

From sequence to 3D
structure of proteins

Metodi per la determinazione sperimentale della struttura di macromolecole

Un po' di storia...

1866	Mendel scopre i geni
1944	Il DNA è il materiale genetico
1951	Prima sequenza di una proteina (insulina)
1953	Struttura del DNA
1959	Struttura della mioglobina
1960s	Delucidazione del codice genetico
1977	Avvento del sequenziamento del DNA
1975-79	Primi clonaggi di geni umani
1986	Sviluppo di un sistema di seq. aut. del DNA
1995	Primo genoma completo (<i>H. influenzae</i>)
1997	Genoma di <i>E. coli</i>
1999	Primo cromosoma umano(Chr #22)
2000	<i>Drosophila</i> /Arabidopsis genomi
2001	Genomi dell'uomo e di topo

Kendrew determinò la struttura a bassa risoluzione (6 Å) della **mioglobina** nel 1957 ed una struttura ad alta risoluzione (2 Å) nel 1959. Perutz completò la struttura a bassa risoluzione (5.5 Å) dell'**emoglobina** lo stesso anno ed una ad alta risoluzione (2.8 Å) nel 1968.



Perutz e Kendrew condivisero il premio Nobel in Chimica nel 1962

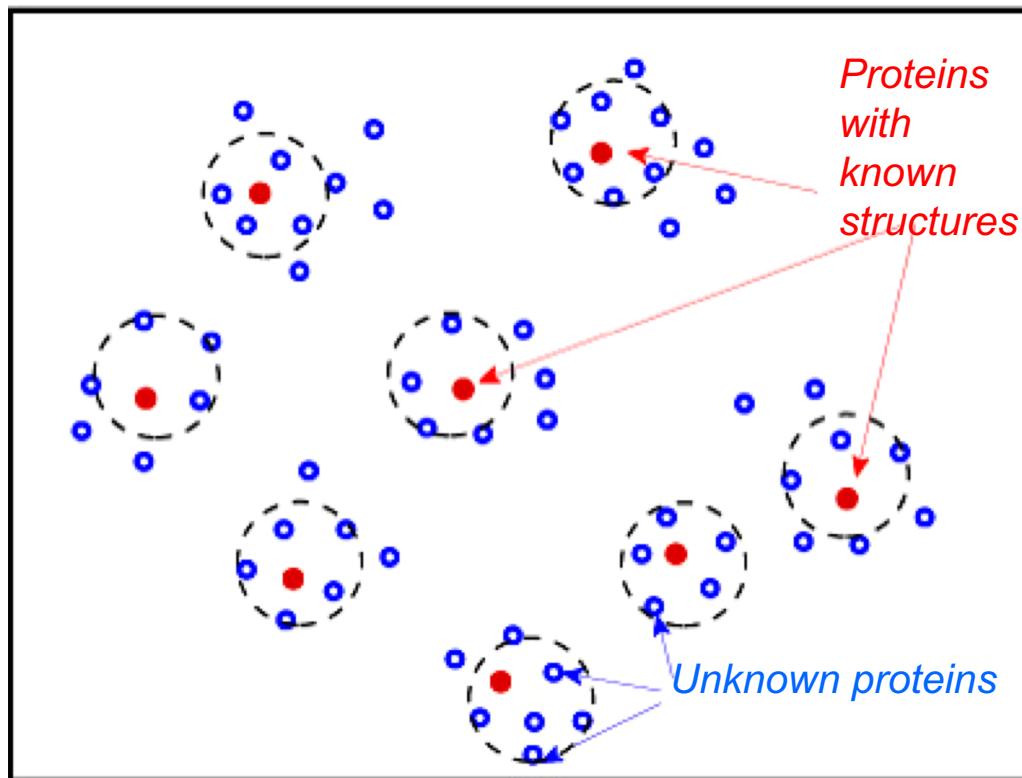
La determinazione sperimentale di strutture di macromolecole si rivelò un processo lungo e piuttosto complicato

Questo ha comportato che le strutture siano state per anni determinate sulla base di specifiche necessità (e.g. di rilevanza medica) e solo dopo che molti dati sperimentali su funzione, localizzazione, interattori, etc. erano già noti

Tra il 1995 e il 2000 viene dato inizio ad un progetto detto di **genomica strutturale**

Genomica strutturale

Determinazione della struttura di proteine i cui fold rappresentino il completo spazio dei fold presenti in natura



La struttura delle restanti proteine può essere predetta basandosi sulla similarità di sequenza

Questo pone un quesito....

Lo spazio dei fold è discreto o continuo?

Metodi per la determinazione della struttura di macromolecole

Esiste solo un ristretto numero di metodi per la determinazione di strutture proteiche complete



Metodi basati sulla **diffrazione** o **scattering** di particelle subatomiche o di onde elettromagnetiche

Metodi spettroscopici. Si basano su variazioni di stati energetici degli atomi delle proteine dovute alla loro interazione con radiazione elettromagnetica a diverse frequenze

Metodi per la determinazione della struttura di macromolecole

Metodi di scattering/difrazione

Diffrazione a raggi X

Scattering di neutroni

Micoscopia elettronica

Metodi spettroscopici

NMR (Nuclear Magnetic Resonance)

EPR (Electron Paramagnetic Resonance)

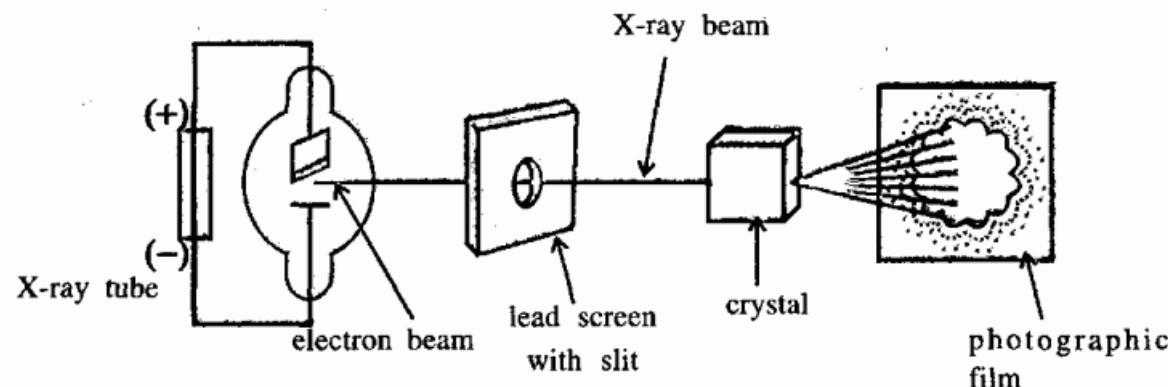
Metodi computazionali

Approccio fisico (ab initio)

