

Proteinová krystalografie

Jaromír Marek,
Laboratoř funkční genomiky
proteinů, MU Brno

Historický úvod

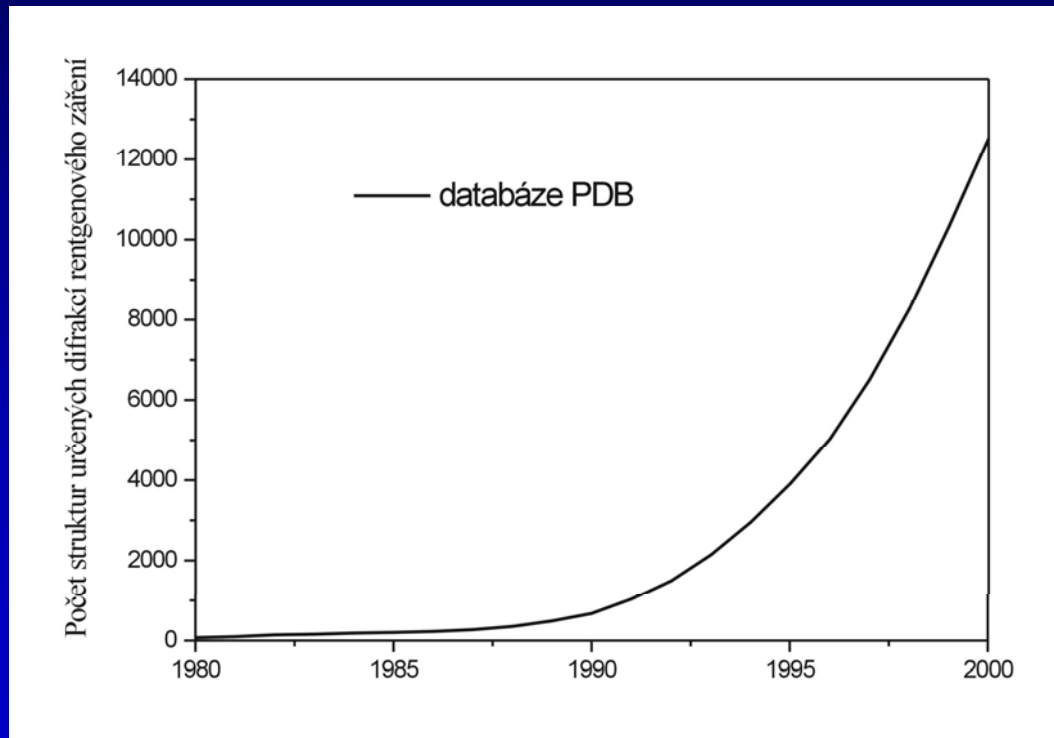
- Monokrystalová strukturní analýza - studium 3-D struktur „molekul“ difrakčními technikami
- Potřeba vhodné „sondy“ – například rentgenového záření (1895, N.c. za fyziku 1901 – W. C. Röntgen)
- 1912 - průkaz vlnové povahy rentgenového záření jeho difrakcí na krystalu (N.c. za fyziku 1914 – M. von Laue)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

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
- 1988 - chemie - J.Deisenhofer, R.Huber & H. Michel - membránové proteiny
- 2006 - chemie – Roger Kornberg – RNA polymeráza
- 2009 - chemie - Venkatraman Ramakrishnan, Thomas A. Steitz, Ada E. Yonath - ribosome

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

90. léta - plošné detektory,
synchrotronové záření, PCR
biologické makromolekuly
(databáze PDB)



Postgenomická éra biologie ?

- J.C. Venter *et al.*: The Sequence of the Human Genome, *Science* **2001** February 16; 291: 1304-1351.
- The genome international sequencing consortium: Initial sequencing and analysis of the human genome, *Nature* **409**, 860-921; 15 February 2001.
- The *Arabidopsis* genome initiative: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815; 14 December 2000.

Studium genetických informací

- Prohledávání velkých databází a hledání jednotlivých genů (lidský genom:25-28000 genů?)
- Určování funkce jednotlivých genů
- Struktury všech proteinů:*Human proteome project*

Techniky studia 3D struktur proteinů

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

Očekává se, že převážná většina struktur globulárních proteinů (proteinů s dobře určenou terciární strukturou) bude určována difrakcí rtg. (resp. synchrotronového) záření i v budoucnu.

Teoretické základy difrakčního studia 3-D struktur

- „sonda“ vhodné velikosti pro studium atomů – rentgenové záření o vlnové délce v oboru standardních meziatomových vzdáleností ($\sim 1 \text{ \AA}$)
- Foton s látkou interaguje rozptylem nebo absorbcí
- Rozptyl - s energetickými ztrátami – Comptonovský
- beze ztrát energie - Thompsonův

