

Proteinová krystalografie

Jaromír Marek,
Centrum strukturní biologie,
CEITEC MU

Historický úvod

- Monokrystalová strukturní analýza - studium 3-D struktur „molekul“ difrakčními technikami
- Potřeba vhodné „sondy“ – například rentgenového záření (1895, N.c. za fyziku 1901 – W. C. Röntgen)
- 1912 - průkaz vlnové povahy rentgenového záření jeho difrakcí na krystalu (N.c. za fyziku 1914 – M. von Laue)
- 1913-14 - První analýzy struktur krystalů (N.c. za fyziku 1915 – W.H. Bragg & W.L. Bragg)

„Přípravné“ metodiky pro proteinovou krystalografii

- 1926 - chemie - T. Svendberg - příprava koncentrovaných roztoků proteinů - ultracentrifugace, jednoznačná Mr
- 1946 - chemie - J.B. Sumner, J.H. Northrop, W. M. Stanley - purifikace enzymů a virů, enzymy jsou proteiny, enzymy a viry jde krystalizovat
- 1958 - chemie - F. Sanger - primární struktura proteinů
- 1980 - chemie – P. Berg, W. Gilbert, F. Sanger – metody na sekvenování DNA

Další strukturní krystalografové - nositelé Nobelovy ceny

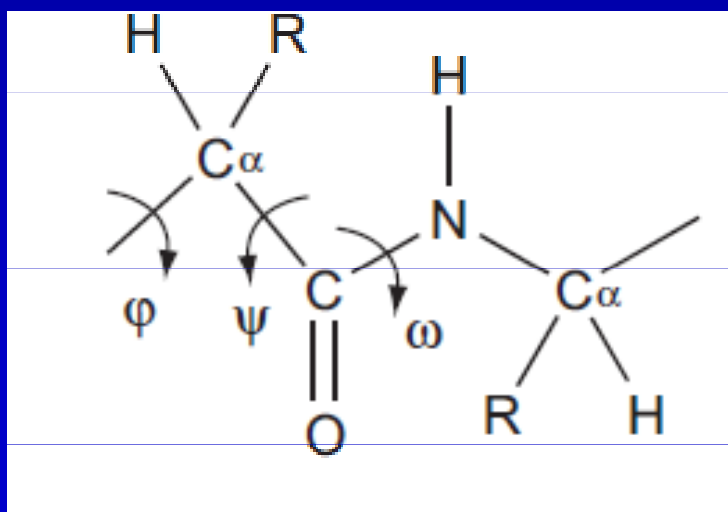
- 1954 - chemie - L. Pauling - „The nature of the chemical bond and the structure of molecules and crystals“
- 1962 - chemie - M.F.Perutz & J. C. Kendrew – první proteiny
- 1962 - medicína - F.H.C.Crick, J.D.Watson, M.H.F.Wilkins-DNA
- 1964 - chemie - D. Crowfoot-Hodgkin - biochemické molekuly
- 1982 - chemie – A. Klug – užití elektronové mikroskopie – struktury virů a komplexů protein-DNA
- 1988 - chemie - J.Deisenhofer, R.Huber & H. Michel - první membránový protein

Strukturní (bio)krystalografové - nositelé Nobelovy ceny po roce 2000

- 2003 - chemie (1/2) - R. MacKinnon - „structural and mechanistic studies of ion channels“
- 2006 - chemie - R. D. Kornberg - „molecular basis of eukaryotic transcription“ (= struktura RNA polymerázy)
- 2009 - chemie - V. Ramakrishnan, T.A. Steitz, A.E. Yonath - „structure and function of the ribosome“
- 2012 – chemie - Robert Lefkowitz, Brian Kobilka - „studies of G-protein–coupled receptors“

Předmět studia strukturní strukturní proteomiky/biologie

- Proteiny – biopolymery, polypeptidy
- Stavební jednotky – levotočivé aminokyseliny
- Primární struktura – posloupnost aminokyselin (NC 1953 – F. Sanger – struktura inzulinu)



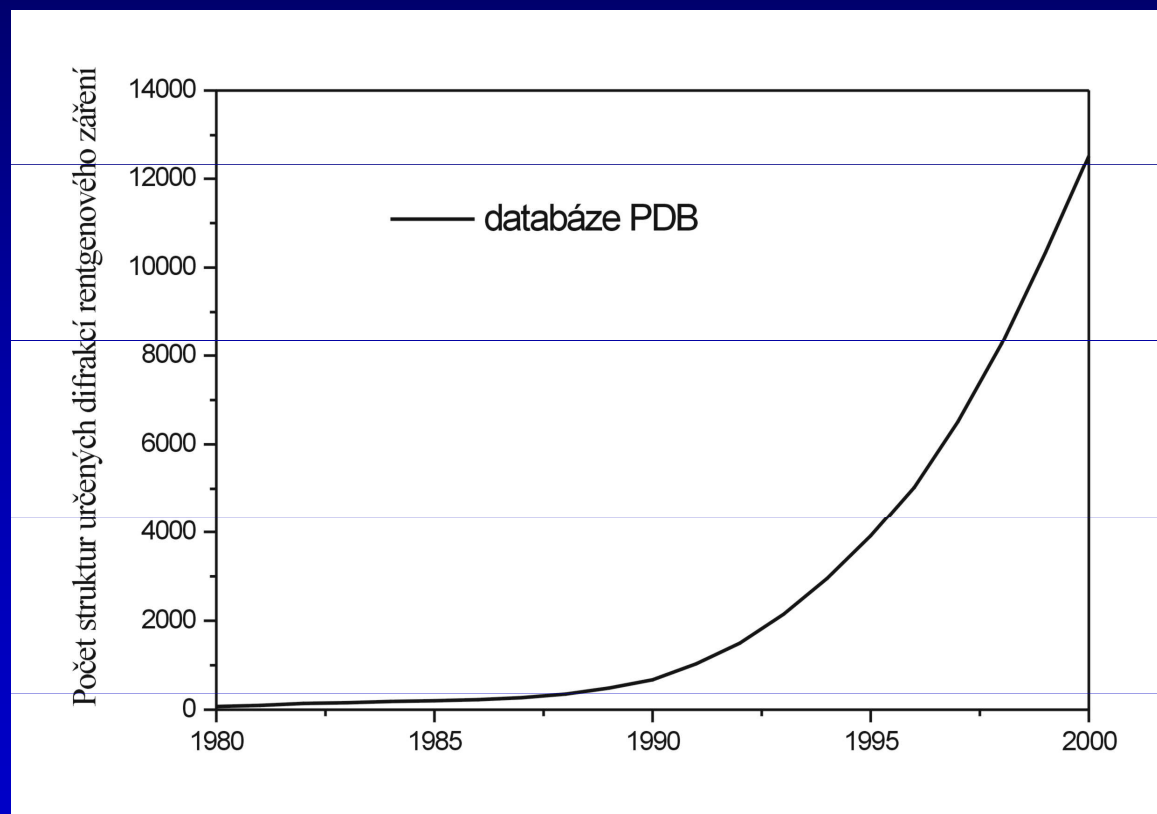
Nárůst užití difrakčních metodik pro určování 3-D struktur krystalů v čase

biologické makromolekuly/
proteiny => databáze PDB

Zlom - 90. léta:

- PCR [objev 1983, (1/2) N.C. za chemii -1993 - Kary B. Mullis]
- plošné detektory RTG záření
- synchrotronové záření
- PC

<https://www.rcsb.org/stats/growth/growth-released-structures>



Techniky studia 3D struktur biomolekul

PDB	<u>Z roztoků</u> NMR	<u>Krystaly</u> Difrakční techniky
2001	: 2 a ¼ tisíce	Přes 12 tisíc
2008	: 7 a ½ tisíce	44 a ½ tisíce
2021	: méně než 14 tisíc	cca 160 tisíc

Očekává se, že převážná většina 3D struktur **globulárních proteinů** (=proteinů s dobře určenou terciární strukturou) bude experimentálně určována difrakcí RTG (resp. synchrotronového) záření i v budoucnu.

Teoretické modely 3D struktur: AlphaFold@DeepMind/Google

Teoretické základy difrakčního studia 3-D struktur

- „sonda“ vhodné velikosti pro studium atomů – např. **rentgenové/el.mag. záření** o vlnové délce v oboru standardních meziatomových vzdáleností ($\sim 1 \text{ \AA}$)
- Foton s látkou interaguje rozptylem nebo absorbcí
- Rozptyl fotonů - se ztrátami – Comptonovský
- **beze ztrát energie - Thompsonův**
- Rozptyl **jiných typů záření** (dualismus vlna/částice) nastává jinými mechanismy
- Elektrony jsou rozptylovány nabitými částicemi
- Rozptyl neutronů - na nukleonech, resp. jádrech atomů

Teoretické základy difrakčního studia 3-D struktur: Thompsonův rozptyl rentgenového záření

Každá nabitá částice se v poli rovinného monochromatického záření stane sekundárním zdrojem elektromagnetického pole

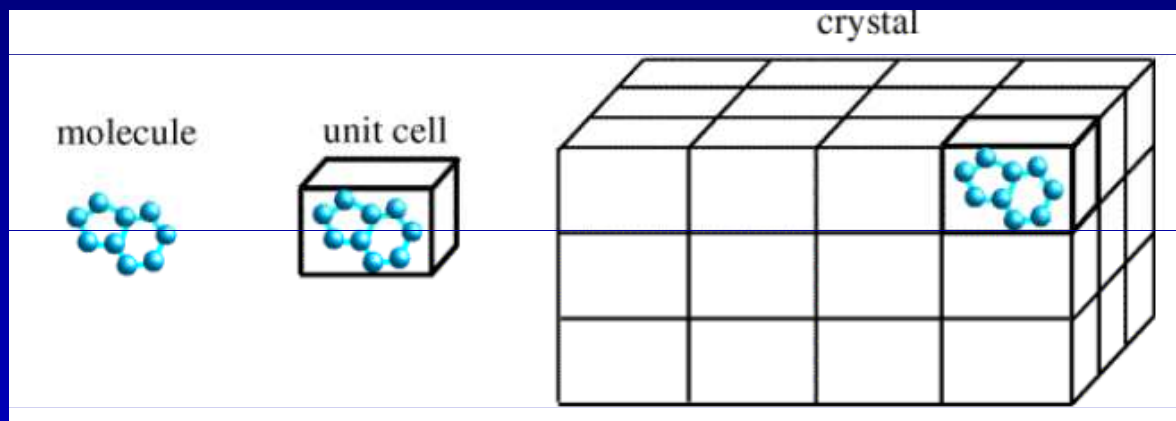
$$I_Q = I_{Oi} \frac{e^4}{m^2 r^2 c^4} \sin^2 \varphi$$

=> rozptyl fotonů na protonech je nevýznamný (1837^2 x slabší než na elektronech), difrakcí RTG záření proto studujeme zejména elektronovou strukturu látky

Základy difrakčního studia 3-D struktur: Rozptyl záření na krystalech

Rozptyl záření na látce: víc (než jedno) rozptylové centrum \Rightarrow prostorem se současně šíří několik (mnoho) sekundárních kulových vln