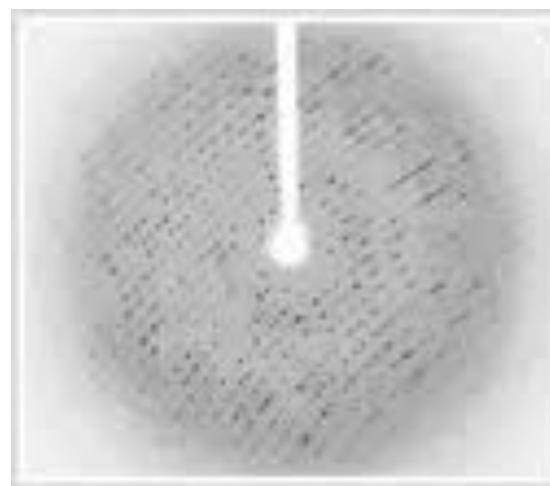
Threading

Modeling per omologia

Metodi di scattering o diffrazione

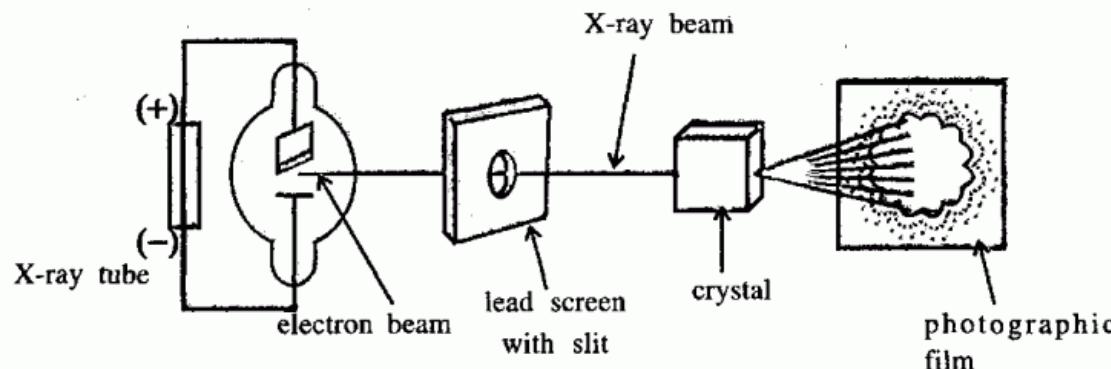


Informazione sulla struttura di una molecola può essere ottenuta osservando come un fascio viene diffratto quando colpisce il cristallo tridimensionale della molecola



Diffrazione a raggi X

E' considerato ad oggi il metodo più accurato per la determinazione della struttura di macromolecole



2.1 A beam of X-rays is diffracted by crystal into a symmetrical pattern

- I raggi X vengono diffratti dalle nubi elettroniche degli atomi del cristallo dando luogo ad un pattern di diffrazione su un materiale sensibile (film)
- Il pattern di diffrazione è una collezione di punti che assomigliano alla superficie di un setaccio
- Essa contiene (quasi) tutta l'informazione per determinare la struttura della molecola
- La conversione del pattern di diffrazione in una struttura molecolare è un passaggio matematico che implica conoscenze fisico-chimiche

Diffrazione a raggi X

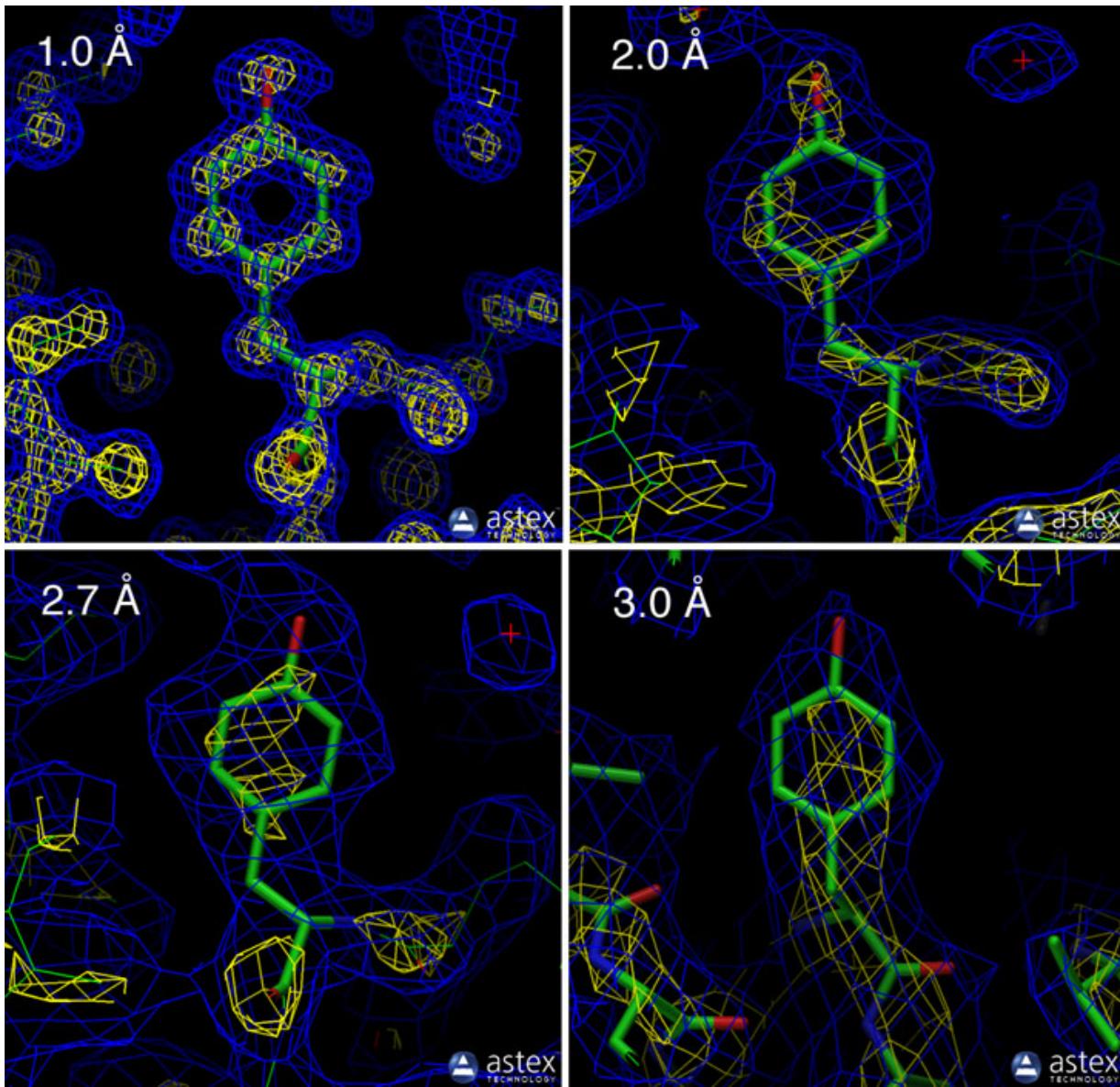
La risoluzione è la minima distanza che un sistema ottico è in grado di distinguere, senza che il fenomeno della diffrazione confonda l'immagine.