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

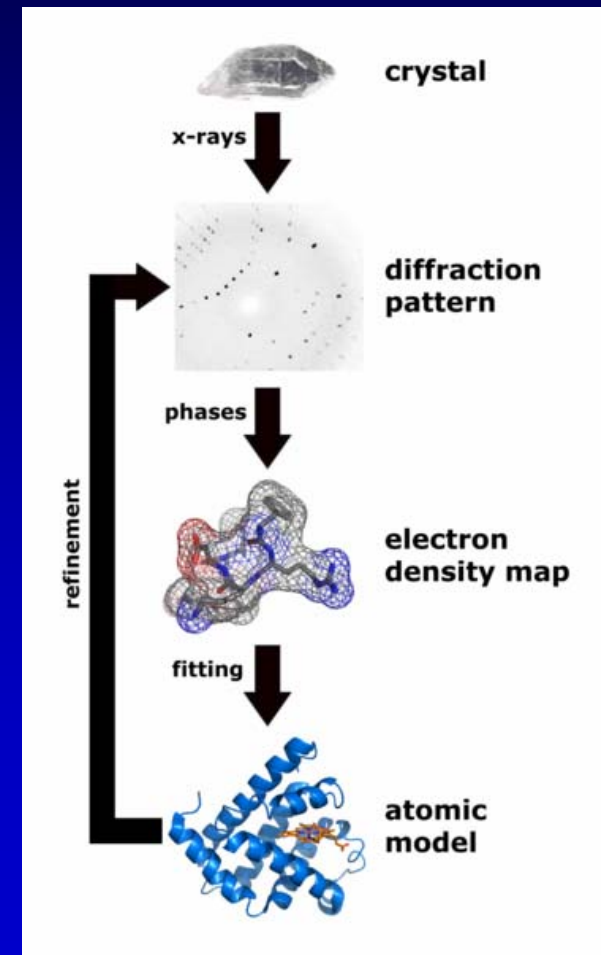
Nabitá částice je v poli rovinného monochromatického záření sekundárním zdrojem elektromagnetického pole

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

Rozptyl na protonech je nevýznamný (je 1837^2 x slabší),
difrakcí RTG záření studujeme elektronovou strukturu látky

Určování struktur enzymů

1. Určení genu
2. Příprava rekombinantního proteinu, čištění, zahušťování...
3. Krystalizace
4. Difrakční experiment
5. Fázový problém, příprava modelu
6. Zpřesňování modelu



Studium genetické informace

- Není dosud přesně známo, kolik genů v lidské DNA je.
- Zatím jen u cca 1/10 lidských enzymů je známa jejich funkce
- Zpracování genetických informací není možná bez počítačů
- Problematické místo - formulace hypotéz o polohách genů a funkcích jejich produktů

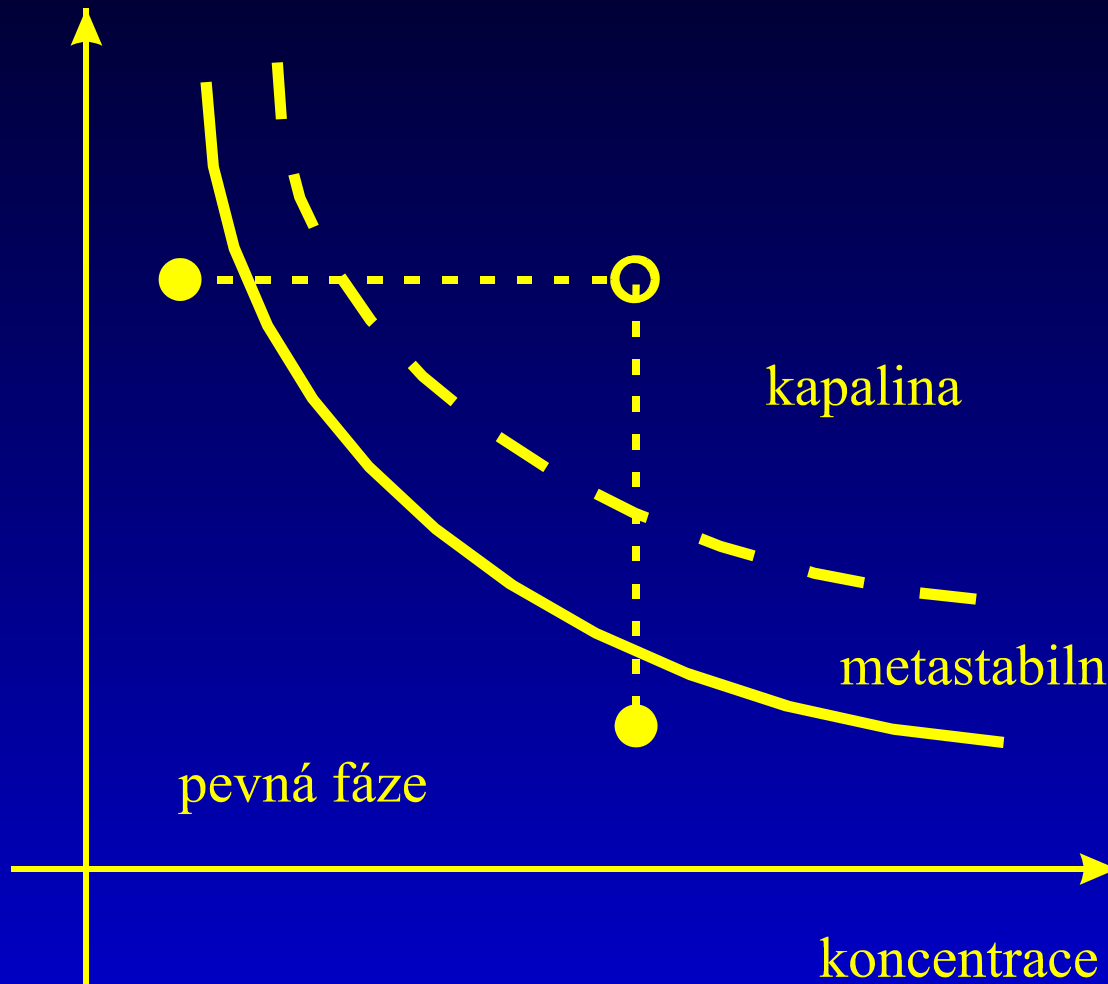
Příprava biologického materiálu

- Naprostá většina v současnosti studovaných proteinů se připravuje biotechnologickými metodami (rekombinantní DNA, nadprodukce v modelovém systému, ...)
- Výhody – „snadnost“ provádění genetických modifikací
- Kritická místa : funkčnost u rekombinantních proteinů
protein musí být rozpustný
čištění