Současné šíření kulových vln \Rightarrow složené vlnění, interference



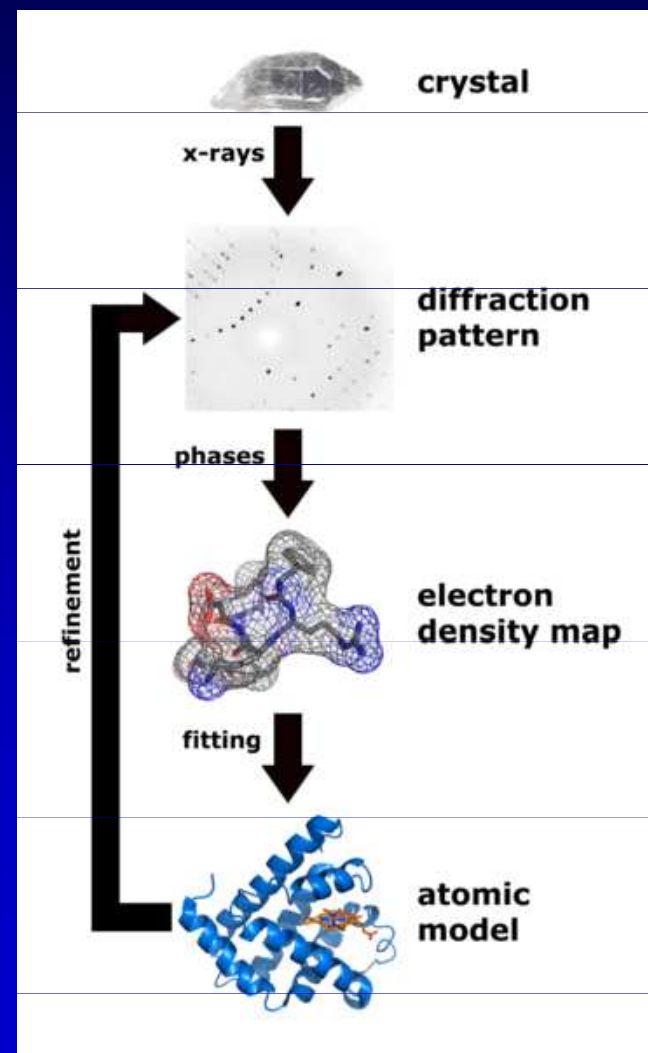
Roztok/amorfní látka: náhodně orientované molekuly \Rightarrow vlny z nesouvisejících počátků a směrů \Rightarrow slabý rozptylový signál

Krystal/periodické prostředí \Rightarrow koherentní záření \Rightarrow možnost zesílení signálu pozitivní interferencí

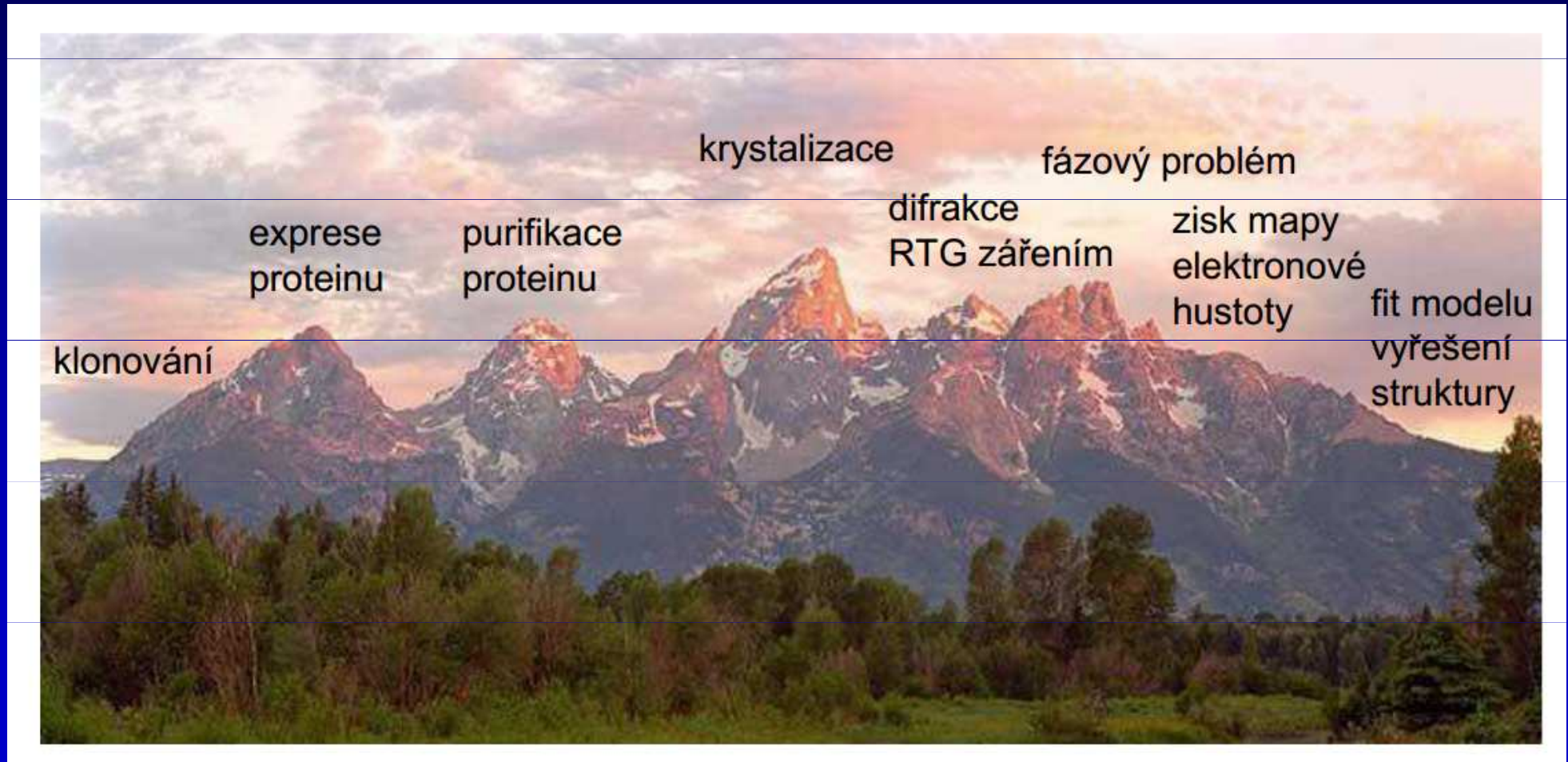
Difrakční podmínky – podmínky pro záření s vyšší (=měřitelnou) amplitudou

Etapy určování struktur enzymů

1. Výběr genu
2. Příprava rekombinantního proteinu, čištění, zahušťování...
3. Krystalizace
4. Difrakční experiment
5. Fázový problém + příprava modelu
6. Zpřesňování modelu
7. Publikace/užití výsledku



Proteinová krystalografie jako pohoří



Volba genu, krystalografické studium genetické informace

- Jakou genetickou informaci strukturně studovat?
- Není dosud přesně známo, kolik genů v lidské DNA je (lidský genom: v průběhu času 25000 \Rightarrow 28000 \Rightarrow \pm 23500 genů/2011)
- Zatím jen u cca 1/10 lidských enzymů je známa jejich funkce
- Strategický cíl – určit struktury všech lidských proteinů (*Human proteome project*)
- Prohledávání velkých databází a hledání jednotlivých genů
- Zpracování - není možné bez počítač. SW + HW
- Problematické místo - formulace hypotéz o polohách genů a funkcích jejich produktů

Příprava biologického materiálu

- Naprostá většina v současnosti studovaných proteinů se připravuje biotechnologickými metodami (rekobinantní DNA, nadprodukce v modelovém systému, ...)
- Výhody – „snadnost“ provádění genetických modifikací
 - jde připravovat Se-Met proteiny (fázový problém)
- Kritická místa : - (ne)funkčnost rekombinantních proteinů
 - produkováný protein musí být rozpustný
 - čištění, (ne)stabilita, tendence k agregacím,...

Krystalizace

- Až doposud nejkritičtější a časově nejnáročnější část určování 3-D struktur makromolekul difrakcí
- Urychlení: využití již získaných zkušeností o krystalování (jiných) proteinů + automatizace/robotizace opakovaně prováděných experimentů
- Empirie + statistické zpracovávání => **krystalizační screens** = množiny roztoků (s různým chemickým složením, pH, aditivy, ...) pokrývající v minulosti úspěšně použité krystalizační podmínky
- Robotika na automatickou krystalizaci: automatické míchání/ředění + pipetování roztoků, pipetování nanolitrových objemů, strojem řízené mikroskopování + počítačové vyhodnocování obrazu)
- Průchodnost: potenciálně až desítky proteinů (desítky tisíc testovacích krystalizací) týdně