Date le opportune lenti, il solo fattore che possa limitare la capacità di risoluzione è la lunghezza d'onda della luce (o di una qualsiasi altra radiazione elettromagnetica) riflessa dall'oggetto che stiamo osservando

$$r \sim \lambda/2$$

La più corta lunghezza d'onda nel visibile è pari a 4000 Å mentre la distanza tra gli atomi nelle molecole è 1-2 Å

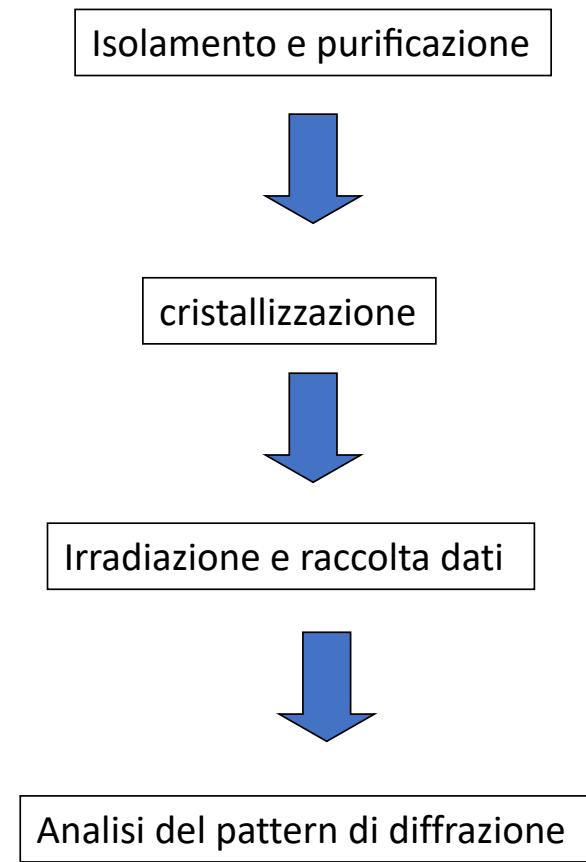
La lunghezza d'onda dei raggi X è $\lambda = 0.1\text{-}100 \text{ \AA}$



Diffrazione a raggi X

- I raggi X non possono essere focalizzati da lenti (no microscopi a raggi X)
- Per via dei molti atomi leggeri (C, N, O, H), le biomolecole deflettono i raggi X debolmente
- Se si aumenta la concentrazione di molecola in soluzione si ottiene un segnale incoerente
- Necessità di un'organizzazione ripetitiva delle molecole nel campione
- Cristalli

Diffrazione a raggi X



Analisi del pattern di diffrazione

L'intensita' misurata per ogni segnale di diffrazione dipende da vari fattori tra cui:

- Intensita' incidente
- Fattori strumentali e geometrici
- Volume del materiale cristallino e suo assorbimento
- Fattore di struttura F_{hkl} del relativo piano di diffrazione

Analisi del pattern di diffrazione

- Dalle intensita' diffratte, fatte le opportune correzioni, possiamo ricavare il **modulo** dei fattori di struttura osservati.
- Se conoscessimo i fattori di struttura in modulo e fase potremmo determinare la struttura dato che le posizioni degli atomi nella cella elementare sarebbero univocamente determinabili.
- Tra reticolo cristallino (reticolo diretto) e reticolo reciproco (fattori di struttura) esiste una precisa relazione.
- Il fattore di struttura e' la trasformata di Fourier della densita' elettronica della cella elementare.

$$F_{hkl}(\vartheta) = \int \rho_{cella}(r) \exp(2\pi i \varphi_{hkl}) d\mathbf{r}$$

Così come è possibile ricostruire il reticolo diretto conoscendo il reticolo reciproco, così l'**anti-trasformata di Fourier del fattore di struttura (se noto in modulo e fase)** consente di ottenere la distribuzione di densità elettronica della cella elementare.

