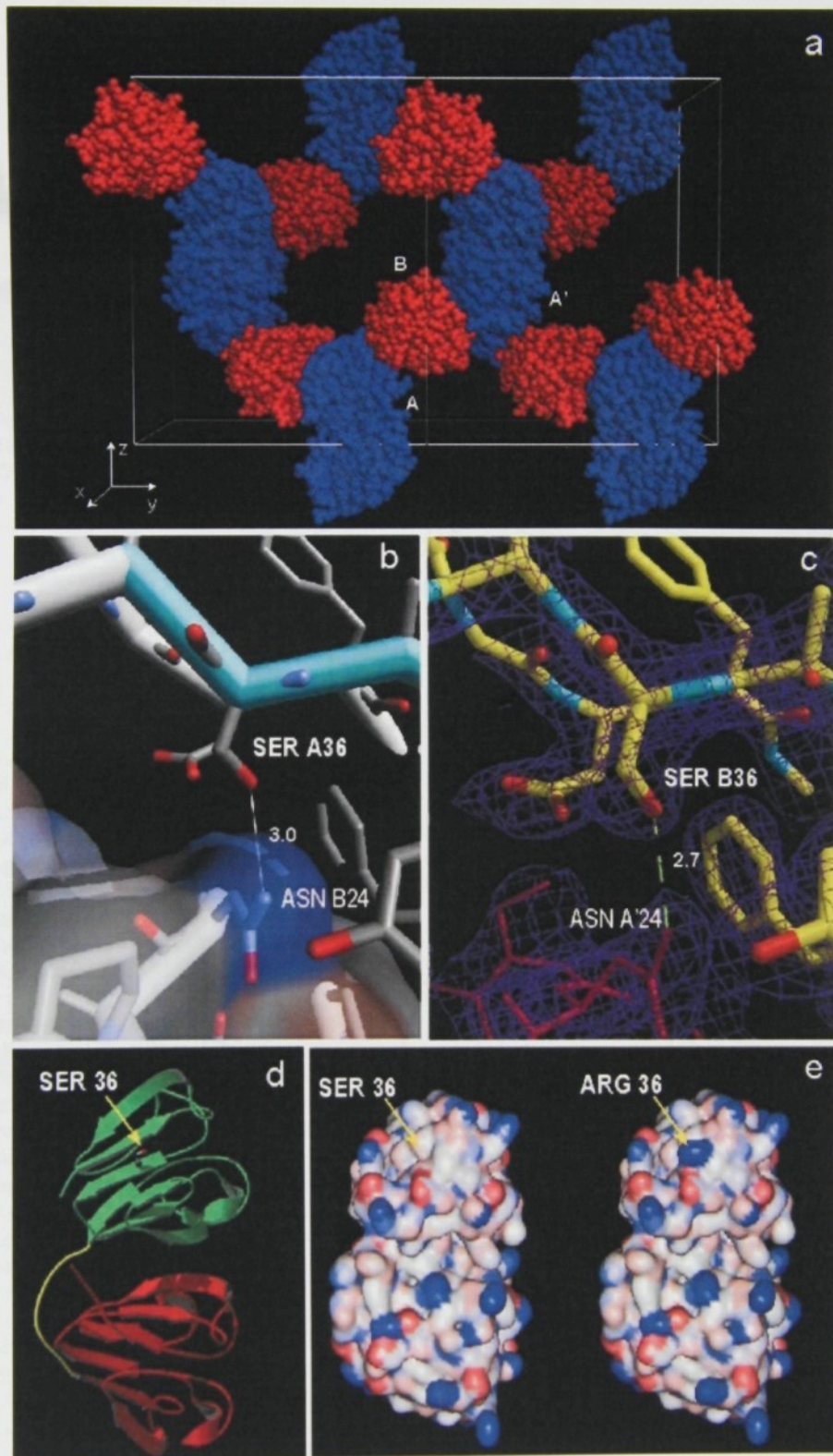
Figure 4. Crystal structure of human R36S γD-crystallin from the lens deposits. (**a**) Crystal packing. Two unit cells are shown; the orientations of their axes are marked at the bottom left. The protein molecules are represented by space-filling models, colored blue for molecules A and the symmetry-related molecules A′ and colored red for molecules B and the symmetry-related molecules B′. (**b**) Detail of interaction between molecule A (top) and molecule B (bottom). The tube representation has, in molecule B, a transparent overlay of the solvent-accessible surface, color coded for partial charges on the atoms (red for negative, blue for positive). (**c**) Detail of interaction between molecule B (top) and the symmetry-related molecule A′ (bottom) shown by electron density contored at 1.5σ. (**d**) Schematic overall view of the main chain fold in the human R36S γD-crystallin. The N-terminal domain is shown in green, the hinge in yellow and the C-terminal domain in red. (**e**) Comparison of the R36S γD-crystallin structure (left) with a molecular model comprising reverse S36→R mutation (right). The solvent-accessible surfaces are shown color coded for partial charges on the atoms (red for negative, blue for positive). The scale and the angle of view are similar to those in (d).

brother and in the control series. The father's DNA sample was not available for analysis; however, his vision was reported by an ophthalmologist to be unaffected and no signs of a cataract were present. As hereditary congenital cataracts tend to be fully penetrative (18), we think that the mutation of the propositus arose *de novo* (*cave! pater semper incertus*).

The predicted R36S replacement in the processed CRYGD was confirmed at the protein level by mass spectrometry of tryptic peptides of the crystal-forming protein (Fig. 3d) showing the expected mass difference in the fragment spanning amino acids 32–58 as well as perfect CRYGD identity in the remaining coverage of ~75% of the polypeptide length. This finding, taken together with high conservation of the said arginine residue (scoring 35 among 36 CRYGD homologs cloned from various species, MAXHOM alignment; http://dodo.cpmc.columbia.edu) provides conclusive evidence for the pathogenic nature of the product of the mutant allele. Mutations in *CRYGD* were recently proved to be cataractogenic, but leading to non-crystalline deposits. The R14C mutation led to non-crystalline punctate opacities (13), whereas the R58H mutation was associated with the aculeiform phenotype (14). Our R36S mutation was uniquely associated with formation of protein crystals *in situ*, whereas all other crystal-forming lens deposits described so far proved to be of uncertain (17) or non-protein (16) nature. The X-ray diffraction measurements with crystals of human mutant R36S CRYGD extracted from the patient's lens enabled us to solve the crystal structure. This structure is the second ever solved for a member of the CRYGD family [the first being that of non-mutated bovine CRYGD (15), used by us as the search model in molecular replacement]. For comparative purposes, results of the protein crystallography study can be divided into two categories: (i) those elucidating the three-dimensional build-up of the mutated human CRYGD molecule; and (ii) those showing the arrangement of the molecules within the crystal lattice. Regarding the former, protein folds in R36S CRYGD and in the related wild-type molecule bear the closest resemblance, as could be expected also from the high homology of primary structures. The distinguishing feature is the crystal packing: in the structure of R36S CRYGD different contacts appear, and this is impossible with wild-type molecules. These contacts, inter alia, make the crystals of the mutant protein somewhat denser. Further, even though there is nothing that would prevent the molecules of mutant R36S CRYGD from forming the same contacts as the related wild-type molecules (15), such crystal packing has not actually occurred in the pathogenesis. All this taken together with consideration of the possible effects of elimination of the R36 surface charge, points to an inherent tendency of the mutant protein molecules to form permanent mutual contacts more easily than their wild-type counterparts. Needless to say, even with solid protein structural knowledge a broad variety of modalities are to be expected to take part in the triggering and progression of crystal growth. Thus, what remains to be explained are the epigenetic phenomena that govern the non-uniform pattern of deposition of the crystals within the cell population (suggesting the importance of local conditions) and the pathogenesis of the accompanying cell degradation. It is highly probable that the mutant CRYGD also exists in an alternative aggregate state represented by the dense amorphous globules,

corresponding to its coprecipitate with both α- and β-crystallins in cold-induced experimental cataract (19).

Apart from this, this specific case permitted an unexpectedly straightforward protein crystallographic interpretation. Such ease may, in a more general context, only strengthen the rationale for the current trend to use X-ray crystallography in conjunction with structural genomics (20).

MATERIALS AND METHODS

Patient

A 5-year-old boy suffering from photophobia and decreased visual acuity (symbol test chart 0.17 in his right eye and 0.15 in his left eye) due to symmetrical crystal deposition and greyish opacities in both lenses (Fig. 1a) underwent uncomplicated bilateral extracapsular cataract extraction by simple aspiration with intraocular lens implantation. There were no other clinical, pathological or biochemical findings, including hair abnormalities. On slit lamp examination there was a normal finding on the conjunctiva, cornea, anterior chamber and on the iris. In both lenses there were numerous transparent longitudinal crystals, some of them reaching 1 mm length with axes randomly oriented. The crystals were evenly distributed throughout the lens with a slightly higher density in the central area. The lens substance in between the crystals and the opacities was clear. The ophthalmoscopic examination of the fundus revealed normal finding. The child underwent uncomplicated successive extracapsular cataract extraction with anterior and posterior circular capsulotomy, anterior vitrectomy and intraocular lens implantation in both eyes, 2 and 5 months after the initial examination. The lens material was aspirated by canula, the anterior chamber being maintained by irrigation with Ringer's solution. Subsequent examinations were carried out 4 and 7 months after the surgery showed quiet eyes with visual acuity 0.66. There were no other clinical findings, including hair abnormalities (see Discussion). Routine biochemical examinations, serum and urinary lactate, pyruvate and 3-OH butyrate concentrations, amino acid and urinary organic acid profiles, and serum galactitol concentration were within control range.

Staining procedures and tissue processing

The smeared lens extract was stained with Giemsa–May–Gruewald, examined under fluorescence and polarization microscopy. Acid phosphatase activity was examined using naphthol ASBI phosphate and hexazonium pararosaniline as coupler (21). Solubility was tested using dried acetone (15 min at room temparature) and chloroform:methanol 2:1 (1 h at room temparature) and 0.1 N HCl (15 min at room temparature). Lipid nature was tested with Sudan Black B. Part of the extracted lens was fixed with 10% paraformaldehyde, dehydrated and embedded into araldite–epon mixture. Semi-thin sections were stained with alkaline Toluidine blue, the thin sections were double contrasted and examined with a JEM 100B microscope.

Crystal isolation

The lens aspirate was methanol-fixed on the microscope slide and the crystals were picked up under an inverse microscope

using a Pasteur pipette and collected in an Eppendorf tube. They were washed with distilled water, solubilized in SDS-containing buffer and analyzed by 10–20% gradient SDS–PAGE. The predominant 20 kDa protein band was electroblotted on a PVDF membrane and microsequenced using Edman degradation on an LF3600D Protein Sequencer (Beckman Instruments, Fullerton, CA). Search in sequence databases was done using the BLAST algorithm (http://www.ncbi.nlm.nih.gov). For mass spectrometry analysis, the band was excised from the SDS–PAGE gel and digested by trypsin. The resulting peptides were extracted into 40% acetonitrile in 1% TFA and analyzed on a MALDI–TOF Bruker Biflexmass spectrometer (Bruker-Franzen, Bremen, Germany), as described (22). Protein identity was elucidated using online resources (http://prowl.rockefeller.edu/cgi-bin/ProFound and http://falcon.ludwig.ucl.ac.uk/mshome3.2.htm).

CRYGD gene amplification and sequencing

The genomic sequence of *CRYGD* gene was obtained from GenBank database (accession nos K03006 and K03005). The following primer pairs containing universal sequencing overhangs (underlined) were used for PCR amplification of two genomic fragments:

exons 1 and 2:

upper primer: 5′-GAAACAGCTATGACCATGGCCCCTTT-TGTGCGGTTCTTGC-3′;

lower primer: 5′-ATACGACTCACTATAGGGCGACTGAT-CGCTACTTCTAATGT-3′

exon 3:

upper primer: 5′-GAAACAGCTATGACCATGCACACTTG-CTTTTCTTCTCTTT-3′;

lower primer: 5′-ATACGACTCACTATAGGGCAAGACA-CAAGCAAATCAGTGCC-3′.

DNA templates were amplified using KlenTaq DNA polymerase (Ab Peptides, St Louis, MO) and standard PCR conditions (94°C for 5 min; 35 cycles of 94°C for 20 s, 60°C for 30 s, 74°C for 30 s; and final extension at 74°C for 10 min). Both fragments were subsequently gel purified and cycle sequenced on both strands using Cy5 5′-labeled T7 and reverse primers (Generi Biotech, Hradec Králové, Czech Republic). The sequences were analyzed on AlfExpress sequencer and read using the ALFWIN software (Pharmacia, Uppsala, Sweden). The sequence analysis was performed on a control, mother, brother and propositus genomic DNA samples isolated from blood (Qiagen, Hilden, Germany).

To confirm the mutation found in the propositus and to be able analyze the family members and controls, the analysis of the mutation destroying restriction site for *Bsh*1236I (MBI, Vilnius, Lithuania) was performed. PCR products of exons 1 and 2 were digested for 3 h according to the manufacturer's instructions. Resulting fragments were analyzed on 12% denaturing polyacrylamide gel and visualized using the ethidium bromide staining (Fig. 3c).

Protein crystallography

Several crystals, ~0.15 × 0.15 × 0.5 mm from the lens tissue extract were transferred to PBS buffer with 25% (v/v) glycerol and flash cooled in 100 K nitrogen stream for measurements. Data were collected to 2.25 Å resolution at ESRF (Grenoble, France), beamline ID 14 EH3 [R_{merge} 6.2% (23% for highest

resolution shell); data completeness 99.6% (99.0%)]. The space group was determined from the systematic extinction of axis reflections and confirmed by Patterson synthesis and molecular replacement. Raw data were processed using the programs Denzo (23) and Scalepack (24) and converted into a unique set of structure factor amplitudes. The initial model was obtained by molecular replacement using the program Epmr (25) and one molecule of bovine CRYGD (26) (PDB code 1elp). Mutations of the bovine to the human amino acid residues and the choice of the side chain rotamers were done in all 23 differing positions: the requirement for substitution was in most cases visible also from the differential electron density map. The structure was refined using the CNS package (27) and Shelxl (28), and non-crystallographic symmetry was included through restraints; manual model building was done using the program O (29).

REFERENCES

1. Hejtmancik, J.F., Kaiser, M.I. and Piatigorsky, J. (1995) Molecular biology and inherited disorders of the eye lens. In Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. (eds), *The Metabolic and Molecular Bases of Inherited Disease*, 7th edn. McGraw-Hill, New York, NY, pp.4325–4349.
2. Hejtmancik, J.F. (1998) The genetics of cataract: our vision becomes clearer. *Am. J. Hum. Genet.*, **62**, 520–525.
3. Yanoff, M. and Fine, B.S. (1996) *Ocular Pathology*, 4th edn. Mosby–Wolfe, London, UK/Baltimore, MD.
4. Chambers, C. and Russell, P. (1991) Deletion mutation in an eye lens beta-crystallin. An animal model for inherited cataracts. *J. Biol. Chem.*, **266**, 6742–6746.
5. Rodriguez, I.R., Gonzalez, P., Zigler Jr, J.S. and Borras, T. (1992) A guinea-pig hereditary cataract contains a splice-site deletion in a crystallin gene. *Biochim. Biophys. Acta*, **1180**, 44–52.
6. Cartier, M., Breitman, M.L. and Tsui, L.C. (1992) A frameshift mutation in the gamma E-crystallin gene of the Elo mouse. *Nature Genet.*, **2**, 42–45.
7. Klopp, N., Favor, J., Loster, J., Lutz, R.B., Neuhauser-Klaus, A., Prescott, A., Pretsch, W., Quinlan, R.A., Sandilands, A., Vrensen, G.F. and Graw, J. (1998) Three murine cataract mutants (Cat2) are defective in different gamma-crystallin genes. *Genomics*, **52**, 152–158.
8. Litt, M., Carrero-Valenzuela, R., LaMorticella, D.M., Schultz, D.W., Mitchell, T.N., Kramer, P. and Maumenee, I.H. (1997) Autosomal dominant cerulean cataract is associated with a chain termination mutation in the human beta-crystallin gene *CRYBB2*. *Hum. Mol. Genet.*, **6**, 665–668.
9. Kannabiran, C., Rogan, P.K., Olmos, L., Basti, S., Rao, G.N., Kaiser-Kupfer, M. and Hejtmancik, J.F.C. (1998) Autosomal dominant zonular cataract with sutural opacities is associated with a splice mutation in the betaA3/A1-crystallin gene. *Mol. Vis.*, **4**, 21–26.
10. Litt, M., Kramer, P., LaMorticella, D.M., Murphey, W., Lovrien, E.W. and Weleber, R.G. (1998) Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene *CRYAA*. *Hum. Mol. Genet.*, **7**, 471–474.
11. Brakenhoff, R.H., Henskens, H.A., van Rossum, M.W., Lubsen, N.H. and Schoenmakers, J.G. (1994) Activation of the gamma E-crystallin pseudogene in the human hereditary Coppock-like cataract. *Hum. Mol. Genet.*, **3**, 279–283.

12. Heon, E., Priston, M., Billingsley, G.D., Lubsen, N. and Munier, F.L. (1999) Clarifying the role of gamma crystallin in congenital CCL cataract. *Am. J. Hum. Genet.*, **65** (suppl.), A19.

13. Stephan, D.A., Gillanders, E., Vanderveen, D., Freas-Lutz, D., Wistow, G., Baxevanis, A.D., Robbins, C.M., VanAuken, A., Quesenberry, M.I., Bailey-Wilson, J. *et al.* (1999) Progressive juvenile-onset punctate cataracts caused by mutation of the gammaD-crystallin gene. *Proc. Natl Acad. Sci. USA*, **96**, 1008–1012.

14. Heon, E., Priston, M., Schorderet, D.F., Billingsley, G.D., Girard, P.O., Lubsen, N. and Munier, F.L. (1999) The gamma-crystallins and human cataracts: a puzzle made clearer. *Am. J. Hum. Genet.*, **65**, 1261–1267.

15. Chirgadze, Y.N. and Tabolina, O.Yu. (1996) Alternating charge clusters of side chains: new surface structural invariants observed in calf eye lens gamma-crystallins. *Protein Eng.*, **9**, 745–754.

16. Duke-Elder, S. (ed.) (1969) *Diseases of the Lens and Vitreous; Glaucoma and Hypotony. System of Ophthalmology*, vol. XI. Henry Kimpton, London, UK, p. 134.

17. de Jong, P.T., Bleeker-Wagemakers, E.M., Vrensen, G.F., Broekhuyse, R.M., Peereboom-Wynia, J.D. and Delleman, J.W. (1990) Crystalline cataract and uncombable hair. Ultrastructural and biochemical findings. *Ophthalmology*, **97**, 1181–1187.

18. Francis, P.J., Berry, V., Moore, A.T. and Bhattacharya, S. (1999) Lens biology: development and human cataractogenesis. *Trends Genet.*, **15**, 191–196.

19. Lo, W.K. (1989) Visualization of crystallin droplets associated with cold cataract formation in young intact rat lens. *Proc. Natl Acad. Sci. USA*, **86**, 9926–9930.

20. Burley, S.K., Almo, S.C., Bonanno, J.B., Capel, M., Chance, M.R., Gaasterland, T., Lin, D., Sali, A., Studier, F.W. and Swaminathan, S. (1999) Structural genomics: beyond the Human Genome Project. *Nature Genet.*, **23**, 151–157.

21. Lojda, Z., Gossaru, R. and Schiebler, T.H. (1979) *Enzyme Histochemistry. A Laboratory Manual*. Springer, Berlin/Heidelberg/New York.

22. Bezouska, K., Sklenar, J., Novak, P., Halada, P., Havlicek, V., Kraus, M., Ticha, M. and Jonakova, V. (1999) Determination of the complete covalent structure of the major glycoform of DQH sperm surface protein, a novel trypsin resistant boar seminal O-glycoprotein related to pB1 protein. *Protein Sci.*, **8**, 1551–1556.

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

24. Collaborative Computational Project No. 4 (1994) The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

25. Kissinger, C.R., Gehlhaar, D.K. and Fogel, D.B. (1999) Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr.*, **D55**, 484–491.

26. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

27. Kissinger, C.R., Gehlhaar, D.K. and Fogel, D.B. (1999) Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr.*, **D55**, 484–491.

28. Chirgadze, Y.N., Driessen, H.P.C., Wright, G., Slingsby, C., Hay, R.E. and Lindley, P.F. (1996) Structure of the bovine eye lens gammaD (gamma-Iiib)-crystallin at 1.95 Angstroms. *Acta Crystallogr.*, **D52**, 712–721.

29. 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 and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.*, **D54**, 905–921.

30. Sheldrick, G.M. and Schneider, T.R. (1997) SHELXL: high-resolution refinement. *Methods Enzymol.*, **277**, 319–343.

31. 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 Crystallogr.*, **A47**, 110–119.

Structural Basis of HIV-1 and HIV-2 Protease Inhibition by a Monoclonal Antibody

Pavlina Rezacova,[1,5,6] Julien Lescar,[2,3,6]
Jiri Brynda,[1] Milan Fabry,[1] Magda Horejsi,[1]
Juraj Sedlacek,[1] and Graham A. Bentley[4,5]
[1]Department of Gene Manipulation
Institute of Molecular Genetics
Academy of Sciences of the Czech Republic
Flemingovo nam. 2
166 37 Prague 6
Czech Republic
[2]Centre de Recherches sur les Macromolécules
   Végétales (affiliated with Université Joseph
   Fourier)-CNRS BP53
F38041 Grenoble Cedex
France
[3]European Synchrotron Radiation Facility
BP220
F38043 Grenoble Cedex
France
[4]Unité d'Immunologie Structurale
   (URA 1961 CNRS)
Département d'Immunologie
Institut Pasteur
75724 Paris Cedex 15
France

Summary

Background: Since the demonstration that the protease of the human immunodeficiency virus (HIV Pr) is essential in the viral life cycle, this enzyme has become one of the primary targets for antiviral drug design. The murine monoclonal antibody 1696 (mAb1696), produced by immunization with the HIV-1 protease, inhibits the catalytic activity of the enzyme of both the HIV-1 and HIV-2 isolates with inhibition constants in the low nanomolar range. The antibody cross-reacts with peptides that include the N terminus of the enzyme, a region that is highly conserved in sequence among different viral strains and that, furthermore, is crucial for homodimerization to the active enzymatic form.

Results: We report here the crystal structure at 2.7 Å resolution of a recombinant single-chain Fv fragment of mAb1696 as a complex with a cross-reactive peptide of the HIV-1 protease. The antibody-antigen interactions observed in this complex provide a structural basis for understanding the origin of the broad reactivity of mAb-1696 for the HIV-1 and HIV-2 proteases and their respective N-terminal peptides.

Conclusion: A possible mechanism of HIV-protease inhibition by mAb1696 is proposed that could help the design of inhibitors aimed at binding inactive monomeric species.

[5]Correspondence: rezacova@img.cas.cz (P.R.), bentley@pasteur.fr (G.A.B.)
[6]These authors contributed equally to this work.