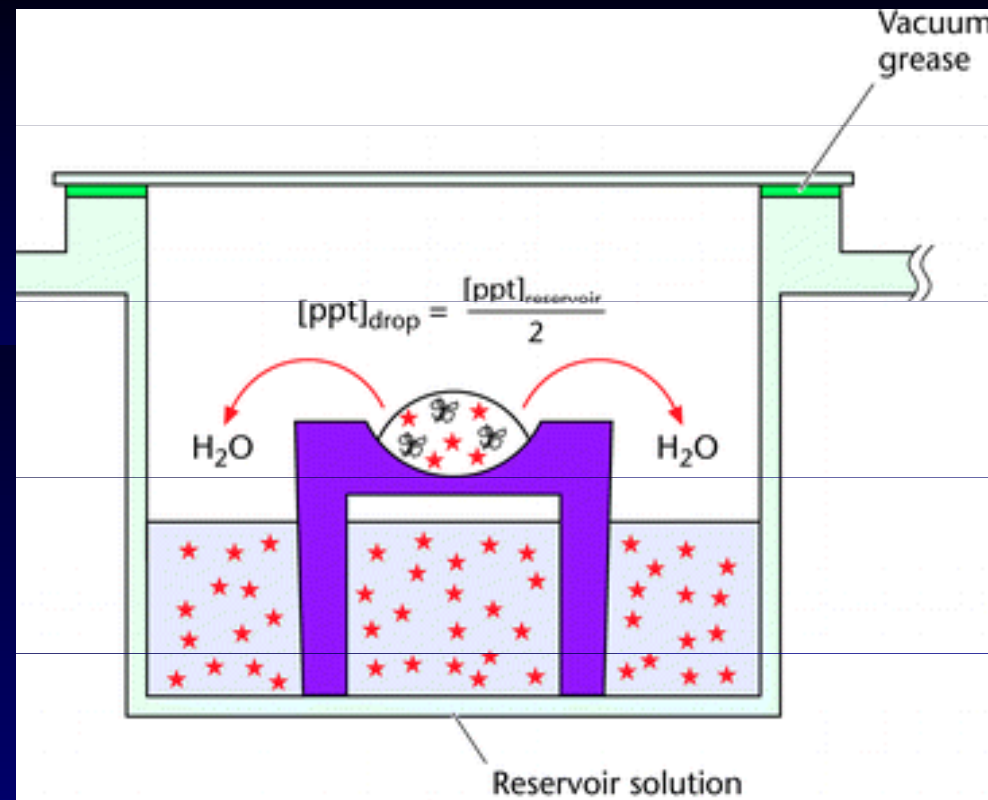
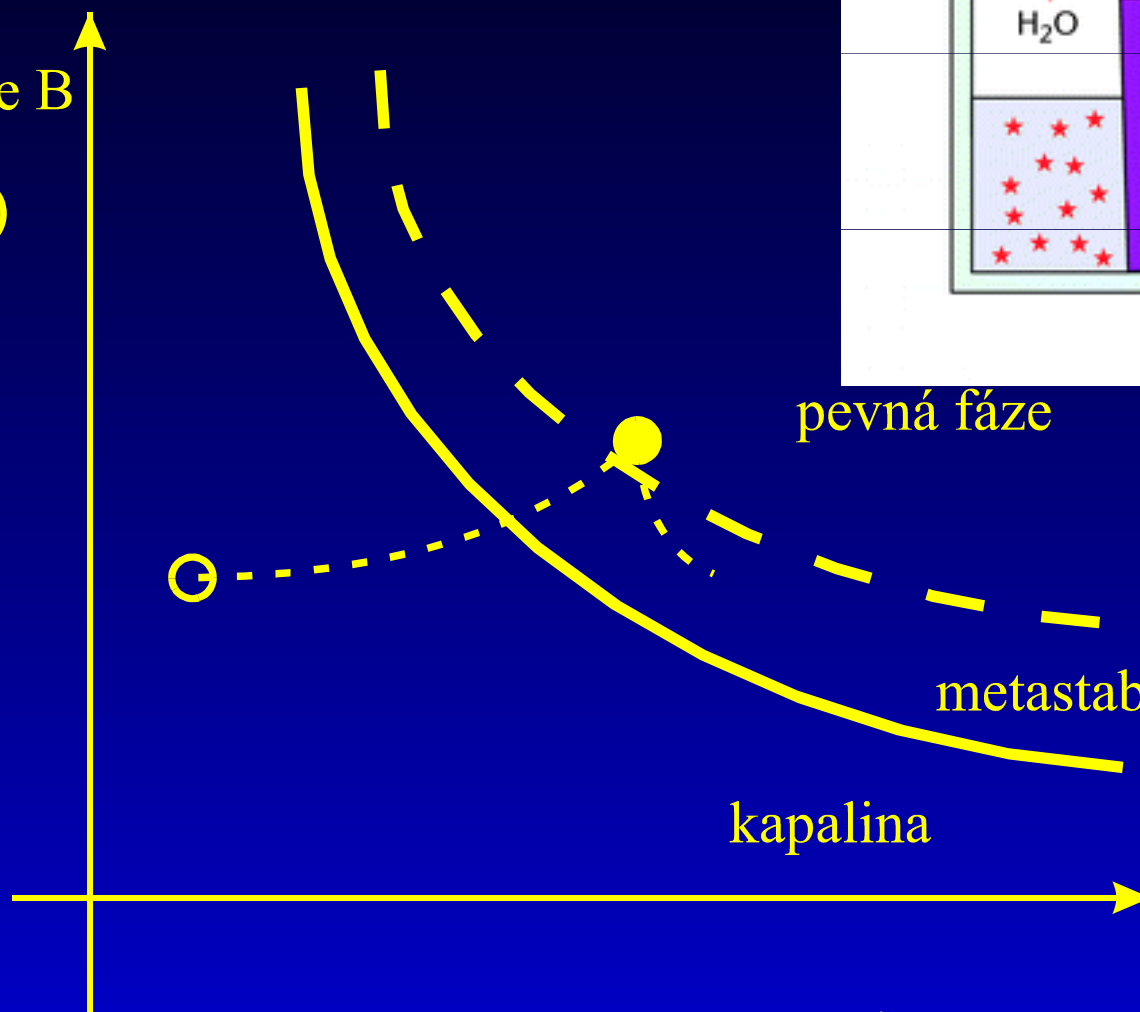
http://en.wikipedia.org/wiki/X-ray_crystallography#Crystallization

Difúzní krystalizační techniky

- Růst proteinových krystalů je relativně pomalý => pomalé změny koncentrací
- Možný „pomalý“ řídicí děj: difúzní vyrovnávání koncentrací (rozpouštědla = vody)
- Metoda sedící/visící kapky (sitting/hanging drop)
- Krystalizace ze směsí látek: **voda/rozpouštědlo**, **srážedlo/precipitant**, pufry (stabilizace pH), **krystalovaná makromolekula**, ionty/soli, malé molekuly a aditiva, ...
- Kapka + zásobník/rezervoár: různě velké objemy, rozdílné složení
- Velikost objemu x doba potřebná k difúznímu ustanovení rovnováhy

Metoda sedící kapky

koncentrace B
(protein)



koncentrace A (srážedlo)

Automatizace pro pěstování proteinových krystalů

- Pipetování mililitrových objemů (pro automatizovanou přípravu krystalizačních „screens“/roztoků)
- Pipetování mikrolitrových objemů (např. příprava roztoků do zásobníků/reservoirs pro difúzní techniky)
- Pipetování nanolitrových objemů (např. příprava kapek pro metody sedící kapky/hanging drop)
- Skladování + manipulace s krystalizačními deskami + automatizovaný mikroskopický monitoring změn v krystalizačních kapkách

<http://www.rigakuautomation.com/desktop>

Automatizace pro pěstování proteinových krystalů



Difrakční experiment pro určení 3-D struktury: Generování rentgenového záření

Konvenční laboratorní zdroje RTG záření:

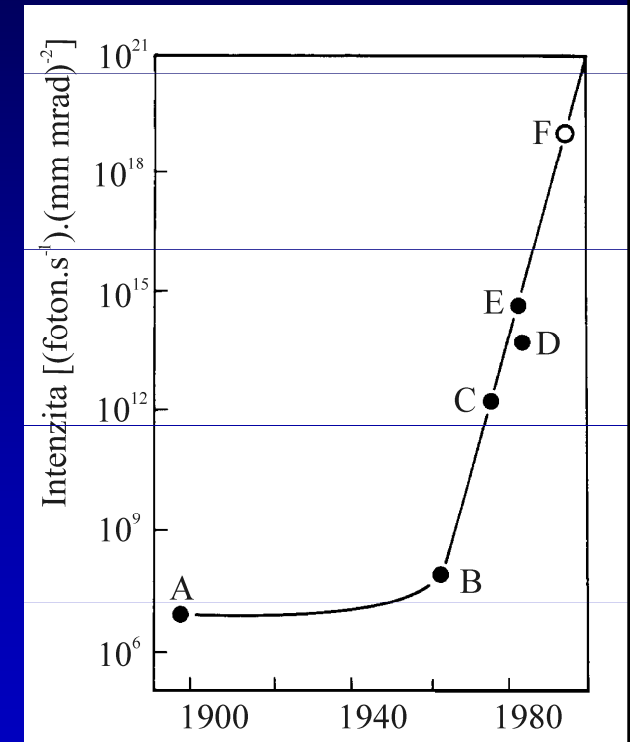
- brzdné záření (nízká účinnost při generování RTG)
- rentgenová lampa (A)
- rotační anoda (B) – X-ray CF – svítivost $2 \cdot 10^9$

Limitace konvenčních zdrojů:

- fixní vlnová délka (charakteristické RTG záření)
- „bodový“ zdroj kulových vln

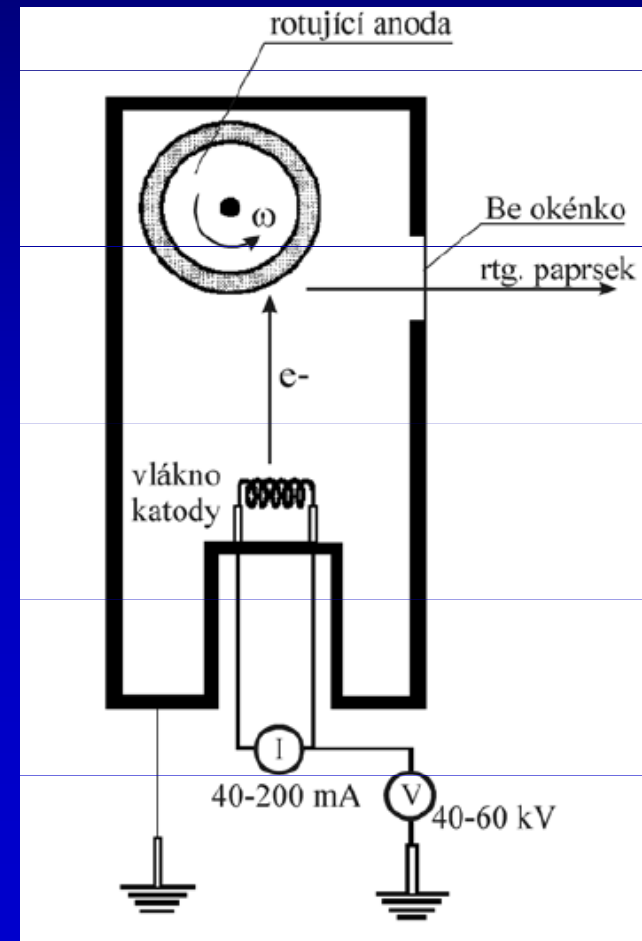
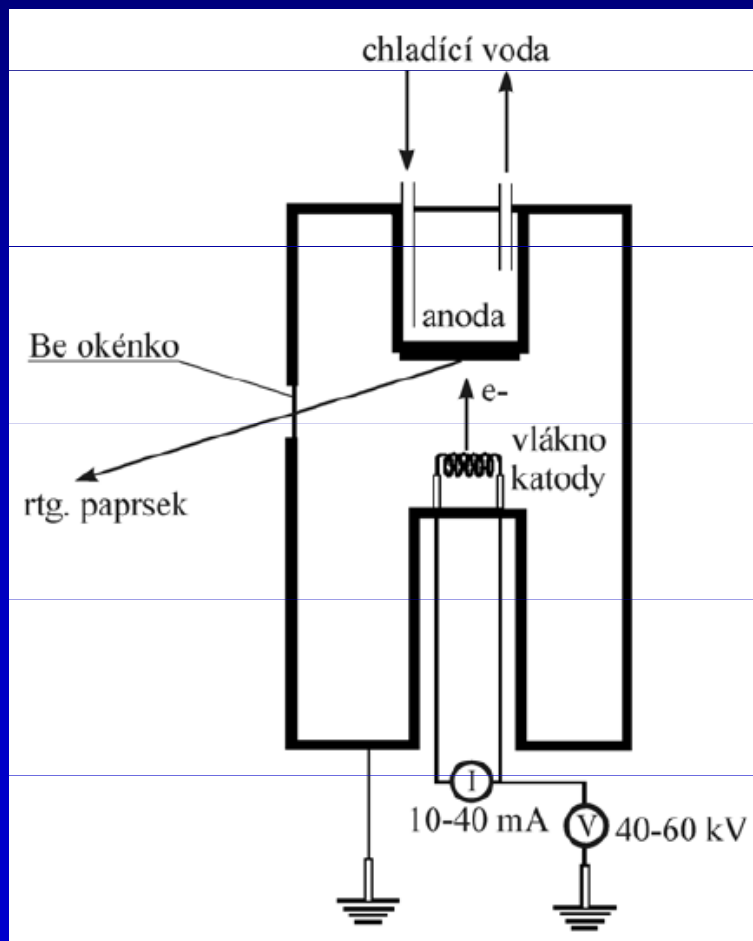
Synchrotrony (C až F)