Krystalizace

- Až doposud nejkritičtější a časově nejnáročnější část určování 3-D struktur
- Využívání již existujících zkušeností o krystalování jiných proteinů (statistické zpracovávání „řidkých“ mnohazměrných množin empirických dat => krystalizační sety pokrývající širokou škálu chemických podmínek)
- Proteomický projekt - první testy s robotem na automatickou krystalizaci (automatické míchání a pipetování roztoků, nanolitrové objemy, strojem řízené mikroskopování a na FT založené vyhodnocování výsledků) . Zpracování až desítek proteinů (desítky tisíc testovacích krystalizací) týdně.

teplota



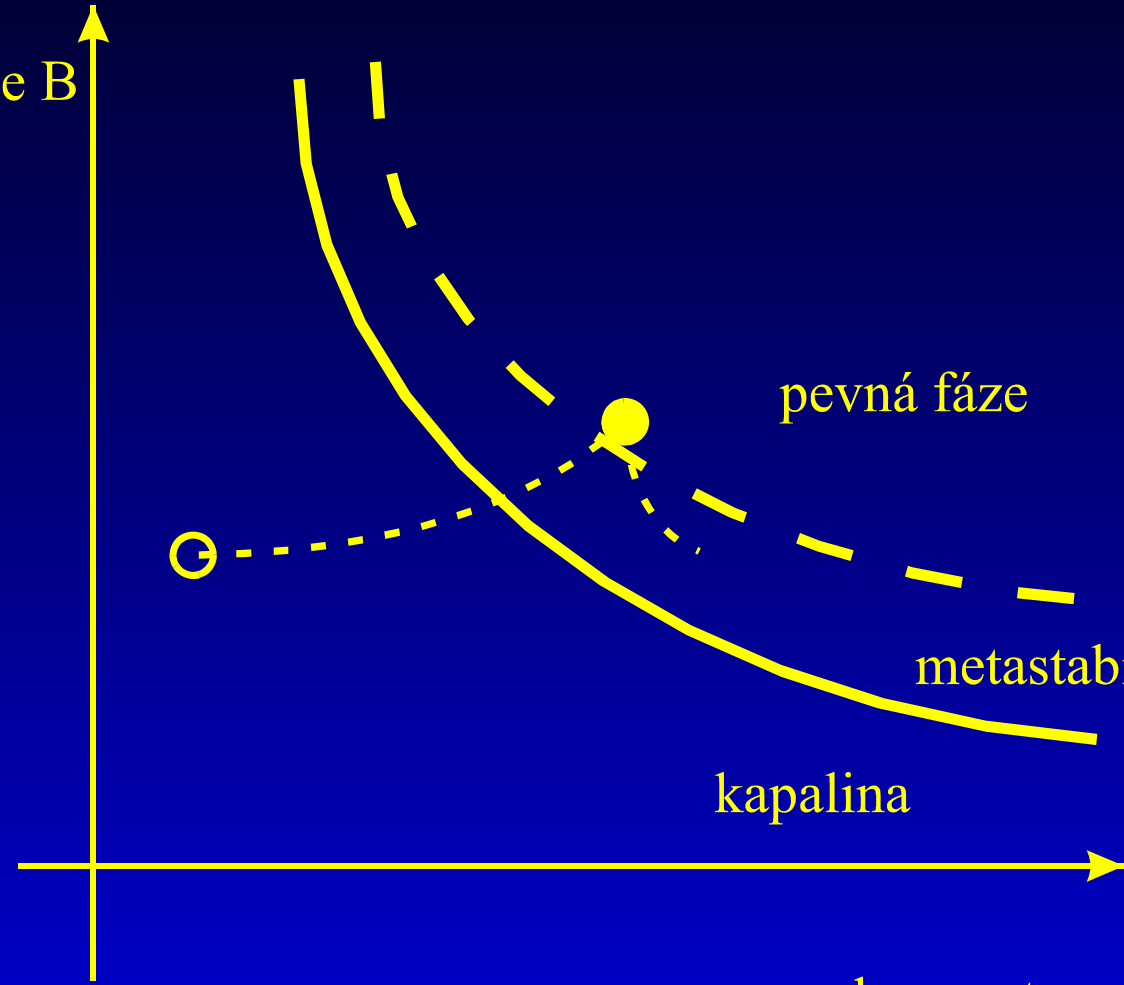
kapalina

metastabilní oblast

pevná fáze

koncentrace

koncentrace B



pevná fáze

metastabilní oblast

kapalina

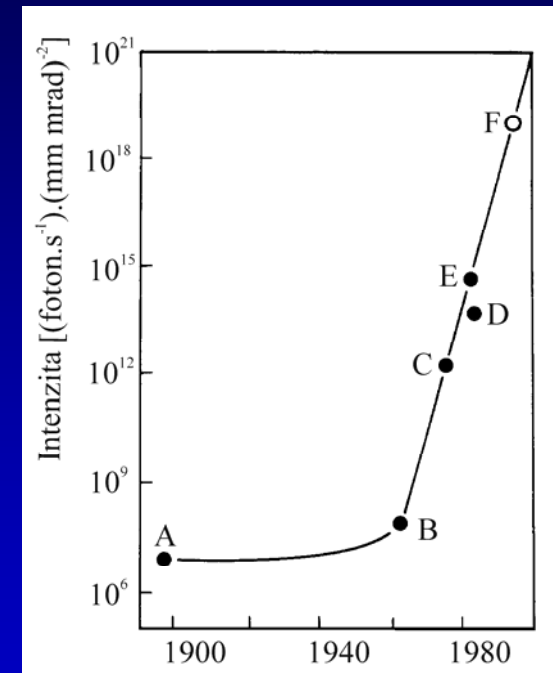
koncentrace A

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

Konvenční laboratorní zdroje **charakteristického**
rentgenového záření – rentgenová lampa a rotační
anoda

Limitace – „bodový“ zdroj kulových vln s omezenými