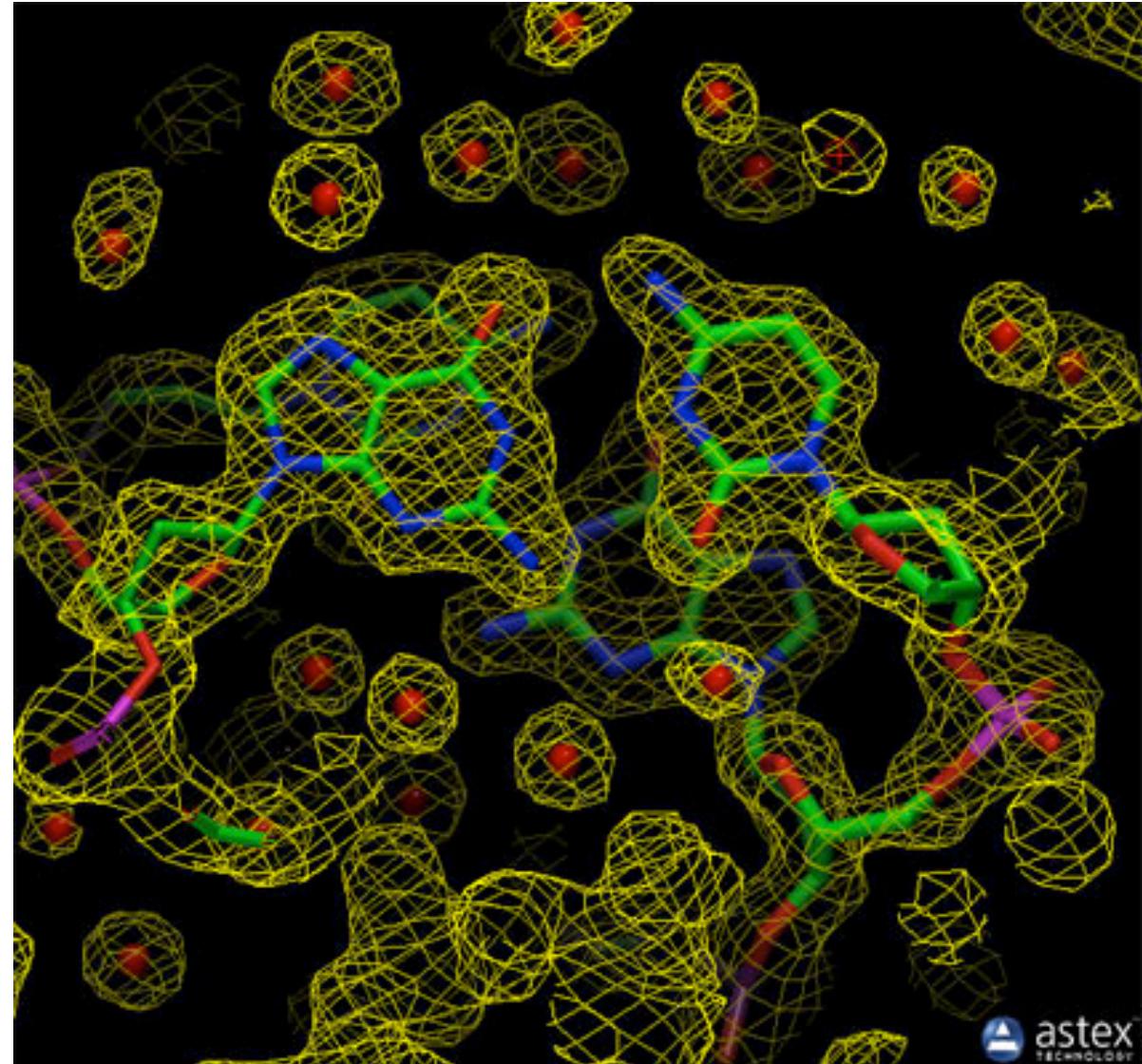
Analisi del pattern di diffrazione

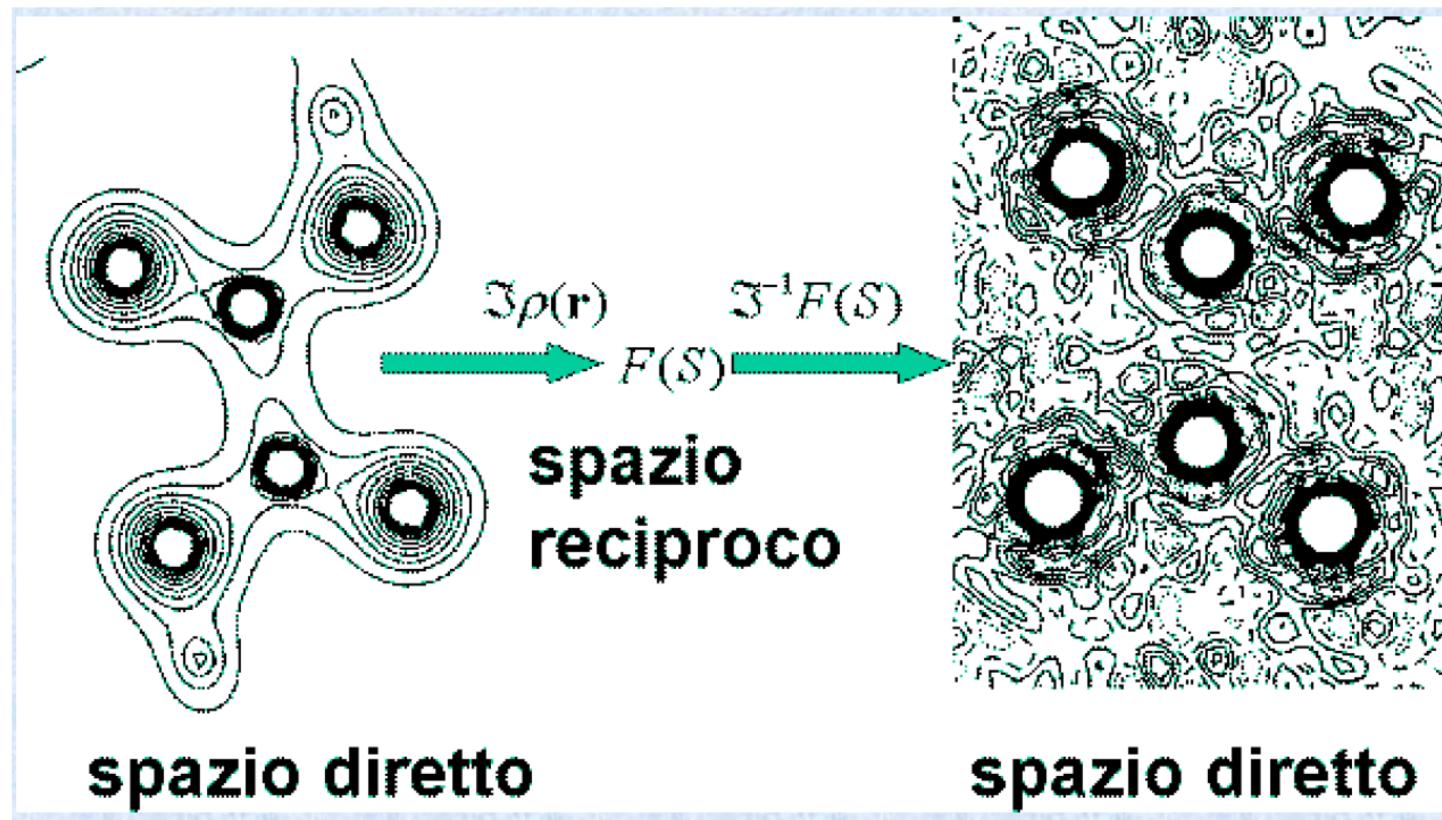
- I dati di diffrazione includono solo l'**intensità** dei punti
- L'intensità dei punti è proporzionale al quadrato dell'ampiezza dei fattori di struttura ma non contiene informazione sulla fase.
- Resta dunque il problema della **fase**
- I fattori di struttura F sono numeri complessi $F=|F| e^{i\phi}$ e gli esperimenti forniscono il modulo di F ma non la fase ϕ .
- Quando si fa l'anti-trasformata di Fourier, l'ampiezza da' una misura della dimensione degli atomi che le diffaggono mentre la fase fornisce informazioni sulla loro posizione relativa.

- Ci sono vari metodi per recuperare l'informazione mancante delle fasi
- Nel caso del **molecular replacement** il cristallografo usa una struttura **omologa nota** per dedurre la struttura ignota
- Il molecular replacement prova a trovare il modello che vada meglio con le intensità sperimentali in strutture note
- R-factor: differenza tra i dati misurati e quelli predetti dal modello
- Il risultato di ottimizzazione è una mappa di densità elettronica
- Essa è una mappa delle regioni in cui la densità elettronica è sufficientemente alta per indicare la presenza di un atomo

The experimental electron density from a structure of DNA (PDB entry [196d](#)), along with the atomic model that was generated based on the data.

The contours surround regions with high densities of electrons, which correspond to the atoms in the molecule.





Poiché il numero di punti del reticolo reciproco accessibili è comunque limitato la ricostruzione della densità elettronica risulta un pò distorta