- zdroje vysoce intenzivního záření (krátký experiment)
- spojité spektrum (fázový problém: Se proteiny + MAD)



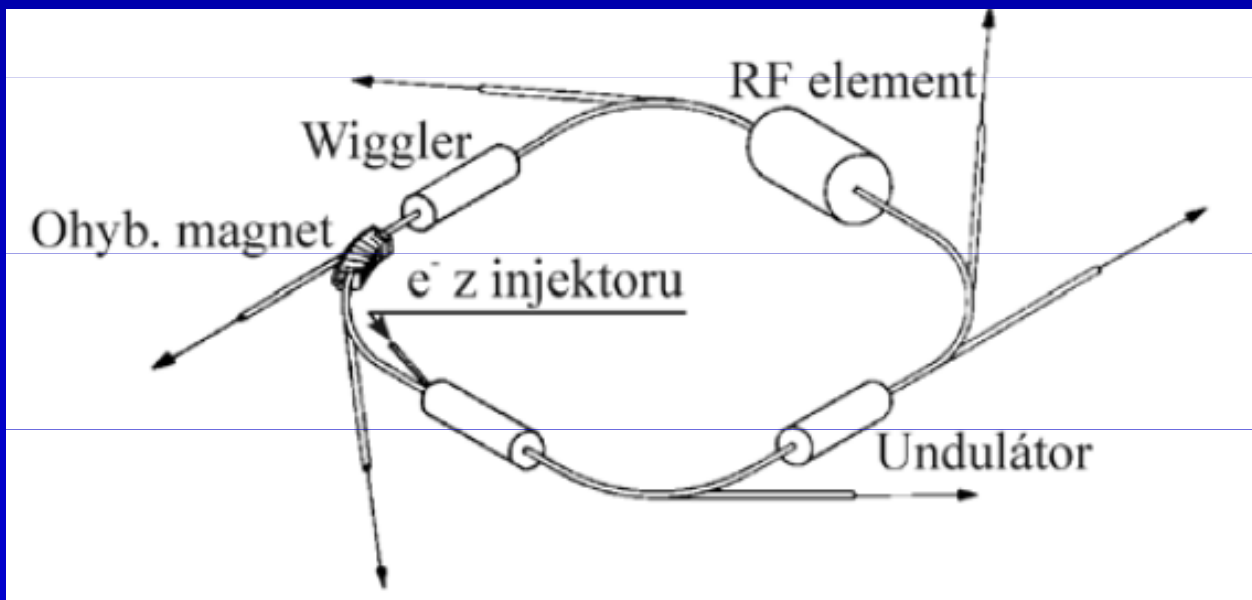
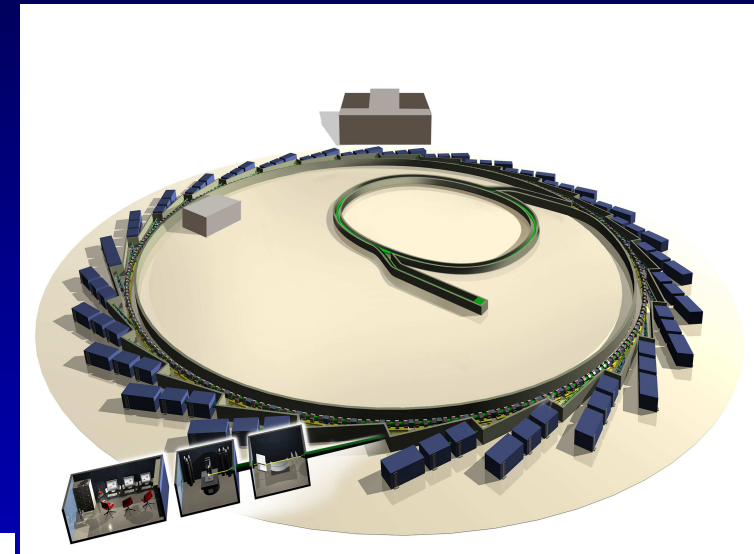
Difrakční experiment pro určení 3-D struktury: Generování konvenčního RTG záření

Brzdné RTG záření => vlastní RTG – jen jednotky % dodané energie
– hlavní faktor limitující výkon – je nutné odvádět teplo



Difrakční experiment : synchrotron jako „nekonvenční“ zdroj RTG záření

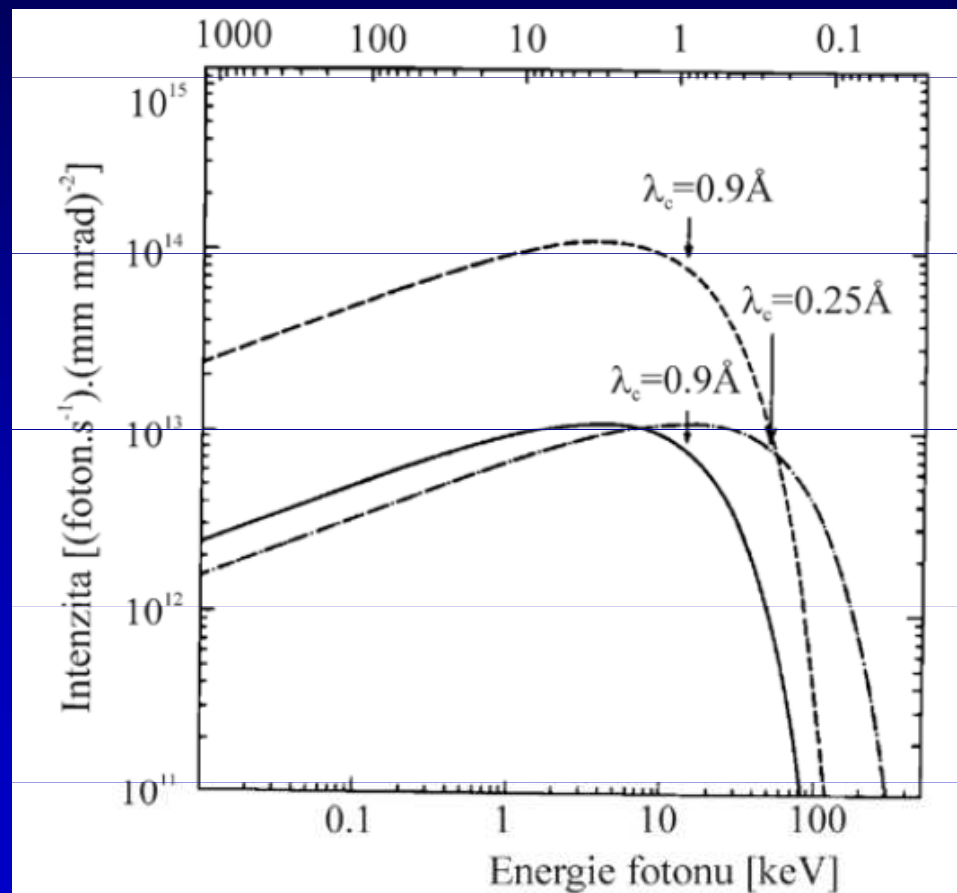
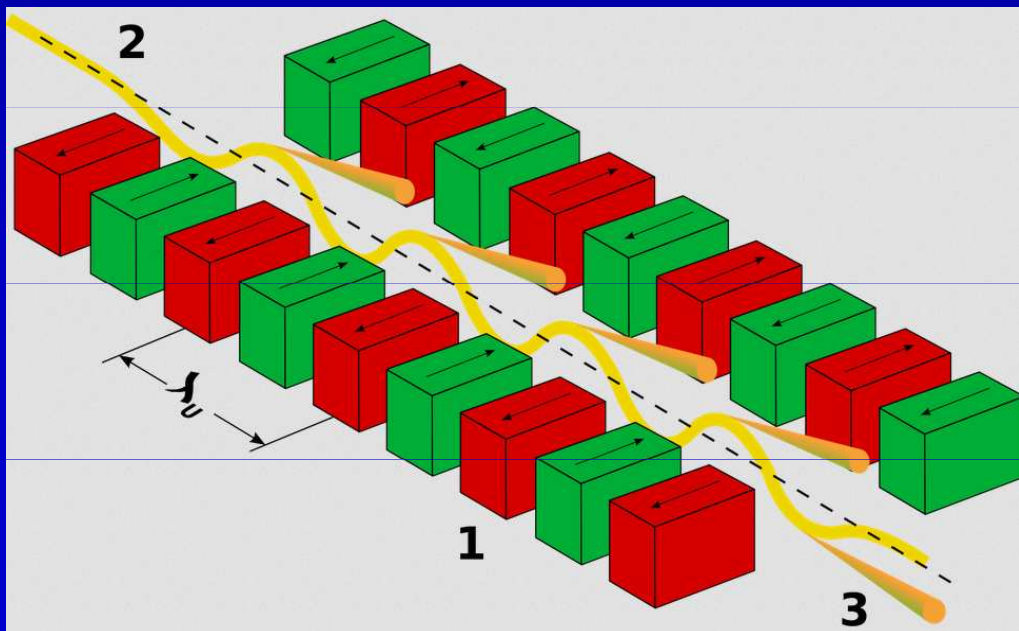
- záření od nabitých částic na zakřivené dráze
- relativistické rychlosti, pohyb ve vakuu, ...
- ohyb dráhy pomocí dipólových magnetů
- **jeden ohybový (bending) magnet**
- **sady magnetů: undulátory a wigglery**



Difrakční experiment : ohybový magnet, undulátor, wiggler

wiggler – ohybové magnety jsou od sebe „daleko“, intenzity záření vystupující z jednotlivých ohybových magnetů se sčítají, $I \sim N$ (N je počet magnetů)

undulátor – ohybové magnety jsou k sobě „blízko“, záření z nich proto interferuje (=sčítají se jejich amplitudy), $I \sim N^2$



Obrázek 22 Spektrální rozložení a kritické vlnové délky pro ohybový magnet (plná čára), wiggler - posouvač vlnových délek (čerchovaná čára) a undulátor (čárkovaná čára) na 6GeV synchrotronu v ESRF (the

Difrakční experiment pro určení 3-D struktury: Plošný detektor s polovodičovými CCD prvky