možnostmi zvyšování vyzářeného výkonu

Synchrotrony – zdroje vysoce intenzivního spojitého
spektra – řádově kratší experiment (řádově hodiny)



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

- Rentgenový film
- Scintilační detektor
- Plošný mnohadrátový proporciální detektor
- Detektor typu „obrazová deska“
- CCD detektor

Difrakční experiment pro určení 3-D struktury: Detektor typu „obrazová deska“ (image plate)

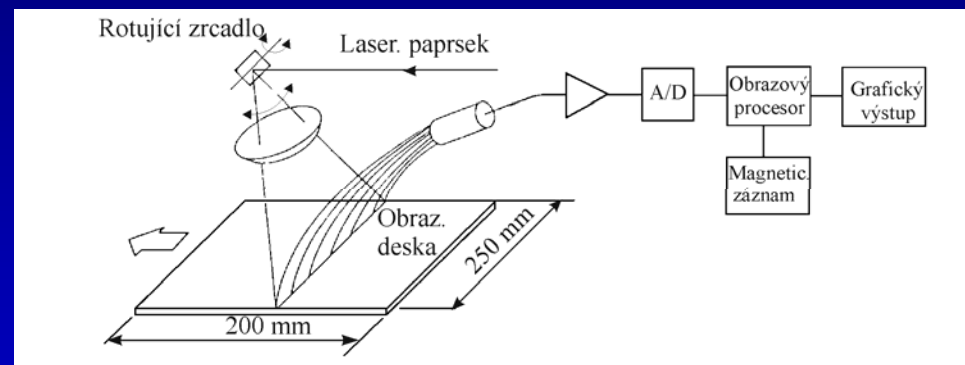
Elektronicky zpracovatelný analogie
klasického filmu

Integrovaný detektor zaznamenávající
virtuální obraz (ionty Eu^{3+})

Klady: - 50x citlivější než film, nízký
šum

+/-: - velké rozměry celého zařízení
- rozlišovací schopnost cca 0.1 mm
daná zrníky krystalů detekční látky

Zápory: - omezená linearita
- „vyvolávání“ je relativně dlouhé



Difrakční experiment pro určení 3-D struktury: Plošný detektor s polovodičovým CCD prvkem

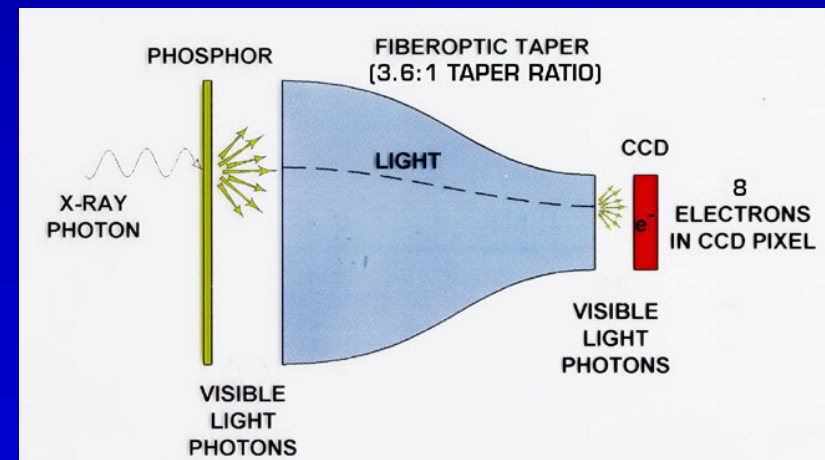
Detektor pracuje s podobnými polovodičovými prvky jako digitální fotoaparát.