Introduction

Mature viral proteins of the human immunodeficiency virus (HIV) are obtained by posttranslational cleavage of the Gag, Gag/Pol, and Env viral polyprotein precursors. The Gag and Gag/Pol polyproteins, in particular, are cleaved by a viral protease, which is itself contained within the Gag/Pol polyprotein chain. Since the demonstration that the human immunodeficiency virus protease (HIV Pr) is essential in the viral life cycle [1–3], this enzyme has become one of the primary targets for antiviral drug design. This has led to the development of many active-site inhibitors, some of which are currently in use as therapeutic agents for AIDS treatment [4]. In order to be proteolytically active, two protease monomers must assemble into a homodimer, with each subunit contributing residues to the substrate binding pocket, including residue Asp-25, which is directly involved in the catalysis [5]. The requirement for protease dimerization to achieve proteolytic activity has led several authors to propose alternative noncompetitive inhibitors that would provoke the dissociation of the active homodimeric assembly [6–8].

With the objective of probing the structural stability of HIV Pr and the eventual design of potential inhibitors of the enzyme that are directed to regions other than the active site, we have examined the effects of anti-HIV-1 Pr monoclonal antibodies (mAbs) on the catalytic activity of the protease [9, 10]. We have recently reported a preliminary study of mAb1696, which, although raised against the HIV-1 Pr, inhibits the catalytic activity of both the HIV-1 and HIV-2 enzymes with inhibition constants of 0.6 nM and 1.5 nM, respectively, at pH 7.4 [11]. This study also showed that mAb1696 cross-reacts with peptides containing the N terminus of the HIV protease. The N-terminal region accounts for a large percentage of the interface between the two HIV Pr monomers because it interdigitates with the C-terminal segment from the other monomer, thus forming a 4-stranded intermolecular β-pleated sheet in the active homodimer. We proposed that mAb1696 inhibits HIV Pr by perturbing the native structure of the enzyme at the dimer interface [11]. In addition, a clustering of negatively charged residues at the antigen binding site was observed in the unliganded Fab crystal structure, which suggested that electrostatic forces play an important role in the interaction between mAb1696 and HIV Pr.

To investigate further the mechanism of HIV Pr inhibition by mAb1696, we have now expressed a single-chain Fv fragment (scFv) in *E. coli*, which contains both 1696 variable domains joined by a flexible linking peptide. Such scFv constructs, which retain the full antigen binding capacities, are the object of very active research. Firstly, they are of interest for structural studies because they usually yield crystals diffracting to higher resolution than the corresponding Fab fragments on
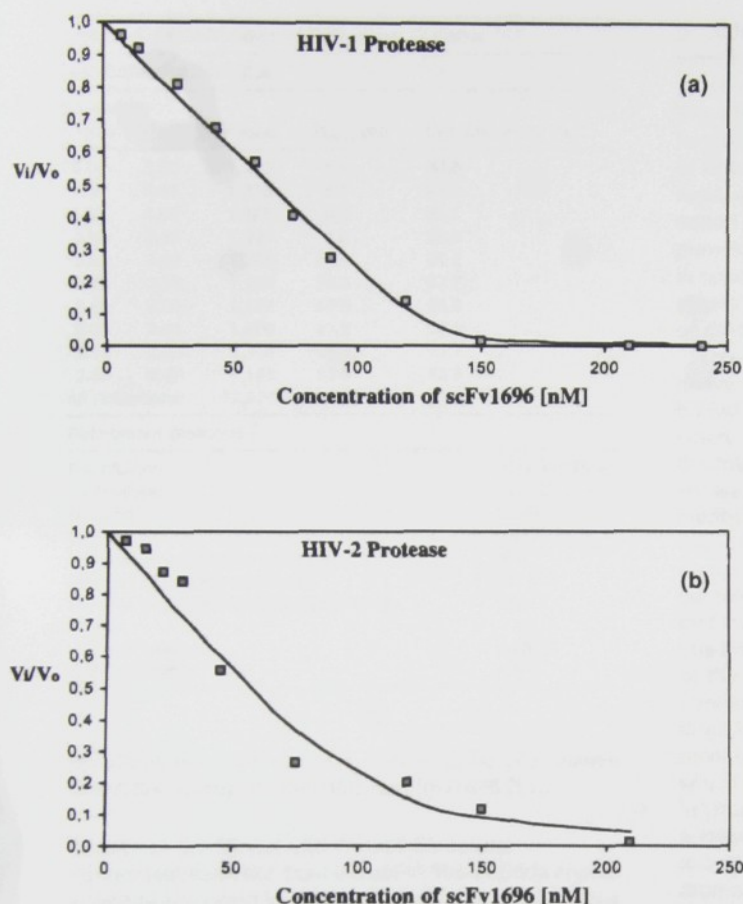
Figure 1. Inhibition of HIV-1 Pr and HIV-2 Pr Activity by scFv1696
(a) The inhibition of HIV-1 Pr activity.
(b) The inhibition of HIV-2 Pr activity.
The initial reaction velocity (v_i) of HIV-1 Pr and HIV-2 Pr, normalized to the initial reaction velocity in the absence of scFv1696 (v_0), is plotted as a function of scFv1696 concentration. The curves show the best fit of the tight binding model [13] to the experimental data (see text for details).

account of their smaller size and decreased interdomain flexibility. Secondly, since they are expressed in bacteria, they are also well suited for binding, mutagenesis and protein engineering studies [12]. In this paper, we report the three-dimensional structure at 2.70 Å resolution of a complex formed between scFv1696 and peptide PQITLWQRR, whose sequence derives from the N terminus of the HIV-1 Pr. (The additional arginine residue was added to the C terminus to increase solubility.) The structure of the complexed scFv1696 is compared with the free Fab1696 structure reported earlier [11]. On the basis of the interactions seen in the complex, the cross-reactivity between mAb1696 and the HIV-1 and HIV-2 protease and their N-terminal peptides can be explained. In addition, a mechanism of HIV Pr inhibition by mAb-1696 is proposed that may help the design of alternative HIV protease inhibitors, aimed at dissociating the homo-dimeric viral enzyme.

Results and Discussion

Inhibition of HIV-1 Pr and HIV-2 Pr by scFv1696
The recombinant single-chain fragment scFv1696 was found to have essentially the same inhibitory capacities as those previously found for Fab1696 or for the complete antibody molecule [11]. Inhibition kinetics of the protease in the presence of scFv1696 were measured by

following the rate of substrate cleavage using reverse-phase HPLC (see Experimental Procedures). The kinetic data were used to estimate the initial reaction velocities of HIV-1 Pr and HIV-2 Pr for each concentration of inhibiting scFv1696. These were normalized to the initial reaction velocity in the absence of inhibitor and were plotted as a function of scFv1696 concentration (Figure 1). Inhibition constants were evaluated by fitting the following equation, which describes the tight binding inhibition model [13], to the experimental points of the graph:

$$v_i/v_0 = 1 - \{[E] + [I_i] + K_{inh} - \sqrt{(([E] + [I_i] + K_{inh})^2 - 4[E][I_i])})/(2[E])\}, \quad (1)$$

where [E] and [I_i] are the protease concentration (held constant) and the scFv1696 concentration, respectively, for the ith run; K_{inh} is the apparent inhibition constant, and v_0 and v_i are the initial reaction velocities for scFv1696 concentrations 0 and [I_i], respectively.

The inhibition constants, K_{inh}, thus obtained for the scFv1696 were 0.6 nM for HIV-1 Pr and 5.5 nM for HIV-2 Pr. The inhibition curves can also be used to obtain the concentration of scFv1696 equivalent to the total amount of active sites present in the reaction mixture. Thus, the calculated stoichiometry of the complexes are 0.9:1 for the scFv1696/HIV-1 Pr and 1.0:1 for the

Table 1. Data Collection and Refinement Statistics

Data Collection Statistics

Resolution Shell (Å)		Unique	R_{merge} (%)	Completeness (%)
30.00	5.80	1,285	5.4	94.8
5.80	4.61	1,296	5.4	98.0
4.61	4.03	1,285	6.2	98.1
4.03	3.66	1,291	9.0	98.2
3.66	3.40	1,285	10.4	98.5
3.40	3.20	1,259	13.3	97.7
3.20	3.04	1,237	16.5	96.8
3.04	2.91	1,260	21.2	96.8
2.91	2.80	1,254	26.3	95.7
2.80	2.70	1,168	32.3	92.8
All reflections		12,620	9.7	96.7

Refinement Statistics

Resolution	30.0 Å–2.70 Å
Reflections	11,202
R factor	0.229
R_{free}	0.287
Number of amino acid residues	476
Number of solvent molecules	53
Number of protein nonhydrogen atoms	794
Rmsd	
Bond lengths	0.007 Å
Bond angles	1.44°

scFv1696/HIV-2 Pr complex, which agree with values found previously for Fab1696 and IgG1696 [11].

Quality of the Model and Overall Structure

Refinement statistics from the scFv1696-peptide crystal structure are given in Table 1. The final model includes 113 residues for the light chain and 120 residues for the heavy chain for each of the two scFv1696 molecules in the asymmetric unit. The two crystallographically independent complexes were tightly restrained during refinement by imposing the noncrystallographic symmetry relating them (see Experimental Procedures). All residues from the model fall into favorable or allowed regions of the Ramachandran energy space, with the exception of Ala-51L, since this residue forms a γ turn [14], which is characteristic of the unique canonical conformation of CDR-L2. Residues belonging to the hypervariable loops were well defined in the electron density map, and the bound peptide could be unambiguously traced for six of its nine residues from Pro-P1 to Trp-P6 in both complexes present in the asymmetric unit (Figure 2). Although some weak density is visible for main chain atoms of Gln-P7, this residue is not included in our current model. Other missing residues from our model belong to the flexible (Gly₄Ser)₃ linker, which connects the C terminus of the light chain to the N terminus of the heavy chain. A total of 53 well-defined water molecules have been placed using difference maps. Of these, 32 are conserved between the two molecules.

Peptide-Antibody Interactions

In the scFv1696-peptide complex, the antigenic peptide adopts an extended conformation in a shallow groove that runs across the antigen binding site. This binding groove is partially closed at the N-terminal end of the

peptide but is more open to the solvent at its C terminus, where the peptide, although not visible in this region, presumably extends out from the antigen binding site. The total molecular surface area buried upon complex formation is 498 Å² for the peptide (representing 72% of its total accessible surface area of the six N-terminal residues) and 443 Å² for the antibody. Overall, the interaction is reminiscent of that seen in MHC-peptide complexes, with a few residues from the antigen being buried in specificity pockets, serving as anchor residues, and others pointing toward the solvent, making only a few or no contacts [15].

The antigenic peptide interacts primarily with the heavy chain of the antibody. This is reflected in the buried surface area per CDR upon complex formation, where that from the heavy chain accounts for 70% of the total buried surface. The hypervariable loops CDR-H1 and CDR-H2 of the antibody establish the largest number of van der Waals contacts (Figure 3) and account for eight of the nine hydrogen bonds seen in the complex (Table 2). As seen in other antibody-peptide complexes, hypervariable region CDR-L2 does not contact the antigen [16]. With the exception of peptide residue Thr-P4, whose side chain points toward the exterior of the antigen binding groove and does not make any contact with the antibody, the remaining ordered residues from the peptide establish several contacts. The most extensive polar interactions are made by Gln-P2, which is buried in a pocket formed by residues protruding from CDR-H1 and CDR-H2, losing 76% of its solvent-accessible surface upon binding. As confirmed by the low temperature factors for its main and side chain atoms, this residue is well stabilized by several polar interactions with the antibody (Table 2). Two other residues, Pro-P1 and Trp-P6, lose 68% and 81%, respectively, of accessible surface area upon complex formation. Pro-P1 and Trp-P6 are buried in pockets formed by acidic residues Asp-H31 and Asp-H98, respectively, and by hydrophobic residues Phe-L94 and Tyr-H47, respectively. The formation of a hydrogen bond between the secondary amino group of the N-terminal Pro-P1 and the carbonyl group of Asp-H31 is consistent with our observation that mAb1696 recognizes the mature processed form of HIV-1 Pr, but not a precursor form containing 20 amino acids upstream from the cleavage site. In the latter unprocessed form, the main chain secondary amide group of the proline residue is not capable of hydrogen bond formation. Furthermore, the presence of a partially closed pocket containing negatively charged residues could neutralize the positive charge of the N terminus of the peptide or the mature protease. Toward its C-terminal end, by contrast, recognition does not seem to be restricted by the length of the peptide, as shown by the broad reactivity between scFv1696 and longer peptides like P1–P13 or the entire protease. The rather small number of well-ordered peptide residues in the binding groove compared with other antibody-peptide complexes is probably compensated by the large percentage of buried surface area upon complex formation. Furthermore, the large hydrophobic interaction could offset a high entropic cost due to the high mobility of the peptide in solution [17].
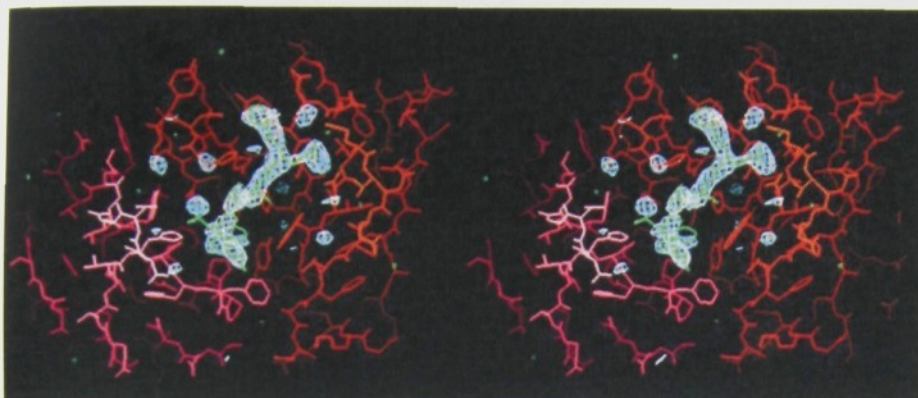
Figure 2. Stereo View from above the Antigen Binding Site

An omit-electron density map calculated with coefficients (F_o-F_c) and phases from the refined model without the peptide residues, contoured at 2.5 rms level. The V_H domain is in red, the V_L domain is in violet, and the peptide and solvent molecules are in green.

Comparison between the Complexed scFv1696 and the Free Fab1696 Structures

To evaluate the contribution of conformational changes within the antigen binding site upon formation of the antibody-peptide complex, the scFv1696-peptide complex was compared with the variable dimer of the nonliganded Fab1696 structure [11]. The crystallographic asymmetric unit of the scFv1696 complex contains two molecules that have been tightly restrained by the noncrystallographic symmetry relating them. Therefore, only one molecule of scFv1696 is used here for comparison.

The overall rms difference in the α-carbon positions of the variable dimers is 1.04 Å after superposition. Part of this deviation is due to a difference in the relative orientation between the V_H and V_L domains of the complexed and nonliganded states of 1696. Thus, after superimposing the V_L domains of the two forms, the V_H domain of the free 1696 variable dimer required an additional 4.3° rotation about its centroid (and a 0.2 Å translation) to give optimum superposition onto V_H of the complexed form. Superimposing the V_H and V_L domains individually gives an rms difference in α-carbon positions of 1.20 Å and 0.51 Å, respectively. These differences result mainly from the adaptation of the CDRs to the antigen upon complex formation (see Figure 4). The most conspicuous differences occur within CDR-H3, which undergoes a large conformational change upon complex formation. When CDR-H3 is excluded from the same pairwise comparison, the rmsd is 0.49 Å. Indeed, all α-carbon atoms of the segment comprising residues Arg-H96–Glu-100A move by more than 2.0 Å between their antigen-bound and free states, with the α-carbon of Tyr-H99 deviating by 7.0 Å (Figure 4). Two of these residues, His-H97 and Asp-H98, establish direct van der Waals contacts with Pro-P1 from the antigenic peptide (Figure 3). This large conformational change thus results in the formation of a binding site with significantly different topology from that of the free Fab1696 (Figures 4 and 5). Hence, the recognition of the antigenic HIV-1 Pr peptide by mAb1696 provides a clear example of an
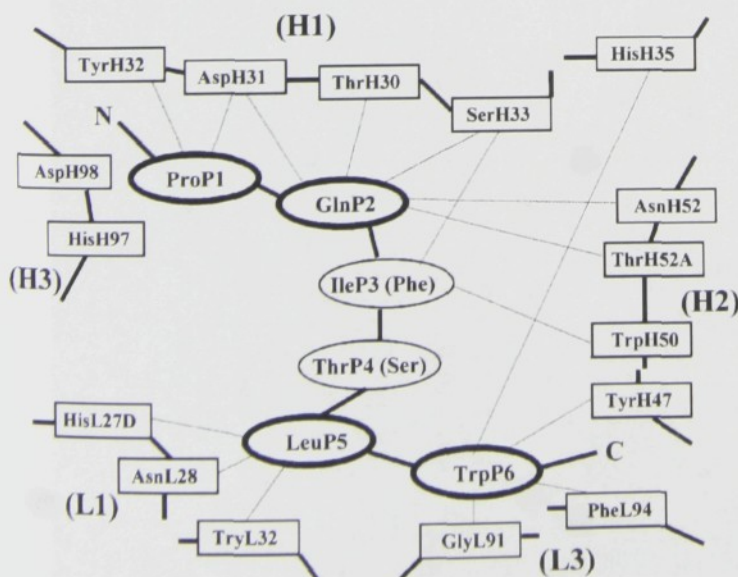

Figure 3. Schematic View of the scFv1696-Peptide Complex

All scFv1696 residues in contact with the peptide are indicated.

Table 2. Polar Contacts in the scFv1696-HIV-1 Pr-Peptide Complex

Peptide Atom	ScFv1696 Atom	Distance scFv1 (Å)	Distance scFv2 (Å)
Pro P1 N	Asp H31 O	2.40	2.79
Gln P2 Nε2	Thr H30 O	2.72	2.82
Gln P2 N	Asp H31 O	3.24	3.23
Gln P2 Oε1	Ser H33 Oγ	2.76	2.96
Gln P2 Oε1	Asn H52 Oδ1	3.21	3.07
Gln P2 Oε1	Thr H52A N	2.89	2.71
Gln P2 Nε2	Thr H52A Oγ1	2.98	3.22
Ile P3 N	Ser H33 Oγ	3.17	3.13
Trp P6 Nε1	Gly L91 O	2.73	2.81

induced fit mechanism for antibody-antigen recognition. Other examples include the recognition of a hemagglutinin peptide by antibody 17/9 [18] and the peptide 36–46 from the HIV-1 Pr [10], where large conformational changes in CDR-H3 are also associated with peptide binding. Of note, the presence of glycine at position 100 in CDR-H3 of mAb1696 (as with Gly-H100B in mAb 17/9 and Gly-H100C in mAb F11.2.32) allows for a larger flexibility of this hypervariable loop.

Conformation of Peptide
The six well-ordered residues from the antigenic peptide adopt a rather extended structure, with several of its backbone torsion angles (residues P2, P3, P4) lying in the β region of Ramachandran space. In contrast to several other peptide-antibody complexes, where the peptide ligand adopts a more compact conformation such as β turns [17], the HIV-1 Pr peptide has no intramolecular interactions that would stabilize the observed conformation in solution. The conformation adopted by the bound peptide is roughly similar to that seen in the native protease, since the 20 main chain atoms of residues P1–P6 can be superimposed with the corre-

sponding atoms from the HIV-1 Pr with an overall rmsd of 1.8 Å (Figure 6). Nevertheless, the detailed conformation of the segment P1–P6 is different in the two structures. In particular, the orientations of the side chains of the N- and C-terminal residues bear no similarity to those in the corresponding segment in the protease.

Cross-Reactivity with HIV-2 Pr N-Terminal Peptide PQFSLWKR
Although mAb1696 was obtained by immunization with the entire HIV-1 Pr, it is nevertheless able to cross-react with HIV-2 Pr with a 3-fold increase in affinity [11]. In addition, peptides containing the N-terminal region of HIV-1 Pr, such as P1–P7 or P1–P13, are able to effectively compete with the protease in binding to the antibody.

Even though we have not succeeded so far in crystallizing the complex formed between mAb1696 and the protease, it is of interest nevertheless to interpret the cross-reactivity in light of our current crystal structure. In the crystal structure of the scFv1696-HIV-1 Pr peptide complex, several residues of the antigenic peptide are deeply buried in complementary pockets formed by antibody residues. This would impose constraints on sequence variability at these positions in order for antigen recognition to be maintained. The cross-reactivity of 1696 with the HIV-2 Pr peptide thus probably stems from the invariance between the two isolates of the four residues situated at positions P1, P2, P5, and P6 of the protease sequence and the conservative substitutions Ile-P3 → Phe and Thr-P4 → Ser, all of which lose more than 60% of their accessible surface area upon complex formation. The substitution Gln-P7 → Lys and further substitutions beyond position 7 would probably not affect the interaction, although they could affect the unbound conformation of the antigen. Other amino acid substitutions at position 4 should be tolerated since the
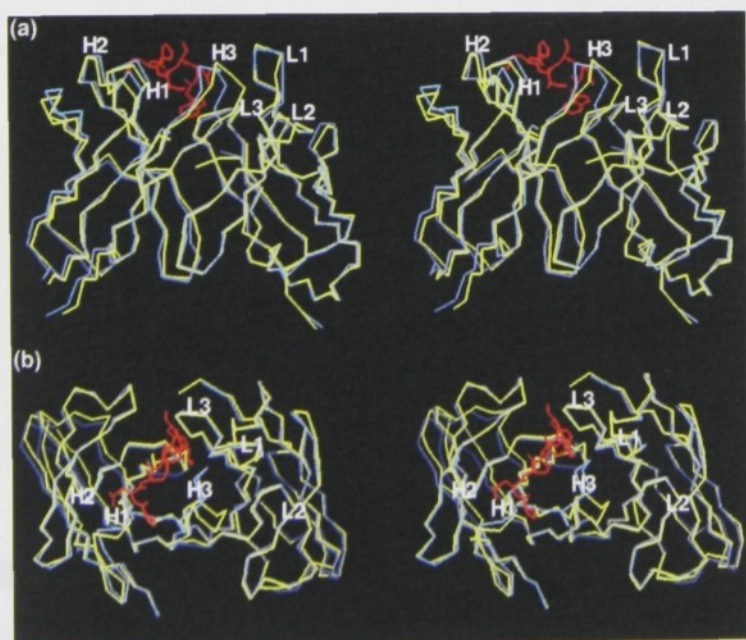

Figure 4. Comparison of the Variable Domains of Noncomplexed Fab1696, PDB Entry 1c17, with the scFv1696 Peptide Complex

(a) A side view.
(b) Viewed from above the antigen binding site.
The noncomplexed state is shown in blue, the complexed state is shown in yellow, and the peptide is shown in red. Superpositions were made by optimizing the correspondence between the V_L domains.
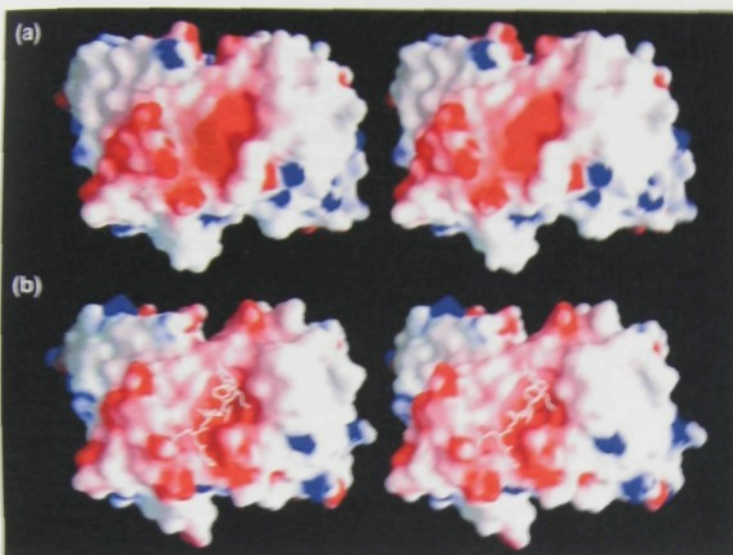
Figure 5. An Overhead View of the Topology of the Antigen Binding Site

(a) The nonliganded Fab1696 scFv1696 complex.
(b) The liganded scFv1696 complex
The surfaces are color coded for electrostatic potential in the range of -10 k_BT(red) to $+10$ k_BT (blue), where k_B is the Boltzmann constant and T is the absolute temperature. This figure was prepared with GRASP [41].

side chain of this residue does not make direct contacts with the antibody. The substitution Ile-P3 → Phe retains the hydrophobic character but must accommodate a slightly larger side chain volume; the phenyl ring could, however, provide favorable aromatic stacking interactions with Trp-H50.