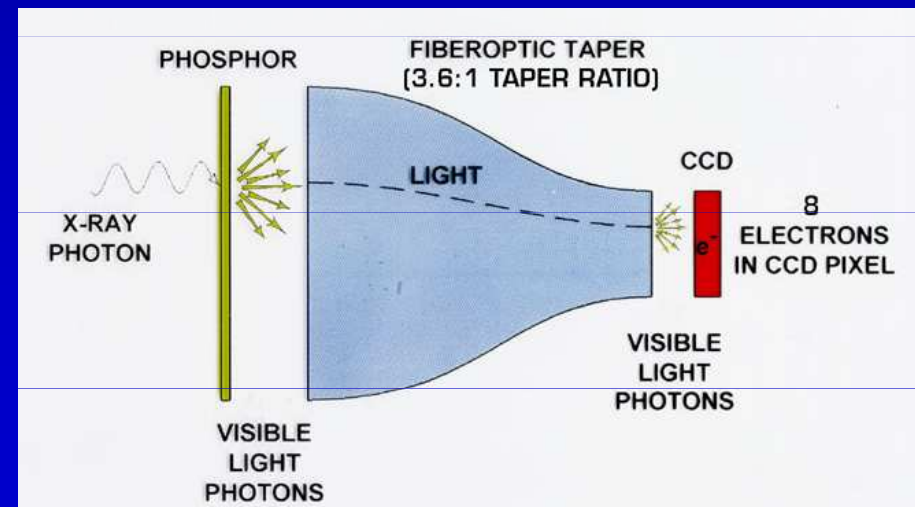
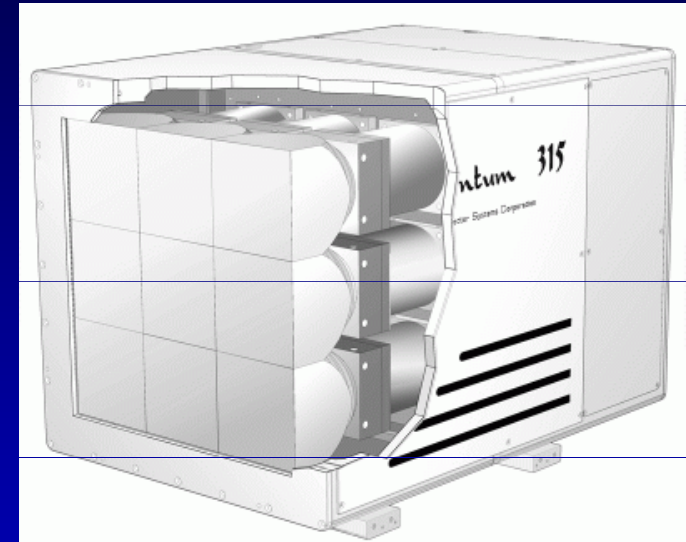
Detekce RTG: scintilační krystal (= „phosphor“, např. z NaI legovaného Tl) převádí RTG záření do pásma viditelného světla

Viditelné světlo se detekuje podobnými mikroelektron. prvky, jaké obsahuje např. digitální fotoaparát

Klady CCD: rychlost

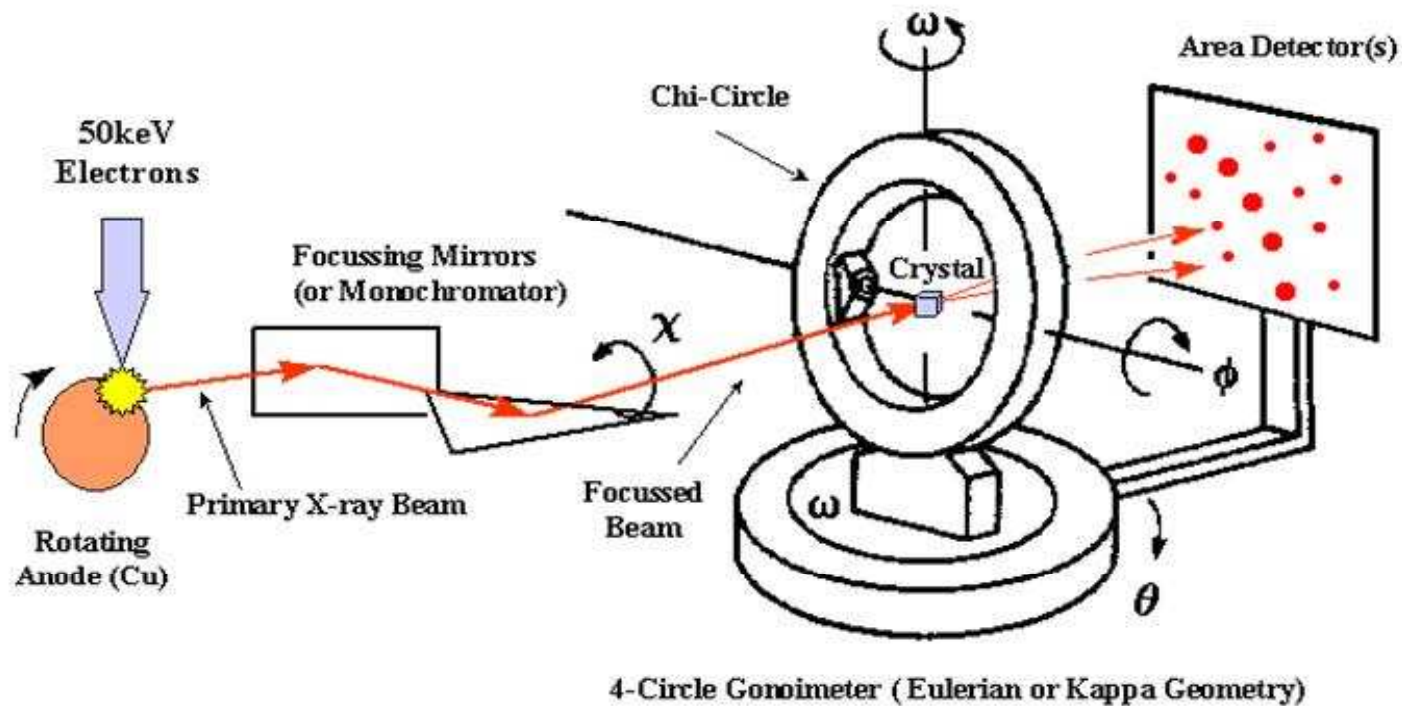
+ -: velikost CCD prvku – osvětluje se zužovacím segmentem ze světlovodivých vláken

Největší zápor: vlastní teplotní šum



Difrakční experiment: automatický difraktometr-schéma

Principle of an X rays diffraction experiment



Experiment: omezené množství dat

- Geometrická limitace počtu měřitelných dat – měříme data jen do určitého (max.) difrakčního úhlu (limitace: prostorové, časové, ...)
- Limit pro max. délku experimentu: radiační poškození

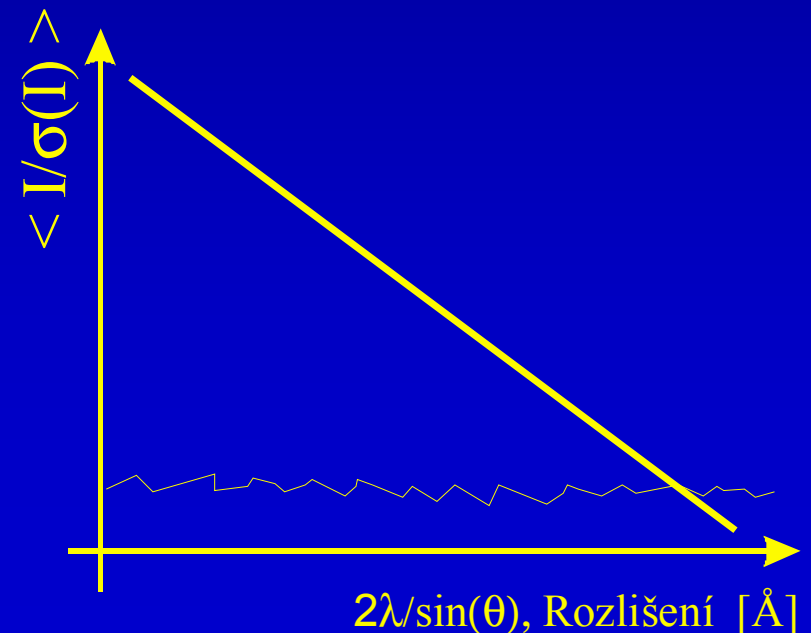
- Braggova rovnice $2d_H \sin \theta = \lambda$

- Difrakční experiment a jeho limitované rozlišení (resolution):

$$\theta_{\max} \Rightarrow d_{\min}$$

- Nedokonalost („mosaicita“) krystalů

- Kryokrystalografie



Difrakční experiment

- Naprostá většina proteinových struktur je v současnosti určována ze synchrotron. difrakčních dat (hlavní klady: intenzita záření, rychlost experimentu, možnost optimalizace vlnové délky) a za kryopodmínek (dlouhodobá stabilita krystalu před experimentem + zvýšená odolnost vůči radiačnímu poškození při měření, lepší difrakční schopnosti, ...).
- „Domácí“ laboratoře: testování difrakční kvality krystalů, „ladění“ kryoexperimentů, předzmrázování krystalů.
- Sběr úplných difrakčních synchrotron. dat – řádově desítky minut, často i výrazně méně
- Automatizovaná výměna a měření zmražených (kryo)vzorků
- Obrovský „boom“ Se-proteinů & MAD/SAD experimentů

Teoretické základy difrakčního studia 3-D struktur: Strukturní faktor, elektronová hustota a fázový problém

Strukturní faktor – popisuje amplitudu difraktované vlny

$$\begin{aligned} F_M(\vec{r}^*) &= \int_V \sum_{j=1}^N \rho_j(\vec{r} - \vec{r}_j) \exp(2\pi i \vec{r}^* \cdot \vec{r}) d\vec{r} = \\ &= \sum_{j=1}^N \int_V \rho_j(\vec{R}_j) \exp\left[2\pi i \vec{r}^* \cdot (\vec{r}_j + \vec{R}_j)\right] d\vec{R}_j = \sum_{j=1}^N f_j(\vec{r}^*) \exp(2\pi i \vec{r}^* \cdot \vec{r}_j) \end{aligned}$$

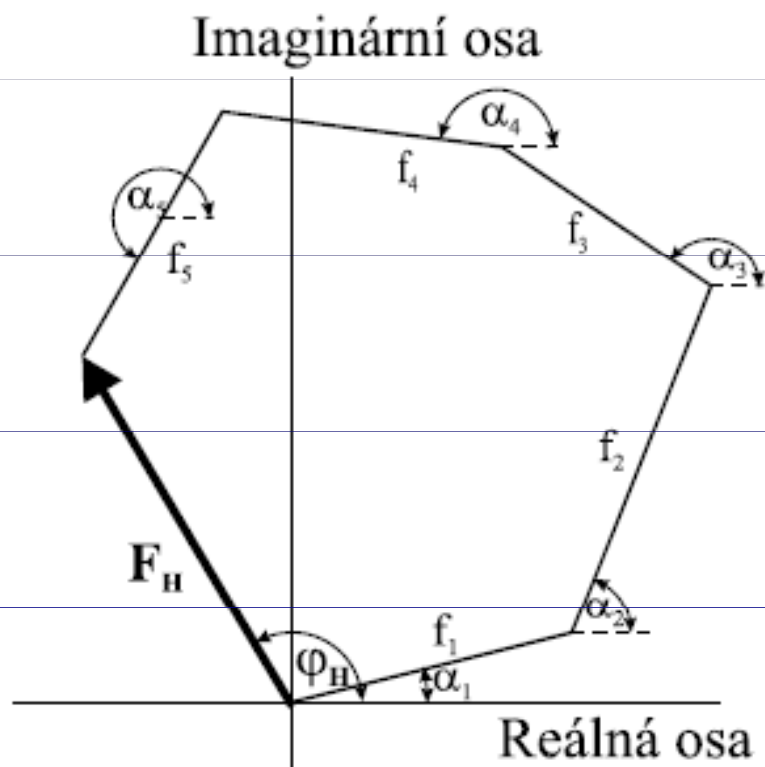
Uvedená suma/integrál = Fourierova transformace (FT). Analogicky k tomu je krystalová elektronová hustota obráceným Fourierovým obrazem strukturních amplitud

$$\rho(\vec{r}) = \int_{V^*} F(\vec{r}^*) \exp(-2\pi i \vec{r}^* \cdot \vec{r}) d\vec{r}^* = \frac{1}{V} \sum_{h,k,l=-\infty}^{+\infty} F_{hkl} \exp[-2\pi i(hx + ky + lz)]$$