Scintilační krystal převádí rentgenové záření do pásma viditelného světla

Klady: - rychlost

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

Zápor: - vlastní teplotní šum polovodičů



Difrakční experiment

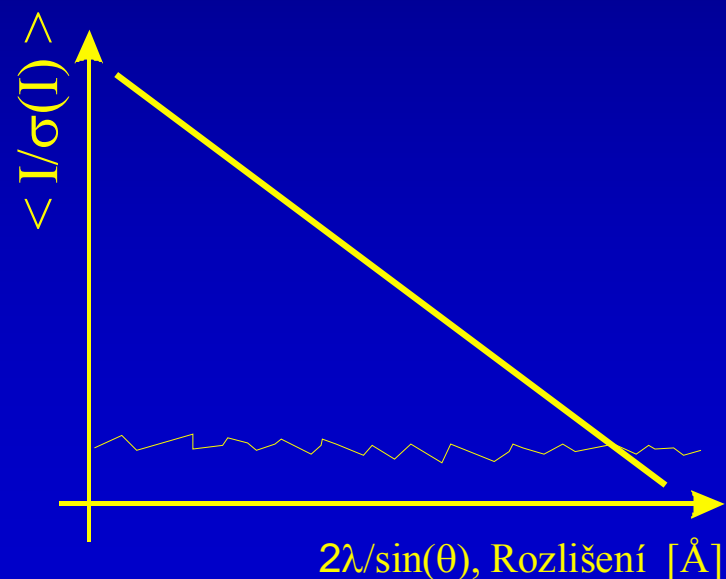
- Naprostá většina proteinových struktur je nyní určována ze synchrotron. difrakčních dat (hlavní klady: intenzita záření, rychlost experimentu, možnost optimalizace vlnové délky) a za kryopodmínek (stabilita krystalu, lepší difrakční schopnosti).
- „Domácí“ laboratoře – testování difrakční kvality krystalů, „ladění“ kryoexperimentů, předzmrázování krystalů.
- Sběr úplných difrakčních synchrotron. dat – řádově desítky minut.
- Automatizovaná výměna a měření zmražených vzorků
- Obrovský „boom“ Se-proteinů & MAD/SAD experimentů

Experiment: omezené množství dat

- Geometrická limitace počtu měřitelných dat – Braggova rovnice

$$2d_H \sin \theta = \lambda$$

- Difrakční experiment a jeho rozlišení (Resolution)



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

Strukturní faktor – popisuje amplitudu difraktované vlny

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

Krystalová elektronová hustota je 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

Fázový problém

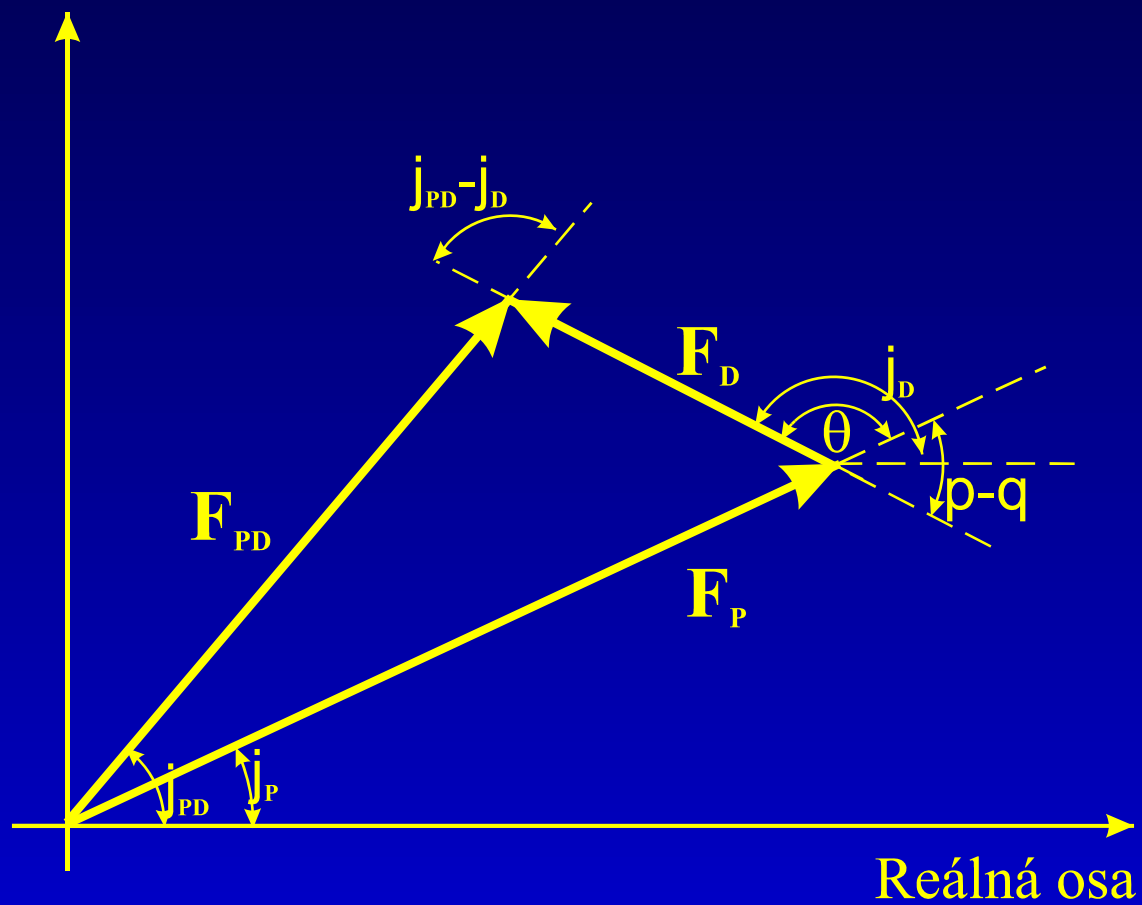
- cíl – zjistit 3-D model studované (makro)molekuly
- prostředek – určit při difrakčním experimentu ztracenou informaci o fázích strukturních amplitud a FT poté získat mapy elektronových hustot
- nejjednodušší metoda – fázový problém vůbec neřešit, využít podobnost studovaného systému se systémem s již známou 3-D strukturou (MR, molecular replacement).
- nutná je poměrně velmi vysoká podobnost mezi modelem a studovaným systémem (AA identita cca 30% a lépe, AA podobnost 50% a lépe)

