

Ottimizzazione del modello molecolare

Laurea Magistrale in Biotecnologie Mediche
Curriculum Nanobiotecnologie

A.A. 2020-21

Il modello strutturale

La risoluzione del problema della fase unito all'applicazione dei metodi di density modification fornisce una mappa di densità elettronica $\rho(r)$ intellegibile.

Inoltre l'utilizzo di procedure di model building automatico può fornire un primo modello molecolare, talvolta parziale, coerente con la densità elettronica osservata.

Il passo successivo consiste nello scegliere la tipologia di modello più opportuno e quindi **migliorare e completare il modello strutturale, coerentemente con i dati sperimentali** (fattori di struttura e densità elettronica)

Il modello strutturale (molecolare) viene migliorato cercando di aumentare l'accordo tra dati sperimentali (fattori di struttura osservati: F_o) e dati calcolati (Fattori di struttura calcolati: F_c) a partire dal modello strutturale.

Più dati sperimentali abbiamo (fattori di struttura F_o) e più preciso sarà il processo di ottimizzazione del modello stesso.

Ottimizzazione del modello strutturale

$$F_c = \sum_{j=1}^N f_j(r^*) \exp \left\{ [2\pi i(hx_j - ky_j - lz_j)] - B_j \left(\frac{\sin\theta}{\lambda} \right)^2 \right\}$$

Scegliere la tipologia di modello più opportuno significa scegliere il modello fisico che meglio possa descrivere il nostro sistema molecolare compatibilmente con la quantità e qualità dei dati acquisiti (essenzialmente con la risoluzione)

Più è grande la quantità di dati osservati (numero di Fo) più il nostro modello strutturale potrà essere dettagliato.

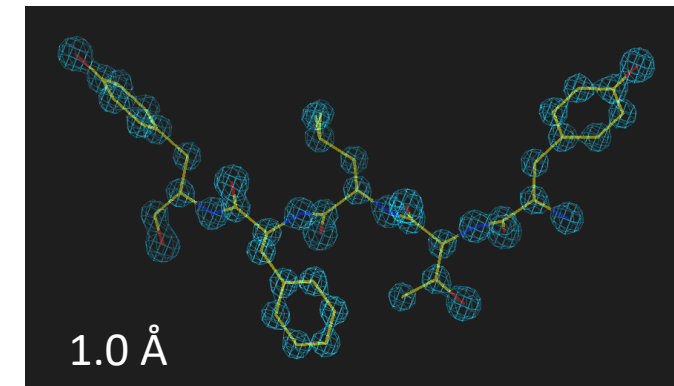
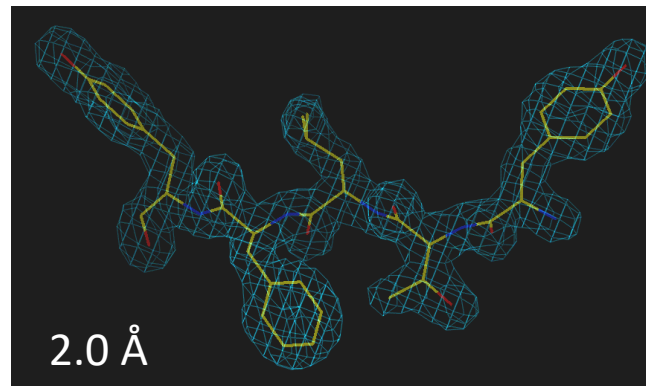
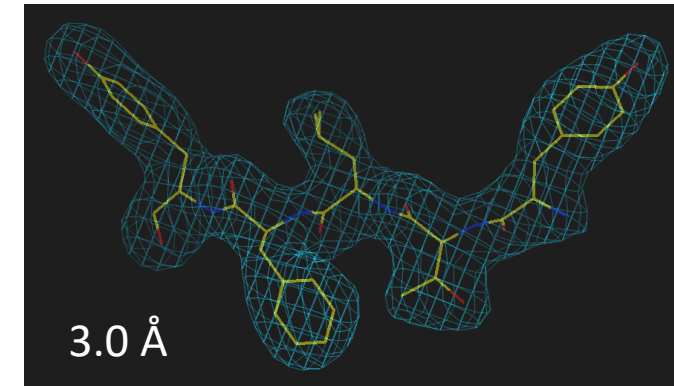
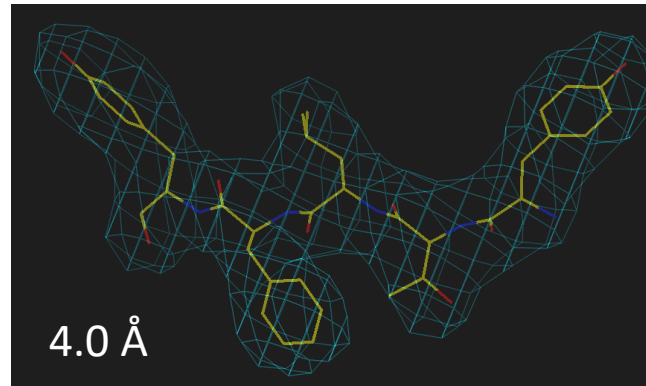
Risoluzione cristallografica e modello strutturale

Risoluzione e informazione strutturale

La risoluzione cristallografica della struttura, ovvero il valore minimo di d_{hkl} , determina l'entità di informazione strutturale ottenibile dai nostri dati sperimentali. **Più la risoluzione cristallografica è alta, più è grande il numero dei fattori di struttura e maggiore è l'informazione strutturale contenuta nei dati di diffrazione.**

La risoluzione cristallografica della struttura, in ultima analisi, determina il modello molecolare più adatto per il completamento del modello strutturale.

Più è elevata la risoluzione cristallografica della struttura e più complesso potrà essere il modello strutturale utilizzabile.



Risoluzione e dettaglio della struttura

Protein Feature	Resolution (Å)
Helix	9
Sheet	4
Main chain	3.7
Aromatic sidechains	3.5
Small sidechains	3.2
Sidechain conformations	2.9
Carbonyl, peptide	2.7
Ordered waters	2.7
Central dimple of aromatic ring	2.4
Correct stereochemistry at Ile CB	2.2
Proline pucker	2.0
Individual atoms	1.5

Nucleic Acid Feature	Resolution (Å)
Double helix	20
Single strand	12
Stacked base pairs	4
Phosphates	3.5
Purine or pyrimidine?	3.2
Individual bases	2.7
Ribose pucker	2.4
Individual atoms	1.5

Orbitali e legami (oltre 1Å)!

Risoluzione e quantità di informazione

A prescindere dalla qualità della misura è chiaro che

umentando la risoluzione (d_{hkl} più piccolo), il numero di fattori di struttura disponibili aumenta:

ad una risoluzione pari a 2.30 Å ho più del doppio dei fattori di struttura che ho ad una risoluzione di 4.86 Å.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION													
RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano	
6.83	23492	2440	2449	99.6%	3.4%	3.3%	23492	60.39	3.6%	99.9*	29*	1.097	1076
4.86	44542	4382	4382	100.0%	4.3%	4.1%	44542	47.29	4.5%	99.9*	17*	1.001	2044
3.97	57703	5635	5635	100.0%	4.4%	4.3%	57703	46.09	4.6%	99.9*	1	0.835	2668
3.45	69584	6675	6675	100.0%	6.7%	6.4%	69584	31.48	7.0%	99.9*	-2	0.826	3191
3.08	77663	7550	7550	100.0%	11.9%	11.8%	77663	18.08	12.5%	99.7*	0	0.812	3628
2.82	87215	8336	8336	100.0%	24.9%	26.3%	87215	9.05	26.2%	98.5*	-2	0.757	4022
2.61	93311	9094	9096	100.0%	47.3%	50.7%	93311	4.82	49.8%	94.4*	0	0.734	4398
2.44	102507	9770	9770	100.0%	78.5%	83.9%	102507	2.93	82.6%	85.8*	-1	0.727	4737
2.30	105417	10314	10331	99.8%	126.3%	139.3%	105391	1.73	133.0%	69.7*	0	0.689	4989
total	661434	64196	64224	100.0%	9.1%	9.3%	661408	17.55	9.6%	99.9*	1	0.787	30753

Poiché la procedura di ottimizzazione della struttura non è altro che un fit del modello molecolare contro i dati sperimentali, **più elevato** è il numero di dati osservati e **più sofisticato** potrà essere il modello utilizzato per il fit.