Fázový problém – neměříme strukturní amplitudy, ale intenzity difrakcí, neznámé amplitudy (resp. fáze amplitud) se musí určit/dopočítat

Fázový problém

- hlavní cíl – zjistit 3-D model studované (makro)molekuly
- mezicíl – určit při difrakčním experimentu ztracenou informaci o fázích strukturních amplitud a poté pomocí FT spočítat mapy elektronových hustot
- nejjednodušší metoda – fázový problém vůbec neřešit, využít podobnost studovaného systému se systémem s již známou 3-D strukturou=modelem (metoda MR, Molecular Replacement, molekulové nahrazení) a fáze počítat na základě tohoto modelu
- stačí jedno měření
- je nutná relativně vysoká 3-D podobnost mezi modelem a studovaným systémem (indikátor podobnosti – 1-D podobnost: identita AA cca 30% a lépe, podobnost AA 50% a víc)
- 2021: AlphaFold, <https://alphafold.ebi.ac.uk/>

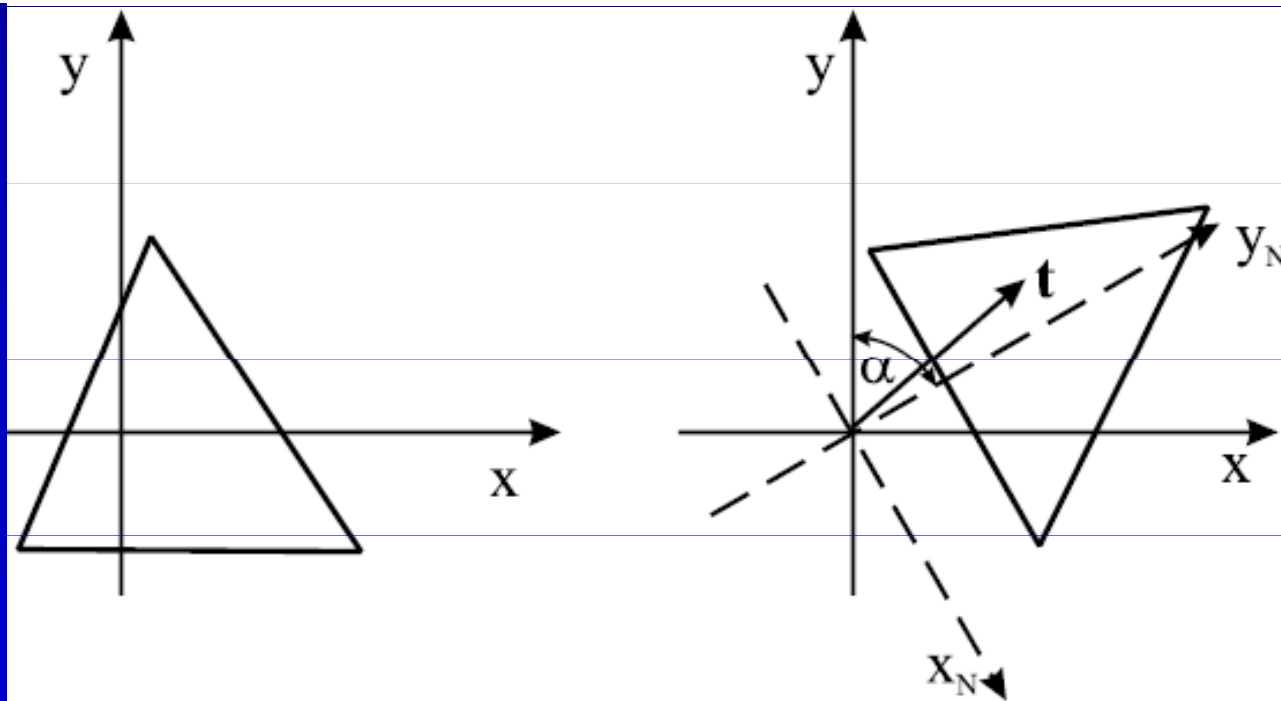


F_H – strukturní amplituda = komplexní vektor

φ_H – fáze = úhel vektoru F_H vůči reálné ose

Schéma MR

Nehledáme jednotlivé fázové úhly φ_H , ale pozici modelu v krystalu = 3 rotace + 3 posuvy modelu



Fázový problém a deriváty proteinů

V proteinových krystalech existují makroskopicky dlouhé „kanály“ rozpouštědla (krystalograficky neuspořádané vody) umožňující difúzi cizích „malých“ molekul do celého objemu krystalu

Terciální struktury globulárních proteinů mohou být až natolik stabilní, že je chemická interakce proteinů s „malými“ molekulami „příliš“ neovlivní

Dobrá „derivující“ molekula – silný vliv na strukturní amplitudu („těžké“ atomy, kovy), malý vliv na 3-D strukturu proteinu (=isomorfní deriváty)

Hledání dobrých protein. derivátů je experimentálně pracné.

Je nutné znát odpovědi mj. na otázky:

Jak koncentrovaný roztok použít?

Jaké molekuly s jakým centr. atomem?

Jak dlouho to má na krystal působit?

...

Těžký atom	Příklady sloučenin
Pt	K_2PtCl_4 , $K_2Pt(CN)_4$, $K_2Pt(NO_2)_4$, $Pt(NH_3)Cl_2$, $Pt(NH_3)Br_2$, $PtCl_4$, $PtCl_2$
Hg	$HgCl_2$, $HgAc_2$
U	$UO_2(NO_3)_2$
Lanthanoidy	$EuCl_3$, $SmAc_3$

Tab. 14

Příklady sloučenin nejčastěji používaných k přípravě kovových derivátů proteinů. Ac = acetyl.

Fázový problém: metody používající deriváty proteinů

Jeden (single) derivát:

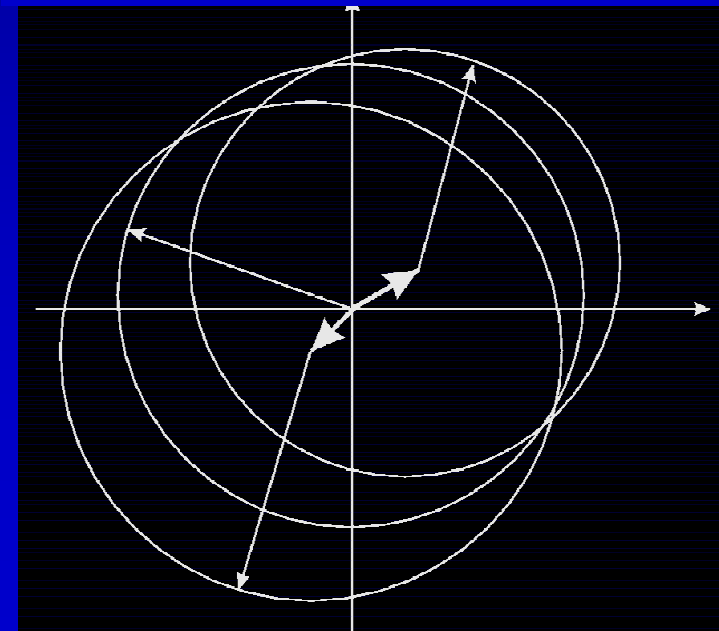
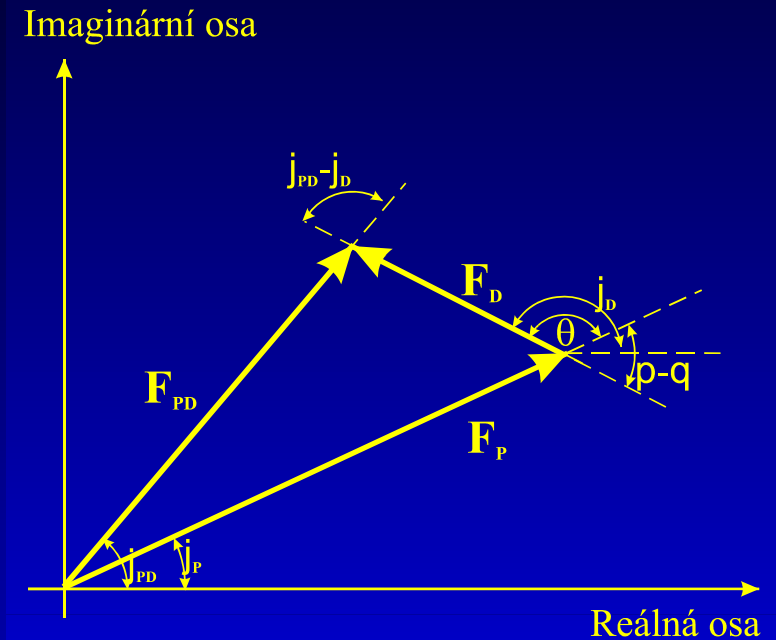
SIR, Single Isomorphous Replacement

= jednoduché isomorfní nahrazení

Vektorové znázornění:

Známe velikost strukturní amplitudy pro nativní a derivovaný protein F_P a F_{PD} , můžeme určit/spočítat velikost i směr F_D (příspěvku těžkého kovu), hledáme neznámé úhly φ_P

Jednoznačné řešení - několik (Multiple) derivátů a několik měření: MIR, Multiple Isomorphous Replacement = vícečetné isomorfní nahrazení



MAD/SAD, Multi/Single wavelength Anomalous Dispersion

MAD/SAD - experimentálně relativně jednoduché řešení fázového problému
Opakované měření na jediném krystalu využívající spektrální závislosti
(=disperze) rozptylových vlastností prvků.