Deriváty proteinů

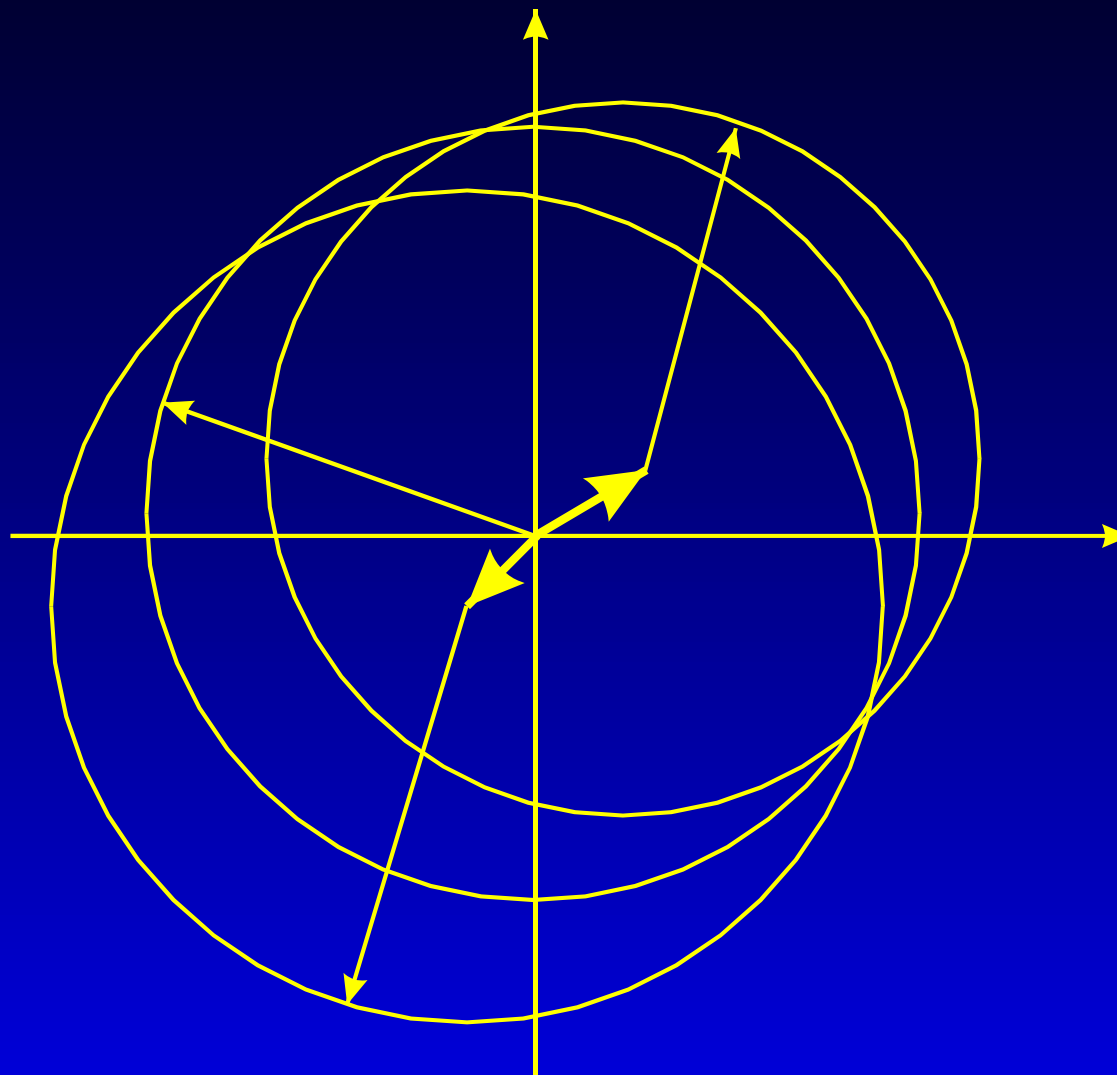
- kanály rozpouštědla (krystalograficky neuspořádané vody)
- relativní stabilita terciální struktury globulárních proteinů při interakci jejich interakci s „malými“ molekulami
- nutnost opakovaných měření s různými dobře difraktujícími izomorfními deriváty
- podobnost struktur proteinů s jejich Se-Met analogy
- disperzní závislost reálné i imaginární složky f_{Se}