Modelli che contemplano molte variabili non sono compatibili con dati di diffrazione a bassa risoluzione, al contrario dati ad altissima risoluzione permettono l'utilizzo di modelli molto sofisticati

Risoluzione e dati: Calmodulina

Calmodulina (gallo)

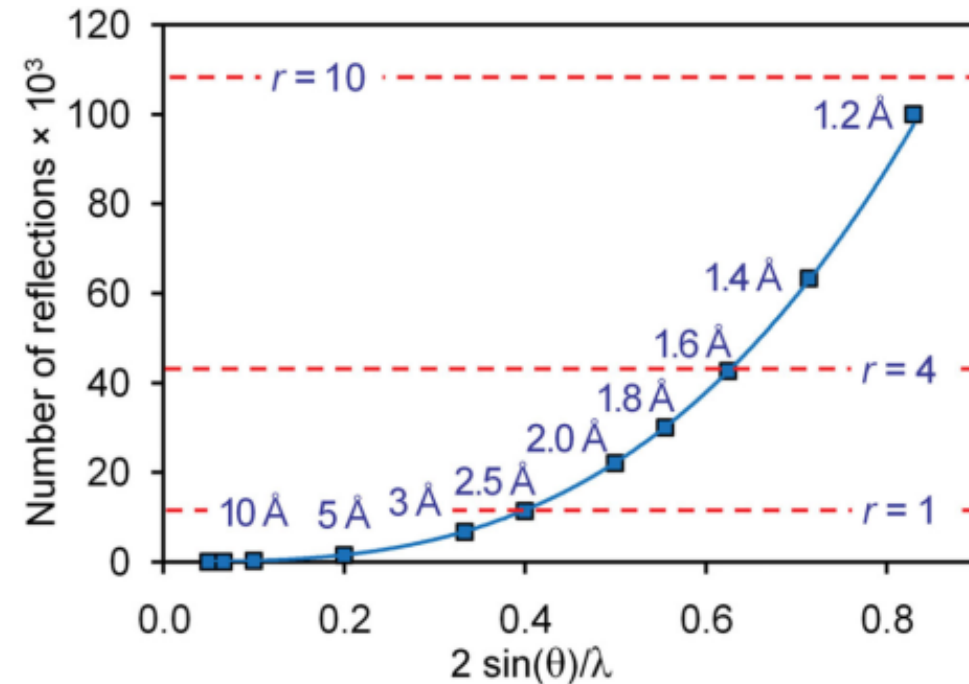
PDB: 1UP5

Gruppo Spaziale: P1

Cella Unitaria: $a = 59.7$, $b = 53.1$, $c = 24.6$ (Å)
 $\alpha = 93.2$, $\beta = 96.7$, $\gamma = 89.2$ (°)

Numero atomi totali ottimizzati: 2567

Numero parametri ottimizzati: circa 10200



r indica il rapporto tra dati osservati e parametri ottimizzati

Risoluzione e qualità dell'informazione

La qualità dei dati è indirettamente legata alla risoluzione massima.

Due set di dati diversi possono avere qualità molto diversa, alla medesima risoluzione.

All'intero di uno stesso set di dati (la totalità dei fattori di struttura acquisiti) i dati a più alta risoluzione sono più deboli e quindi affetti da un errore maggiore.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION													
RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano	
6.83	23492	2440	2449	99.6%	3.4%	3.3%	23492	60.39	3.6%	99.9*	29*	1.097	1076
4.86	44542	4382	4382	100.0%	4.3%	4.1%	44542	47.29	4.5%	99.9*	17*	1.001	2044
3.97	57703	5635	5635	100.0%	4.4%	4.3%	57703	46.09	4.6%	99.9*	1	0.835	2668
3.45	69584	6675	6675	100.0%	6.7%	6.4%	69584	31.48	7.0%	99.9*	-2	0.826	3191
3.08	77663	7550	7550	100.0%	11.9%	11.8%	77663	18.08	12.5%	99.7*	0	0.812	3628
2.82	87215	8336	8336	100.0%	24.9%	26.3%	87215	9.05	26.2%	98.5*	-2	0.757	4022
2.61	93311	9094	9096	100.0%	47.3%	50.7%	93311	4.82	49.8%	94.4*	0	0.734	4398
2.44	102507	9770	9770	100.0%	78.5%	83.9%	102507	2.93	82.6%	85.8*	-1	0.727	4737
2.30	105417	10314	10331	99.8%	126.3%	139.3%	105391	1.73	133.0%	69.7*	0	0.689	4989
total	661434	64196	64224	100.0%	9.1%	9.3%	661408	17.55	9.6%	99.9*	1	0.787	30753

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION													
RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano	
3.74	26347	7170	7248	98.9%	3.6%	3.8%	26339	33.14	4.2%	99.7*	55*	1.057	3430
2.65	39902	12580	13057	96.3%	3.8%	3.9%	39898	27.45	4.6%	99.7*	63*	1.435	6022
2.16	50356	16433	16875	97.4%	4.9%	4.4%	50023	21.65	6.0%	99.5*	58*	1.451	7636
1.87	65790	19665	20023	98.2%	6.8%	6.1%	65254	16.06	8.1%	99.4*	47*	1.225	9236
1.68	77979	22538	22722	99.2%	11.3%	11.0%	77290	9.47	13.3%	98.8*	27*	0.893	10530
1.53	74738	24732	25105	98.5%	20.8%	20.7%	74062	5.00	25.4%	96.2*	2	0.710	11545
1.42	88088	26983	27149	99.4%	33.1%	38.4%	87452	2.96	39.7%	91.9*	-7	0.567	12784
1.33	97942	29056	29284	99.2%	46.6%	60.6%	96970	1.89	55.2%	86.9*	-9	0.496	13653
1.25	102145	30550	31260	97.7%	71.4%	97.2%	100972	1.17	84.7%	80.0*	-11	0.459	14253
total	623287	189707	192723	98.4%	5.7%	5.9%	618260	9.29	6.7%	99.8*	29*	0.818	89089

L'ottimizzazione del modello molecolare

$$F_o = |F_o| \exp(i\varphi_c)$$

$$F_c = \sum_{j=1}^N f_j(r^*) \exp[2\pi i(hx_j + ky_j + lz_j)]$$

L'ottimizzazione del modello molecolare si traduce nell'individuare i parametri che permettono il migliore accordo tra il dato sperimentale (F_o) e il dato calcolato (F_c)

Il calcolo del fattore di struttura F_c prevede la conoscenza di

- **Coordinate**
- **Fattori termici**

Per **TUTTI** gli atomi contenuti nell'unità asimmetrica

Il modello molecolare: coordinate

Modelli per le coordinate

$$F_c = \sum_{j=1}^N f_j(\mathbf{r}^*) \exp[2\pi i(\mathbf{h}\mathbf{x}_j + \mathbf{k}\mathbf{y}_j + \mathbf{l}\mathbf{z}_j)]$$

Per una data molecola si possono scegliere modelli per le coordinate diversi:

Il modello più intuitivo e più usato è quello che associa ad ogni atomo 3 variabili posizionali indipendenti ovvero le 3 coordinate x, y, z.

In genere con una risoluzione migliore di 3.2-3.5 Å si utilizza questo modello per descrivere le coordinate.

Si possono fare scelte diverse in relazione alla quantità di dati sperimentali a disposizione (risoluzione).

Modelli a corpo rigido

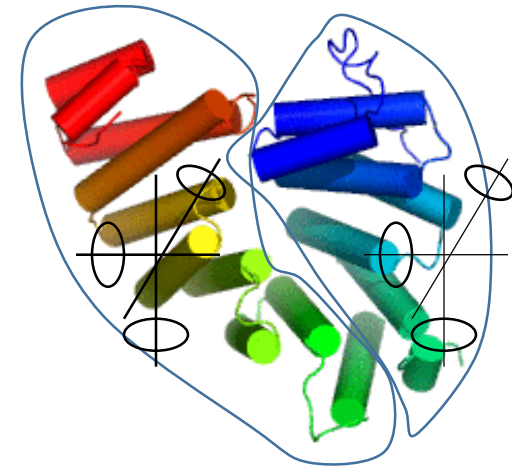
Nel caso in cui il numero di dati osservati sia ridotto (bassa risoluzione), si possono utilizzare modelli semplificati per i parametri posizionali:

- Associare **3 coordinate (baricentro) all'intero aminoacido**
- Associare **3 coordinate al gruppo $\text{NC}\alpha\text{C}(=\text{O})$ e 3 coordinate alla catena laterale**

Un modello che utilizza un numero di variabili decisamente inferiore è quello del **corpo rigido**: un certo numero di atomi (per esempio un dominio strutturale) sono considerati rigidamente collegati, quindi ho solo 6 variabili indipendenti (3 rotazioni e 3 traslazioni) per tutti gli N atomi costituenti il dominio, che viene 'spostato' rigidamente.