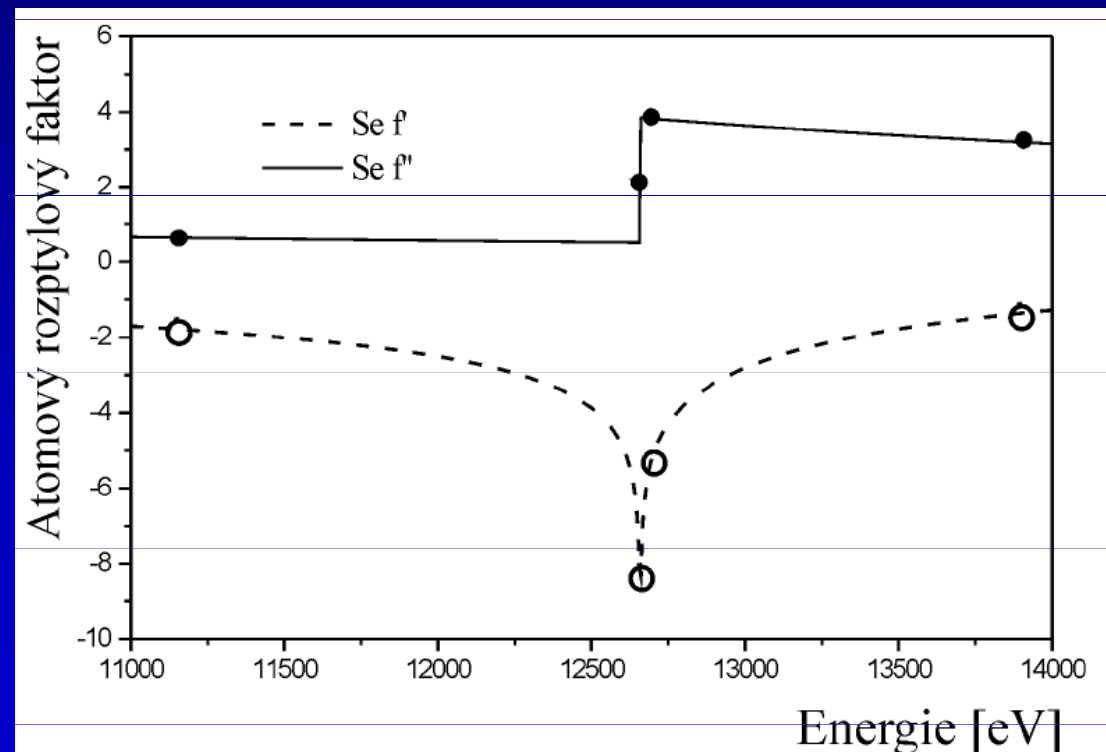
Změna vlnové délky = synchrotron

Vhodný prvek:

Se (SeMet),

S (S-S můstky),

kovy v metaloproteinech, ...



Obrázek 88 Graf teoretické závislosti f'' a f' selenu na vlnové

Zpřesňování proteinů: omezené množství dat

Geometrická limitace počtu měřitelných dat – měříme jen do určitého (max.) difrakčního úhlu

Rozlišení [Å]	Počet nezáv. reflexí	Poměr počtů	
		reflexí a {x,y,z}	proměnných {x,y,z,B}
40.0-3.0	3500	0.8	-
40.0-2.5	6800	1.6	1.2
40.0-1.9	13500	3.1	2.3
40.0-1.5	29800	6.8	5.1
40.0-1.2	58800	13.3	10.0
40.0-1.0	81300	18.5	13.8

- Protein s 182 AA, 40% solventu a 1468 atomy
- +/- 4500 souřadnic, 6000 proměnných včetně faktorů B
- Parametr B - teplotní kmity (=započtení neurčitosti/neuspořádanosti)

Zpřesňování proteinových struktur

- Experimentální proměnné – difrakce
- Modelová funkce – strukturní amplituda
- Startovní strukturní model (z MR) a/nebo mapa el. hustoty (získaná po vyřešení fázového problému)
- Kritérium správnosti – R faktory

$$R = \frac{\sum_H \left| |F_H^o| - K |F_H^c| \right|}{\sum_H |F_H^o|}$$

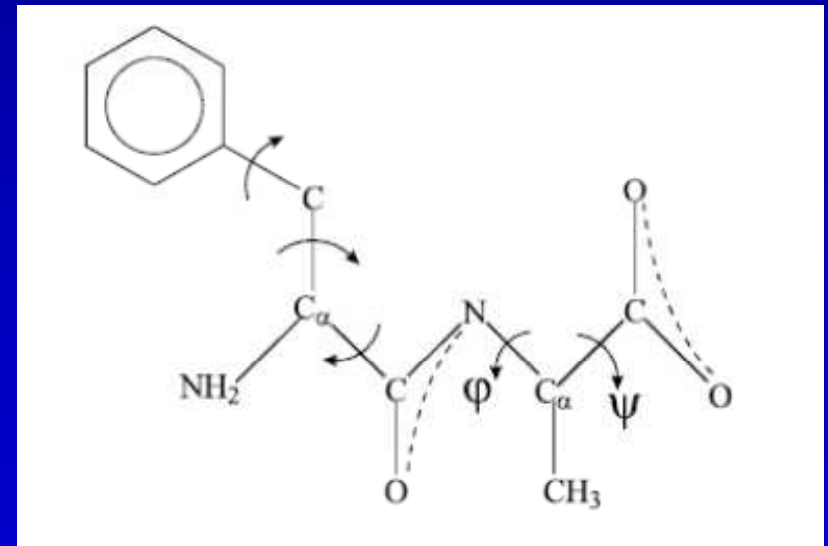
Možné problémy:

- Limitovaný počet pozorování daný rozlišením experimentu
- Nelineárnost problému – iterativnost, konvergence
- Lokální vs. globální minima
- „Přefitování“ (možnost biologicky nesmyslného „řešení“ obsahujícího (jen) molekuly H₂O)

Limit. počet dat: snížení počtu proměnných

- „constrained“ minimalizace: $\vec{X} = C \cdot \vec{X}' + c$
- „Tvrdé“ vazební podmínky
- Triviální aplikace – operace symetrie pro S.G. vyšší než P1
- „rigid body refinement“ + volné proměnné popisující AA

- příklad: 17 atomový fragment fenylalanin-alanin
- 51 vs 11 parametrů
- aplikace – zpřesňování el. hustoty



Limit. počet dat: zvýšení počtu „pozorování“

- „měkké“ vazební podmínky
- „restrained“ minimalizace – využití nekryystalografických dat
- popis pomocí „tolerancí“

$$\text{dist}(Atom_1, Atom_2) = D \pm \sigma(D)$$

- chemická „energie“ popisující vzdálenosti, úhly, planarity, ...

$$S_D = W_D \sum_j w_j \left(d_{j,IDEAL} - d_{j,MODEL} \right)^2$$

- zdroj dodatečných. údajů – strukturní databáze, spektrální data, QM výpočty
- 2 stupňové zpřesňování – generování chem. informace + minimalizace rozdílů model/experiment
- váhové koeficienty, současné zpřesňování RTG dat + molekulární mechanika/molekulární dynamika

Zpřesňování proteinových struktur: minimalizace

- Krystalografie malých molekul – metoda nejmenších čtverců

$$S^1 = \sum_H w_H \left(|F_H|_o - |F_H|_c \right)^2 \qquad S^2 = \sum_H w_H \left(|F_H|_o^2 - |F_H|_c^2 \right)^2$$

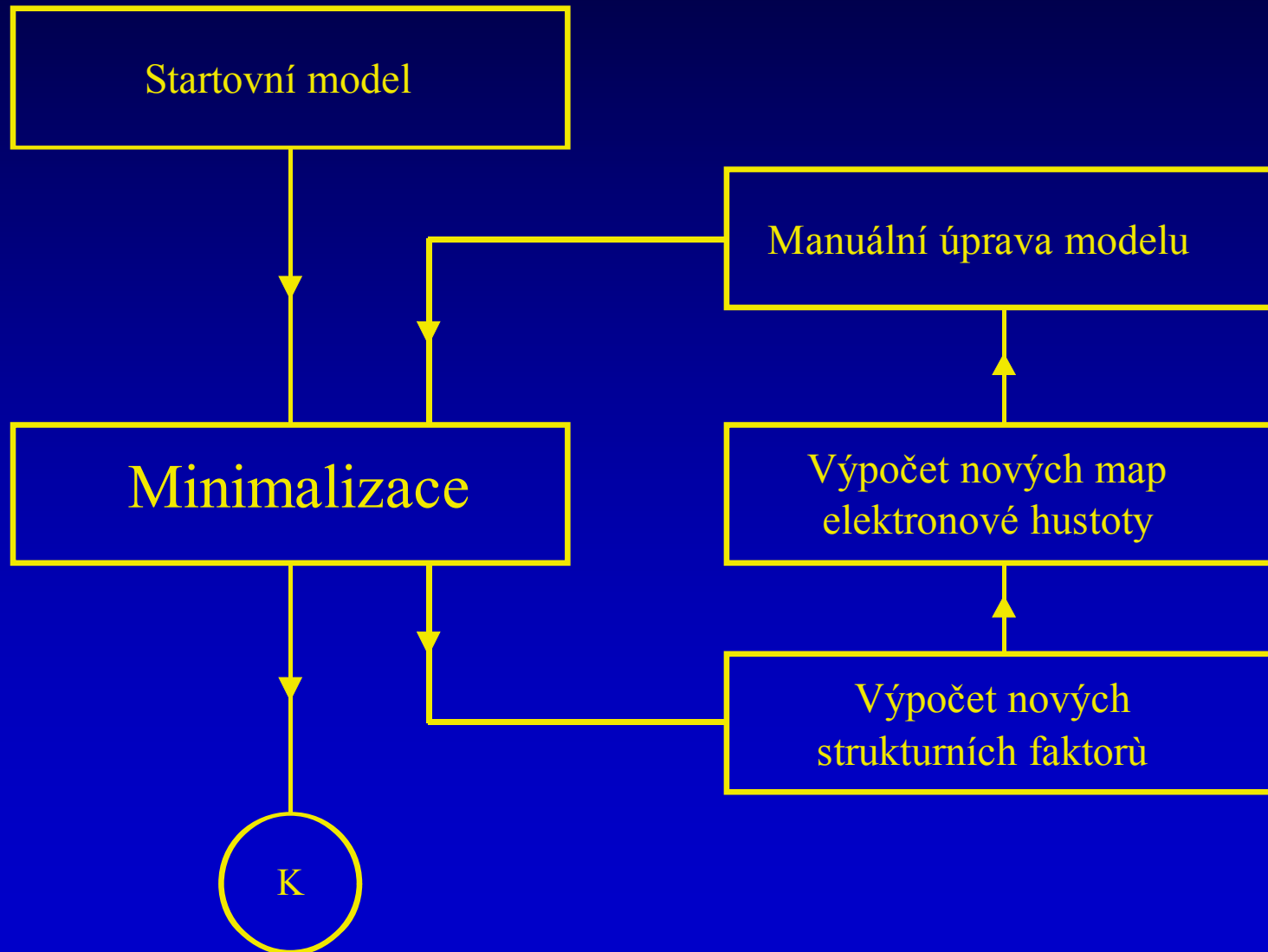
- Taylorův rozvoj kolem minima se členy 1. řádu vede na soustavu lineárních rovnic

$$S^i = \sum_H w_H \left(\Delta F_H^i - \sum_k \frac{\delta |F_H|_c^i}{\delta x_k} \Delta x_k \right)^2 = (\mathbf{F} - \mathbf{AX})^{-1} \mathbf{W} (\mathbf{F} - \mathbf{AX})$$

Zdroje nelinearity:

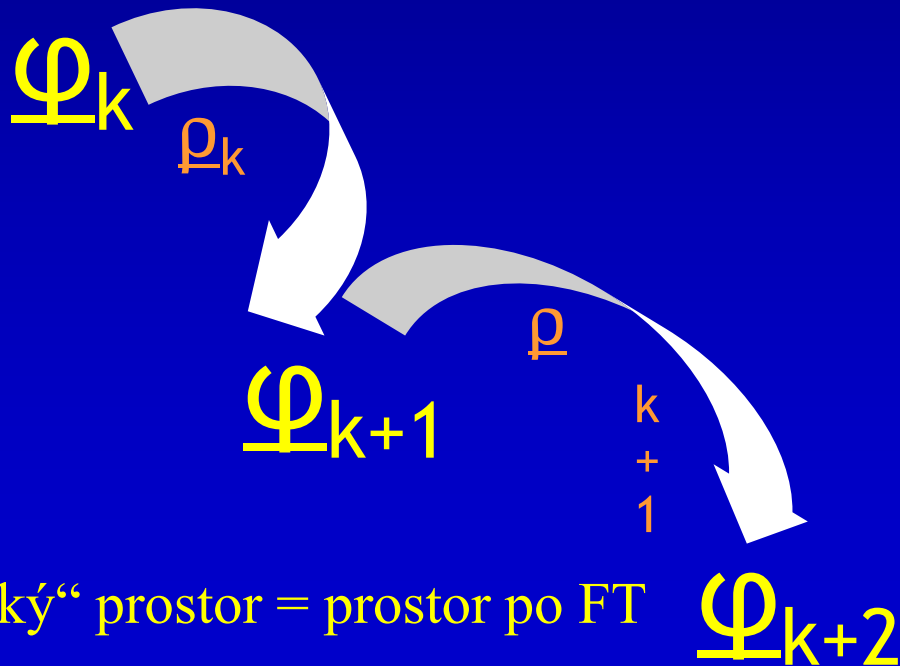
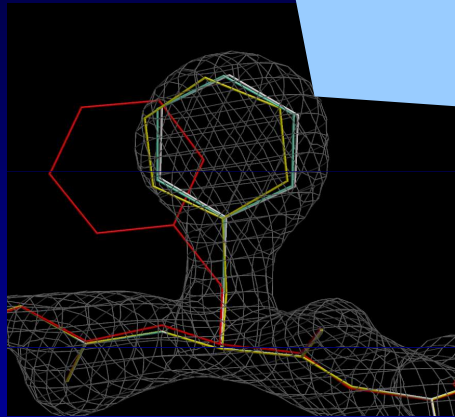
- Proteiny – jde doopravdy o rozvoj kolem globálního minima?
- Špatně určené startovní fáze => pomalá konvergence ke správnému řešení

Iterativní zpřesňování proteinů

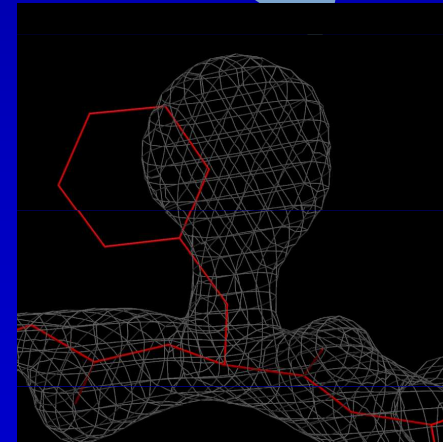


Iterativní zpřesňování proteinů

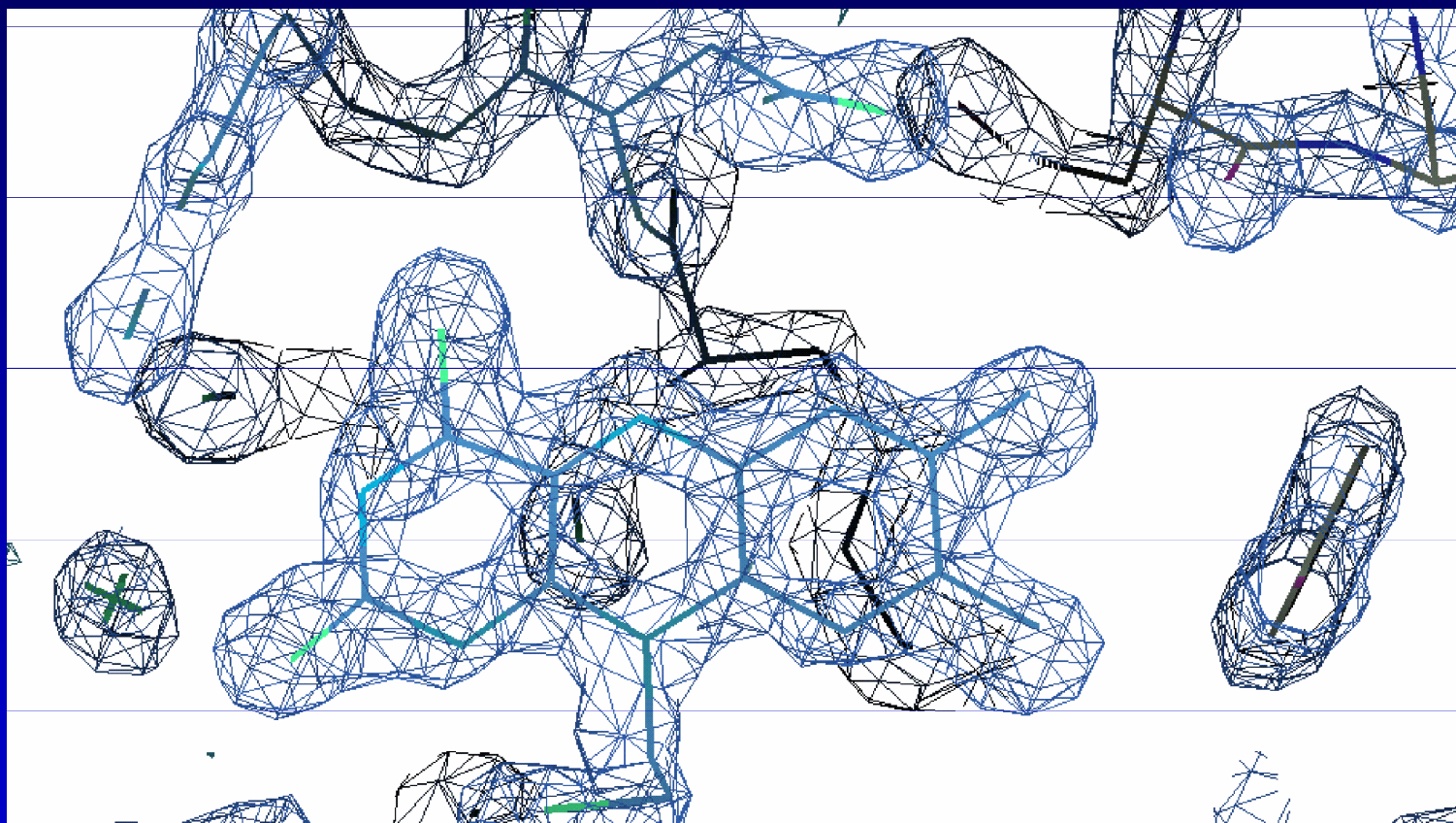
„Přímý“ prostor, prostor atomových souřadnic + elektron. hustot



„Reciprokový“ prostor = prostor po FT



Elektronová hustota vs 3-D model

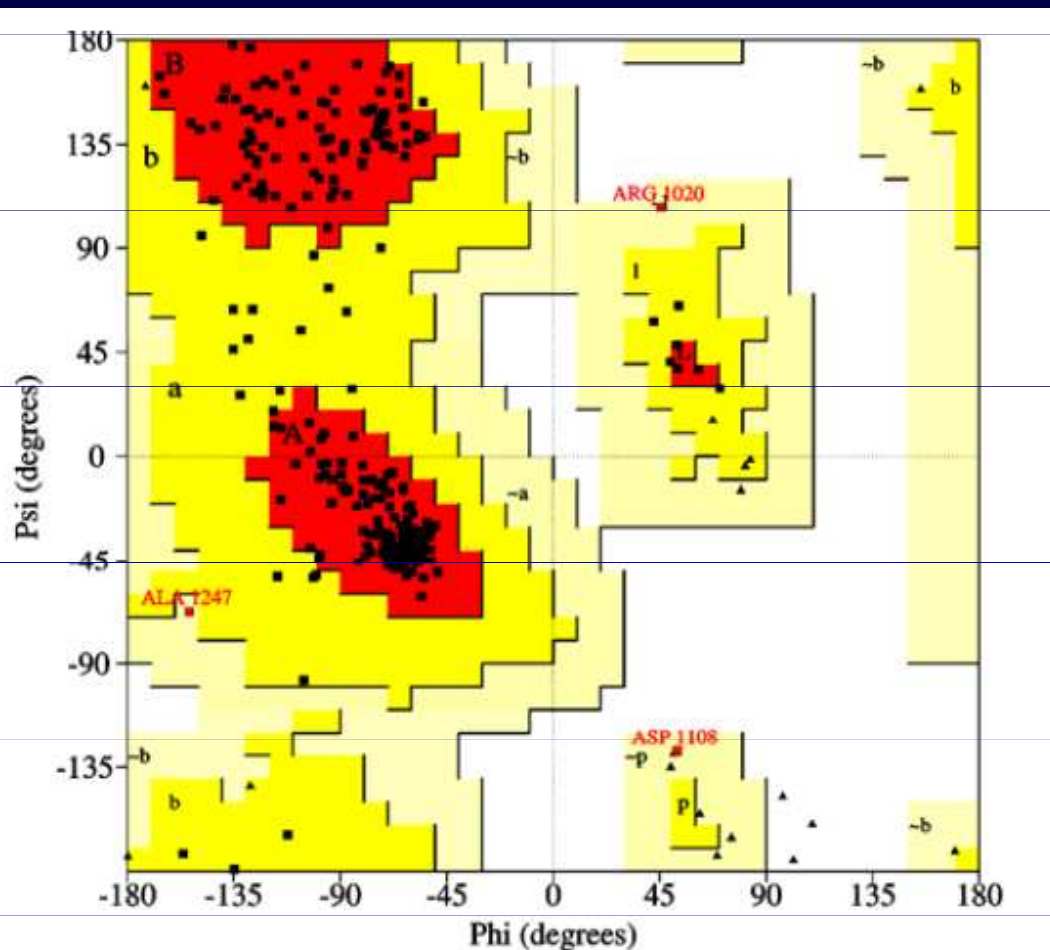


Kritéria správnosti 3-D modelů

$$R = \frac{\sum_H \left| |F_H^o| - K |F_H^c| \right|}{\sum_H |F_H^o|}$$

$$F_H^c = \sum_{m=1}^M f_m \exp(2\pi i H \vec{r}_m)$$

$$R_{FREE} = \frac{\sum_{H \in T} \left| |F_H^o| - |F_H^c| \right|}{\sum_{H \in T} |F_H^o|}$$



Obrázek 96 Ramachandranův graf dobře vypřesněného peptidického řetězce haloalkan dehalogenasy LinB. Glycin je značen trojúhelníčky (▲), ostatní rezidua čtverečky (■).