Izomorfní nahrazení

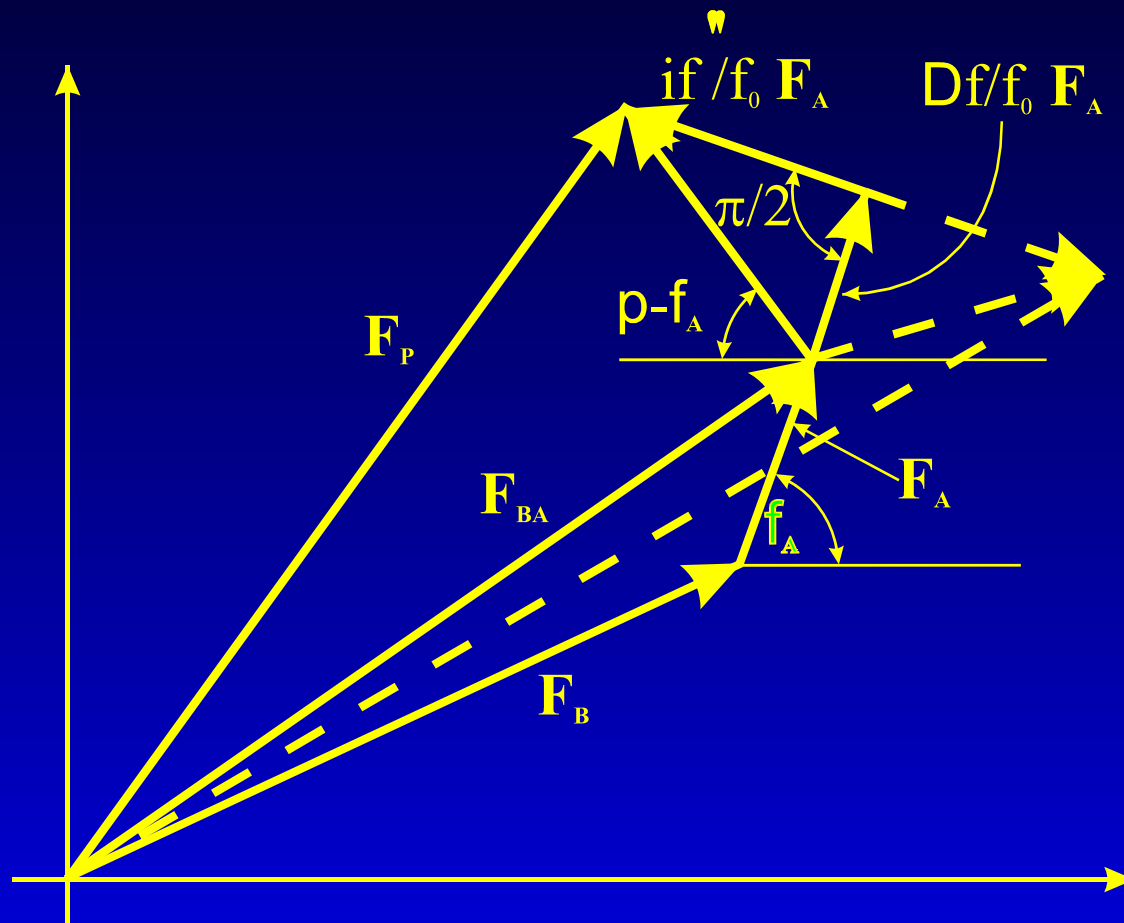
Imaginární osa



Vícenásobné nahrazení (MIR)



Vícenásobná anomální disperse (MAD)



Zpřesňování proteinů: omezené množství dat

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 B
- Teplotní kmity

Zpřesňování proteinových struktur: možné problémy

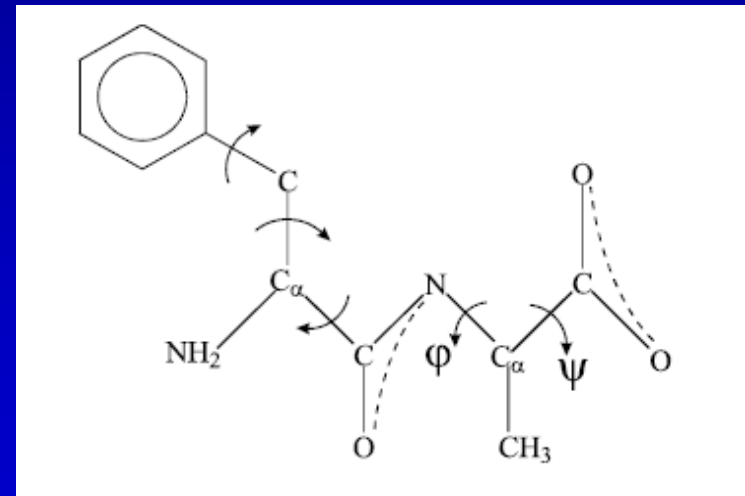
- Experimentální proměnné – difrakce
- Modelová funkce – strukturní amplituda
- Startovní strukturní model – MR, fáz. problém + mapa el. hustoty
- Kritérium správnosti – R faktory

- Limitovaný počet pozorování daný rozlišením experimentu
- Nelineární problém – iterativnost, konvergence
- Lokální vs. globální minima
- „Přefitování“

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 nekrystalografických dat
- popis pomocí „tolerancí“

$$\text{dist}(\text{Atom}_1, \text{Atom}_2) = D \pm \sigma(D)$$

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

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

- dodatečné údaje – strukturní databáze, spektrální data, QM výpočty
- váhové koeficienty
- 2 stupňové zpřesňování – generování chem. informace + minimalizace