Il modello a corpo rigido è utilizzato solo nelle fasi preliminari dell'ottimizzazione, specialmente se le fasi sono state determinate con il metodo del Molecular Replacement

Questo modello è anche utilizzato in fase di ottimizzazione se ho dati sperimentali a risoluzione molto bassa (4 Å o peggiore)



Modelli per il moto termico

Modelli per il moto termico

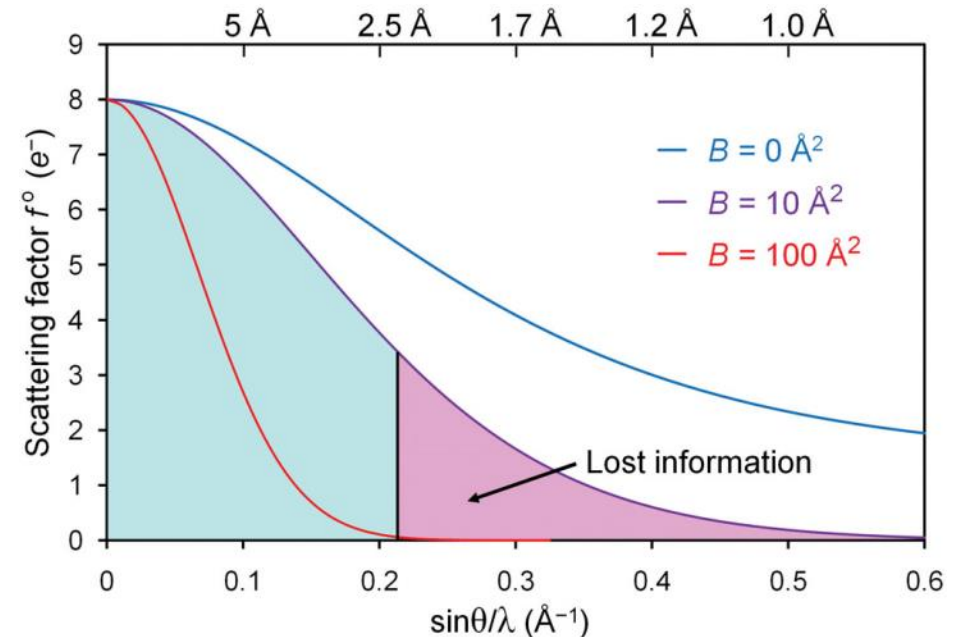
$$F_c = \sum_{j=1}^N f_j(r^*) \exp \left\{ [2\pi i(hx_j - ky_j - lz_j)] - B_j \left(\frac{\sin\theta}{\lambda} \right)^2 \right\}$$

Gli atomi si muovono nel cristallo e il moto termico influenza la risoluzione poiché causa un'attenuazione delle intensità diffratte.

Il moto termico viene modellato come un termine di 'smorzamento' denominato **fattore termico**

I fattori termici sono tra i parametri da modellare nel corso del procedimento di ottimizzazione del modello.

Come per le coordinate si possono fare varie scelte, più o meno accurate, a seconda della risoluzione della struttura ovvero del numero di dati osservati.



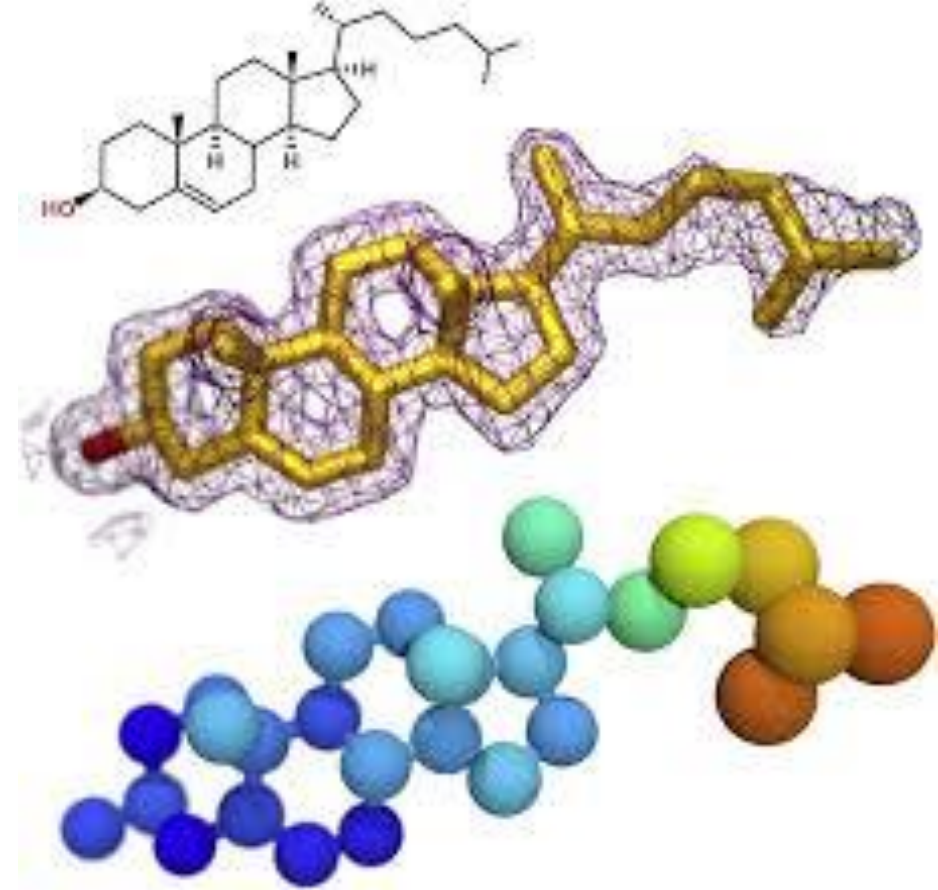
© Garland Science 2010

Modelli per il moto termico isotropo

Anche per il moto termico si possono utilizzare modelli diversi:

Moto termico uguale in tutte le direzioni (**isotropo**), viene utilizzato un solo fattore termico (**B**, **fattore di Debye-Waller**). Lo si utilizza per risoluzione medio-alta (3.0 – 1.4 Å).

E' il modello più utilizzato e associa un parametro termico ad ogni atomo

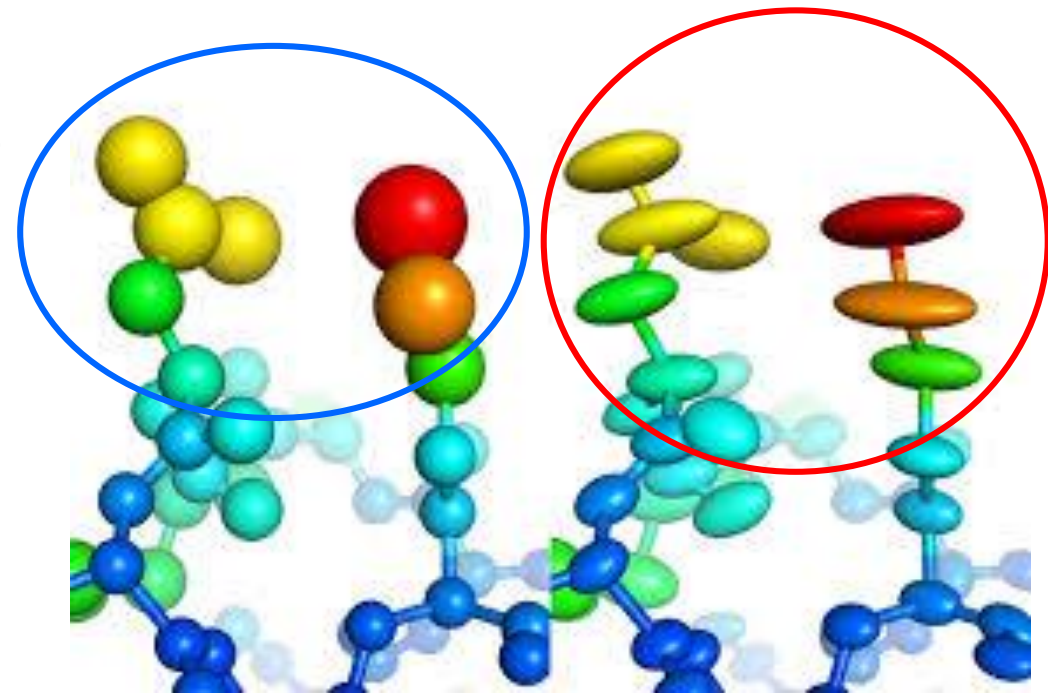


Modelli per il moto termico anisotropo

Moto termico anisotropo, diverso nelle tre direzioni, è descritto da un ellissoide (6 parametri) e non da una sfera (1 parametro).

E' Utilizzato ad alta risoluzione (< 1.4 Å).

Per risoluzioni molto basse (< 3.5 Å) si utilizzano modelli di moto termico più 'semplici' come per esempio un singolo fattore termico per tutti gli atomi della catena laterale di un aminoacido e un fattore termico per gli atomi della catena principale dello'amminoacido.



Moto Termico
Isotropo

Moto Termico
Anisotropo

Procedura di raffinamento (minimi quadrati)

Approccio generale

Una volta determinata le fasi iniziali dei fattori di struttura, eventualmente migliorate per mezzo della density modification, possiamo calcolare i fattori di struttura (F_{calc}) e la densità $\rho_{obs}(r)$.

$$F_{calc}(h, k, l) = \sum_{j=1}^N f_j \exp \left\{ [2\pi i (hx_j - ky_j - lz_j)] - B_j \left(\frac{\sin \theta}{\lambda} \right)^2 \right\}$$

Le quantità calcolate sono delle funzioni dei parametri molecolari (x, y, z e B per ogni atomo) e devono essere in accordo con i dati sperimentali ($|F_{obs}|$ o più semplicemente $|F_o|$).

$$\rho_{obs}(r) = \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} |F_{obs}| \exp[-2\pi i (hx + ky + lz) - i\varphi_{calc}]$$

La fase di ottimizzazione del modello, definita come (r)**affinamento della struttura (*structural refinement*)**, consiste nel **determinare i parametri che massimizzano l'accordo tra dato sperimentale e dato calcolato**.

Nel corso del processo di affinamento strutturale, in genere, **il modello viene modificato** e migliorato, in modo opportuno, affinché l'accordo tra dati osservati e calcolati sia ottimale.