It is not clear from the crystal structure why mAb1696 suffers a loss of affinity toward HIV-2 Pr at lower pH while that for HIV-1 is maintained [11], since the disordered residue P7 (Gln → Lys) cannot be in close contact with any acidic residue of the antigen binding site. The different behavior as a function of pH could arise from the difference in isoelectric points (theoretical values obtained from the ExPASy server are 8.8 and 5.3 for HIV-1 Pr and HIV-2 Pr, respectively) and/or a difference in pH dependence of the dimerization constant [19].

Mechanism of inhibition

In our previous report, we showed, by using size-exclusion chromatography, that mAb1696 was able to form specific molecular complexes with both the HIV-1 and HIV-2 proteases [11]. In addition, the stoichiometry of the complex was evaluated as one protease monomer per 1696 binding site. Since the structural similarity between the bound antigenic peptide and the corresponding segment of the native protease is only approximate (Figure 6), it is not possible to make an unequivocal proposition for docking the protease itself onto the antigen binding site of scFv1696. Nevertheless, it is interesting to use the orientation taken by the peptide within the antigen binding site to generate possible models for the HIV-protease in complex with the antibody. These models show that binding of the segment 1–6 by mAb 1696 is not possible in the active homodimer, mainly because of steric hindrance with the adjacent C-terminal strand in the 4-stranded β-pleated sheet. This is in agreement with the measured stoichiometry of the complex and also with binding studies in the presence of an active-site inhibitor, which suggest that 1696 binds to the monomer [11].
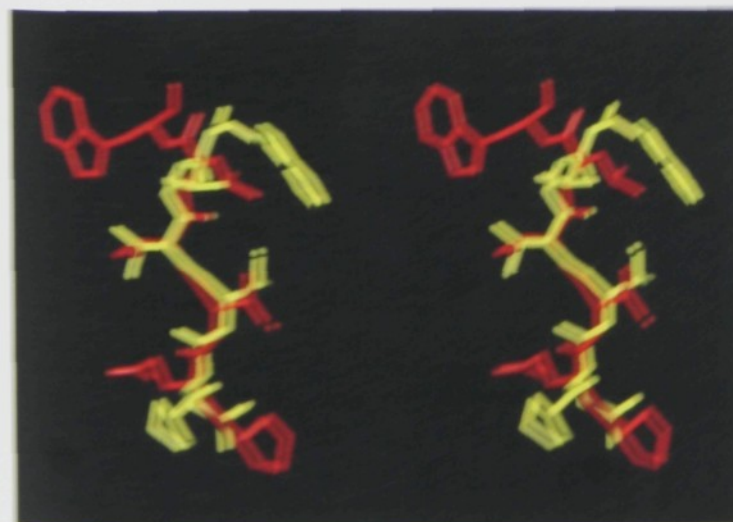

Figure 6. Comparison of the Structures Adopted by the Segment P1–P6 of HIV-1 Pr
The complex with scFv1696 is shown in yellow, and in the native protease (PDB entry 9hvp), it is shown in red. The superposition was made by minimizing the rms difference in the main chain atomic positions of the P1–P6 segment (rms difference 1.2 Å).

Although docking one subunit only of the native protease also generates steric hindrance, the N- and C-terminal segments are probably disordered in the absence of the 4-stranded intermolecular antiparallel β sheet, which maintains the dimeric structure. Indeed, a mature protease monomer lacking residues 1–4 has been shown by NMR to be significantly unfolded in the absence of an active-site inhibitor [20]. It is possible that, during immunization, the protease was presented in a partially denatured monomeric form and that mAb1696 was elicited against a nonnative structure of the enzyme. This possibility has also been proposed for the anti-HIV-1 Pr mAb F11.2.32 [10] and the anti-polymerase Taq antibody TP7 [21], where the authors suggest that TP7 would trap a conformer within a set of thermodynamically accessible conformations of the enzyme.

Two hypotheses thus emerge that could explain the inhibition of HIV-Pr activity by mAb1696. First, by binding to the N-terminal region of the native homodimeric enzyme, which is located at the external surface in the homodimeric enzyme structure and is thus relatively accessible to antibody molecules, mAb1696 could disrupt the dimeric interface of the enzyme, leading to its dissociation. Second, the dimer-monomer equilibrium of the enzyme could be shifted toward inactive monomeric species by capture of the latter by mAb1696. Although scFv1696 inhibits HIV-1 Pr nine times more efficiently than it inhibits HIV-2 Pr, measurements of affinity constants by indirect ELISA, made under buffer conditions identical to those of inhibition, show that mAb1696 and Fab1696 have a 3-fold higher affinity for HIV-2 Pr in comparison to HIV-1 Pr [11]. This apparent difference may be due to unaccounted systematic errors in either of the two kinds of experimental measurements. Nonetheless, the two experiments are different, one being a kinetic measure of inhibition and the other being a binding study under equilibrium conditions. Moreover, inhibition might conceivably begin before complete dissociation of the protease dimer has occurred if the initial interaction with the antibody distorts the structure sufficiently. This interpretation would be more in line with the first hypothesis for the inhibition mechanism.

Biological Implications

In this study, we have determined the three-dimensional structure of a bacterially expressed scFv of the anti-HIV-1 protease monoclonal antibody 1696 in complex with a peptide fragment that contains the linear epitope of the viral enzyme. The broad reactivity of mAb1696, which allows it to bind to both HIV-1 Pr and HIV-2 Pr with high affinity, derives from invariant amino acids or conservative substitutions within the first six residues of the N-terminal segment of the two isolates. Most of these residues are deeply buried at the antibody-antigen interface and establish extensive contacts to the antibody. Using the peptide as a guide, a docking complex of a whole protease monomer was generated that supports the hypothesis that mAb1696 inhibits HIV-Pr by favoring the dissociation of the active homodimer.

For viruses such as HIV, which express polyprotein precursors, a careful regulation of the enzymatic proteolysis of their polypeptide chains is required to ensure correct virus assembly. By coupling mAb1696 with a fluorescent dye or protein, one could use it as a probe to elucidate the pathway that leads to the formation of a mature virus within an infected cell. Since mAb1696 inhibits the HIV-Pr from both the HIV-1 and HIV-2 isolates with inhibition constants in the low nanomolar range, it is therefore also a good candidate for studies aimed at controlling the viral infection with intracellularly expressed antibodies [22].

Experimental Procedures

Inhibition of HIV-1 and HIV-2 Pr by scFv1696
Inhibition of HIV Pr activity by scFv1696 was measured from the cleavage rate of the substrate KARVNle-F(NO_2)EANle following methods described in [23, 24]. Reactions were carried out in PBS buffer (pH 7.4) with 1% (v/v) Tween 20, 1% (w/v) BSA, 0.02% (v/v) β-mercaptoethanol, and 1 mM EDTA. Native HIV-2 Pr and the Q7K/L33I/L63I HIV-1 Pr mutant [25], which is highly resistant to autolysis while still retaining all enzyme kinetic properties, were used for inhibition measurements. A fixed quantity of HIV Pr (70 nM and 125 nM for HIV-1 Pr and HIV-2 Pr, respectively) was first incubated with varying amounts of scFv1696 for 20 min at 37°C. The reaction was then started by adding a constant amount of substrate solution (final concentration 0.135 nM), and aliquots of the reaction mixture, which were removed at fixed time intervals over a total period of 1 hr, were stopped by the addition of 20% (v/v) TFA. The quantity of cleavage product from each aliquot was determined by reverse-phase HPLC (Merck) using a VYDAC C18 column (4.5 mm × 150 mm) and a methanol/water gradient. Kinetic data were analyzed with ENZFITTER (version 1.0.3, R.L. Latherbarrow, Elsevier-Biosoft, 1987).

Cloning, Expression, Purification, and Complex Preparation
The coding sequences for V_L and V_H variable domains of mAb1696 were obtained by RT-PCR from the total mRNA of hybridoma 1696 cells, as had already been reported [11]. A corresponding scFv construct contains the 113 N-terminal residues from V_L and the 120 residues from V_H, covalently linked together by the 15-residue $(Gly_4Ser)_3$ linker [26, 27]. For cloning, an intermediary vector based on pBluescript was first assembled by sequential insertion of oligonucleotide duplexes, introducing an initiator ATG codon (as part of the NdeI restriction site) and the codon for the first V_L residue, the linker sequence flanked by the last two V_L residues and the first three V_H residues, and unique restriction sites allowing cassette-like insertion of the V_L (EcoRV-XhoI) and V_H (PstI-EcoRI) domains. Prior to the V_H domain insertion, the C terminus of this fragment was modified by PCR so that the terminator codon and an EcoRI site could be introduced. Finally, the scFv1696-encoding region was cloned as the NdeI-EcoRI fragment into a T7 promoter-driven expression plasmid.

BL21(DE3) E. coli cells, when transformed with the above expression plasmid, accumulated the scFv1696 product as insoluble inclusion bodies after induction with 1 mM IPTG. A 1-liter bacterial culture yielded typically about 10 g wet cells and about 1 g washed inclusion bodies. Solubilization and refolding protocols were adapted from [28]: inclusion bodies were solubilized in 50 mM Tris-HCl (pH 8.0) with 2% N-lauroyl sarcosine (SLS, Sigma). The solution was vigorously stirred overnight at room temperature in the presence of 50 μM $CuSO_4$ to oxidize SH groups by air. Urea was then added to 6 M concentration, and the SLS was removed by retention on anion-exchange resin (Dovex). Refolding was performed by diluting the denatured protein and by dialysing into 10 mM borate buffer (pH 8.0). Purification included three chromatography steps, using DEAE-Sephacel, Sephadex G-75, and Mono-Q 5/5 HR columns (Pharmacia). The fraction content was checked throughout the purification by SDS-PAGE on a 15% acrylamide gel and analytical HPLC gel filtration using a Superose 12 column (Pharmacia). Monomeric and dimeric forms of scFv1696 in equilibrium were observed. The purified protein was transferred by dialysis into 10 mM Tris-HCl (pH

7.5), 5% glycerol and was stored at −25°C. The yield was about 6 mg of the purified scFv protein from 1 liter of bacterial culture.

Crystallization and Data Collection
The complex was prepared by mixing scFv1696 with a 3-fold molar excess of peptide PQITLWQRR. Monomeric and dimeric forms of the complex were separated by FPLC using a Mono-Q 5/5 HR column (Pharmacia). In contrast to free scFv1696, these complexed forms are stable and do not interchange. The complex was concentrated by ultrafiltration on Centricon-10 (Millipore) to 14 mg/ml. Crystallization trials were performed by the hanging-drop technique using the sparse matrix method [29] as well as with a series of different PEG precipitants. Crystals were obtained under the following conditions: 1 μl of the precipitant (0.05 M tri-sodium citrate, 0.1 M sodium phosphate (pH 5.5), 24% PEG 3400, 0.2 M ammonium sulfate) was mixed with 2 μl of the solution containing the complex of scFv1696 and the peptide at a concentration of 14 mg/ml and was allowed to equilibrate over a reservoir of 1 ml of the precipitating solution at 18°C. After several days, needle-shaped crystals appeared (the smallest dimension was 20 μm).

Crystals were transferred for a few seconds to the reservoir buffer with 25% (v/v) glycerol and were cryo-cooled to 100°K in a nitrogen gas stream (Oxford Cryosystem). Measurements were made at an X-ray wavelength of 0.93 Å on the ID14-1 beam-line at the European Synchrotron Radiation Facility (ESRF) (Grenoble) using one crystal. Diffraction intensities, recorded on a MarCCD system, were integrated and scaled with the HKL suite of programs [30]. The space group is P2₁, with unit cell dimensions a = 45.50 Å, b = 57.05 Å, c = 91.03 Å, β = 97.10° and two scFv1696-peptide complexes per asymmetric unit. Data collection statistics are summarized in Table 1.

Structure Solution and Refinement
AMoRe [31] was used for molecular replacement calculations. Rotation function calculations were performed between the resolution limits of 20.0 Å and 4.0 Å using the variable dimer of Fab1696 (PDB code 1lc7) as a search model and a Patterson integration radius of 20 Å. Rotation searches returned two unambiguous solutions (correlation coefficients of 34.1% and 27.5% compared to 16.0% for the highest noise peak), which, in turn, gave clear results in the translation function. A 2-body translation search confirmed these solutions and placed the two independent scFv fragments on a common crystallographic origin. Rigid body refinement of the two molecules converged to a correlation coefficient of 70.3% and an R factor of 31.9% for 3636 reflections between 9.0 Å and 4.0 Å resolution. For control, a self-rotation calculation using data between 20.0 Å and 4.0 Å and an integration sphere of 15 Å diameter yielded a peak (height was 31% of the origin peak) corresponding to the rotation needed to superimpose the two crystallographically independent molecules.

All the CDRs were omitted from the initial model, which was subjected to alternate cycles of molecular dynamics refinement with CNS [32], using the parameter set of Engh and Huber [33], and manual model building using QUANTA (Molecular Simulations) and O [34]. R_{free} [35] was used to monitor the progress of the refinement by omitting 10% of the data (1286 reflections), randomly chosen, throughout the procedure. The main chain tracing was carried out using electron density maps calculated with weighted Fourier coefficients [36]. Using this procedure, the CDRs were successively built as polyalanine chains, and identifiable side chains were modeled accordingly as refinement progressed. During the entire refinement procedure, tight noncrystallographic symmetry restraints were applied to the scFv1696 complexes (with a constant of 300 kcal/mol). Once the model for the scFv1696 molecules had been completed, connected density visible at the antigen binding site could be interpreted as the N-terminal proline and glutamine residues of the bound peptide. Subsequent cycles of refinement and model building allowed residues P1–P6 of the peptide to be fitted in difference electron density maps. The final statistics are given in Table 2. The quality of the model was controlled with PROCKECK [37]. Superposition of the various molecular components (PDB code 3hvp) was carried out using LSQKAB from the CCP4 program suite [38]. Buried surface areas were calculated using MS [39], using a probe radius of 1.7 Å

and water molecules excluded from the calculation. Antibody residues are numbered using the Kabat convention [40], preceded by L or H for light and heavy chain, respectively.

References

1. Kohl, N.E., et al., and Sigal, I.S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. Proc. Nat. Acad. Sci. USA 85, 4686–4690.

2. Rosé, J., Babé L., and Craik, C. (1995). Defining the level of human immunodeficiency virus type I (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. J. Virol. 69, 2751–2758.

3. Katz, R.A., and Skalka, A.M. (1994). The retroviral enzymes. Ann. Rev. Biochem. 63, 133–173.

4. Rana, K.Z., and Dudley, M.N. (1999). Human immunodeficiency virus protease inhibitors. Pharmacotherapy 19, 35–59.

5. Wlodawer, A., and Erickson, J. (1993). Structure-based inhibitors of HIV-1 protease. Ann. Rev. Biochem. 62, 543–585.

6. Babé, L.M., Pichuantes, S., and Craik, C.S. (1991). Inhibition of HIV protease activity by heterodimer formation. Biochemistry 30, 106–111.

7. Babé, L.M., Rosé, J., and Craik, C.S. (1992). Synthetic interface peptides alter dimeric assembly of the HIV1 and 2 proteases. Protein Sci. 1, 1244–1253.

8. Schramm, H.J., Nakashima, H., Schramm, W., Wakayama, H., and Yamamoto, N. (1991). HIV-1 reproduction is inhibited by peptides derived from the N and C-termini of HIV-1 protease. Biochem. Biophys. Res. Comm. 179, 847–851.

9. Lescar, J., et al., and Bentley, G.A. (1996). Preliminary crystallographic studies of an anti-HIV-1 protease antibody which inhibits enzyme activity. Protein Sci. 5, 966–968.

10. Lescar, J., et al., and Bentley, G.A. (1997). Three-dimensional structure of an Fab-peptide complex: structural basis of HIV-1 protease inhibition by a monoclonal antibody. J. Mol. Biol. 267, 1207–1222.

11. Lescar, J., et al., and Bentley, G.A. (1999). Inhibition of the HIV-1 and HIV-2 proteases by a monoclonal antibody. Protein Sci. 8, 2686–2696.

12. Dall'Acqua, W., and Carter, P. (1998). Antibody engineering. Curr. Opin. Struct. Biol. 8, 443–450.

13. Williams, J.W., and Morrison, J.F. (1979). The kinetics of reversible tight-binding inhibition. Methods Enzymol. 63, 437–467.

14. Milner-White, E.J., Ross, B.M., Ismail, R., Belhady-Mostefa, K., and Poet, R. (1988). One type of gamma turn rather than the other gives rise to chain reversal in proteins. J. Mol. Biol. 204, 777–782.

15. Madden, D.R. (1995). The three-dimensional structure of peptide-MHC complexes. Ann. Rev. Immunol. 13, 587–622.

16. Wilson, I.A., and Garcia, K.C. (1997). T-cell receptor structure and TCR complexes. Curr. Opin. Struct. Biol. 6, 839–848.

17. Stanfield, R.L., and Wilson, I.A. (1995). Protein peptide interactions. Curr. Opin. Struct. Biol. 5, 103–113.

18. Rini, J.M., Schulze-Gahmen, U., and Wilson, I.A. (1992). Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science 255, 959–965.

19. Darke, P.L. (1994). Stability of dimeric retroviral proteases. Methods Enzymol. 241, 104–127.

20. Louis, J.M., Clore, G.M., and Gronenborn, A.M. (1999). Autoprocessing of HIV-1 protease is tightly coupled to protein folding. Nat. Struct. Biol. 6, 868–875.

21. Murali, R., Sharkey, D.J., Daiss, J.L., and Murthy, H.M.K. (1998). Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: the Fab is directed against an intermediate in the helix-coil dynamics of the enzyme. Proc. Nat. Acad. Sci. USA 95, 12562–12567.

22. Rondon, I.J., and Marasco, W.A. (1997). Intracellular antibodies (intrabodies) for gene therapy of infectious diseases. Ann. Rev. Microbiol. 51, 257–283.

23. Urban, J., et al., and Strop, P. (1992). Reduced-bond tight-binding inhibitors of HIV1 protease. FEBS Lett. 298, 9–13.

24. Majer, P., et al., and Strop, P. (1993). Specificity mapping of HIV-1 protease by reduced bond inhibitors. Arch. Biochem. Biophys. 304, 1–8.

25. Mildner, A.M., et al., and Tomasselli, A.G. (1994). The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. Biochemistry 33, 9405–9413.

26. Huston, J.S., et al., and Oppermann, H. (1988). Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. Proc. Natl. Acad. Sci. USA 85, 5879–5883.

27. Johnson, S., and Bird, R.E. (1991). Construction of single-chain Fv derivatives of monoclonal antibodies and their production in Escherichia coli. Methods Enzymol. 203, 88–98.

28. Kurutz, I., Titus, J.A., Jost, C.R., and Segal, D.M. (1995). Correct disulphide pairing and efficient refolding of detergent-solubilized single-chain Fv proteins from bacterial inclusion bodies. Mol. Immunol. 32, 1443–1452.

29. Jancarik, J., and Kim, S.-H. (1991). Sparse matrix: a screening method for crystallization of proteins. J. Appl. Crystallogr. 24, 409–411.

30. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326.

31. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallogr. A 50, 157–163.

32. Brunger, A.T., et al., and Warren, G.L. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D 54, 905–921.

33. Engh, R.A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr. A 47, 392–400.

34. 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 Crystallogr. A 47, 110–120.

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

36. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A 42, 140–149.

37. Laskowski, R.A., McArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.

38. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 50, 760–763.

39. Connolly, M.L. (1983). Solvent accessible surfaces of proteins and nucleic acids. Science 221, 709–713.

40. Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., and Foeller, C. (1991). Sequences of proteins of immunological interest. In U.S. Department of Health and Human Services, 5th edn. (Bethesda MD: National Institutes of Health), pp. xvi.

41. Nicholls, A., Sharp, K.A., and Honig, B. (1991). Proteins folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins Struct. Funct. Genet. 11, 281–296.

Accession Numbers

Atomic coordinates have been deposited in the Protein Data Bank under entry code 1jp5.

Structure of a single-chain Fv fragment of an antibody that inhibits the HIV-1 and HIV-2 proteases

Julien Lescar,[a]† Jiri Brynda,[b] Milan Fabry,[b] Magda Horejsi,[b] Pavlina Rezacova,[b] Juraj Sedlacek[b] and Graham A. Bentley[c]*

[a]European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble, France, [b]Department of Gene Manipulation, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 166 37 Prague 6, Czech Republic, and [c]Unité d'Immunologie Structurale (URA 2185 CNRS), Département de Biologie Structurale et Chimie, Institut Pasteur, 25 Rue du Dr Roux, 75724 Paris CEDEX 15, France

† Present address: School of Biological Sciences, Nanyang Technological University, 1 Nanyang Walk Blk 5 Level 3, Singapore 637616.

Correspondence e-mail: bentley@pasteur.fr

The monoclonal antibody 1696, which was raised against the HIV-1 protease, inhibits the catalytic activity of the enzyme from both the HIV-1 and HIV-2 strains. The antibody cross-reacts with peptides containing the N-terminus of the enzyme, which is highly conserved between these strains. The crystal structure of a single-chain Fv fragment of 1696 (scFv-1696) in the non-complexed form, solved at 1.7 Å resolution, is compared with the previously reported non-complexed Fab-1696 and antigen-bound scFv-1696 structures. Large conformational changes in the third hypervariable region of the heavy chain and differences in relative orientation of the variable domains are observed between the different structures.

1. Abbreviations and symbols

CDR, complementarity-determining region; L1, L2, L3, first, second and third CDR, respectively, of the light chain; H1, H2, H3, first, second and third CDR, respectively, of the heavy chain; HIV, human immunodeficiency virus; mAb, monoclonal antibody; r.m.s.d., root-mean-square deviation; scFv, single-chain Fv fragment; V_H, heavy-chain variable domain; V_L, light-chain variable domain.

Antibody residues are numbered using the Kabat convention, preceded by L or H for light and heavy chain, respectively.