Zpřesňování proteinových struktur: minimalizace

- Krystalografie malých molekul – metoda nejmenších čtverců

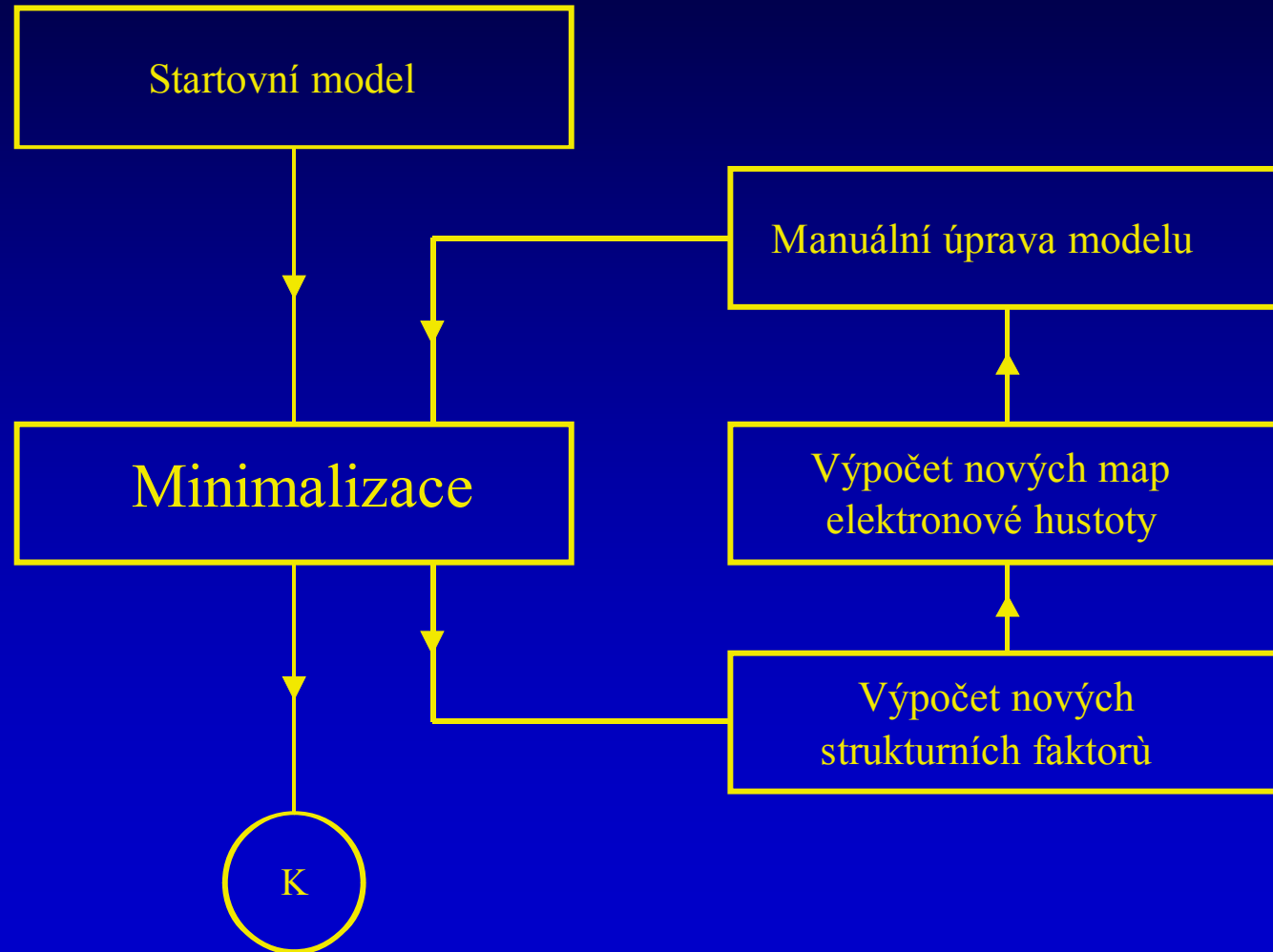
$$S^1 = \sum_H w_H \left(|F_H|_o - |F_H|_c \right)^2 \qquad S^2 = \sum_H w_H \left(|F_H|_o^2 - |F_H|_c^2 \right)^2$$

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

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

- Proteiny – jde doopravdy o rozvoj kolem globálního minima?
- Špatně určená fáze – pomalá konvergence (minimalizace rozdílů mezi mapami elektron. hustoty)

Iterativní zpřesňování proteinů



Stavba a zpřesňování modelu

- na empirii založené úpravy map elektronové hustoty
- knihovny fragmentů
- zpřesňování pomocí maximalizace entropie vs. konvenční minimalizace nejmenších čtverců (LS)
- eliminace nedostatku exp. dat využíváním chem. informace
- „brute force“ přístup: doplňování a ubírání molekul vody & hledání chemické interpretace nového modelu

Zpřesňování : maximalizace pravděpodobnosti

- Pannu & Read (1996) – alternativní přístup – nejpravděpodobnější řešení (maximum likelihood method, ML)

$$S_{LS} = \sum_H w_H \left(|F_H|_o - |F_H|_c \right)^2 \quad S_{ML} = \sum_H \frac{1}{\sigma_{ML}^2} \left(|F_H|_o - \langle |F_H| \rangle^{CV} \right)^2$$

- podobnost vztahů pro LS a ML
- metoda nejmenších čtverců – speciální případ maximalizace pravděpodobnosti
- problémy – váhy, implementace „chyb“, „přefitování“
- „kross-validace“, R-free

Proteomický projekt

- automatizovaný přístup k řešení proteinových struktur
- požadovaný veliký výstup & obrovské finance na vstupu
- limitovaná účinnost