Spazio reciproco o Spazio reale

$$F_{calc}(h, k, l) = \sum_{j=1}^N f_j \exp \left\{ [2\pi i (hx_j - ky_j - lz_j)] - B_j \left(\frac{\sin\theta}{\lambda} \right)^2 \right\}$$
$$\rho_{obs}(r) = \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} |F_o| \exp[-2\pi i (hx + ky + lz) - i\varphi_{calc}]$$

L'affinamento strutturale può essere condotto:

- Nello spazio reciproco, si cerca di migliorare l'accordo tra $|F_o|$ e $|F_{calc}|$ (indicato come $|F_c|$)
- Nello spazio reale, migliorando $\rho_{obs}(r)$

In generale, nell'affinamento cristallografico, si utilizza l'approccio nello spazio reciproco, viene quindi cercato il migliore accordo tra $|F_o|$ e $|F_c|$

Nella pratica dell'affinamento cristallografico, una certa parte dell'affinamento è comunque svolto nello spazio reale utilizzando la **grafica molecolare** per migliorare delle parti del modello.

Nell'ottimizzazione dei modelli in Cryo-EM si opera generalmente nello spazio reale.

Il metodo dei minimi quadrati

Si tratta di minimizzare la funzione cercando i parametri migliori x_p che diano il valor minimo della funzione $M(x_p)$

$$M(x_1, x_2, \dots, x_p) = \sum_{i=1}^N [y_i^o - y_i^c(x_1, x_2, \dots, x_p)]^2$$

Cosa che viene fatta derivando rispetto ai parametri x_p

$$\left\{ \begin{array}{l} \frac{\partial M(x_1, \dots, x_p)}{\partial x_1} \\ \dots \\ \frac{\partial M(x_1, \dots, x_p)}{\partial x_p} \end{array} \right.$$

La soluzione del sistema di equazioni fornisce i valori di x_p che massimizzano l'accordo tra il dato osservato e quello calcolato.

Il metodo dei minimi quadrati in cristallografia

$$Q_{xray} = \sum_{i=1}^N [|F_{o,i}| - k|F_{c,i}(x_j, y_j, z_j, B_j)|]^2$$

Nel corso dell'affinamento cristallografico viene minimizzata la differenza (elevata al quadrato) tra i fattori di struttura osservati e i fattori di struttura calcolati a partire dal modello disponibile (coordinate, fattori termici).

La funzione che esprime il fattore di struttura calcolato **non è lineare**, di conseguenza **ciò che viene determinata è la variazione che devo apportare ai valori iniziali dei parametri per ottenere il valore minimo della funzione minimizzata.**

$$\frac{\partial Q_{xray}(x_1, x_2, \dots, x_p)_0}{\partial x_p} \quad \longrightarrow \quad \Delta x_1, \Delta x_2, \dots, \Delta x_p$$

I nuovi valori dei parametri, dopo la minimizzazione, saranno:

$$(x_1)_0 + \Delta x_1, (x_2)_0 + \Delta x_2, \dots, (x_p)_0 + \Delta x_p$$

Il fattore R (*R-Factor*)

La minimizzazione viene eseguita in modo iterativo per un certo numero di volte fino a che i valori trovati di Δx_p , sono sufficientemente piccoli (condizione di convergenza).

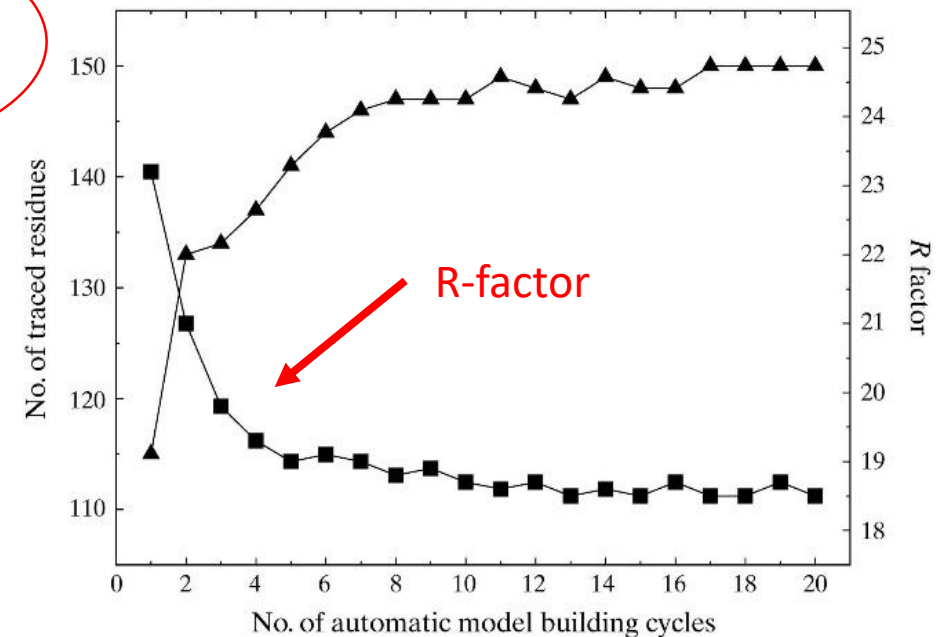
Il buon andamento della minimizzazione è verificato controllando **il fattore R (*R-factor*)**

Differenza tra valori
osservati e calcolati

$$R = \frac{\sum_{hkl} |F_{\text{obs}} - F_{\text{calc}}|}{\sum_{hkl} |F_{\text{obs}}|}$$

Se il modello è migliorato, R deve diminuire!

Il modello si potrà dire definitivo quando il fattore R non è ulteriormente migliorabile.



Minimi quadrati con vincoli (restraints)

Limiti dei minimi quadrati

I minimi quadrati sono un metodo efficiente per determinare i migliori parametri cristallografici (coordinate, fattori termici).

Attualmente sono utilizzati metodi di natura più generale e complessa (metodi basati sulla Maximum Likelihood) di cui i minimi quadrati sono un caso particolare. Concettualmente non cambia molto.

I minimi quadrati sono facilmente applicabili all'affinamento cristallografico, soffrono però di due limiti propri del metodo:

- Il rapporto tra dati osservati e parametri da 'affinare'
- Incapacità di determinare il minimo assoluto della funzione minimizzata

Minimi quadrati e dati osservati

Perché il metodo dei minimi quadrati non-lineari possa dare risultati attendibili **il rapporto tra il numero di dati osservati (Fo) e il numero di parametri da minimizzare, deve essere di 5-10**. Inoltre, se il rapporto osservate/parametri è inferiore a 1 il metodo non è applicabile (sistema non-determinato)
Nel caso delle macromolecole biologiche, ho molti atomi e di conseguenza molti parametri (x, y, z, B per ogni atomo del modello).

Ho bisogno di molti dati, non sempre disponibili alla risoluzione cristallografica ottenibile dal cristallo! (scegliere il modello opportuno!)

Calmodulina (PDB_id: 1up5)

Gruppo spaziale: P1

Cella: a = 59.696, b = 53.073, c = 24.572

$\alpha = 93.22$, $\beta = 96.86$, $\gamma = 89.19$

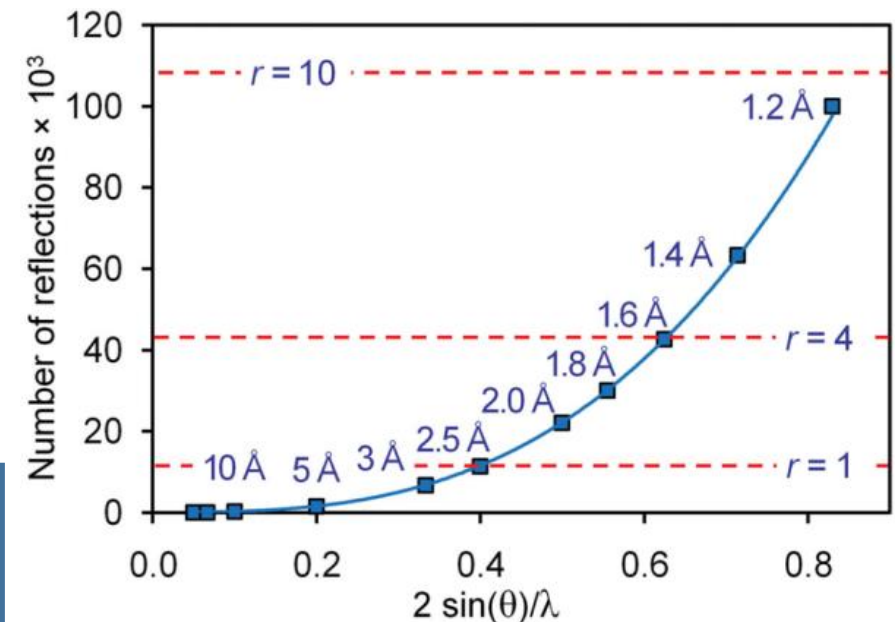
Risoluzione: 1.9 Å

N° atomi: 2567

N° parametri: **10108** (x, y, z, B)

N° F_{obs} : **23558**

$$\frac{N^{\circ} F_{obs}}{N^{\circ} parametri} = 2.34$$



© Garland Science 2010

Minimi Quadrati con restraints

Nella grande maggioranza dei casi, ovvero per le risoluzioni più comunemente osservate ($< 1.5 \text{ \AA}$), il numero di fattori di struttura non è sufficiente ad assicurare un rapporto osservate/parametri adeguato ad un corretto funzionamento del metodo dei Minimi Quadrati non-lineari.