2. Introduction

Proteins from HIV are expressed in the form of polyprotein precursors that are cleaved by a protease encoded in the viral genome. Because it plays an essential role in the infectious cycle of the virus, HIV protease is one of the primary targets for antiviral drug design (Wlodawer & Gutschina, 2000). We have produced monoclonal antibodies raised against HIV-1 protease in order to study the enzyme's activity and function and to aid in the design of novel inhibitors targeted to regions other than the active site. Two anti-HIV-1 protease mAbs, F11.2.32 and 1696, which inhibit the enzymatic activity of the protease have been characterized (Lescar et al., 1996, 1997, 1999). We have recently reported the crystal structure at 2.7 Å resolution of a cross-reaction complex formed between scFv-1696 and a peptide, PQITL-WQRR, whose sequence is derived from the N-terminus of the HIV-1 protease (Rezacova et al., 2001). This structure supports the hypothesis that mAb-1696 inhibits the HIV

protease by favouring the dissociation of the active homodimer. The crystal structure of the ligand-free scFv-1696 molecule that we describe here reveals a plasticity in the topology of the antigen-binding site of this mAb. Crystals of non-complexed scFv-1696 were obtained under acidic conditions (pH 3.5). Here, we compare this structure, determined at 1.7 Å resolution, with that of non-complexed Fab-1696 crystallized at pH 7.5 (Lescar et al., 1999) and the antigen-bound scFv-1696 crystallized at pH 5.5 (Rezacova et al., 2001).

3. Experimental

3.1. Crystallization and data collection

Needle-shaped crystals of approximate dimensions $0.15 \times 0.01 \times 0.01$ mm were obtained by the hanging-drop technique using the following conditions: 5 µl of precipitant (0.1 M sodium citrate pH 3.5, 17.5% PEG 4000, 0.175 M ammonium sulfate) was mixed with an equal volume of a solution of scFv-1696 at a concentration of 5 mg ml^{-1} in 10 mM Tris–HCl pH 7.3. The drop was suspended over a reservoir containing 1 ml of the precipitant solution at 291 K and allowed to equilibrate. For data collection, crystals were briefly transferred to the reservoir buffer containing 25%(v/v) glycerol before being cryocooled to 100 K in a nitrogen-gas stream. Measurements were made with a MAR Research 345 imaging-plate system using an X-ray wavelength of 0.990 Å on the ID2 beamline at the ESRF (Grenoble). Owing to the small size of the crystals, it was not possible to obtain a complete data set from one crystal alone without incurring severe radiation damage. Three crystals were therefore used to obtain an acceptably complete

short communications

Table 1

Data set and refinement statistics.

Values in parentheses are for the highest resolution shell (1.79–1.70 Å).

Space group	$P2_12_12_1$
Unit-cell parameters	
a (Å)	126.93
b (Å)	61.21
c (Å)	37.30
Resolution range (Å)	20.0–1.70
Unique reflections	45422 (5930)
Data redundancy	4.89 (3.34)
Completeness (%)	91.0 (84.2)
$I/\sigma(I)$	34.3 (6.6)
R_{merg} (%)	9.0 (36.0)
V_M (Å³ Da⁻¹)	2.07
R_{work} (%)	22.7
R_{free} (%)	25.6
No. of protein atoms	3656
No. of heteroatoms	2
No. of solvent molecules	530
R.m.s.d. bond lengths (Å)	0.005
R.m.s.d. bond angles (°)	1.34
Ramachandran plot	
Core region (%)	87.8
Additionally allowed regions (%)	11.2
Generously allowed regions (%)	0.5
Disallowed region (%)	0.5
Average B factors, all protein atoms (Å²)	28.1
R.m.s.d. B factors, all protein atoms (Å²)	1.7

data set. Integration, scaling and merging of the intensities were carried out using the program *XDS* (Kabsch, 1988), with data frames from the three crystals being scaled and merged simultaneously. Details of crystal parameters and data-collection statistics are given in Table 1.

3.2. Solution and refinement of the structure

A preliminary model of the structure was found by molecular replacement using the program *AMoRe* (Navaza, 1994). The rotation-function calculation was performed between resolution limits of 20.0 and 4.0 Å using the variable dimer (V_H/V_L) of Fab-1696 (PDB code 1lc7) as the search model and a Patterson integration radius of 20 Å. Two strong peaks in the rotation function (correlation coefficients of 0.173 and 0.171, compared with 0.138 for the third highest peak) clearly indicated the presence of the two anticipated molecules in the asymmetric unit. The two scFv molecules were then placed in the unit cell using one- and two-body translation searches. Rigid-body refinement of the two molecules yielded a correlation coefficient of 0.443 and an R factor of 0.450 for 3592 reflections in the resolution range 9.0–4.0 Å resolution.

The structure was refined using the program *CNS* (Brünger *et al.*, 1998). After omitting the hypervariable loops from the molecular-replacement model, the atomic parameters were refined using all measured

data between resolution limits of 20 and 1.7 Å, with the exception of 2280 randomly chosen reflections (5% of the data) used to monitor the refinement progress by the free R factor (R_{free}). As refinement proceeded, the hypervariable loops were progressively built into the electron-density maps as poly-alanine chains and identifiable side chains were modelled accordingly. Tight non-crystallographic symmetry restraints were applied during the initial phases of the refinement, but were relaxed during the final stages. Once the model for the scFv-1696 molecules had been completed, water and solute molecules were fitted to difference electron-density maps. Structure-refinement statistics are summarized in Table 1.

4. Results

4.1. Description of the structure

The final model includes 113 residues for the light chain and 120 residues for the heavy chain for both scFv-1696 molecules. All residues in the model fall into favourable or allowed regions of the Ramachandran plot except ValL51, which is contained in the γ-turn defining the unique canonical structure of CDR-L2. The flexible $(Gly_4Ser)_3$ linker connecting the C-terminus of the heavy chain to the N-terminus of the light chain was not visible for either molecule. Residues belonging to the hypervariable loops are generally well defined in the electron density. Several residues of CDR-H3 of molecule 1, however, have high temperature factors and their positions are less well defined than those of molecule 2, although the path of the main chain is unambiguous. A total of 530 well defined water molecules and two chloride ions were placed.

A strong positive peak in the difference electron density was located at the apex of the CDR-H3 loop of the antibody (Fig. 1). First interpreted as a water molecule, this density was subsequently modelled as a chloride ion (from the Tris–HCl buffer of the protein solution), coordinated by four nitrogen amide hydrogen-bond donors from CDR-H3 (N—Cl⁻ distances < 3.5 Å) and one solvent molecule. The presence of a chloride ion could thus strongly influence the conformation of CDR-H3 observed in this crystal form. In addition, one intermolecular interaction between HisH97 N$^{\epsilon 2}$

Table 2

Pairwise comparisons of the structures of the variable domains of non-complexed scFv-1696 (this work, two molecules), ligand-free Fab-1696 (PDB code 1cl7) and scFv-1696 in complex with the HIV-1 peptide (PDB code 1jp5).

For each structural comparison, the r.m.s. differences between the main-chain atom positions of equivalent residues of the variable domains together ($V_L + V_H$), the V_L domain only and the V_H domain only, respectively, are given in Å. Superpositions were carried out using program *LSQKAB* from the *CCP4* program suite (Kabsch, 1988; Collaborative Computational Project, Number 4, 1994).

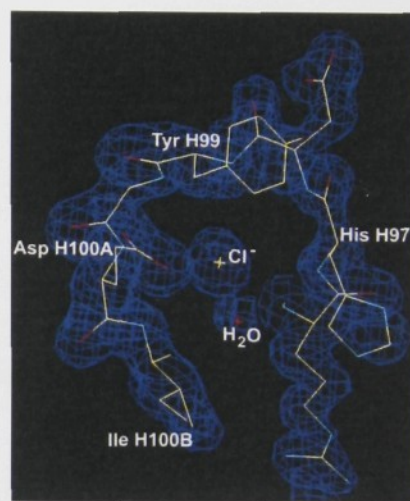|	scFv-1696-2	Fab-1696	scFv-1696 complex
scFv-1696-1	0.61, 0.40, 0.42	1.00, 0.74, 0.93	1.24, 0.55, 1.47
scFv-1696-2		1.22, 0.74, 0.86	1.39, 0.57, 1.44
Fab-1696			1.05, 0.56, 1.21

Figure 1

View of CDR-H3, with electron-density map, calculated with $(3F_o - 2F_c)$ coefficients and phases from the refined model, contoured at the 1.2 r.m.s. level. A prominent solvent peak, at hydrogen-bonding distances from four nitrogen amides of CDR-H3 and one water molecule, has been modelled as a chloride ion. Those residues of CDR-H3 for which the amide nitrogen is less than 3.5 Å from the putative chloride ion, together with the coordinating water molecule, are indicated.

in the CDR-H3 of scFv-1696 molecule 2 and the carboxylic group of GluL79 of scFv-1696 molecule 1 could also stabilize the conformation of CDR-H3 of scFv-1696 molecule 2.

4.2. Comparison of the ligand-free scFv-1696 at acid pH with ligand-free Fab-1696 and complexed scFv-1696 at neutral pH

After superimposing the V_L domains of the different forms, additional rotations and translations are required in order to optimize the superposition of V_H domains (Table 2). These differences in relative orientation of the variable domains give rise to antigen-binding sites with quite different topologies: the ligand-free scFv-1696 binding site is more open than that of the
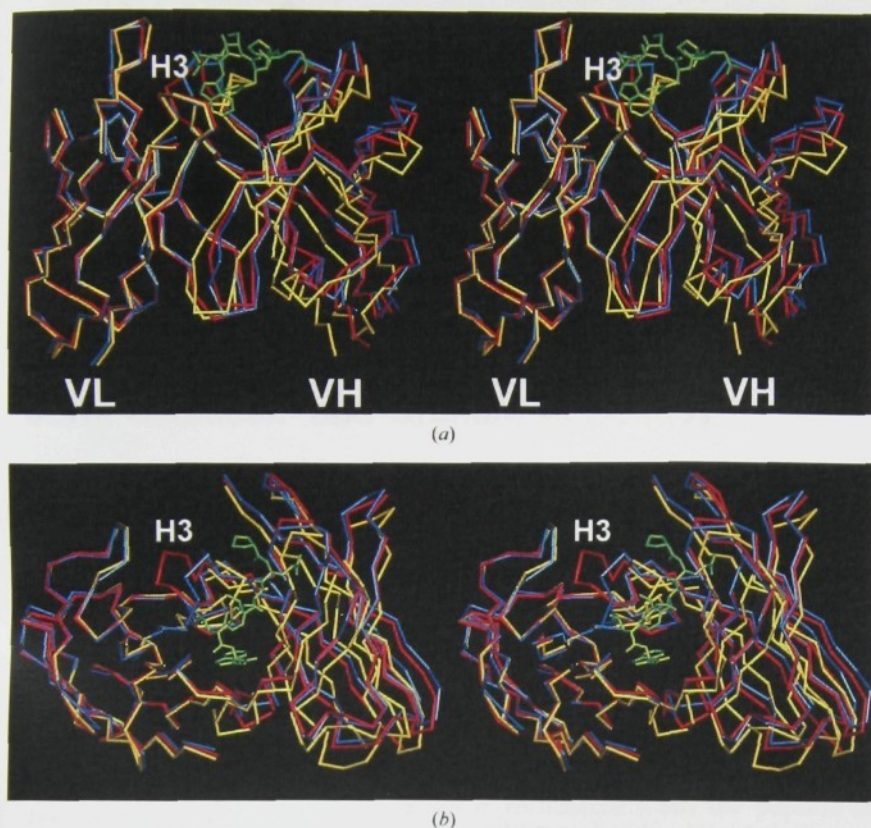
CDR-H3 with CDR-H1 in the Fab-1696
structure at neutral pH.

5. Conclusion

The structural changes observed in the
binding site of mAb-1696 result in the
formation of antigen-binding pockets with
quite different shapes. The conformational
changes observed between the three
different crystal forms result from rigid-
body movements of variable domains
accompanied by large conformational
changes in CDR-H3. Three different
conformations for CDR-H3 of mAb-1696
have been identified: the antigen-bound
conformation of CDR-H3 appears to be
induced by the presence of the antigenic
peptide, whereas the two alternative struc-
tures of CDR-H3 that we have observed in
the non-complexed forms of 1696 point to
the possible roles of pH and solute mole-
cules in influencing the conformation.

Figure 2
Comparison of the variable domains of non-complexed Fab-1696 (PDB code 1cl7, blue), molecule 1 of the non-
complexed scFv-1696 (this work, yellow) and complexed scFv-1696 (PDB code 1jp5, red). The peptide is shown in
green. Superposition has been made by optimizing the correspondence between the main-chain atoms of the V_L
domains of the three structures. Fv1696 is shown from the side of (a) and from above (b) the antigen-binding site.
The V_L domain is shown on the left of the Fv fragment, while the V_H domain is on the right. The third CDR of V_H
(H3) is indicated.

References

Brünger, 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., Read, R. J., Rice, L. M., Simonson,
T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–
921.
Collaborative Computational Project, Number 4
(1994). *Acta Cryst.* D**50**, 760–763.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**,
392–400.
Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
Lescar, J., Brynda, J., Rezacova, P., Stouracova, R.,
Riottot, M. M., Chitarra, V., Fabry, M., Horejsi,
M., Sedlacek, J. & Bentley, G. A. (1999). *Protein
Sci.* **8**, 2686–2696.
Lescar, J., Stouracova, R., Riottot, M. M.,
Chitarra, V., Brynda, J., Fabry, M., Horejsi, M.,
Sedlacek, J. & Bentley, G. A. (1996). *Protein Sci.*
5, 966–968.
Lescar, J., Stouracova, R., Riottot, M. M.,
Chitarra, V., Brynda, J., Fabry, M., Horejsi, M.,
Sedlacek, J. & Bentley, G. A. (1997). *J. Mol.
Biol.* **267**, 1207–1222.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Rezacova, P., Lescar, J., Brynda, J., Fabry, M.,
Horejsi, M., Sedlacek, J. & Bentley, G. A.
(2001). *Structure*, **9**, 887–895.
Wlodawer, A. & Gutschina, A. (2000). *Biochim.
Biophys. Acta*, **1477**, 16–34.

ligand-free Fab-1696, while the complexed
scFv-1696 structure adopts a position inter-
mediate between the two (Fig. 2). It is
notable that the closed conformation
adopted by the ligand-free Fab-1696 binding
site (but not that of the ligand-free scFv1696
molecule) would introduce some steric
hindrance between its CDR-H1 and the
N-terminal end of the antigenic peptide, as
seen in the complexed structure with
scFv-1696.

4.3. Structure and flexibility of CDR-H3

The most conspicuous changes in the
tertiary structure of mAb-1696 occur in
CDR-H3. By contrast, the backbone
conformations of the five other CDRs are
quite conserved. The residues situated
between ArgH96 and GluH100A, at the
solvent-exposed end of CDR-H3, adopt
strikingly different positions. In the peptide
complex of scFv-1696, CDR-H3 moves

towards the V_L domain, giving rise to a
shallow groove at the centre of the antigen-
binding site where the peptide is buried
(Rezacova et al., 2001). In the ligand-free
scFv-1696 crystal structure obtained at pH
3.5, however, CDR-H3 folds back towards
the V_H domain (Fig. 2). This position
occludes the antigen-binding pocket by
bringing residues AspH98 and TyrH99 into
the space taken up by GlnP2 and IleP3 of
the antigenic peptide. In the ligand-free
Fab-1696 structure crystallized at pH 7.5, the
observed conformation of CDR-H3 is
maintained in part by an ion pair between
AspH95 and ArgH96 (Lescar et al., 1999).
This salt bridge is lost in the ligand-free
scFv-1696 structure of CDRH3, possibly
because of protonation of the carboxylic
group of AspH95 at pH 3.5. The acidic pH
could also account for the disruption of
the hydrogen-bond interaction between
AspH95 and HisH35, which associates

JMB

Three-dimensional Structure of an Fab-Peptide Complex: Structural Basis of HIV-1 Protease Inhibition by a Monoclonal Antibody

Julien Lescar[1], Renata Stouracova[1,2], Marie-Madeleine Riottot[1]
Véronique Chitarra[1], Jiri Brynda[1,2], Milan Fabry[2], Magda Horejsi[2]
Juraj Sedlacek[2] and Graham A. Bentley[1*]

[1]*Unité d'Immunologie Structurale (URA 1961 CNRS), Département d'Immunologie, Institut Pasteur, 75724 Paris Cedex 15 France*

[2]*Department of Gene Manipulation, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo nam. 2 166 37, Prague 6 Czech Republic*

*Corresponding author

F11.2.32, a monoclonal antibody raised against HIV-1 protease ($K_d = 5$ nM), which inhibits proteolytic activity of the enzyme ($K_{inh} = 35(\pm 3)$ nM), has been studied by crystallographic methods. The three-dimensional structure of the complex between the Fab fragment and a synthetic peptide, spanning residues 36 to 46 of the protease, has been determined at 2.2 Å resolution, and that of the Fab in the free state has been determined at 2.6 Å resolution. The refined model of the complex reveals ten well-ordered residues of the peptide (P36 to P45) bound in a hydrophobic cavity at the centre of the antigen-binding site. The peptide adopts a β hairpin-like structure in which residues P38 to P42 form a type II β-turn conformation. An intermolecular antiparallel β-sheet is formed between the peptide and the CDR3-H loop of the antibody; additional polar interactions occur between main-chain atoms of the peptide and hydroxyl groups from tyrosine residues protruding from CDR1-L and CDR3-H. Three water molecules, located at the antigen-antibody interface, mediate polar interactions between the peptide and the most buried hypervariable loops, CDR3-L and CDR1-H. A comparison between the free and complexed Fab fragments shows that significant conformational changes occur in the long hypervariable regions, CDR1-L and CDR3-H, upon binding the peptide. The conformation of the bound peptide, which shows no overall structural similarity to the corresponding segment in HIV-1 protease, suggests that F11.2.32 might inhibit proteolysis by distorting the native structure of the enzyme.

© 1997 Academic Press Limited

Keywords: antibody cross-reactivity; crystal structure; enzyme inhibition; Fab fragment; HIV-1 protease

Correspondence address: G. A. Bentley, Unité d'Immunologie Structurale (URA 1961 CNRS), Département d'Immunologie, Institut Pasteur, 75724 Paris Cedex 15, France; or e-mail: bentley@pasteur.fr

Introduction

The genome of human immunodeficiency virus type 1 (HIV-1) encodes three enzymes, each with an essential function in the complete infectious cycle of the pathogen (reviewed by Katz & Skalka, 1994). One of these enzymes is a protease that specifically cleaves the Gag and Gag/Pol polyprotein precursors to form the structural proteins (matrix, capsid and nucleocapsid) and viral enzymes (including the protease itself) of HIV-1, as well as subsidiary polypeptides p1, p2 and p6 (Ratner *et al.*, 1985; Henderson *et al.*, 1992). The HIV-1 protease is a homodimeric enzyme belonging to the family of aspartyl proteases (Pearl & Taylor, 1987). The monomer comprises 99 amino

© 1997 Academic Press Limited

acid residues containing the triplet Asp25-Thr26-Gly27, which is located near the dimer interface. Thus, in the functional homodimer, the two amino acid triplets are adjacent to one another, forming a pepsin-like catalytic site at the bottom of a hydrophobic cavity (Navia *et al.*, 1989; Wlodawer *et al.*, 1989). The catalytic site is covered by two flap regions, one contributed by each subunit (residues 44 to 57), which undergo substantial movements upon binding substrate (Fitzgerald & Springer, 1991). Dynamic studies by NMR (Nicholson *et al.*, 1995) and molecular simulation (Harte *et al.*, 1992; Collins *et al.*, 1995) show that the flap regions are indeed highly mobile, a property which would facilitate access to the active site by substrate and the subsequent release of product. Since the presence of inhibitors, or mutations that compromise proteolytic activity, leads to the assembly of immature, non-infectious viral particles (Kohl *et al.*, 1988; Ashorn *et al.*, 1990; Rosé *et al.*, 1995), HIV-1 protease has been the object of considerable study, particularly in the development of inhibitors with therapeutic application.

Our objective has been to study structural changes in HIV-1 protease that are important for enzyme function by analysing the effect of protease-inhibiting antibodies. Accordingly, we have raised murine monoclonal antibodies against the protease and selected those which cause significant inhibition of catalytic activity for further study by crystallographic analysis. From analysis of the three-dimensional structure of antibody-antigen complexes, we aim to correlate conformational restraints or modifications, imposed on the protease by the bound antibody, to the inhibition of enzyme function.

One of the monoclonal antibodies thus obtained, designated F11.2.32, binds to HIV-1 protease with an affinity of 5 nM at pH 5.2 (Lescar *et al.*, 1996), as measured with the BIAcore system (Pharmacia Biosensor), and induces significant inhibition of proteolytic activity. In addition, epitope mapping has identified the peptide P36-P46 (MSLPGRWKPKM) as the shortest segment of HIV-1 protease sequence capable of effectively competing with the native enzyme in binding to F11.2.32 (Lescar *et al.*, 1996). While the region, 36 to 46, is an immunodominant epitope in the Armenian hamster (Croix *et al.*, 1993), it is only weakly recognised by polyclonal sera from HIV-infected individuals (Björling *et al.*, 1992), suggesting a variable immune response to HIV-1 protease that depends on species or the mode of antigen presentation.

Although we have not yet succeeded in crystallising a complex of the Fab fragment of F11.2.32 with the protease, we have obtained crystals of a complex formed with the peptide segment P36-P46 of the enzyme. We have also crystallised Fab F11.2.32 in the unliganded state. Crystallisation conditions, diffraction measurements (made at cryogenic temperature) and the determination of preliminary models by molecular replacement pro-

cedures for both crystal forms have been reported in an earlier publication (Lescar *et al.*, 1996). The resolution of the diffraction data reported at that time was 2.1 Å for Fab F11.2.32-peptide(P36-P46) and 2.9 Å for the free Fab. We have since obtained an improved diffraction data set for the free Fab extending to 2.6 Å resolution, also measured at cryogenic temperature. This paper describes the refinement of Fab F11.2.32 in the free state and as a complex with the peptide P36-P46, and gives a detailed description of the two structures. Possible mechanisms for the inhibition of HIV-1 protease activity by F11.2.32 are discussed in the light of these structural results.

Results and Discussion

Inhibition of HIV-1 protease by F11.2.32

Initial reaction velocities, determined from kinetic runs of substrate cleavage by HIV-1 protease in the presence of differing concentrations of Fab F11.2.32, were normalised to the initial reaction velocity in the absence of inhibitor and plotted as a function of Fab concentration (Figure 1). These data were evaluated using the tight-binding inhibition model by fitting the following equation (Williams & Morrison, 1979) to the experimental points on the graph, as described in Materials and Methods:

$$v_i/v_o = 1 - 1/(2[E])\{[E] + [I_i] + K_{inh}$$

$$- \sqrt{([E + [I_i] + K_{inh})^2 - 4[E][I_i]}\}$$

where [E] and [I_i] are the protease concentration (held constant) and the Fab concentration for the *i*th experiment, respectively; K_{inh} is the apparent inhibition constant; v_o and v_i are the initial reaction velocities for Fab concentrations 0 and [I_i], respectively.

The best fit of the equation to these data, shown by the curve in Figure 1, gives K_{inh} as 35($\pm$3) nM and the concentration of epitope, [E], as 82($\pm$8) nM. When estimated by titration with an active-site inhibitor, the protease concentration was found to be 50($\pm$18) nM for the dimer (data not shown). These results could thus be compatible with a stoichiometry of Fab to active protease dimer being 2:1.