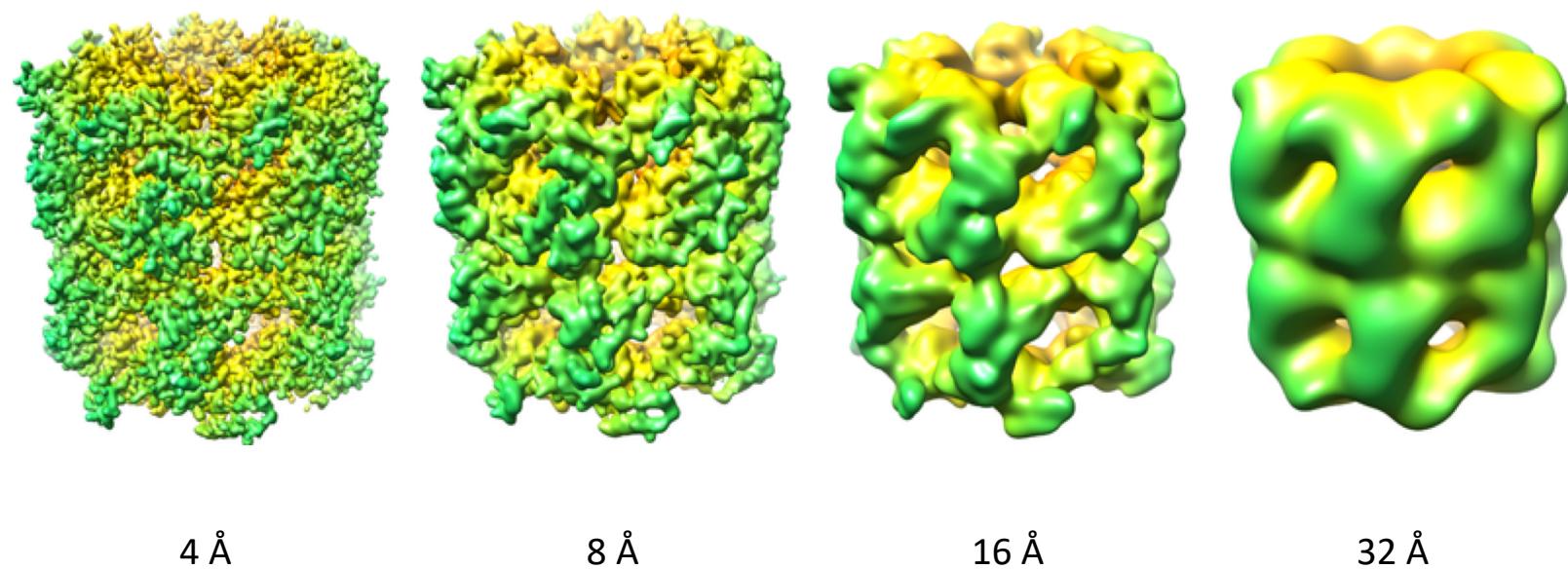
- Ma che tipo di atomo?
- Per questo passaggio è fondamentale la conoscenza e l'esperienza del cristallografo
- Possono anche essere usati calcoli di tipo energetico o geometrico
- La composizione della proteina ha un ruolo importante nella qualità del modello finale
- Le regioni ordinate sono più semplici da determinare
- Le regioni disordinate o flessibili (in particolare i loops) sono difficili
- Alcune regioni possono rimanere irrisolte

Informazione ottenuta dalla cristallografia

- **Struttura molecolare**
- **Deviazioni** indotte dalle vibrazioni di ciascun atomo intorno alla sua posizione media: B-factor o fattore di temperatura
 - 20 < B-factor < 45: atomi ordinati
 - B-factor > 60: atomi flessibili
- **Risoluzione:** la più piccola separazione tra piani di atomi che diffraggono i raggi X
 - 4.0-5.0 Å: forma della proteina e posizione delle strutture secondarie
 - 3.5 Å: traccia della catena polipeptidica
 - 2.0-2.5 Å: mostra le catene laterali degli amino acidi
 - 1.0-1.2 Å: mostra gli atomi individuali eccetto gli H
 - 0.8 Å: mostra gli atomi di H e la valenza di legami covalenti.

La posizione degli atomi di H è particolarmente importante perché permette la determinazione dello stato di protonazione dei residui ionizzabili e l'orientazione delle molecole di acqua intorno e dentro la proteina

Caratteristiche strutturali di proteine a vari gradi di risoluzione



Problemi del metodo

Preparazione della proteina

cristallizzazione (pH, temperatura, salinità, etc.)

over-espressione in organismi semplici (no post-translational processing)

proteine di membrana (flessibili e circondate da lipidi)

Raccolta dati

Dati indiretti

problema della fase

I raggi X possono danneggiare le proteine

Qualità dei risultati

Condizioni innaturali

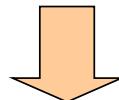
La proteina cristallizzata non è in soluzione

cambiamento conformazione della proteina (non-nativo)

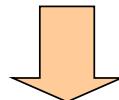
perdita di comportamento dinamico

Metodi spettroscopici

Interazione di molecole con radiazione elettromagnetica

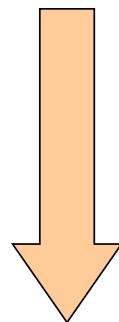
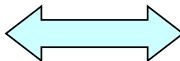


Cambiamento del livello di energia delle molecole



Emissione di energia da parte delle molecole irradiate

La dimensione e/o
frequenza dell'emissione
dipende dall'ambiente
chimico locale degli atomi
che emettono energia



Informazione strutturale sulla molecola

Metodi spettroscopici

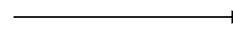
I vari metodi spettroscopici si distinguono per la diversa frequenza delle radiazioni elettromagnetiche che utilizzano

NMR (Nuclear Magnetic Resonance)



onde radio

EPR (Electron Paramagnetic Resonance)



micronde

FTIR (Fourier transform IR o Spettroscopia di Raman)



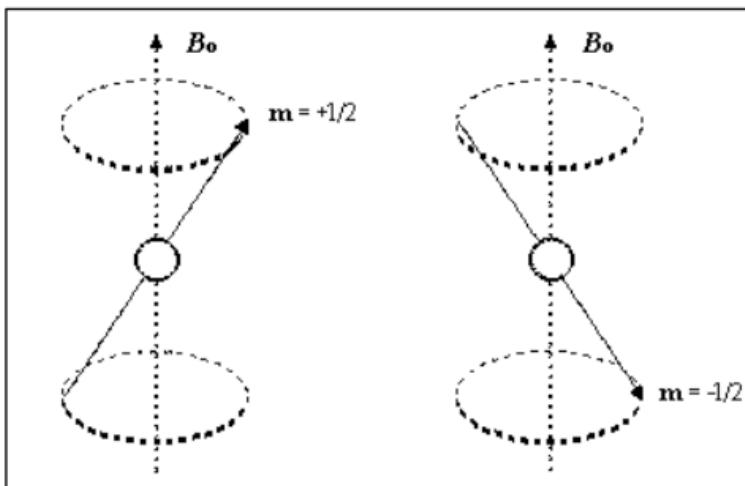
radiazione
infrarossa

NMR (Nuclear Magnetic Resonance)

- 1) Introduzione del campione all'interno di un forte campo magnetico B_0 per orientare gli spin nucleari nella posizione allineata con il campo o contro il campo.
- 2) Applicazione di un impulso di radiofrequenza per produrre un eccesso di nuclei eccitati con lo spin in opposizione al campo (durata: alcuni microsecondi).
- 3) Registrazione del FID, il segnale emesso dai nuclei mentre gli spin nucleari ritornano alla situazione di equilibrio (durata: circa un secondo).
- 4) Elaborazione matematica dei dati al computer applicando la Trasformata di Fourier per ottenere lo spettro NMR in funzione delle frequenze.