In Assenza di un corretto rapporto tra osservate e parametri, si può incorrere in risultati sbagliati e non realistici.

Il problema del basso rapporto osservate/parametri può essere risolto introducendo dei vincoli 'fisici' nella funzione da minimizzare (restraints).

In sostanza i restraints si comportano come dati 'sperimentali' che noi aggiungiamo alla nostra funzione.

Restraints geometrici e stereochimici

Nel caso delle strutture macromolecolari noi abbiamo informazioni sulla geometria e sulla stereochimica della molecola, ovvero sappiamo quali valori dovremmo aspettarci per alcuni parametri geometrici/stereochimici come (non è un elenco esaustivo...) :

- Distanze di legame
- Angoli di legame
- Angoli di Torsione
- Planarità di anelli (tirosine e fenilalanine)
- Chiralità degli aminoacidi

In sostanza per ognuno dei legami o angoli di legame (e angoli di torsione, etc...) potremo scrivere una relazione del tipo:

$$(d_{obs} - d_{calc})^2$$

d_{obs} è noto dalla letteratura
 d_{calc} è il valore nel mio modello

Posso così scrivere una funzione da minimizzare del tipo:

$$Q_{geom} = \sum (d_{obs} - d_{calc})^2$$

Minimi quadrati con restraints geometrici

Minimizzerò la funzione:

$$Q_{total} = w_{xray} Q_{xray} + Q_{geom} = w_{xray} \sum_{\mathbf{h}} \frac{(F_{\mathbf{h}}^{obs} - F_{\mathbf{h}}^{calc})^2}{(\sigma_{\mathbf{h}}^{obs})^2} + \sum_{i=1}^{restraints} \frac{(r_i^{obs} - r_i^{calc})^2}{(\sigma_{r(i)}^{obs})^2}$$

$$Q_{total} = w_a Q_{xray} + Q_{geom}$$

with

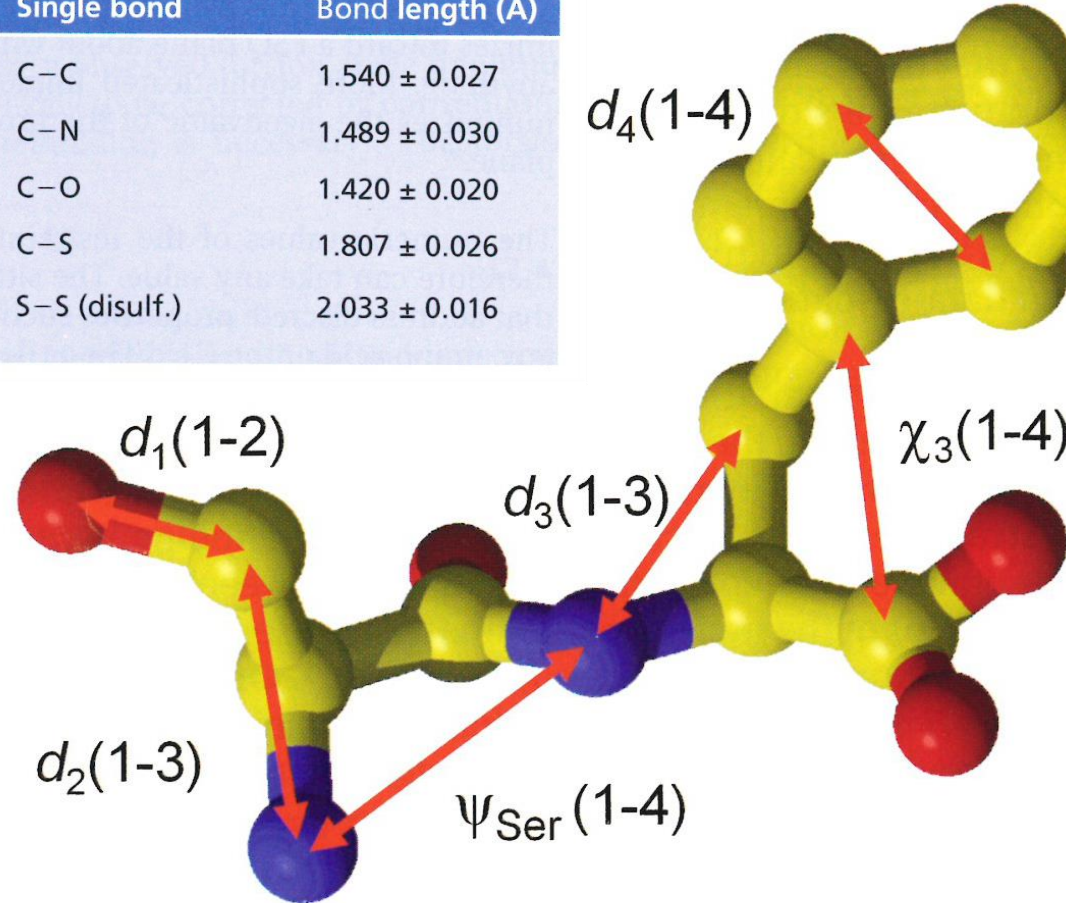
$$Q_{geom} = Q_D + Q_P + Q_{CHIV} + Q_T + Q_{NB} + Q_B + Q_{NCS}$$

Q_{geom} è la somma di tanti restraints, di natura geometrica e/o stereochimica

Restraints geometrici: distanze e angoli

Peptide bond	Bond length (Å)	Single bond	Bond length (Å)
C α -C	1.525 \pm 0.026	C-C	1.540 \pm 0.027
C-N	1.336 \pm 0.023	C-N	1.489 \pm 0.030
N-C α	1.459 \pm 0.020	C-O	1.420 \pm 0.020
C=O	1.229 \pm 0.019	C-S	1.807 \pm 0.026
C α -C β	1.530 \pm 0.020	S-S (disulf.)	2.033 \pm 0.016

Per gli aminoacidi i parametri geometrici e stereochimici sono quantità note!

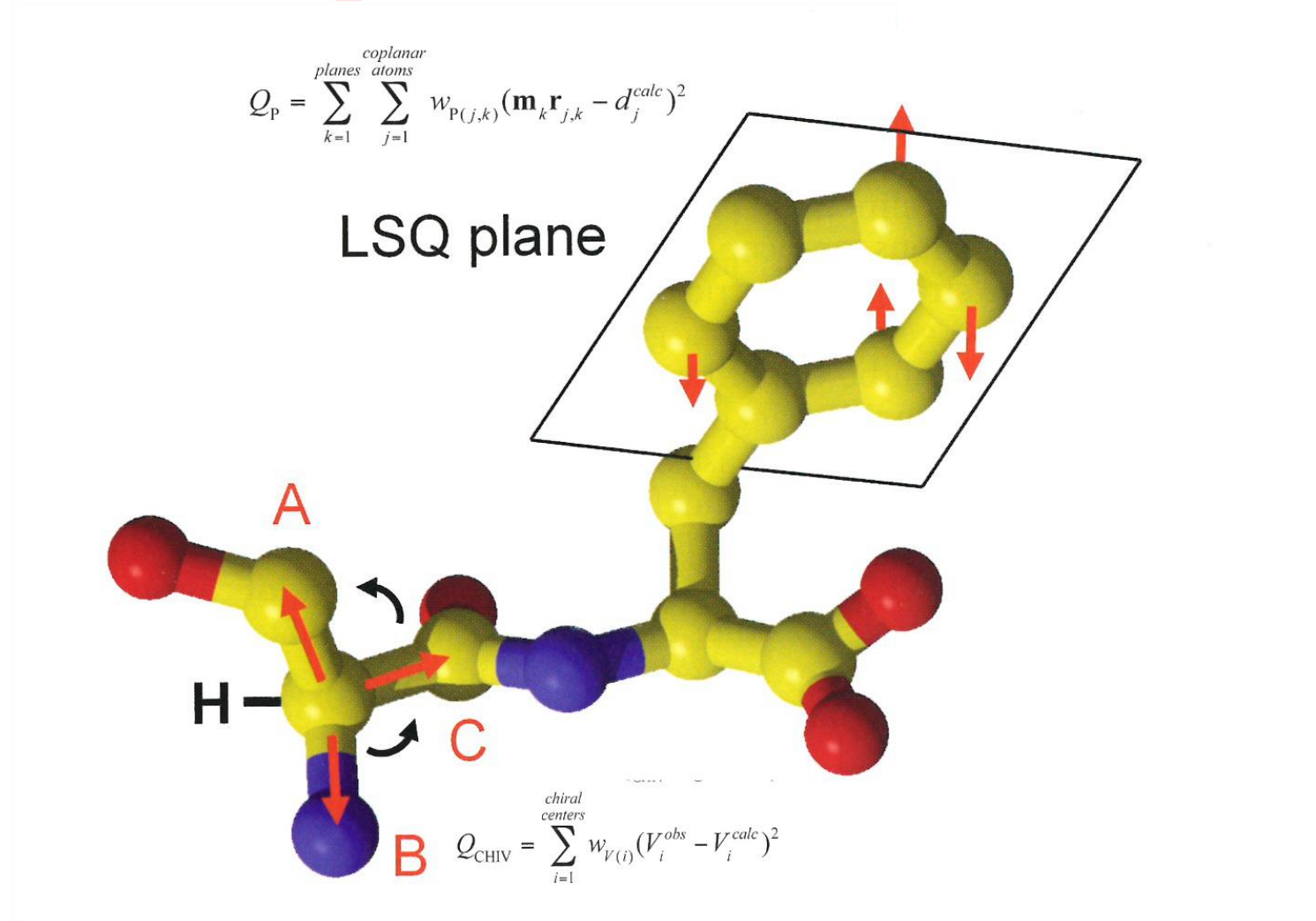


$$Q_D = \sum_{j=1}^{dist} w_{d(j)} (d_j^{obs} - d_j^{calc})^2$$

$$Q_T = \sum_{t=1}^{torsions} w_{\chi(t)} (\chi_t^{obs} - \chi_t^{calc})^2$$

...