General quality of the structures

Crystal parameters are given in Table 1 and statistics from the refinement of the crystal structure of the complexed and free forms of F11.2.32 are summarised in Table 2. Ramachandran diagrams of the main-chain dihedral angles are shown in Figure 2. Each crystal form contains two molecules in the asymmetric unit.
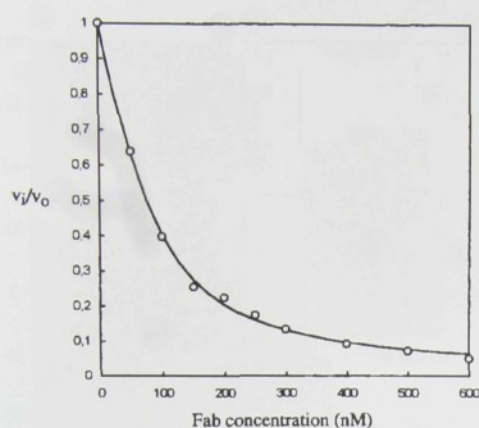
Figure 1. Inhibition of HIV-1 protease activity by Fab F11.2.32. The initial reaction velocity (v_i), normalised to the initial reaction velocity in the absence of Fab (v_o), is plotted as a function of Fab concentration. The protease was held at constant concentration (50 nM). The curve shows the best fit of the tight-binding model (Williams & Morrison, 1979) to the experimental data, from which the deduced value of K_{inh} is 35($\pm$3) nM (see Materials and Methods for details).

Table 2. Refinement statistics

	Fab F11.2.32 (complex)	Fab F11.2.32 (free)
Resolution (Å)	7.0–2.2	8.0–2.6
Reflections used in refinement	35,173 ($F > 3\sigma$)	24,344 ($F > 2\sigma$)
R-factor (with no σ cut-off)	0.198 (0.217)	0.206 (0.208)
R_{free} (with no σ cut-off)	0.289 (0.307)	0.317 (0.318)
No. of residues	903	872
No. of solvent molecules	400	235
r.m.s. deviations		
Bond lenths (Å)	0.011	0.011
Bond angles (°)	1.7	1.7
Mean B-factor (Å^2)		
Main-chain	19	16
Side-chain	21	17
Solvent	26	20

Atomic coordinates have been deposited in the Brookhaven Protein Data Bank as entry 2hrp (r2hrpsf for structure factors) for Fab F11.2.32-peptide(P36-P46) and as entry 1mf2 (r1mf2sf for structure factors) for Fab F11.2.32.

Fab F11.2.32-peptide(P36-P46)

Residues missing from the Fab fragments in the final model of the complex are located at the carboxy-terminal ends of the polypeptide chains. These are residues L212 to L214 for the light chain (L) of molecule 1 in the asymmetric unit, and close to the site of the papain cleavage on the heavy chain, (H) beyond residues H225 and H223 for molecules 1 and 2, respectively. Electron density is weak for the regions H42 to H43 and H130 to H133 of both heavy chains; these are located in loops exposed to solvent. In addition, the side-chains of residues Arg-H100, Tyr-H100A and Asp-H100B from CDR3-H of both heavy chains are in weak density; these are exposed to the solvent, making no contacts with the peptide, and are mobile ($B > 50$ Å^2). The side-chain of Tyr-H100A in neither Fab was visible in the electron density and was therefore not modelled beyond C$^\beta$. The electron density of the hypervariable regions is otherwise well-defined. The bound peptide in both complexes of the asymmetric unit has been unambiguously traced for ten of the 11 residues from Met-P36 to Lys-P45. Of these, only the side-chain of Lys-P45 was not completely built. The structure

of complex between Fab F11.2.32 and the longer peptide, P36-P57, (Lescar *et al.*, 1996) similarly shows no electron density beyond peptide residue P45 (data not shown). The average temperature factor for the main-chain, side-chain and solvent atoms is 19 Å^2, 21 Å^2 and 26 Å^2, respectively.

Free Fab F11.2.32

The carboxy-terminal ends of the light and heavy chains are also disordered in both Fab molecules in the asymmetric unit; residues beyond L212 and H223 are not located in either molecule. Heavy-chain residues H128 to H134 of both Fabs were not built in the final model because of poorly defined electron density in these regions. Hypervariable regions CDR1-L and CDR3-H are very mobile in the free Fab, as evidenced by their high thermal parameters; side-chains for CDR1-L residues Tyr-L27D and Tyr-L28, and CDR3-H residues Tyr-H100A and Asp-H100B, were not built beyond C$^\beta$. The average temperature factor for the main-chain, side-chain and solvent atoms is 16. Å^2, 17 Å^2 and 20 Å^2, respectively.

Structure of Fab F11.2.32

The asymmetric unit of both crystal structures contains two molecules of Fab related by a non-crystallographic dyad. In each case, the dimer of Fabs (Figure 3) is maintained by almost identical contacts from two symmetric sets of hydrogen bonds (between Ser-H25 and Ser-H136', and Thr-H153 and Ser-H163'), two symmetric sets of salt bridges (between Arg-L61 and Glu-L79', and Arg-L108 and Glu-H1') and numerous van der Waals interactions. The buried surface area between the two antibody fragments for the free and com-

Table 1. Unit cell parameters

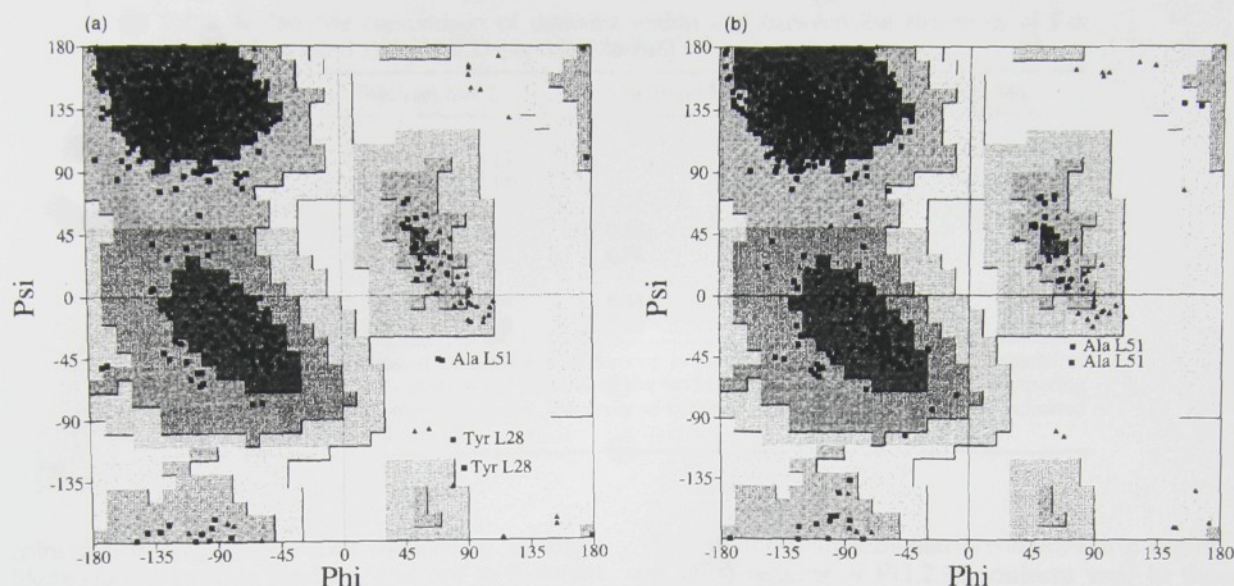| Fab F11.2.32 (free)	Fab F11.2.32 (complex)
$P2_1$	$P2_12_12_1$
$a = 67.6$ Å	$a = 82.4$ Å
$b = 94.7$ Å	$b = 96.3$ Å
$c = 70.3$ Å	$c = 105.8$ Å
$\beta = 105.8°$	
$Z = 4$	$Z = 8$
$V_M = 2.2$ Å^3/Da	$V_M = 2.1$ Å^3/Da

Figure 2. Ramachandran diagrams for (a) free Fab F11.2.32 and (b) Fab F11.2.32-peptide(P36-P46). Residues in unfavourable regions are named; glycine residues are shown by triangles. (Produced by the program PROCHECK, Laskowski *et al.*, 1993).

plexed forms is 1851 Å² and 2186 Å², respectively. The elbow angles (the angle subtended by the pseudo dyads of the variable and constant domain dimers) are correspondingly close to each other: 151° for the free Fab and 154° for the peptide complex.

The main-chain atoms of the Fabs within the dimer superimpose on each other with an r.m.s. deviation of 0.31 Å for all 434 equivalent residues for the complex, and 0.20 Å for the unliganded form. This difference derives from the refinement
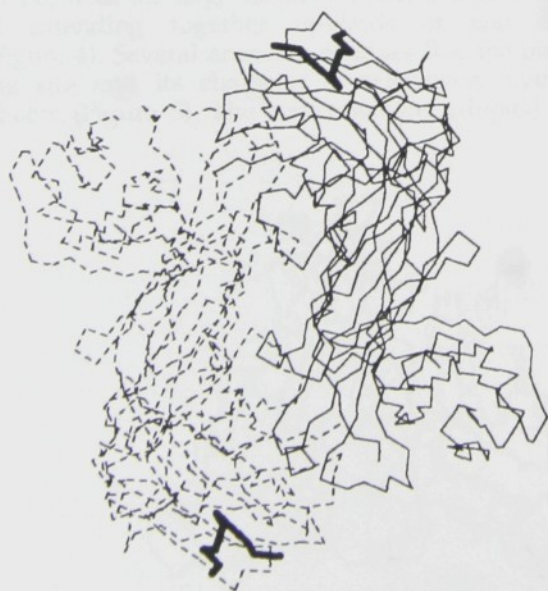

Figure 3. The Fab dimer in the asymmetric unit of the complex viewed along the non-crystallographic dyad. The two Fabs are differentiated by continuous and broken thin lines, while the bound peptide is drawn in thick continuous line.

protocol (see Structure refinement in Materials and Methods) where the non-crystallographic symmetry has been less tightly constrained for the complexed Fab because of the higher resolution of the data (2.2 Å *versus* 2.6 Å). A comparison of the variable (framework residues only) and constant domains within and between crystal structures is presented in summary in Table 3. Although the variable light chain domain (V_L) framework residues from the free and complexed forms of F11.2.32 coincide within the expected experimental error (Luzzati, 1952), a greater deviation is observed between the variable heavy chain domain (V_H) domains; this is due to a structural difference in the segment from Ala-H40 to Gly-H44 of the unliganded F11.2.32 when compared to the complexed Fab or, indeed, other reported Fab structures. In particular, Glu-H42 and Lys-H43 have high temperature factors in both crystal forms, indicating considerable mobility in this region. While the cause of this difference is not clear, it might be related to the effect of Fab dimerisation on the V_H/V_L interface, since V_L and V_H of one Fab make important contacts with V_L and the C_H^1-C_L pair, respectively, of the other Fab of the dimer. Consequently, a tension might exist at the V_H/V_L interface that could render the conformation of the segment H40-H44 sensitive to the elbow angle since the generally conserved polar interaction across the interface between Gln-H39 (at one end of this segment) and Gln-L42, which is present in both forms of F11.2.32, forms a fixed point of contact between the variable domains.

The variable domains show a small difference in relative orientation between the antigen-bound and unbound states; after superimposing the V_L domains of each form, an additional rotation of 3.8° is required to bring the V_H domains into optimal

Table 3. Pairwise comparison of domains within and between the structures of Fab F11.2.32 (free) and Fab F11.2.32-peptide(P36-P46)

	Fab(free) Mol.1		Fab(free) Mol.2		Fab(complex) Mol.1	
Fab(free)						
Mol.2	0.15	0.15			V_L	C_L
	0.13	0.13			V_H	C_H^1
Fab(complex)						
Mol. 1	0.36	0.46	0.36	0.46		
	0.78	0.35	0.79	0.35		
Fab(complex)						
Mol.2	0.33	0.46	0.34	0.46	0.23	0.26
	0.62	0.39	0.63	0.39	0.31	0.32

The results are expressed as the r.m.s. difference in Å between main-chain atom positions of equivalent residues. Only framework residues of the variable domains were used when comparing between the free and complexed forms. The order of comparisons is V_L, C_L, V_H, C_H^1 as indicated in the top right-hand cell.

coincidence (Figure 4). This difference is most likely due to peptide binding, and not to different crystal contacts, since the quaternary conformation is effectively identical for the two independent Fab molecules in the asymmetric unit within each crystal structure. The relative rotation of the variable domains falls into the lower range of values found from pairwise V_L/V_H domain comparisons analysed by Stanfield *et al.* (1993), and agrees with their empirical observation that a larger V_L/V_H interface correlates with a smaller movement upon binding antigen; the total surface area buried at the variable dimer interface of complexed F11.2.32 is 1612 Å², of which 372 Å² in contributed by the 16 residue-long CDR3-H.

Antigen-binding site

The antigen-binding site is concave in overall shape, with the large loops of CDR1-L and CDR3-H extending together upwards at one side (Figure 4). Several aromatic residues line the binding site and its character is extensively hydrophobic (Figure 5). The conformations adopted by

five of the complementarity determining region's (CDRs) regions of F11.2.32 conform well to those predicted by the canonical structure hypothesis of Chothia *et al.* (1989) (CDR1-L: canonical structure 3; CDR2-L: unique; CDR3-L: 1; CDR1-H: 1 and CDR2-H: 3). The dihedral angles adopted by Ala-L51 (Figure 2), similar in the unbound and complexed states where the conformation of CDR2-L is clearly defined, are characteristic of a γ-turn (Milner-White *et al.*, 1988) and of this canonical structure. Its conformation seems to be influenced by an extensive network of main-chain hydrogen bonds with buried residues L33 and L34. In the uncomplexed Fab structure, Tyr-L28, which belongs to an energetically unfavourable region of the Ramachandran plot, is located in a flexible part of the molecule.

Hypervariable regions CDR1-H, CDR2-H and CDR3-L, being less exposed, display little change in their main-chain and side-chain positions upon binding antigen. The most conspicuous changes occur at the mobile ends of the two long loops of CDR1-L and CDR3-H, which undergo important
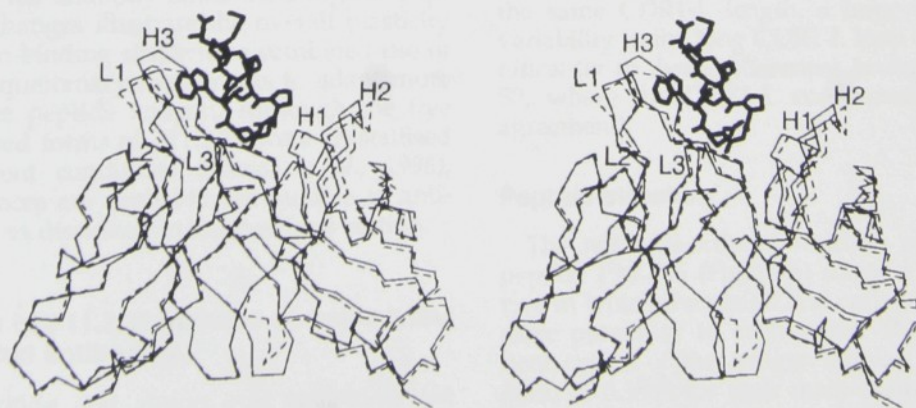

Figure 4. Stereo view of Fab F11.2.32 in the free and complexed states after superposition of the α-carbon atoms of the V_L domains, emphasizing the conformational changes that occur upon binding the antigenic peptide. The variable doamins are drawn in continuous and broken thin lines for the complexed and free forms, respectively, while the complete peptide is drawn in thick line.

Figure 5. View of the interaction between the peptide and the molecular surface of the antigen-binding site, colour coded for electrostatic potential: red for negative and blue for positive; prepared with program GRASP (Nicholls *et al.*, 1991).

conformational changes in both the main-chain and side-chain structure. In particular, two continuous stretches, L27C-L30 (CDR1-L) and H99-H100D (CDR3-H), deviate by as much as 1.0 and 2.7 Å, respectively, between the α-carbon positions of the antigen-bound and free states. There is a 180° rotation of the Tyr-L28 peptide bond when the hydroxyl group of this residue moves in to form a hydrogen bond to the amino terminus of the peptide. Residue Asp-H100B moves by 2.7 Å from its unliganded structure in order to make a hydrogen bond to Lys-P45. As described above, a small relative rotation of the variable domains also occurs when the antibody binds the antigenic peptide. These changes illustrate the overall plasticity of the antigen-binding site, with a combined use of tertiary and quaternary movements to adapt more closely to the peptide antigen. Although the free and complexed forms of F11.2.32 were crystallised under different conditions (Lescar *et al.*, 1996), these differences are probably attributable to antigen-binding, as discussed in the previous section.

Comparison of F11.2.32 variable domains with closely related antibodies

The nucleotide and amino acid sequences are given in Table 4. The heavy chain of F11.2.32 bears a close sequence homology to the anti-oxazolone mAb NQ10/12.5 (Alzari *et al.*, 1990) and the catalytic mAb CNJ206 (Golinelli-Pimpaneau *et al.*, 1994; Brookhaven Protein Data Bank PDB

(Bernstein *et al.*, 1977), entry 1kno), which, like F11.2.32, are both derived from V$_H$ genes of the V$_H$-MOPC21 subgroup (Kabat *et al.*, 1991). For the V$_H$ domain, F11.2.32 has 103 residues identical to NQ10/12.5 and 106 residues identical to CNJ206. Ninety-eight α-carbon atoms from the NQ10/12.5 and CNJ206 V$_H$ domains (excluding residues from CDR3-H and FR4) can be superimposed upon V$_H$ of the F11.2.32 peptide complex with an r.m.s deviation of 0.43 Å and 0.50 Å, respectively. Only the CDR3-H loop differs widely between the three V$_H$ domains, conferring a different shape, and thus specificity, on each antigen-binding site. This is achieved by recruiting different diversity and joining gene segments, leading to the formation of a hydrophobic pocket for the NQ10/12.5 (CDR3-H of five residues), a groove for CNJ206 (CDR3-H of ten residues) and a large protruding loop for F11.2.32 (CDR3-H of 16 residues which uses the long J$_{H4}$ gene segment (Kabat *et al.*, 1991)).

The light chain of F11.2.32 (subgroup VIII-MOPC70 and Jκ$_1$ joining gene; Kabat *et al.*, 1991) is homologous to mAb 59.1 (Ghiara *et al.*, 1994; PDB entry 1acy) and mAb 50.1 (Rini *et al.*, 1993; PDB entries 1ggb, 1ggc, 1ggi) which are both specific for epitopes on the V3 loop of the gp120 envelope protein of HIV-1. For residues coded by the V$_κ$ gene, F11.2.32 has 78 and 75 residues identical with 50.1 and 59.1, respectively. Of note, the CDR1-L region of 59.1 displays only one amino acid difference to that of F11.2.32 (Tyr-L27D → Ser). Nevertheless, the way these three monoclonal antibodies recognise their respective peptide antigens is different. A slightly weaker homology is found with the light chain of the anti-digoxin mAb 40-50 (Jeffrey *et al.*, 1995; PDB entry 1ibg, 73 identical residues encoded by the V$_κ$ gene) and the mAb N10, (Bossart-Whitaker *et al.*, 1995; PDB entry 1nsn, 71 identical residues encoded by the V$_κ$ gene) specific for staphylococcal nuclease. Superposition of the complete V$_κ$ domain of F11.2.32 (111 residues) upon that of 50.1, 59.1, 40-50 and N10 gives an r.m.s. difference in α-carbon positions of 0.63 Å, 0.71 Å, 0.60 Å and 0.89 Å. While all five V$_κ$ domains share the same CDR1-L length, a large conformational variability in the long CDR1-L loop contributes significantly to these differences (except for mAb 40-50, where the CDR1-L conformation is in close agreement).

Peptide structure

The ten well-ordered residues of the antigenic peptide P36-P46 (Figure 6) adopt a compact structure in which the amino and carboxy termini are in close proximity to each other. The two independent copies of the antigen display an r.m.s. deviation of 0.35 Å for their main-chain atoms. Taking into account the higher mobility of the terminal ends of the peptide, these values are comparable to the mean error expected for the atomic positions for this structure (Luzzati, 1952). Residues Leu-P38, Pro-P39, Gly-P40 and Arg-P41 form a type II

Table 4. Nucleotide and deduced amino acid sequences of the V_L and V_H domains of F11.2.32

Heavy Chain

```
                              GAG TCT GGG GGA GGC TTA GTG CAA CCT GGA GGG TCC CGG AAA CTC TCC
          Asp Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly Ser Arg Lys Leu Ser
           1   2   3   4   5   6   7   8   9   10  11  12  13  14  15  16  17  18  19  20  21


TGT GCA GCC TCT GGA TTC ACT TTC ATG AGG TTT GGA ATG CAC TGG GTT CGT CAG GCT CCA GAA AAG GGG CTA
Cys Ala Ala Ser Gly Phe Thr Phe Met Arg Phe Gly Met His Trp Val Arg Gln Ala Pro Glu Lys Gly Leu
 22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45
                                        <------ H1 ------>

GAG TGG GTC GCA TAC ATT AGT AGT GGC AGT AGT ACC ATC TAC TAT GCA GAC ACA GTG AAG GGC CGA TTC ACC
Glu Trp Val Ala Tyr Ile Ser Ser Gly Ser Ser Thr Ile Tyr Tyr Ala Asp Thr Val Lys Gly Arg Phe Thr
 46  47  48  49  50  51  52  A   53  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68
                     <-------------------- H2 -------------------------------->

ATC TCC AGA GAC AAT CCC AAA AAC ACC CTG TTC CTG CAA ATG ACC AGT CTA AGG TCT GAG GAC ACG GCC CTG
Ile Ser Arg Asp Asn Pro Lys Asn Thr Leu Phe Leu Gln Met Thr Ser Leu Arg Ser Glu Asp Thr Ala Leu
 69  70  71  72  73  74  75  76  77  78  79  80  81  82  A   B   C   83  84  85  86  87  88  89

TAT TAC TGT GCA AGA AGT GGA GGG ATT GAG AGG TAC GAC GGT ACG TAC TAT GTT ATG GAC TAC TGG GGT CAA
Tyr Tyr Cys Ala Arg Ser Gly Gly Ile Glu Arg Tyr Asp Gly Thr Tyr Tyr Val Met Asp Tyr Trp Gly Gln
 90  91  92  93  94  95  96  97  98  99 100  A   B   C   D   E   F   G   H  101 102 103 104 105
                 <---------------------------- H3 ---------------------------->

GGA ACC TCA GTC ACA GTC TCC TCA
Gly Thr Ser Val Thr Val Ser Ser
106 107 108 109 110 111 112 113
```

Light Chain

```
                          CTG ACC CAG TCT CCA GCT TCT TTG GCT GTG TCT CTA GGG CAG AGG GCC ACC
          Asp Thr Val Leu Thr Gln Ser Pro Ala Ser Leu Ala Val Ser Leu Gly Gln Arg Ala Thr
           1   2   3   4   5   6   7   8   9   10  11  12  13  14  15  16  17  18  19  20


ATC TCC TGT AGA GCC AGC GAA AGT GTG GAT TAT TAT GGC AAG AGT TTT ATG AAC TGG TTC CAA CAG AAA CCA
Ile Ser Cys Arg Ala Ser Glu Ser Val Asp Tyr Tyr Gly Lys Ser Phe Met Asn Trp Phe Gln Gln Lys Pro
 21  22  23  24  25  26  27  A   B   C   D   28  29  30  31  32  33  34  35  36  37  38  39  40
                 <-------------------------- L1 -------------------------->

GGA CAG CCA CCC AAA CTC CTC ATC TAT GCT GCA TCC AAC CAA GGA TCC GGG GTC CCT GCC AGG TTT AGT GGC
Gly Gln Pro Pro Lys Leu Leu Ile Tyr Ala Ala Ser Asn Gln Gly Ser Gly Val Pro Ala Arg Phe Ser Gly
 41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61  62  63  64
                                 <----------- L2 --------->

AGT GGG TCT GGG ACA GAC TTC AGC CTC CAC ATC CAT CCT ATG GAG GAG GAT GAT TCT GCA ATG TAT TTC TGT
Ser Gly Ser Gly Thr Asp Phe Ser Leu His Ile His Pro Met Glu Glu Asp Asp Ser Ala Met Tyr Phe Cys
 65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88

CAG CAA AGT AAG GAG GTT CCG TGG ACG TTC GGT GGA GGG ACC AAG CTG GAG ATC
Gln Gln Ser Lys Glu Val Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
 89  90  91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106
<-------------- L3 -------------->
```

Nucleotides corresponding to the first five amino acid residues of the V_H domain and the first three residues of the V_L domain have not been sequenced; residues for these positions were obtained by amino acid microsequencing. The V_L gene was not sequenced beyond codon 106 since nucleotides beyond this point were removed by cleavage at a *Bgl*II site. Residue numbering and indicated hypervariable regions are according to Kabat *et al.* (1991).