NMR (Nuclear Magnetic Resonance)

Quando un nucleo dotato di spin viene immerso in un campo magnetico, il nucleo è sottoposto ad una coppia di forze che lo fanno ruotare per allineararlo con il campo magnetico esterno



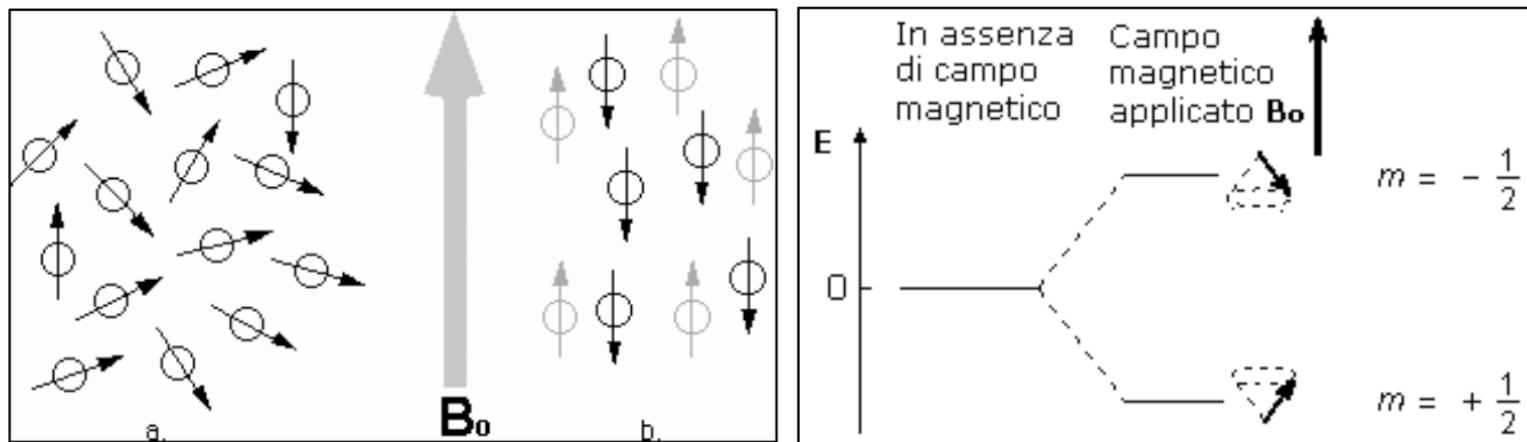
Frequenza di Larmor

$$\nu = \gamma \frac{B_o}{2\pi} \quad (\text{in Hertz})$$

γ = momento giromagnetico.
Dipende dal nucleo in esame.

NMR (Nuclear Magnetic Resonance)

In un campo magnetico esterno, stati di spin diversi hanno energie diverse (minore quello con spin allineato al campo esterno)



All'aumentare del campo, aumenta la frequenza di Larmor e quindi aumenta la differenza di energia tra i due livelli

Lo squilibrio tra i momenti magnetici paralleli e antiparalleli crea una piccola polarizzazione degli spin che da' luogo ad una magnetizzazione macroscopica netta

NMR (Nuclear Magnetic Resonance)

- Il campione viene irradiato con un impulso (oscillante) di radiofrequenza che contiene anche la frequenza di Larmor dei nuclei in esame
- Si verifica un'interazione della componente magnetica della radiazione con i momenti magnetici nucleari
- I nuclei assorbono energia e subiscono una transizione di spin, ossia un cambiamento di orientazione dello spin nucleare
- Gli spin sono in risonanza con la radiazione applicata
- Quando si interrompe l'irradiazione, i nuclei eccitati emettono per qualche istante un segnale di radiofrequenza
- Questo è il segnale che viene raccolto ed amplificato dalla strumentazione

NMR (Nuclear Magnetic Resonance)

Ogni nucleo ha il suo spettro di risonanza unico

Data l'**intensità** del campo magnetico esterno, la **frequenza** di eccitazione ed il **tipo** di nucleo, il segnale emesso da un certo nucleo isolato può essere facilmente identificato dallo spettro NMR

Quando il nucleo fa parte di una molecola (ossia è circondato da altri nuclei), il segnale che esso emette è diverso che se fosse isolato

Questo fenomeno si chiama **chemical shift (spostamento chimico)** ed è dovuto al campo magnetico indotto degli **elettroni** degli atomi adiacenti al nucleo il quale lo scherma dalla piena forza del campo magnetico esterno.

Nuclei **schermati** (de-schermati) sentono un campo magnetico più **debole** (forte), subiscono transizione a frequenze **minori** (maggiori), e hanno spostamenti chimici **minori** (maggiori)

Le condizioni di risonanza si hanno quando si irradia il nucleo con una radiofrequenza

$$v_s = (\gamma/2\pi) B_s,$$

dove B_s è il valore di campo magnetico sentito dal nucleo.

Il *campo sentito* B_s è minore del *campo applicato* B_0 , poichè l'insieme dei nuclei e delle nuvole elettroniche che circondano il nucleo (*intorno chimico*) si comporta come uno schermo magnetico.

$$B_s = B_0 (1 - k),$$

dove k è una costante (*costante di schermo*) che dipende dall'intorno chimico del nucleo.