Restraints geometrici: Planarità e chiralità



Restraints come indicatori di qualità

Il numero di parametri geometrici da utilizzare come 'dati osservati' è in generale elevato e comunque sufficiente per una corretta applicazione del metodo dei minimi quadrati.

Laddove la risoluzione lo consenta, è possibile non utilizzare i restraints, ottenendo una struttura priva di ogni vincolo pre-definito e quindi più **accurata**. Questo accade per risoluzione molto elevata (< 1.3 Å)

In genere, oltre al fattore R, che rende conto dell'accordo tra fattori di struttura calcolati e osservati, vengono anche riportate le deviazioni quadratiche medie per i vari restraints geometrici. Devono essere piccoli.

Rmsd(bond distance) < 0.1 Å

Rmsd (bond angle) < 2°

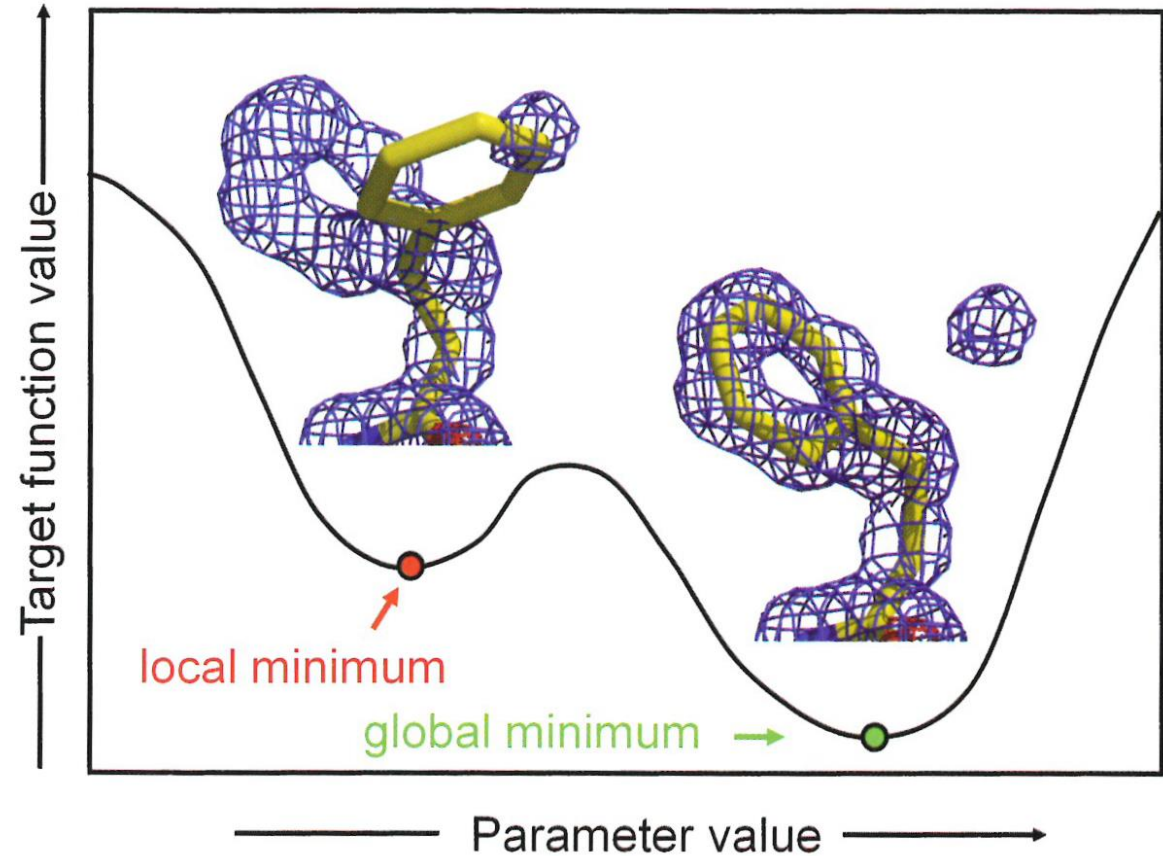
Grafica Molecolare

Convergenza dei minimi quadrati

Il metodo dei minimi quadrati funziona molto bene nel determinare il minimo locale dei parametri da affinare.

Il minimo locale non sempre coincide con il minimo assoluto della funzione!

Per ovviare a questo problema si fa ricorso sia a **protocolli manuali** (*grafica molecolare*), che ad **algoritmi di tipo numerico** (*Rigid-Body, Simulated Annealing*)

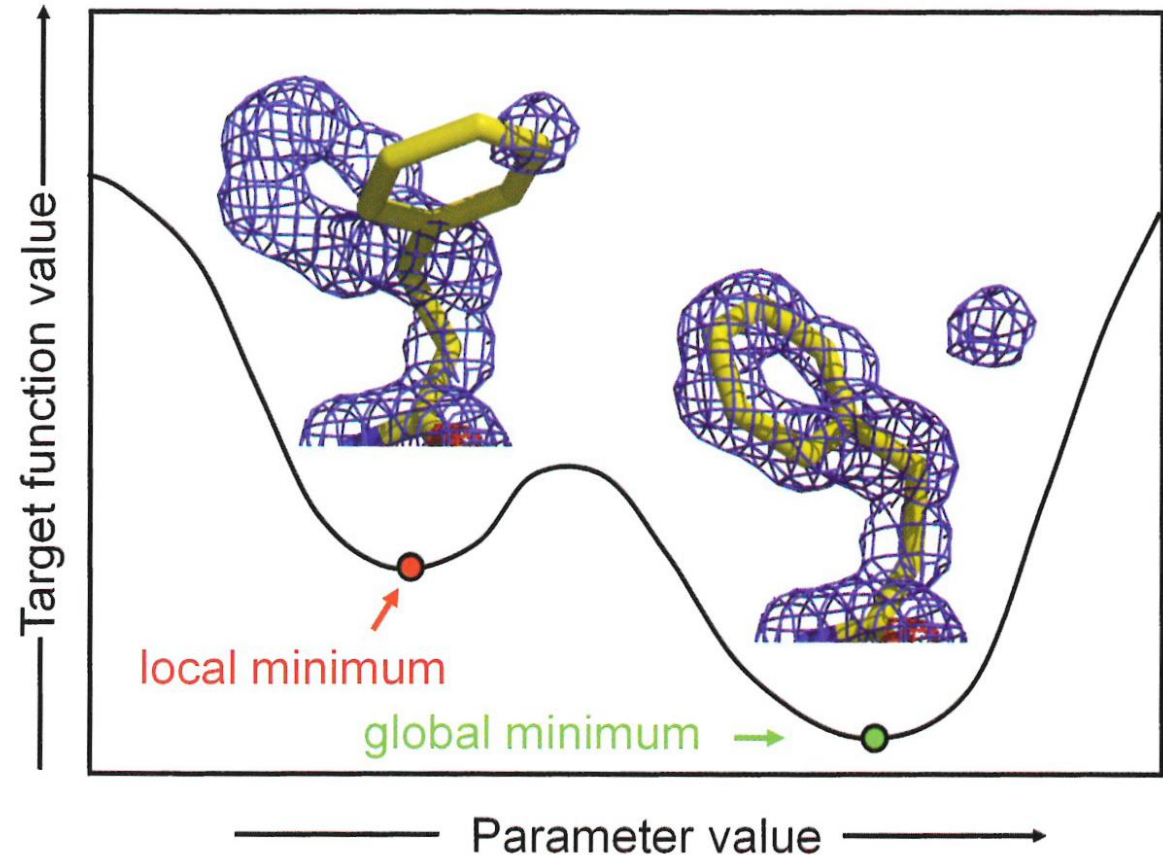


Grafica Molecolare

L'idea di base è quella di osservare, con un sistema di grafica computerizzata, sia il modello che la mappa di densità elettronica $\rho(r)$, tra loro sovrapposti.

L'eventuale disaccordo tra mappa e modello, non risolto dai minimi quadrati per via del limitato *raggio di convergenza*, può essere **corretto** 'ruotando' e/o 'spostando' parti di molecola, 'aggiustando' così il modello.

(nel caso in figura basta ruotare la catena laterale della fenilalanina).



Mappe di differenza

Per semplificare le cose e rendere più facile l'individuazione delle parti del modello problematiche, vengono calcolate e comparate con il modello due tipi particolari di mappe:

Mappa Fo-Fc (detta **mappa differenza**: ΔF)

$$\frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} (F_{obs} - F_{calc}) \exp[-2\pi i (hx + ky + lz)]$$

Mappa 2Fo-Fc

$$\frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} (2F_{obs} - F_{calc}) \exp[-2\pi i (hx + ky + lz)]$$

La mappa ΔF indica gli errori presenti nel modello, sarà positiva laddove il modello richiede 'più materia' (atomi mancanti), sarà invece negativa dove c'è troppa materia (atomi posizionati nel posto sbagliato).