β turn with Pro-P39 at position $(i + 1)$ and Gly-P40 at position $(i + 2)$. The peptide conformation is stabilised by several intramolecular interactions involving both side-chain and main-chain atoms. In particular, main-chain hydrogen bonds maintain a β hairpin-like form for the peptide; the carbonyl oxygen atom of Leu-P38 makes a bifurcated hydrogen bond with the amide nitrogen atoms of Arg-P41 and Trp-P42, and the amide group of Leu-P38 interacts with the carbonyl group of Trp-P42. The side-chain of Arg-P41 contributes additional stability to the peptide conformation by folding back to form a charged hydrogen bond with the carbonyl

oxygen atom of Pro-P39. It is not known if peptide(P36-P46) is well structured in solution, but the presence of residues Pro-P39 and Gly-P40 could favour a type II β turn. β-Turn motifs have been frequently observed for peptides in solution (Dyson & Wright, 1991) as well as in complexes with antibodies (Wilson & Stanfield, 1994, Stanfield & Wilson, 1995).

Globally, the compact β-hairpin conformation of the bound peptide does not resemble the more open structure of this segment in HIV-1 protease (Figure 7). Although the φ-φ angles of peptide segments P36-P39 and P42-P45 fall in the β region of
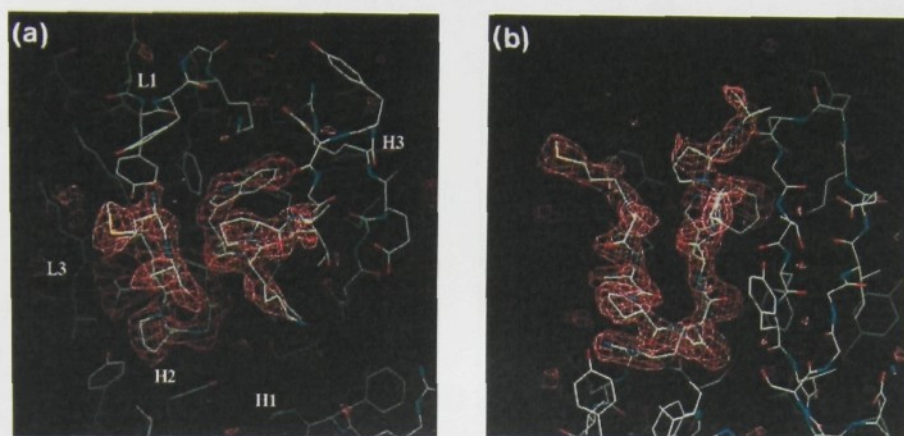
Figure 6. Difference electron density (contoured at 3σ) in the region of the antigen-binding site in which peptide(P36-P46) was omitted from the phase calculation. (a) View from above the antigen-binding site. The hypervariable regions in contact with the antigen are indicated. (b) View from the side to emphasise the β-turn structure adopted by the peptide in complex and the extensive interactions made with CDR3-H.

the Ramachandran plot, like their counterparts in the cognate protein, the main-chain conformation differs in the central region at P40 and P41 because of the type II β turn (Figure 7b). Thus, while the α-carbon atoms of peptide segments P36-P40 and P42-P45 each separately superimpose well on HIV-1 protease (Figure 7c and d), there is no overall correspondence to the native structure of the protein.

Peptide-antibody interactions

The peptide forms direct interatomic contacts with CDR1-L, CDR3-L, CDR2-H and CDR3-H (Figures 6 and 8). Hypervariable regions CDR2-L and CDR1-H, located at the periphery of the antigen-binding site, are too distant to interact directly with the compact antigen; indeed, they have been noted generally for making no, or only few, interactions with other Fab-bound peptides (Stanfield & Wilson, 1993). Upon formation of the complex, 555 Å^2 of solvent-accessible area for Fab F11.2.32 and 470 Å^2 for the peptide (i.e. 55% of the total peptide surface) become buried. Three water molecules, which bridge contacts between the antibody and the bound peptide, are completely buried at the interface. The presence of these solvent molecules significantly improves the complementarity between antibody and antigen; when included in the calculation, the buried surface is 567 Å^2 for Fab F11.2.32 and 530 Å^2 for the hydrated peptide.

The most extensive interactions from the antibody are made by CDR1-L and CDR3-H, which together account for nearly two-thirds of the buried surface of the Fab (146 Å^2 and 203 Å^2, respectively). The antigen-binding site includes several aromatic side-chains in direct contact with the peptide: six tyrosine residues (L27D, L28, H50, H58, H100E and H100F), one phenylalanine residue (L32) and one tryptophan residue (L96) (Figures 8

and 9). As generally observed in other Fab-peptide complexes, the hypervariable regions of F11.2.32 form a hydrophobic cavity, in contrast to the more flat topology often found at the binding site of anti-protein antibodies (Stanfield & Wilson, 1993; MacCallum *et al.*, 1996).

Nine hydrogen bonds are formed between F11.2.32 and the peptide, with additional hydrogen bonds contributed by the three water molecules buried at the antibody-antigen interface (Table 5 and Figure 8). Most polar interactions involve main-chain atoms of the peptide. In particular, a four-stranded anti-parallel β sheet is formed between the peptide and CDR3-H through Arg-P41, Lys-P43 and Lys-P45 of the antigen and Tyr-H100F, Thr-H100D and Asp-H100B of the hypervariable loop (Figure 6b). The only polar contacts from side-chain atoms of the peptide involve the guanidinium group of Arg-P41, which forms hydrogen bonds to Tyr-H58 and Lys-L92. An interaction between the charged amino group of Lys-L30 and the aromatic ring of Trp-P42 could also contribute some polar character.

Peptide residues losing more than half of their solvent-accessible surface area upon binding to F11.2.32 also have lower temperature factors (Figure 9). They form the continuous segment from Leu-P38 to Trp-P42 (which includes the type II β turn) located at the centre of the antigen-binding site in contact with CDR3-L, CDR2-H and CDR3-H. The three residues Gly-P40, Arg-P41 Trp-P42, which essentially lose complete contact with the solvent, account for nearly 55% of the lost accessible surface of the peptide in the complex. The most numerous contacts contributed by the peptide come, in fact, from Arg-P41 and Trp-P42; Arg-P41 forms five hydrogen bonds to F11.2.32, one of them being mediated by a water molecule to O$^\gamma$ of Ser-L93, while Trp-P42, whose aromatic side-chain is sandwiched between CDR1-L and CDR3-H, makes exclusively van der Waals contacts with the
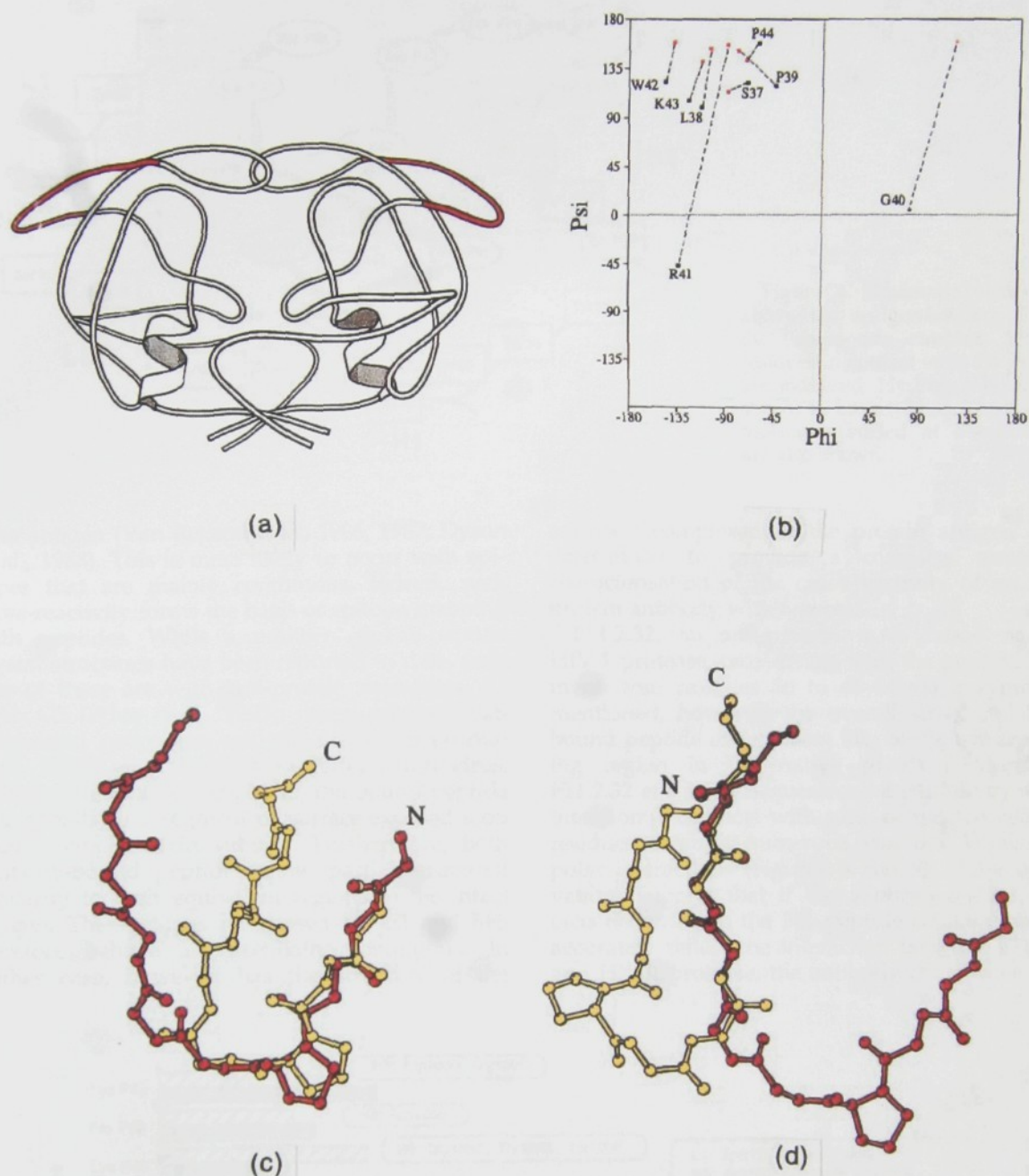
(a)

(b)

(c)

(d)

Figure 7. Comparison between the structure adopted by peptide(P36-P46) in the complex with Fab F11.2.32 and the conformation adopted by this segment in the native protease (PDB entry code 3hvp). (a) A schematic view of the protease in which the location of the epitope recognised by F11.2.32 is shown in red. (b) Ramachandran graph indicating the differences in φ-φ angles of the segment 36 to 46 in the bound peptide (black) and the protease (red); residues are named at the positions for the bound peptide and broken lines connect equivalent residues of the peptide and the protease. Comparison of residues 36 to 45 of the protease (red) and bound peptide (yellow): (c) after superimposing residues 36 to 40 and (d) after superimposing residues 41 to 44. The Figure was prepared with program MOLSCRIPT (Kraulis, 1991).

antibody. Residue Ser-P37, on the other hand, makes no interactions with the Fab molecule since it is directed towards the solvent. The carboxy-terminal region of the peptide extends beyond the long CDR3-H loop and is completely disordered for Met-P46.

Cross-reactivity of F11.2.32 with peptide(P36-P46) and possible mechanisms of HIV-1 protease inhibition

It is not uncommon that anti-protein antibodies cross-react with peptide fragments from the cog-
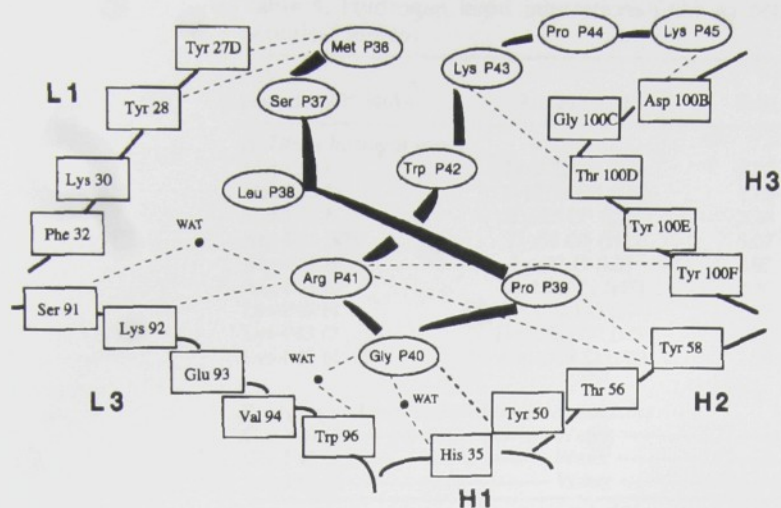
Figure 8. Schematic view from above the antigen-binding site in the Fab-peptide complex. All Fab residues in contact with the peptide are indicated. Hydrogen bonds are shown as broken lines. Three water molecules buried at the interface are also shown.

nate antigen (Van Regenmortel, 1986, 1987; Dyson *et al.*, 1988). This is most likely to occur with epitopes that are mainly continuous; indeed, such cross-reactivity forms the basis of epitope mapping with peptides. While a number of Fab-peptide crystal structures have been reported to date, only two of these are with anti-protein antibodies: (1) mAb C3 (Wien *et al.*, 1995), raised against heat-inactivated poliovirus, and (2) mAb 8F5 (Tormo *et al.*, 1994), raised against the human rhinovirus. In both Fab-peptide complexes, the bound peptide corresponds to a segment of surface-exposed loop from a viral protein subunit. Furthermore, both antibody-bound peptides show partial structural similarity to their equivalent regions in the intact antigen. The epitopes recognised by C3 and 8F5 therefore behave as essentially continuous. In neither case, however, has the structure of the

antibody complexed to the protein antigen been determined to provide a complete structural characterisation of the cross-reactivity of an anti-protein antibody with a peptide.

F11.2.32, an anti-protein mAb raised against HIV-1 protease, cross-reacts with the peptide fragment from residues 36 to 46 of the enzyme. As mentioned, however, the overall structure of the bound peptide differs from that of the corresponding region in the native protease. Moreover, F11.2.32 efficiently sequesters the peptide by direct interatomic contacts with nine of the ten ordered residues through numerous van der Waals and polar interactions (Figures 8 and 9). These observations suggest that if the antibody-antigen contacts occurring in the Fab-peptide crystal structure accurately reflect the interaction between F11.2.32 and HIV-1 protease, the antibody could have been
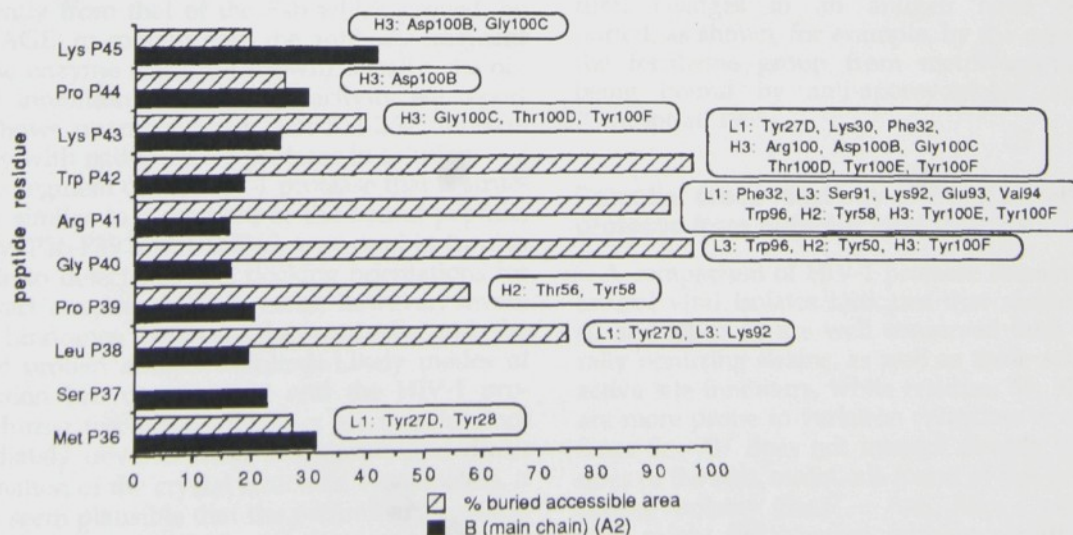

Figure 9. Temperature factors of main-chain peptide atoms (filled bars) and percentage of accessible surface areas of the peptide residues buried upon complex formation with the Fab (hatched bars). The corresponding Fab residues contacting each peptide residue are indicated.

Table 5. Hydrogen bond interactions (<3.5 Å) between F11.2.32 and peptide(P36-P46)

HIV-1 peptide (P36-)46)	Fab F11.2.32	Distance (Å) Fab1	Fab2
A. *Direct hydrogen bonds*			
Met-P36 O	Tyr27D O^η (L1)	2.72	2.91
Met-P36 N	Tyr27D O^η (L1)	2.73	3.51
Met-P36 N	Tyr28 O^η (L1)	3.26	3.44
Arg-P41 $N^{\eta1}$	Tyr58 O^η (H2)	3.07	3.15
Arg-P41 $N^{\eta2}$	Lys92 O (L3)	2.92	2.90
Arg-P41 O	Tyr100F N (H3)	2.90	2.99
Lys-P43 N	Thr 100D O (H3)	2.99	2.88
Lys-P43 O	Thr100D O^η (H3)	2.89	2.91
Lys-P45 N	Asp100B O (H3)	2.98	2.98

B. *Hydrogen bonds mediated by water molecules*

Gly-P40 O ———3.55——— Water ———2.77——— Trp96 (L3) $N^{\epsilon1}$
Gly-P40 O ———2.67——— Water ———2.73——— His35 (H1) $N^{\epsilon2}$
Arg-P41 O ———2.63——— Water ———2.54——— Ser91 (L3) O^γ

In B, distances are displayed for the first complex only

elicited by a non-native form of the viral enzyme. The homodimeric protease has limited stability (Szeltner *et al.*, 1996; Mildner *et al.*. 1994) and we cannot exclude the possibility that the immunisation procedures used to obtain F11.2.32 (see Production of hybridomas in Materials and Methods) might have lead to presentation of the antigen in a form different from its native state. Indeed, the preliminary screening for anti-protease clones was made by direct enzyme-linked immunoabsorbant assay (ELISA) in which the enzyme was coated onto the surface of the wells of the plastic ELISA plate. Such procedures could have favoured selection of those clones secreting antibodies specific for denatured forms of the antigen. ELISA measurements made in the solution phase show, nonetheless, that F11.2.32 does bind to native HIV-1 protease. Moreover, FPLC analysis of Fab/protease mixtures allowed identification of a peak migrating differently from that of the Fab which proved, by SDS-PAGE, to contain both the antibody fragment and the enzyme (data not shown). Finally, the observed inhibition of proteolyic activity we report here shows unambiguously that F11.2.32 interacts directly with native HIV-1 protease in solution.

Each segment of the HIV-1 protease that is structurally similar to a segment of the bound peptide, namely, P36-P39 and P42-P45, was used in turn as a guide to detect possible docking orientations for the intact enzyme. In both cases, however, severe steric hindrance between the antibody and the docked protein antigen resulted. Likely modes of interaction between F11.2.32 and the HIV-1 protease during the initial encounter are therefore not immediately obvious from this simple and direct examination of the crystal structure. Nonetheless, it would seem plausible that the preliminary contacts formed between F11.2.32 and the protease could involve residues 39 to 41 of the antigen, since they occupy the centre of the antigen-binding site in the peptide complex. Indeed, induced conformational changes to this region of the enzyme might readily follow because these residues are exposed in the native enzyme structure and are quite mobile, as indicated by their thermal parameters.

If F11.2.32 were to induce conformational changes in the enzyme, as suggested by the Fab-peptide crystal structure, the distortion of the functionally important flap regions could be sufficient to inhibit proteolytic activity. Furthermore, molecular dynamics simulations suggest that local changes in HIV-1 protease conformation may provoke long-range effects within the monomer or between monomers within the dimer (Harte *et al.*, 1992). Thus, the effect upon binding may be even more extreme; the antibody might induce further structural changes more distant from the epitope which could, for example, lead to a more facile dissociation of the functional dimer to inactive monomers. Instances of antibody-induced structural changes in an antigen have been reported, as shown, for example, by the expulsion of the ferriheme group from metmyoglobin upon being bound by anti-apomyoglobin antibodies (Crumpton, 1965).

Potential cross-reactivity of F11.2.32 with protease from different viral isolates

A comparison of HIV-1 protease sequences from several viral isolates indicates that residues 38 to 40 and 42 to 46 are well conserved both in naturally occurring strains, as well as those selected by active site inhibitors, while residues 36, 37 and 41 are more prone to variation (Winslow *et al.*, 1995). Since Ser-P37 does not interact directly with residues of the Fab, mutations found at this position in several isolates (Ser37 → Asn, Asp, Cys, Ala, or Gln) might be tolerated without significant loss of affinity. On the other hand, the mutation Arg41 → Lys, found in certain isolates and in several clinical strains, could abolish recognition by

Fab F11.2.32, since Arg41 makes numerous contacts upon complex formation. Since most of the polar interactions observed in the Fab F11.2.32-peptide(P36-P46) complex are made by main-chain atoms of the peptide, a significant number of the mutations found in different viral isolates might be tolerated by the antibody. Such a lack of specificity has indeed been observed for certain peptide-binding proteins of known structure. The peptide-binding protein Oppa can accommodate side-chains of different nature in a voluminous hydrated cavity through polar interactions with main-chain atoms of the peptides (Tame et al., 1994). In the case of MHC molecules, disparate peptide sequences can be accommodated in the binding groove of the same MHC molecule (Matsumura et al., 1992, Madden et al., 1993).