Le condizioni di risonanza non avvengono per

$$v_0 = \gamma B_0 / 2\pi,$$

ma per

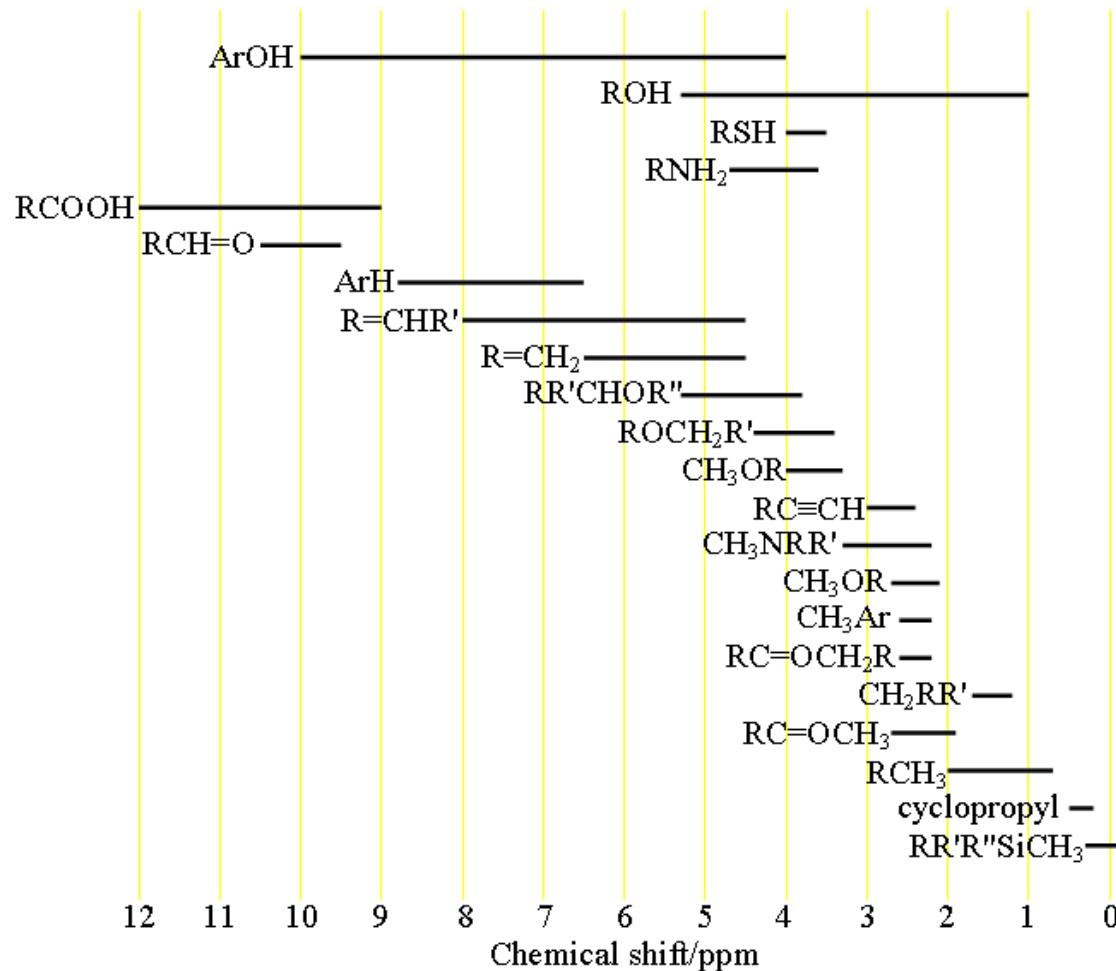
$$v_s = \gamma B_s / 2\pi,$$

dove

$$v_s < v_0 \text{ perch\`e } B_s < B_0.$$

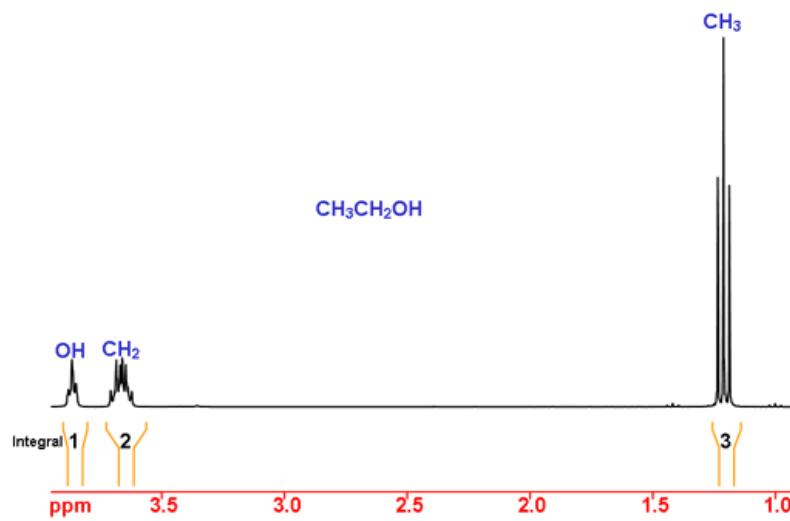
Siccome k è dell'ordine di 10^{-6} , la differenza ($B_0 - B_s$) è dell'ordine dei ppm (parti per milione), così come la differenza $v_0 - v_s$.

Ossia, la frequenza di risonanza di un nucleo è minore di qualche ppm rispetto a quella teorica che si avrebbe non tenendo conto dello schermo magnetico esercitato dall'intorno chimico.

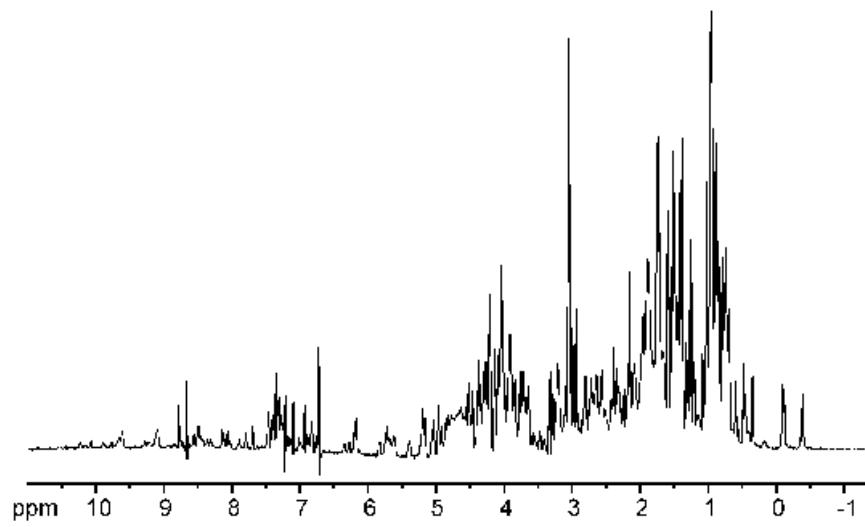


Usando i chemical shift NOTI di nuclei in diversi contesti chimici, è possibile decifrare la struttura di una proteina dal suo spettro NMR

ppm = parti per milione



Spettro NMR dell'etanolo



Spettro NMR di una proteina

La conversione di uno spettro in informazione strutturale si chiama **assegnazione dello spettro**

Analisi di:

- 1) spostamento chimico
 - 2) area dei picchi (integrali, numero di nuclei)
 - 3) accoppiamento (i picchi dello spettro NMR vengono assegnati ai vari nuclei di una proteina trovando l'**accoppiamento** tra nuclei che ne identificano la vicinanza spaziale)
-
- 1) lo **spostamento chimico** ci dice a quale **gruppo funzionale** appartiene il nucleo che produce quel segnale
 - 2) l'**area** ci dice **quanti nuclei** producono quel segnale
 - 3) la **molteplicità** ci dice **quanti nuclei adiacenti** vi sono intorno a quelli che producono il segnale.

$\delta=3,77$: idrogeni vicini all'ossigeno elettronegativo

A=2: un gruppo con 2 idrogeni
forse un CH_2

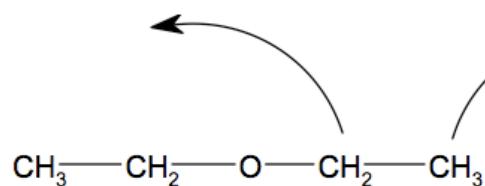
m=4: un gruppo che è vicino a 3 idrogeni ($4=3+1$)