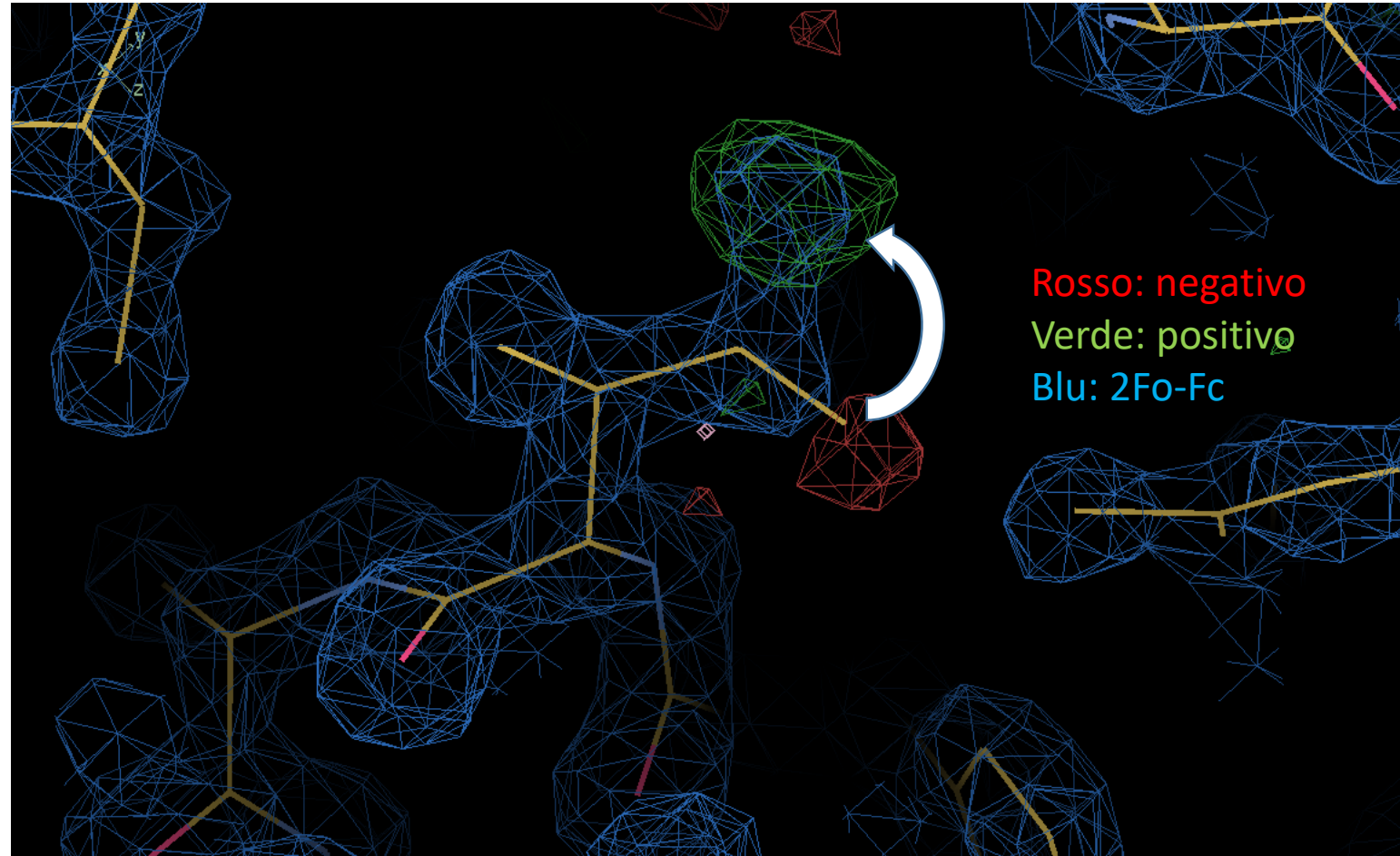
La mappa 2Fo-Fc è invece una mappa di densità $\rho(r)$ corretta per la mappa ΔF (in effetti è una mappa $\rho(r) + \Delta F$). Sarà più intensa laddove mancano atomi, al contrario sarà meno intensa laddove il modello è 'sbagliato'.

Nella letteratura scientifica la mappa ΔF positiva ha in genere colore verde, quella ΔF negativa colore rosso. La mappa 2Fo-Fc ha invece colore blu (sempre positiva).

Mappe differenza

L'utilizzo delle mappe differenza aiuta notevolmente a capire dove esistono problemi o il fit non è soddisfacente.

La grafica molecolare permette una grande flessibilità nell'aggiustare il modello coerentemente con le mappe



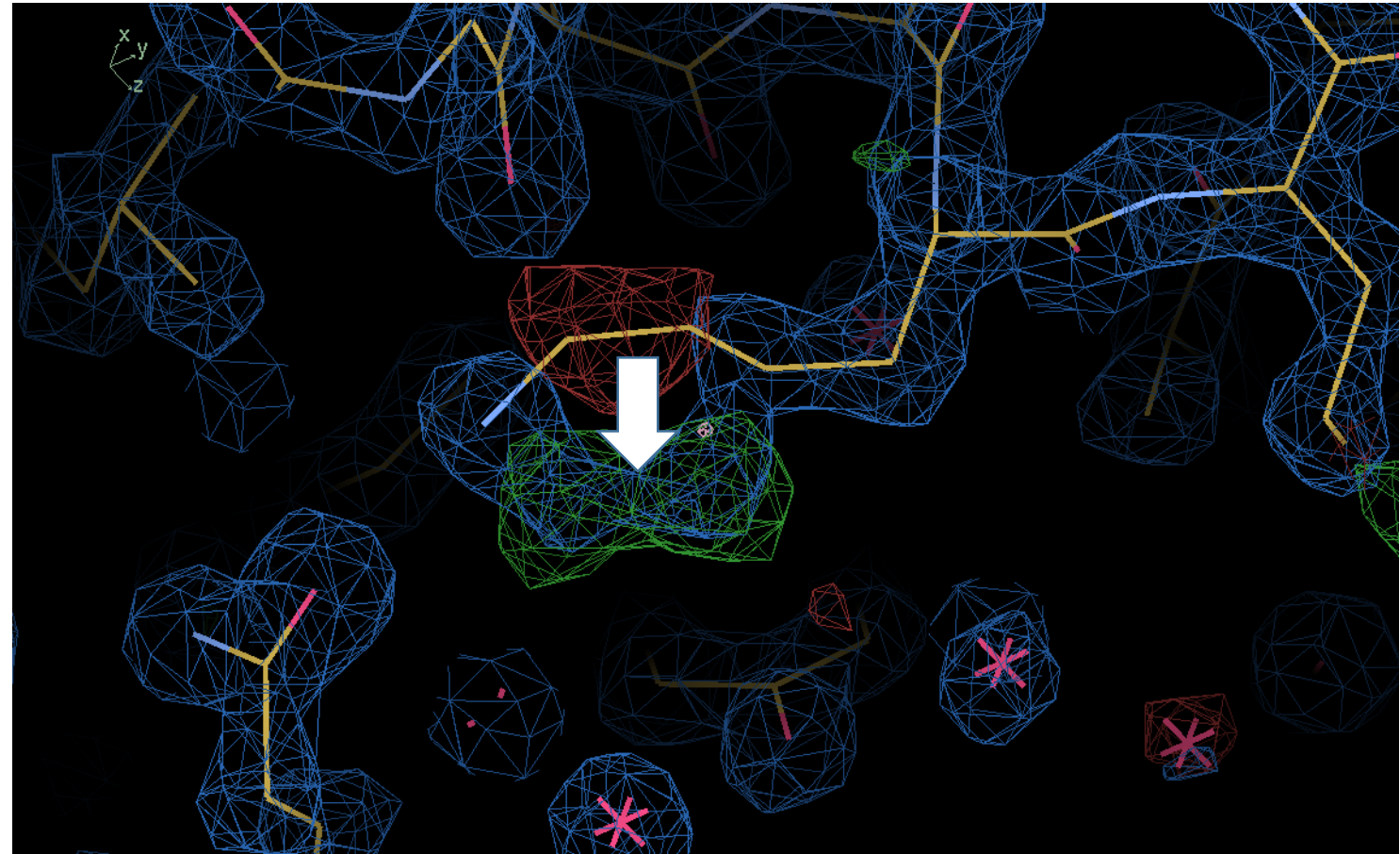
Mappe differenza

Rosso: negativo

Verde: positivo

Blu: 2Fo-Fc

Anche in questo caso è necessario un intervento 'manuale' per ottenere un corretto posizionamento della catena laterale della lisina e quindi un buon fit con la densità elettronica.