Conclusion

In this study, we have determined the three-dimensional structure of the Fab fragment of an anti-HIV-1 protease monoclonal antibody, F11.2.32, both in the unliganded form and as a complex with a peptide fragment from the putative epitope of the viral enzyme. If the observed structure of the Fab-peptide complex closely mimics the interaction between F11.2.32 and HIV-1 protease, the antibody must induce significant conformational changes in the native enzyme upon binding to it. Accordingly, we have identified a region defined between residues 36 and 46 on the surface of HIV-1 protease which appears to be important for its function, since binding by F11.2.32 leads to a significant inhibition of the proteolytic activity.

Loeb et al. (1989) have shown that mutagenesis of HIV-1 protease between residues 36 and 46 often produces an enzyme of reduced activity. Thus, while HIV-1 protease with substitutions at these positions could escape binding by F11.2.32, it might also be less active. Hence, this region next to the flap could be an alternative target for anti-viral agents in the form of intra-cellular antibodies (Chen et al., 1994; Duan et al., 1994) or CDR mimetics (Saragovi et al., 1991) whose design could be guided by the structure presented here.

Materials and Methods

Production of hybridomas

Monoclonal antibodies, raised against recombinant HIV-1 protease with the BH10 isolate sequence (Ratner et al., 1985), were obtained by the hybridoma technique of Köhler & Milstein (1975). BALB/c mice were immunised by administering a 10 μg dose of protease in the presence of Freund's adjuvant, followed by three booster injections of the same dose at 15-day intervals. A final injection of 10 μg of protease without adjuvant was given three days before sacrifice, and the splenocytes thus obtained were hybridised and cloned by limiting dilution. Clones were selected for anti-protease activity as measured by direct ELISA. The mAb F11.2.32 (IgG1,κ),

expressed by one of these cell lines, was purified from ascites fluid as described for other immunoglobulins (Poljak, 1985) and the Fab was obtained by papain digestion.

Inhibition of HIV-1 protease by F11.2.32

Kinetic measurements were performed according to methods described by Urban et al. (1992) and Majer et al. (1993). Rates of cleavage of the substrate KARVNle-F(NO₂)EANle (Richards et al., 1990) were followed in PBS buffer (pH 7.2) with 1% (v/v) Tween 20, 1% (w/v) BSA, 0.02% (v/v) β-mercaptoethanol and 1 mM EDTA. HIV-1 protease at concentration 50 nM was first incubated with varying amounts of Fab F11.2.32 (0 to 600 nM) for 20 minutes at 37°C and the reaction was started by adding the substrate solution (final concentration 0.125 mM). Aliquots of the reaction mixtures were stopped after fixed time intervals by adding 20% (v/v) TFA. The total reaction time for each run was one hour. The substrate cleavage products were measured by reverse-phase HPLC (Merck) using a VYDAC C18 column (4.5 mm × 1500 mm) and an acetonitrile/water gradient. The resulting data were analysed with the program Enzfitter (version 1.03, R.L. Latherbarrow, Elsevier-Biosoft, 1987).

Nucleotide and protein sequencing of F11.2.32

A total mRNA fraction was purified from 2×10^6 hybridoma cells using the Quick Prep Micro mRNA Purification kit (Pharmacia Biotech). The cDNA was produced from 1 μg of mRNA using 200 units of Superscript Reverse Transcriptase (GIBCO BRL) and a random selection of hexanucleotide primers (50 pmol) in a reaction volume of 20 μl. Genes coding for the variable domains were amplified from 1 μl of cDNA in each case by the polymerase chain reaction, using 60 pmol each of V_L-5′ and V_L-3′ primers for V_L, and 250 pmol of V_H-5′ and 60 pmol V_H-3′ primers for V_H. The amplified gene fragments were purified on 2% (w/v) agarose (SeaPlaque, FMC) and digested with PvuII and BglII for the V_L, and with PstI and EcoRI for V_H. Following purification on 2% agarose, the V_L and V_H gene inserts were cloned into pBluescript (cut with EcoRV and BamHI) and pUC18 (cut with PstI and EcoRI) vectors, respectively. The primers were: V_L-5′, 5′-d(GACACTCAGCTGACCCAGTCTC-CA),V_L-3′, 5′-d(GTTAGATCTCCAGCTTGGTCCC), V_H-5′, 5′-d(GACGTSCARCTGCAGGARTCWGG), V_H-3′, 5′-d(ATGGAATTCTGAGGAGACTGT). (Restriction sites, which are underlined, are PvuII, BglII, PstI and EcoRI, respectively, and standard abbreviations have been used for mixed sites: S, C or G; R, A or G; W, A or T.) The V_L-3′ primer is according to Orlandi et al. (1989) while the V_H-3′ primer is according to Ge et al. (1995). The V_L-5′ and V_H-5′ primers were designed from the sequence of the first eight residues of the light and heavy chains, as obtained by Edman degradation on a 470 Applied Biosystems Sequencer. Nucleotide sequencing was performed in both directions by the dideoxy chain termination method on an automatic sequencer and confirmed by comparison of at least three products from separate polymerase chain reactions. Sequences were analysed using the program Blast (Altschul et al., 1990). The complete sequence of V_H and V_L domains is given in Table 4; numbering of the amino acid positions follows the convention of Kabat et al. (1991). The nucleotide sequences of the variable heavy and light chains have

been deposited with the Genbank (accession numbers U62632 and U62633, respectively).

Diffraction measurements from the free Fab F11.2.32

Diffraction measurements and data reduction for the Fab F11.2.32-peptide(P36-P46) complex have been described (Lescar et al., 1996). Data for the free Fab F11.2.32 were measured to 2.6 Å resolution from a single crystal using flash-freezing techniques. The crystal was prepared for cryocooling by soaking in the crystallisation buffer to which glycerol (20% v/v) had been added, then mounted on a fibre loop and maintained at 110 K using the Oxford Cryosystem. X-ray diffraction data were collected on a MARresearch imaging plate using graphite-monochromated Cu-Kα radiation. The unit cell parameters of the unliganded Fab (Table 1) differ from those reported previously (Lescar et al., 1996), perhaps reflecting the variability that can occur with different flash-frozen samples. From a total of 105,230 measured Bragg intensities extending to 2.6 Å resolution, a unique data set of 25,505 reflections (96.8% complete) was obtained with a merging R-factor of 0.106 (55.4% of the data have an intensity greater than 3σ in the highest resolution shell).

Structure refinement

Fab F11.2.32-peptide(P36-P46)

Initial models for the free and complexed Fab were obtained by molecular replacement as previously reported (Lescar et al., 1996). Since data for the peptide complex extended to a higher resolution, this structure was refined first. During the initial stages of the refinement, the sequence of the antibody was not known, except for the first eight amino-terminal residues of the light and heavy chains, as obtained by amino acid microsequencing. With this information, tentative assignments were made for the V_H and V_L families, and a model for each domain was selected accordingly from known antibody structures. Thus, V_L was modelled on AN02 (Brünger et al., 1991, PDB entry 1baf) and V_H on NQ10 (Alzari et al., 1990), but with the three CDR loops of each domain removed and all non-conserved framework residues within the subgroup modelled as alanine. The atomic model was subjected to alternate cycles of refinement by simulated annealing using the program X-PLOR (Brünger et al., 1987) with the molecular parameter set of Engh & Huber (1991), and manual model building to $(2F_o - F_c)$ electron density maps using the programs QUANTA (Molecular Simulations Inc.) and O (Jones et al., 1991). Calculations were made with a set of reflections taken between the resolution limits of 7.0 Å and 2.2 Å with $F > 3σ$ (35,173 reflections). The initial model from molecular replacement had an R-factor of 0.439 and an R_{free} (Brünger, 1992) of 0.448. (1748 data were randomly chosen to monitor R_{free}).

During most of the refinement procedure, only one of the two independent molecules in the asymmetric unit was constructed using real-space averaged electron density maps calculated with the program DM from the CCP4 program suite (CCP4, 1994). After a series of modifications had been made to the model, the second molecule was generated from the first for the subsequent cycle of X-PLOR by applying the non-crystallographic symmetry. Using this procedure, five of the six CDR loops were successively constructed in the following order: CDR2-L, CDR3-L, CDR1-H, CDR1-L and CDR2-H. Some connected density became visible at the centre of the antigen-binding site as the refinement progressed, but no interpretation was attempted yet, since CDR3-H was absent from the model. The model, without the peptide and the CDR3-H loop included, had an R-factor of 0.262, and an R_{free} value of 0.360. At this stage, the sequence of F11.2.32 became available (Table 4) and many of the appropriate side-chains could be successfully fitted to the electron density. The subsequent cycle of X-PLOR returned an averaged electron density map in which the main chain of CDR3-H, comprising 16 residues, could be unambiguously traced, although some side-chains at the extremity of the loop were not visible.

Construction of the bound peptide was then begun. The aromatic ring of Trp-P42 and the guanidinium group of Arg-P43 were clearly identified in the difference map, and an unambiguous tracing of residues P38 to P42 of the peptide readily followed. From this point onwards, the two molecules in the asymmetric unit were refined and built independently. After additional cycles of simulated annealing, the peptide had been traced from Met-P36 to Lys-P45 (although no side-chain could be placed for P45). Solvent molecules were located with the aid of the program ARP (Lamzin & Wilson, 1993), working in the restrained-option mode; in total, 400 solvent molecules were thus placed. (87 pairs of solvent molecules are related by the non-crystallographic dyad.) The final R-factor and R_{free} value were 0.198 and 0.289, respectively. Refinement statistics are summarised in Table 2.

Free Fab F11.2.32

The coordinates of the Fab from the refined peptide complex were superimposed onto the molecular replacement solution of the free Fab (two molecules in the asymmetric unit) to generate the initial model. Structure factors for the refinement were those with $F > 2σ$ lying between the resolution limits 8.0 Å and 2.6 Å (24,431 reflections). Residues L27D to L29 from CDR1-L and residues H96 to H103 from CDR3-H were removed from both molecules, since these correspond to the extremities of long, flexible loops. During the refinement with X-PLOR, non-crystallographic restraints were applied between the two molecules of the asymmetric unit. Initial R-factor and R_{free} were 0.418 and 0.433, respectively (1181 reflections were used to monitor R_{free}). Alternate cycles of refinement and rebuilding to electron density maps led to the successful construction of the main chain of CDR1-L and CDR3-H, together with most of their side-chains, giving an R-factor and R_{free} of 0.251 and 0.345, respectively.

Most of the intermolecular contacts are made between the two Fab molecules in the asymmetric unit which are related by a non-crystallographic dyad (see Results and Discussion). Moreover, side-chains involved in other intermolecular contacts were generally well defined. A cycle of simulated annealing, with non-crystallographic symmetry restraints removed, was nonetheless made at this point; although the R-factor dropped to 0.245 and the R_{free} rose to 0.373. Refinement was therefore continued with restraints applied as before, except for the protruding loops of CDR1-L and CDR3-H, as well as the side-chains of Arg-L18 and Arg-H18 for which the electron density showed clear deviations from non-crystallographic symmetry. A total of 235 solvent molecules, of which 78 pairs were related by the non-crystallographic dyad, were placed in subsequent cycles. Solvent molecules were retained in the model only if

their temperature factor was less than 40 Å^2 and if they had an equivalent solvent molecule in the second Fab, or less than 25 Å^2 if there was no such equivalent solvent. The final model, for which the r.m.s. deviation between all (main and side-chain) atoms of the two Fabs was 0.24 Å after superposition, gave an R-factor of 0.206 and R_{free} of 0.317. Final refinement statistics are given in Table 2.

Structure analysis and modelling

Solvent-accessible surface areas were calculated with a probe radius of 1.7 Å and standard van der Waals radii for protein atoms using the program MS (Connolly, 1983). Electrostatic surface potentials were mapped on the solvent-accessible surface with the program GRASP (Nicholls *et al.*, 1991) using the full charge set and default parameters. Framework residues chosen for superpositions, using the program LSQKAB from the program suite (CCP4, 1994) were taken from the structural alignment proposed by Padlan (1994).

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Alzari, P. M., Spinelli, S., Mariuzza, R. A., Boulot, G., Poljak, R. J., Jarvis, J. M. & Milstein, C. (1990). Three dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. *EMBO J.* **9**, 3807–3814.

Ashorn, P., McQuade, T. J., Thaidrivongs, S., Tomaselli, A. G., Tarpley, W. G. & Moss, B. (1990). An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection. *Proc. Natl Acad. Sci. USA*, **87**, 7477–7476.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O. & Shimanouchi, T. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–523.

Björling, E., Goobar-Larson, L., Utter, G., Norrby, E. & Chiodi, F. (1992). Four distinct antigenic regions are present in the primary structure of HIV-1 and HIV-2 proteinases. *AIDS*, **6**, 157–163.

Bossart-Whitaker, P., Chang, C. Y., Novotny, J., Benjamin, D. C. & Sheriff, S. (1995). The crystal structure of the antibody N10-staphylococcal nuclease complex at 2.9 Å resolution. *J. Mol. Biol.* **253**, 559–575.

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

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic *R* factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Brünger, A. T., Leahy, D. J., Hynes, T. R. & Fox, R. O. (1991). 2.9 Å resolution structure of an anti-dinitrophenyl-spin label monoclonal antibody Fab fragment with bound hapten. *J. Mol. Biol.* **221**, 239–256.

CCP4 (Collaborative Computing Project, No. 4)., (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.

Chen, S., Khouri, Y., Bagley, J. & Marasco, W. A. (1994). Combined intra- and extracellular immunization against human immunodeficiency virus type 1 infection with a human anti-gp120 antibody. *Proc. Natl Acad. Sci. USA*, **91**, 5932–5936.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature*, **342**, 877–883.

Collins, J. R., Burt, S. K. & Erickson, J. W. (1995). Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics. *Nature Struct. Biol.* **2**, 334–338.

Connolly, M. J. (1983). Analytical molecular surface calculation. *J. Appl. Crystallog.* **16**, 548–558.

Croix, D. A., Yeh, H. Y., Sedlacek, J., Luftig, R. B. & Gottlieb, P. D. (1993). A dominant epitope of HIV-1 protease recognised by hamster monoclonal antibodies. *J. Acquir. Immune Defic. Syndr.* **6**, 558–566.

Crumpton, M. J. (1965). Conformational changes in sperm-whale metmyoglobin due to combination with antibodies to apomyoglobin. *Biochem. J.* **100**, 223–232.

Duan, L., Bagasra, O., Laughlin, M. A., Oakes, J. W. & Pomerantz, R. J. (1994). Potent inhibition of human immunodeficiency virus type 1 replication by intracellular anti-rev single chain antibody. *Proc. Natl Acad. Sci. USA*, **91**, 5075–5079.

Dyson, H. J. & Wright, P. E. (1991). Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 519–538.

Dyson, H. J., Lerner, R. A. & Wright, P. E. (1988). The physical basis for induction of protein-reactive anti-peptide antibodies. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 305–324.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A*, **47**, 392–400.

Fitzgerald, P. M. D. & Springer, J. P. (1991). Structure and function of retroviral proteases. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 299–320.

Ge, L., Knappik, A., Pack, P., Freund, C. & Plückthun, A. (1995). Expressing antibodies in *Escherichia coli*. In *Antibody Engineering* (Borrebaeck, C. K., ed.), pp. 229–266. Oxford University Press, New York, Oxford.

Ghiara, J. B., Stura, E. A., Stanfield, R. L., Profy, A. T. & Wilson, I. A. (1994). Crystal structure of the principal neutralization site of HIV1. *Science*, **264**, 82–85.

Golinelli-Pimpaneau, B., Gigant, B., Bizebard, T., Navaza, J., Saludjian, P., Zewel, R., Tawfik, D. S., Eshar, Z., Green, B. S. & Knossow, M. (1994). Crystal structure of a catalytic antibody Fab with esterase like activity. *Structure*, **2**, 175–183.

Harte, W. E., Swaminathan, S. & Beveridge, D. L. (1992). Molecular dynamics of HIV-1 protease. *Proteins: Struct. Funct. Genet.* **13**, 175–194.

Henderson, L. E., Bowers, M. A., Sowder, R. C., Serabyn, S. A., Johnson, D. G., Bess, J. W., Arthur, L. O., Bryant, D. K. & Fenselau, C. (1992). Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings and complete amino acid sequences. *J. Virol.* **66**, 1856–1865.

Jeffrey, P. D., Schildbach, J. F., Chang, C. Y., Kussie, P. H., Margolies, M. N. & Sheriff, S. (1995). Structure and specificity of the anti-digoxin antibody 40-50. *J. Mol. Biol.* **248**, 344–360.

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

Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest. 5th edit.* US Department of Health and Human Services, NIH, Bethesda, MD.

Katz, R. A. & Skalka, A. M. (1994). The retroviral enzymes. *Annu. Rev. Biochem.* **63**, 133–173.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M. & Sigal, I. S. (1988). Active human immunodeficiency virus protease is required for viral activity. *Proc. Natl Acad. Sci. USA,* **85**, 4686–4690.

Köhler, G. & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predetermined specificity. *Nature,* **256**, 495–497.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic photos of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallog. sect. D,* **49**, 129–147.

Laskowski, R. A., Mc Arthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Lescar, J., Stouracova, R., Riottot, M-M, Chitarra, V., Brynda, J., Fabry, M., Horejsi, M., Sedlacek, J. & Bentley, G. A. (1996). Preliminary crystallographic studies of an anti-HIV-1 protease antibody which inhibits enzyme activity. *Protein Sci.* **5**, 966–968.

Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E. & Hutchinson, C. A. (1989). Complete mutagenesis of the HIV-1 protease. *Nature,* **340**, 397–400.

Luzzati, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallog.* **5**, 802–810.

MacCallum, R. M., Martin, A. C. R. & Thornton, J. M. (1996). Antibody-antigen interactions: contact analysis and binding site topography. *J. Mol. Biol.* **262**, 732–745.

Madden, D. R., Garboczi, D. N. & Wiley, D. C. (1993). The antigenic identity of peptide-MHC complexes: a comparison of the conformation of five viral peptides presented by HLA-A2. *Cell,* **75**, 693–708.

Majer, P., Urban, J., Gregorova, E., Konvalinka, J., Novek, P., Stehlikova, J., Andreansky, M., Sedlacek, J. & Strop, P. (1993). Specificity mapping of HIV-1 protease by reduced bond inhibitors. *Arch. Biochem. Biophys.* **304**, 1–8.

Matsumura, M., Fremont, D. H., Peterson, P. A. & Wilson, I. A. (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science,* **257**, 927–934.

Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarich, J. L., Howe, W. J., Tomich, C. C., Smith, C. W., Heinrikson, R. L. & Tomasselli, A. (1994). The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. *Biochemistry,* **33**, 9405–9413.

Milner-White, E. J., Ross, B. M., Ismail, R., Belhady-Mostefa, K. & Poet, R. (1988). One type of gamma turn rather than the other gives rise to chain reversal in proteins. *J. Mol. Biol.* **204**, 777–782.

Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L. & Springer, J. P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature,* **337**, 615–620.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Nicholson, L. K., Yamazaki, T., Torchia, D. A., Grzesiek, S., Bax, A., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Lam, P. Y. S., Jadhav, P. K., Hodge, C. N., Domaille, P. J. & Chang, C.-H. (1995). Flexibility and function in HIV-1 protease. *Nature Struct. Biol.* **2**, 274–280.

Orlandi, R., Güssow, D. H., Jones, P. T. & Winter, G. (1989). Cloning immunoglobulin variable domains for expression by the polymerase chain reaction. *Proc. Natl Acad. Sci. USA,* **86**, 3833–3837.

Padlan, E. A. (1994). Anatomy of the antibody molecule. *Mol. Immunol.* **31**, 169–217.

Pearl, L. H. & Taylor, W. R. (1987). A structural model for the retroviral proteases. *Nature,* **329**, 351–354.

Poljak, R. J. (1985). Crystallisation of immunoglobulins and their fragments for X-ray diffraction studies. *Methods Enzymol.* **116**, 190–200.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong Staal, F. (1985). Complete nucleotide sequence of the AIDS virus HTLV-III. *Nature,* **313**, 277–284.

Richards, A. D., Phylip, L. H., Farmerie, W. G., Scarborough, P. E., Alvarez, A., Dunn, B. N., Hirel, Ph.-H., Konvalinka, J., Strop, P., Pavelickova, L., Kostka, V. & Kay, J. (1990). Sensitive, soluble chromogenic substrates for HIV-1 proteinase. *J. Biol. Chem.* **265**, 7733–7736.

Rini, J. M., Stanfield, R. L., Stura, E. A., Salinas, P. A., Profy, A. T. & Wilson, I. A. (1993). Crystal structure of a HIV-1 neutralizing antibody, 50.1 in complex with its V3 peptideloop antigen. *Proc. Natl Acad. Sci. USA,* **90**, 6325–6329.

Rosé, J., Babé, L. & Craik, C. (1995). Defining the level of Human Immunodeficiency Virus Type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. *J. Virol.* **69**, 2751–2758.

Saragovi, H. U., Fitzpatrick, D., Raktabutr, A., Nakanishi, H., Kahn, M. & Greene, M. I. (1991). Design and synthesis of a mimetic from an antibody complementarity determining region. *Science,* **253**, 792–795.

Stanfield, R. L. & Wilson, I. A. (1993). X-ray crystallographic studies of antibody-peptide complexes. *Immunomethods,* **3**, 211–221.

Stanfield, R. L. & Wilson, I. A. (1995). Protein-peptide interactions. *Curr. Opin. Struct. Biol.* **5**, 103–113.

Stanfield, R. L., Takimoto-Kamimura, M., Rini, J. M., Profy, A. T. & Wilson, I. A. (1993). Major antigen-induced domain rearrangements in an antibody. *Structure,* **1**, 83–93.

Szeltner, Z. & Polgar, L. (1996). Conformational stability and catalytic activity of HIV-1 protease are both enhanced at high salt concentration. *J. Biol. Chem.* **271**, 5458–5463.

Tame, J. R. H., Murshadov, G. N., Dodson, E. J., Neil, T. K., Dodson, G. G., Higgins, C. F. & Wilkinson, A. J. (1994). The structural basis of sequence independent peptide binding by OppA protein. *Science,* **264**, 1578–1581.

Tormo, J., Blaas, D., Parry, N. R., Rowlands, D., Stuart, D. & Fita, I. (1994). Crystal structure of a human rhinovirus neutralizing antibody complexed with a peptide derived from the viral capsid protein VP2. *EMBO J.* **13**, 2247–2256.

Urban, J., Konvalinka, J., Stehlikova, J., Gregorova, E., Majer, P., Soucek, M., Andreansky, M., Fabry, M. & Strop, P. (1992). Reduced-bond tight-binding inhibitors of HIV-1 protease. *FEBS Letters,* **298**, 9–13.

Van Regenmortel, M. H. V. (1986). Which structural features determine protein antigenicity? *Trends Biochem. Sci.* **11**, 36–39.

Van Regenmortel, M. H. V. (1987). Antigenic cross-reactivity between proteins and peptides: new insights and applications. *Trends Biochem. Sci.* **12**, 237–240.

Wien, M. W., Filman, D. J., Stura, E. A., Guillot, S., Delpeyroux, F., Crainic, R. & Hogle, J. M. (1995). Structure of the complex between the Fab fragment of a neutralizing antibody for type 1 poliovirus and its viral epitope. *Nature Struct. Biol.* **2**, 232–243.

Williams, J. W. & Morrison, J. F. (1979). The kinetics of reversible tight-binding inhibition. *Methods Enzymol.* Purich, D. L., ed. **63**, 437–467.