$\delta=0,95$: idrogeni su un normale carbonio primario

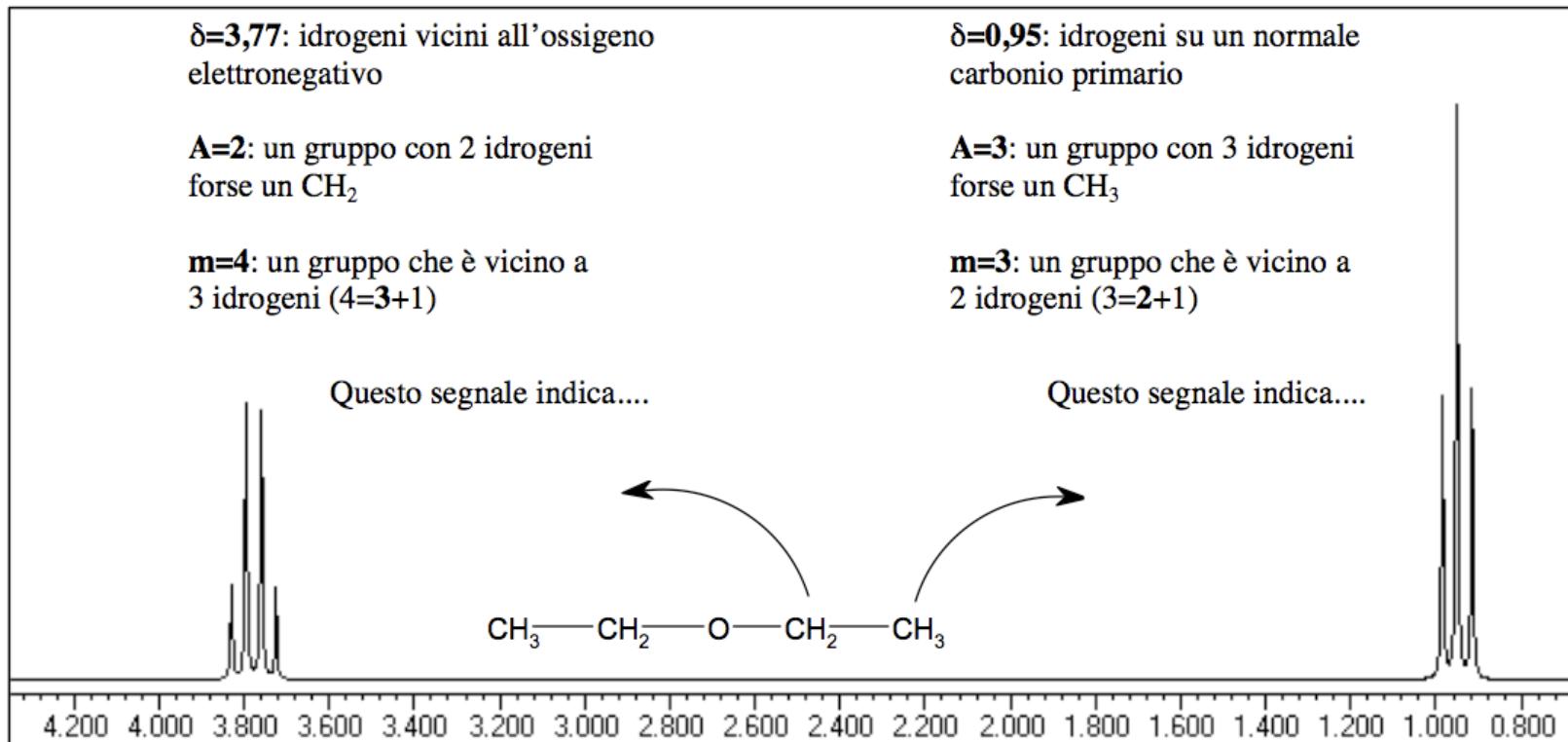
A=3: un gruppo con 3 idrogeni
forse un CH_3

m=3: un gruppo che è vicino a 2 idrogeni ($3=2+1$)

Questo segnale indica....

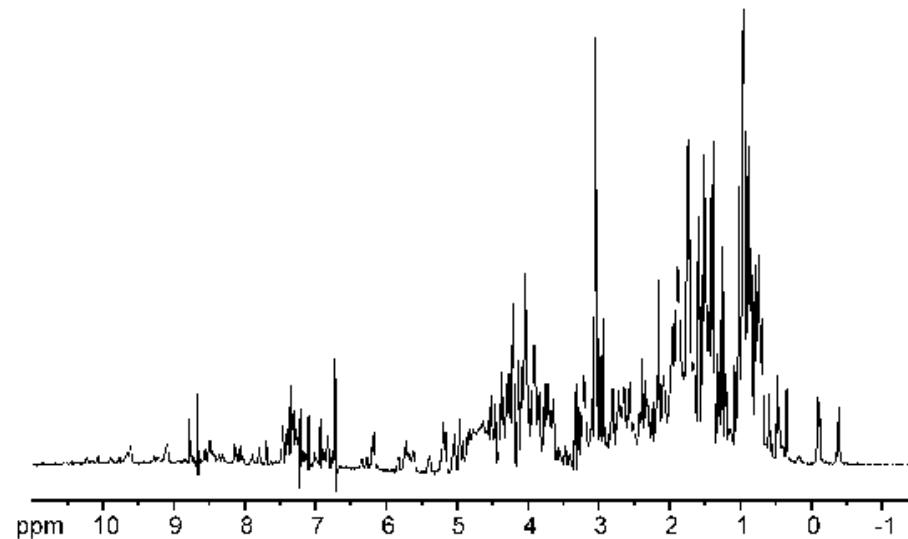


Questo segnale indica....

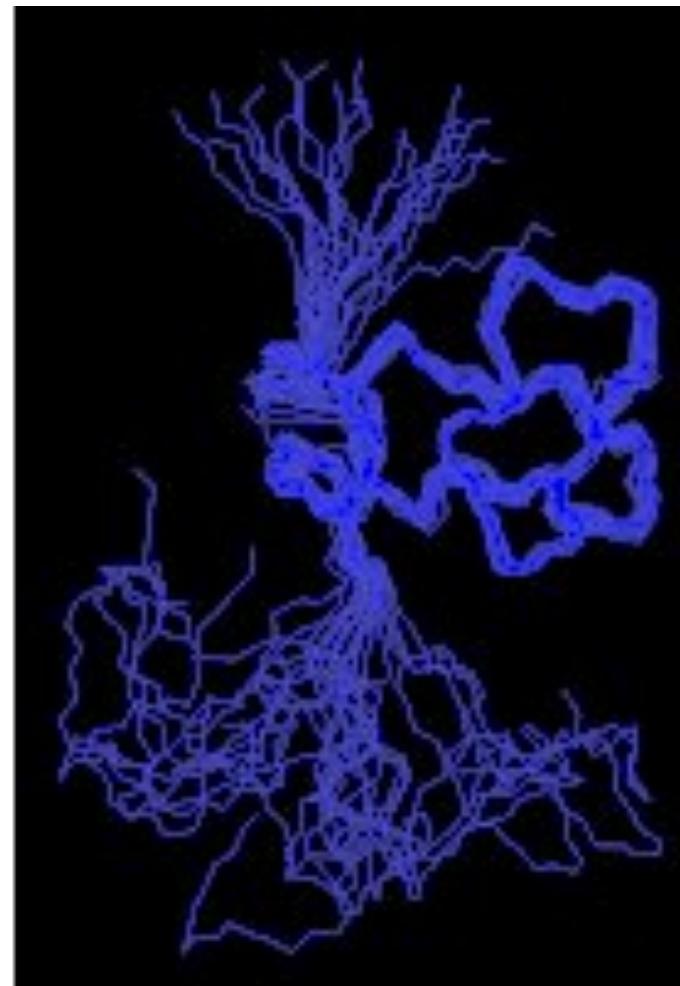


2D-NMR

Uno spettro NMR di una macromolecola può contenere centinaia o migliaia di picchi di risonanza



Per separarli si possono fare misure ripetute in cui le differenze di tempo tra misure consecutive creano delle correlazioni tra nuclei adiacenti





Strumento NMR a 900MHz con magnete di 21.2 Tesla presso l'HWB-NMR
(Birmingham, UK)

Vantaggi e svantaggi del metodo

- Consente di determinare la struttura anche in assenza di cristallo
- Consente di determinare la situazione dinamica della struttura
- Permette una facile analisi di processi di legame (enzima-substrato, enzima-cofattore, legame proteina-ligando, etc.)
- Non danneggia il campione (onde radio)
- NMR dello stato solido consente la determinazione di proteine di membrana e di precipitati amiloidi
- E' limitato a strutture piccole (≤ 40 kDa)
- Richiede più interventi manuali rispetto ad altri metodi