Wilson, I. A. & Stanfield, R. L. (1994). Antibody-antigen interactions: new structures and new conformational changes. *Curr. Opin. Struct. Biol.* **4**, 857–867.

Winslow, D. L., Stack, S., King, R., Scarnati, H., Bincsik, A. & Otto, M. J. (1995). Limited sequence diversity of the HIV type 1 protease gene from clinical isolates and in vitro susceptibility to HIV protease inhibitors. *AIDS Res. Hum. Retroviruses,* **11**, 107–113.

Wlodawer, A., Miller, M., Jakolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L., Clawson, L. M., Schneider, J. & Kent, S. B. H. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science,* **245**, 616–621.

Edited by I. A. Wilson

1. H. Hope, *Annu. Rev. Biophys. Chem.* **19** (1990) 107.

2. K. D. Watenpaugh, *Curr. Opin. Str. Biol.* **1** (1991) 1012.

3. D. W. Rodgers, *Structure* **2** (1994) 1135.

4. C. Nave, *Radiat.Phys. Chem.* **45** (1995) 483.

5. D.J. Haas, M.G. Rosmann, *Acta Cryst.* **B51** (1970) 998.

6. T. Y. Teng, *J. Appl. Cryst.* **23** (1990) 387.

7. H. Hartmann, F. Parak, W. Steigemann, G. A. Petsko, D. Ringe-Ponzi, H. Frauenfelder, *Proc. Natl. Acad. Sci. USA* **79** (1982) 4967.

8. U. F. Thomanek, F. Parak, R. L. Mossbauer, H. Formanek, P. Schwager, W. Hope, *Acta Cryst.* **A29** (1973) 263.

9. D. V. Rodgers, in *International Tables for Crystallography Vol. F* (2001) 202.

10. E. F. Garman, E. P. Mitchell, *J. Appl. Cryst.* **29** (1996) 584.

11. T.-Y. Teng, *J. Appl. Cryst.* **31** (1998) 252.

X-ray protein diffractometer and basis of the rotation method
PROTEINOVÝ DIFRAKTOMETR – MĚŘENÍ ROTAČNÍ METODOU

Jiří Brynda

Department of Gene Manipulation, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 166 37 Prague 6, Czech Republic

Keywords:

protein diffractometer, rotation method, macromolecular crystallography

Abstract

X-ray data collection is central experiment in a crystal structure analysis. Recent advances such as cryogenic cooling and two-dimensional detectors have made data collection technically easier, but it remains fundamental scientific procedure underpinning the whole structural analysis. A brief discussion of the most important aspects of the data collection is given. Reviews on X-ray data collection techniques in macromolecular crystallography can be found in: International Tables for Crystallography, *Volume F,* 177 - 195

Úvod

Měření difrakčních dat je základní experiment v krystalové strukturní analýze. Pro struktury malých molekul jsou běžně měřena data do atomového rozlišení (většinou do 0.8 Å), takže fázový problém může být řešen přímo a pozice atomů mohou být rafinovány plně anizotropním modelem.

Krystaly makromolekul jsou z hlediska měření difrakčních dat mnohem problematičtější. První problém vyplývá z velikosti elementární buňky, díky které jsou průměrné intenzity reflexí mnohem nižší a jejich počet vyšší (tabulka 1). Za druhé, část objemu krystalu tvoří kanály vyplněné neuspořádaným rozpouštědlem, takže dále snižují intenzitu reflexí na vysokém rozlišení a ve většině případů omezují rozlišení na mnohem nižší než atomové. Za třetí, opět především díky obsahu rozpouštědla, jsou krystaly citlivé na radiační poškození. Nedávný technický pokrok v měřících metodách, jako např. používání synchrotronového záření, měření zmražených krystalů a vysoce účinné plošné detektory, usnadnil měření, ale to stále zůstává základem celé strukturní analýzy. Proto je nezbytné tomuto klíčovému kroku věnovat maximální pozornost.

Tabulka 1. Velikost elementární buňky a počet reflexí.

Sloučenina	Mříťový parametr [Å]	Objem buňky [Å]	Počet reflexí	Prům. intenzita
Organická sl.	10	1000	2000	1
Supermolekula	30	27000	30000	1/27000
Protein	100	10^6	10^5	10^{-6}
Virus	400	10^8	10^6	10^{-8}

1. Součásti proteinového difraktometru

Pro sběr dat z makromolekulárního monokrystalu při konkrétní vlnové délce potřebujete difraktometr složený z následujících součástí:

(1) zdroj RTG záření
(2) optickou soustavu pro fokusaci RTG záření na vzorek
(3) monochromátor pro selekci jedné vlnové délky záření
(4) kolimátor pro úpravu rozměrů RTG záření
(5) „závěrku" (shuter) pro kontrolu expozice vzorku
(6) goniostat spojený s držákem vzorku

Dále je velice vhodné mit k dispozici:

(1) zařízení pro měření zmražených krystalů
(2) účinný plošný detektor
(3) software pro měření, ukládání a zobrazování exponovaných dat
(4) software pro zpracování dat, tj. získání intenzit jednotlivých Braggovských difrakcí a jejich chyb.

2. Zdroj RTG záření - rotační anoda

Ve strukturní analýze makromolekul je praktiky vždy používána Cu anoda s fixní vlnovou délkou 1.542 Å (pokud není zdrojem synchrotronové záření). Rotační anoda má proti klasické rentgence výhodu vyšší intenzity RTG záření. Systémy jsou dnes běžně vybaveny grafitovým monochromátorem, nebo fokusačními zrcadly,

nebo vícevrstvou optikou (Osmic), které zajišťují fokusaci a monochromatizaci RTG záření. Velice důležitá je volba apretury kolimátoru, ta by měla odpovídat velikosti krystalu, avšak pro velké krystaly s velkou elementární buňkou je lépe požít kolimator s menší apreturou něž velikost krystalu pro lepší prostorové rozlišení reflexí.
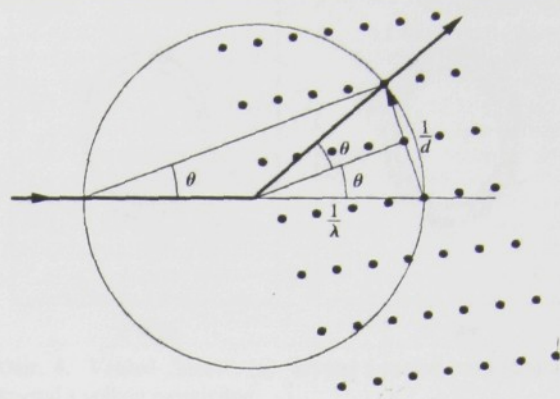
3. Goniostat a detektor

Difrakční podmínka je pro konkrétní reflexi splněna pokud odpovídající bod reciproké mříže protíná plochu Ewaldovy sféry. Pokud je ozařován stacionární krystal RTG zářením, pak pouze některé reflexe splňují difrakční podmínku. Proto pokud chceme změřit vetší množství dat (pokud možno všechny měřitelné, tj. kompletní soubor dat), musíme změnit buď velikost Ewaldovy sféry nebo orientaci krystalu. První možnost, s využitím polychromatického, „bílého" RTG záření je základ Laueho metody. Pokud používáme monochromatické záření vybrané vlnové délky, pak musíme geometrii krystalu během expozice měnit tak, aby byly difrakční podmínky splněny pro více reflexí. Proteinové difraktometry prakticky vždy požívají rotaci krystalu a plošný detektor.

Prvním plošnými detektory byly fotografické filmy automaticky snímané optickými denzitometry. Takto digitalizované difrakční obrazce byly přeneseny na disk. Hlavní nevýhody použiti fotografických filmů byla pracnost, časová náročnost, vysoké pozadí a malý dynamický rozsah. Postupně byly vyvinuty další plošné detektory: drátové proporcionální čítače, „Imaging plates" a konečně CCD detektory (charge-coupled devices). Prakticky všechny současné plošné detektory jsou dnes používány v kombinaci s goniostatem, který zajišťuje rotaci krystalu okolo jedné osy během expozice. Zbylé přístroje využívají kappa geometrie (ω, κ, φ) kolíbky goniometru pro nastavení krystalu do další počáteční polohy, ale během měření taktéž rotují krystal kolem jedné osy.

4. Základy rotační metody

Při použití Ewaldovy konstrukce (obr. 1), která je geometrickou interpretací Braggova zákona, je monochromatické záření representováno povrchem koule o poloměru $1/\lambda$ a krystal reciprokou mřížkou. Reciproká mřížka sestává z bodů ležících na konci vektorů kolmých k difrakčním rovinám a jejich délka je úměrná převrácené hodnotě mezirovinné vzdálenosti $1/d$. Při rotační metodě je krystal otáčen o daný úhel definovaný jako úhel φ.

Pro krystal v jedné orientaci bude pouze malá část reflexí splňovat difrakční podmínku. Počet reflexí bude velice malý pro malé molekuly (v některých orientacích i nulový). Krystaly makromolekul mají však větší elementární buňky (a, b, c v řádu 100 Å), pokud srovnáváme s vlnovou délkou použitého záření, což v geometrické interpretaci znamená, že reciproká mřížka je velice hustě zaplněna vzhledem k velikosti Ewaldovy sféry. Z toho vyplývá, že mnoho reflexí difraktuje současně do různých směrů, jelikož mnoho bodů reciproké mřížky leží současně na povrchu Ewaldovy sféry. To také vysvětluje výhodu velkých plošných detektorů pro měření krystalů s velkou elementární buňkou.

Obr.1. Ewaldova konstrukce. Bod reciproké mříťky leťí na povrchu Ewaldovy sféry pokud je splnena následující trigonometrická podmínka: $1/2d = (1/\lambda) \sin \theta$. Po jednoduché úpravě dostaneme Braggovu rovnici: $\lambda = 2d \sin\theta$. Takťe pokud bod s indexy *hkl* reciproké mříťky leťí na povrchu Ewaldovy sféry, je splněna difrakční podmínka a dojde k difrakci v daném směru.

Body reciproké mříže leží v rovinách, takže protínají Ewaldovu sféru v kružnicích. Odpovídající difraktované reflexe z centra Ewaldovy sféry tak tvoří kužely a jelikož při většině experimentů je detektor umístěn kolmo na směr primárního svazku záření, tvoří reflexe rovin elipsy (viz obr. 2). Pokud některá osa krystalu leží téměř paralelně s primárním zářením, elipsy vytvářejí soustředné kruhy s centrem ve středu detektoru. Všechny reflexe ležící na jedné kružnici (elipse) mají jeden index společný a pro necentrované buňky postupně vzrůstá o jednotku. Vzdálenosti mezi kruhy závisejí na převrácené hodnotě parametru rovnoběžného s primárním zářením.
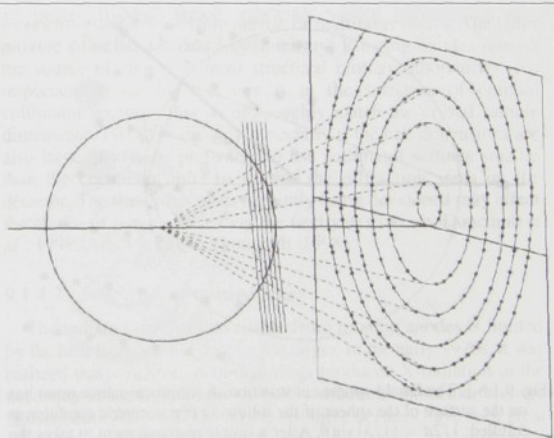
Obr. 2 Rovina reflexí, která je přibliťně kolmá k záření dá vzniknout elipse reflexí na detektoru.

4.1. Mosaicita krystalu a divergence záření

Při Ewaldově konstrukci předpokládáme ideální zdroj záření s totálně paralelním zářením a ideální krystal se všemi elementárními buňkami stejně orientovanými, takže výsledkem by byly bodové, nekonečně ostré Braggovské difrakce. Reálné experimentální podmínky se od tohoto ideálu liší ve třech aspektech. Za prvé, dopadající záření není nikdy dokonale paralelní, za druhé, reálné záření není striktně monochromatické a za třetí, reálné krystaly jsou tvořeny malými mosaikovými bloky, které nejsou vzájem-

(a)

(b)
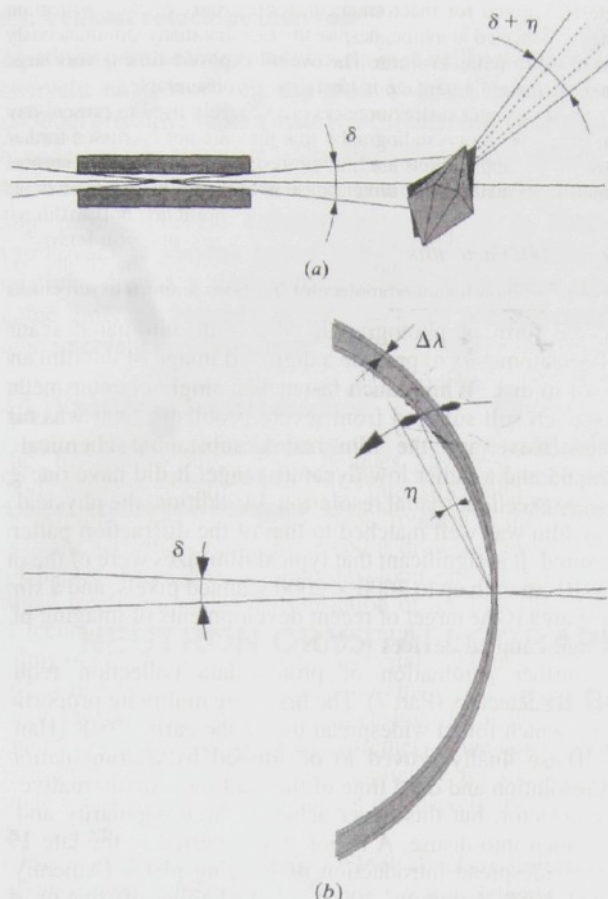
Obr. 3 Schematické znázornění divergence záření δ a mosaicity krystalu η. (a) V přímém prostoru, (b) v reciprokém prostoru, kde je Ewaldova sféra rozšířena díky konečné šířce pásu použité vlnové délky δλ/λ.

ně přesně stejně orientované. Pokud měříme krystaly makromolekul při laboratorní teplotě je rozpětí vzájemné orientace mosaikových bloků (mosaicita) většinou do 0.05°, ale pokud krystaly měříme zmražené můťe mosaicita vzrůst ať na 1° i více. Všechny tyto efekty (schematicky znázorněné na obr. 3) rozšiřují difrakce na větší rozsah rotace krystalu.

4.2. Rotační snímky a „lunes"

Pokud chceme změřit další reflexe, které v danné orientaci neleží na povrchu Ewaldovy sféry, při použití monochromatického záření, musíme krystalem otáčet aby pro další reflexe byly splněny difrakční podmínky. Pokud je krystal otáčen okolo jedné osy během po sobě následujících expozicích, jedná se o rotační metodu. V praxi je rotační osa skoro vždy kolmá k dopadajícímu záření, tak aby byla zachována symetrie mezi oběma polovinami kompletního snímku. Toto je nejčastěji používaná metoda při měření difrakčních dat krystalů makromolekul. Pokud je krystal rotován během expozice, elipsy popsané pro fixní pozici krystalu mění svoji pozici. Výsledkem je, že snímek vzniklý během expozice obsahuje všechny difrakce mezi dvěmi limitními pozicemi elips na začátku a konci dané rotace, které tvoří tzv. „půlměsíce" (lunes).
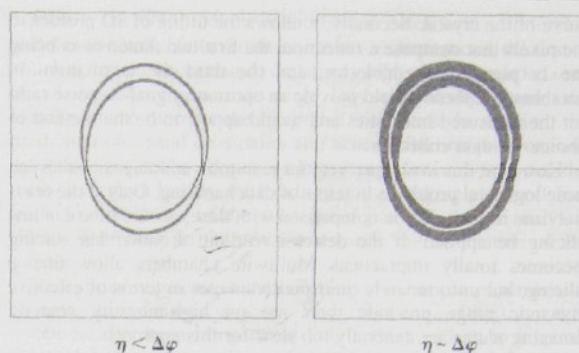
Obr. 4. Vzhled „lunes": (a) krystal s malou mosaicitou, (b) krystal s velkou mosaicitou.

4.3. Částečně a úplně změřené difrakce

Při rotační metodě vznikají „lunes" jež obsahují difrakce mezi elipsami vzniklými mezi místy, která byla na Ewaldově sféře proťata danou rovinou reciproké mřížky mezi začátkem a koncem rotačního intervalu Δφ. Pokud byla Ewaldova sféra proťata daným bodem reciproké mříže kompletně, je změřena úplně a naopak, pokud bod reciprokého prostoru prošel jen z části Ewaldovou sférou, dokončí průchod při další rotaci a daná reflexe je na jednotlivých rotačních snímcích registrována po částech. Vzhledem k velké hustotě reciproké mříže je nemožné změřit všechny reflexe naráz během jedné velké rotace, jelikož by docházelo k překrývání difrakcí. Na šířku „lunes" má zásadní vliv mosaicita krystalu (vliv divergence a mosaicity nelze jednoduše separovat, navíc divergence je prakticky konstantou pro daný zdroj a je většinou zanedbatelná ve srovnání s mosaicitou). Pokud je mosaicita malá jsou „lunes" úzké, naopak velká mosaicita značně rozšiřuje plochu „lunes" (viz obr. 4).
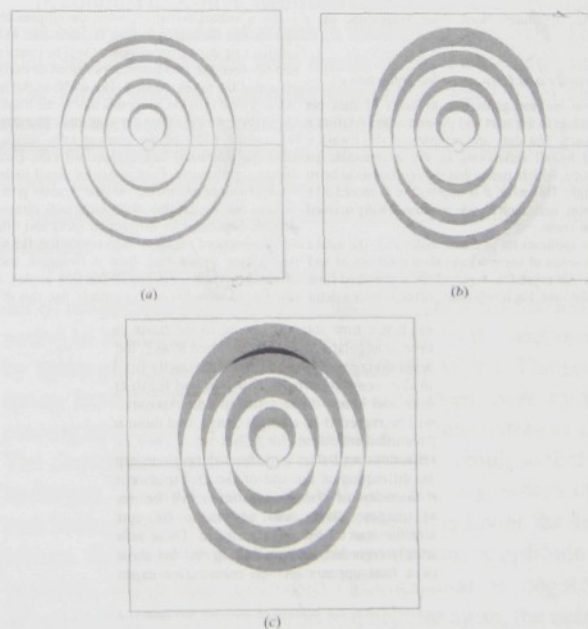
Obr. 5. Šířka „lunes" je úměrná Δφ (rotačnímu úhlu na jednu expozici). Velikost Δφ (rotační úhel na jednu expozici) roste od (a) do (c), ať u (c) dochází k překryvu.

4.4. Velikost rotačního intervalu

Zásadním parametrem rotační metody je velikost rotačního intervalu na jednotlivou expozici. V praxi existují dva základní přístupy: velké a malé rotační intervaly (v originále wide and fine slicing φ). Obě metody mají své výhody i limitace. Jemné vzorkování minimalizuje pozadí, umožňuje přesné fitování difrakčního profilu ve φ. Hrubé vzorkování je vhodné pokud doba čtení z detektoru je relativně dlouhá a dále většina reflexí je změřena jako úplně změřené. Na obr.5 můžete vidět vliv velikosti rotačního intervalu na šířku „lunes".

4.5. Vzdálenost krystal detektor

Dalším zásadním parametrem rotační metody je vzdálenost od krystalu k detektoru (d). Na všech difraktometrech pro měření makromolekulárních krystalů je tato vzdálenost měnina, většinou je minimální vzdálenost limitována možnou srážkou detektoru s goniometrem nebo chlazením a maximální vzdálenost je dána konstrukční volbou výrobce. Čím kratší je d, do tím většího rozlišení je možno měřit. Jelikož je však rozlišení téměř vždy limitováno difrakční kvalitou vzorku, je vhodné nastavit vzdálenost tak, aby byla vhodně využita celá plocha detektoru (jak praví klasici proteinové krystalografie: „Je důležité měřit intenzity, nikoli indexy".). S d souvisí i prostorové rozlišení jednotlivých difrakcí. Další faky je fokusace svazku záření, pokud je záření fokusováno na určitou vzdálenost, pak jsou difrakce nejlépe prokresleny v této vzdálenosti.

Literatura

International Tables for Crystallography, Volume F, 177 - 195.

NEUTRON CRYSTALLOGRAPHY OF BIOLOGICAL MOLECULES

Eva Buchtelová

Department of Biochemistry, Faculty of Science, Charles University, 128 43 Prague, Czech Republic

Keywords

neutron diffraction, protein, biological macromolecules, single-crystal diffraction, hydrogen, neutron detector, deuterium

Abstract

In biocrystallography, neutron diffraction is used mainly to determine the position of hydrogen atoms, the position and orientation of solvent water bound to the biomolecules and the position and low-resolution structure of other molecules rich in hydrogen, if this information cannot be obtained otherwise. The low flux of neutron beams limits this method to exceptionally large crystals or to non-crystalline material. This report focuses on single-crystal structure determination. Experimental setup at LADI instrument at ILL including a high-resolution position sensitive detector for Laue diffraction is described. Deuterisation of the sample as a way to increase the signal-to-noise ratio and the possibilities it creates for selective labeling are discussed.

Advantages of neutron diffraction - why to use neutrons

Neutron diffraction is used much less than X-ray diffraction for protein structure determination. This is because the flux of neutron beams is much lower than the flux of X-ray beams, and also because of low availability of neutron sources. Nevertheless some problems can be solved only by diffraction of neutrons [1]. The main advantage of neutrons is their relatively high elastic scattering by hydrogen. Hydrogen atoms can be seen in maps obtained with neutrons even if they are not visible with X-rays. Neutron diffraction is therefore used to determine the positions of important hydrogen atoms, such as atoms involved in enzymatic reactions or in important hydrogen bonds and to study hydratation of biomolecules. Neutron diffraction can tell us whether a particular acidic group is dissociated or has a hydrogen atom bound to it, it can discriminate between water and hydroxyl anion in the active site of an enzyme, it can determine the orientation of a water molecule. Even if only low resolution data can be obtained, neutron diffraction can give us information on the position of solvent or other hydrogen containing molecule. In such cases labelling of the molecule of interest by replacing its hydrogen atoms by deuterium is used.

In contrast to X-rays, neutrons are not scattered by electrons but by the nuclei of atoms in biological materials. The interactions of neutrons with nuclei depend on strong nuclear forces and the scattering amplitudes vary from element to element in a non systematic way, which means atoms of similar atomic mass can be distinguished which would seem the same with X-rays. Even different isotopes or nuclei with different spin have different scattering amplitudes.

The scattering amplitude of hydrogen is of the same order of magnitude as the amplitudes of other atoms predominating in biological molecules (Table 1.), so the scattering by hydrogen is not overridden by heavier atoms. The scattering from hydrogen can be increased even more by replacing hydrogen ^1H by deuterium (^2H, often written as D). The deuterium amplitude is almost twice as high as that of hydrogen ^1H. Moreover incoherent scattering, which creates background in the diffraction image, is lower for deuterium than for hydrogen. The deuterium amplitude is positive, while the hydrogen ^1H amplitude is negative. When hydrogen atom is bound to a heavier atom, the neighbouring positive and negative densities can partially cancel out. This does not happen with deuterium. The difference between hydrogen and deuterium scattering can be used for labelling only some atoms and/or computing difference in