Convergenza e procedure di raffinamento

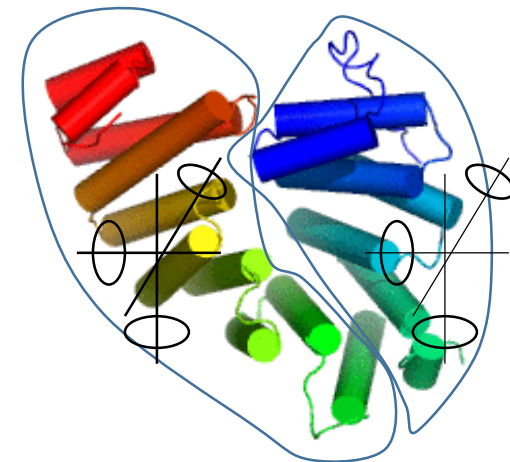
Corpo Rigido (Rigid Body)

Un approccio di tipo algoritmico è quello dell'affinamento a corpo rigido (Rigid Body).

In questo caso la molecola (proteina) viene suddivisa in tanti sotto-domini (al limite anche tutta la proteina) e ad essi sono associati tre coordinate (centro di massa) e tre angoli.

Nel corso dell'affinamento le coordinate dei domini, come anche i loro angoli sono modificati in modo da migliorare l'R-factor.

E' un metodo con maggiore raggio di convergenza (non si limita a minimi locali), ma applicato in genere solo nelle fasi iniziali dell'affinamento (non è molto preciso).



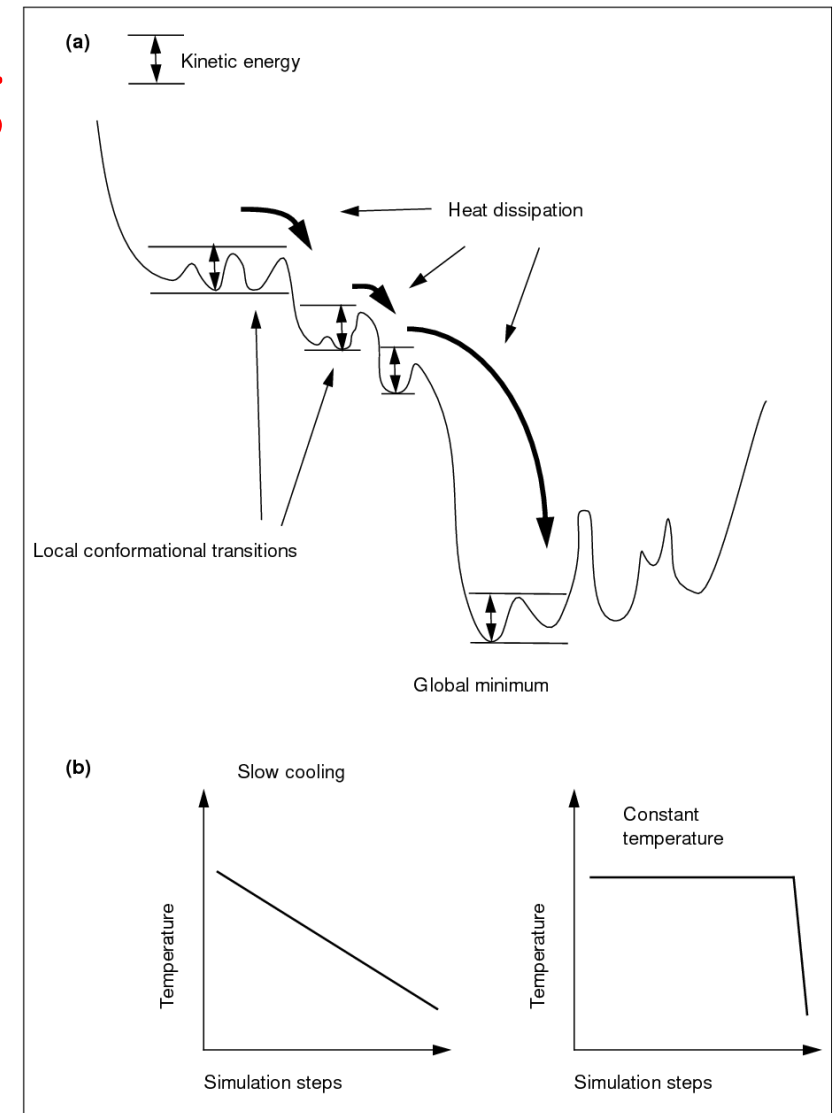
Simualted Annealing

Il problema dei minimi locali può essere visto come l'impossibilità di superare una barriera energetica per passare da una conformazione sbagliata a quella corretta.

In un sistema fisico, una particella che ha una quantità sufficiente di energia cinetica può superare barriere energetiche anche rilevanti (dipende dalla temperatura).

Nel metodo del Simulated Annealing viene simulato un 'riscaldamento' della molecola, ovvero viene conferita agli atomi la capacità di muoversi più liberamente, nel corso dell'affinamento, in modo da poter superare barriere 'energetiche' anche di una certa entità.

Il sistema è dapprima 'riscaldato', per permettere agli atomi di muoversi, e poi 'raffreddato', in modo da far posizionare gli atomi in posizioni nelle nuove posizioni di minimo. In questo modo si facilita il passaggio da un minimo locale ad un minimo assoluto

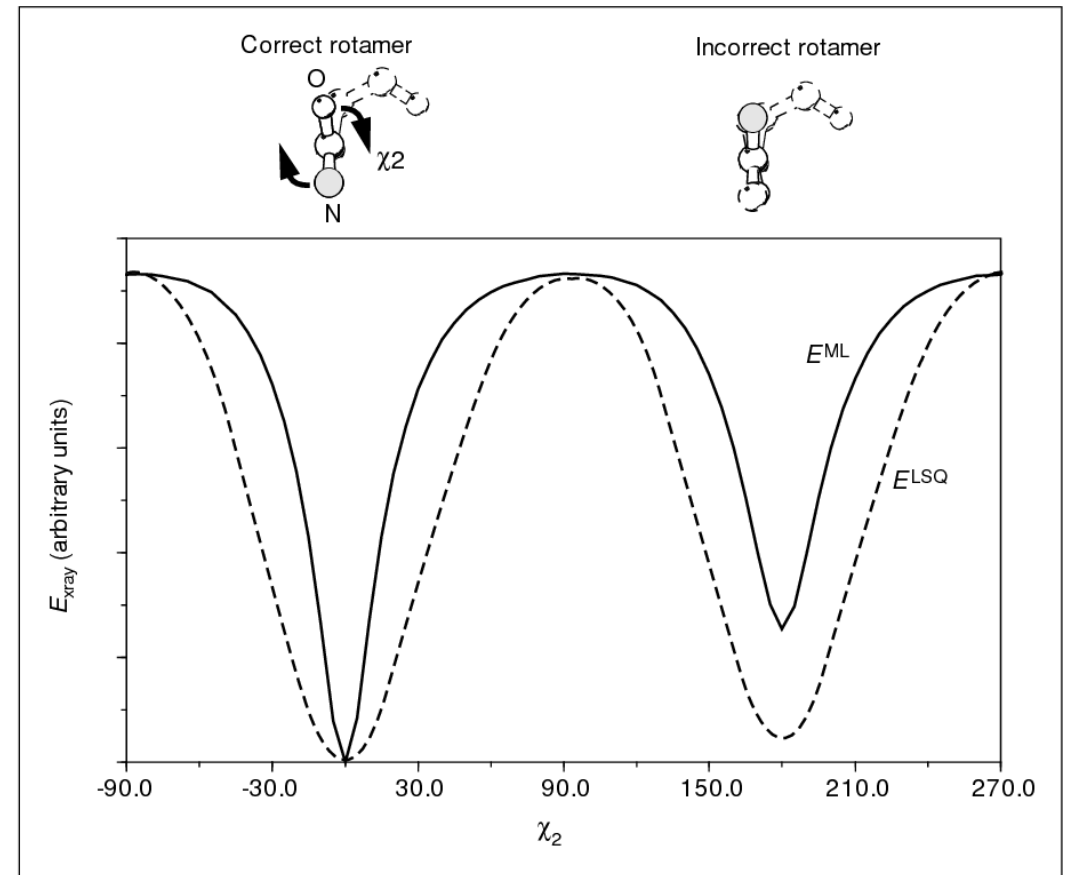


Simulated Annealing

Il metodo del Simulated Annealing ha un buon raggio di convergenza, ovvero permette di superare i minimi locali, tuttavia non è molto preciso.

In genere lo si utilizza nelle fasi iniziali dell'affinamento, specialmente se il problema della fase è stato risolto con il metodo del Molecular Replacement

E' un metodo computazionalmente piuttosto intensivo (si basa sulla Dinamica Molecolare) ma non in modo esagerato.



Simulated Annealing e rimozione del Bias

Sappiamo come la fase dei fattori di struttura giochi un ruolo importante nella ricostruzione della densità elettronica.

Nel metodo del Molecular Replacement, il modello utilizzato può influenzare le fasi dei fattori di struttura in modo erroneo, lasciando una sua 'impronta' nella densità elettronica.

Questo effetto va sotto il nome di *bias* e si manifesta con la presenza di segmenti di densità elettronica dovuti al modello (utilizzato nel Molecular Replacement) e non alla proteina target (quella di cui sto ottimizzando la struttura).

Il Simulated Annealing è molto efficace nel rimuovere il bias dovuto a parti 'sbagliate' del modello, per questo è molto usato dopo il Molecular Replacement **o per 'ricostruire' parti del modello probabilmente erronee.**



Il modello Finale

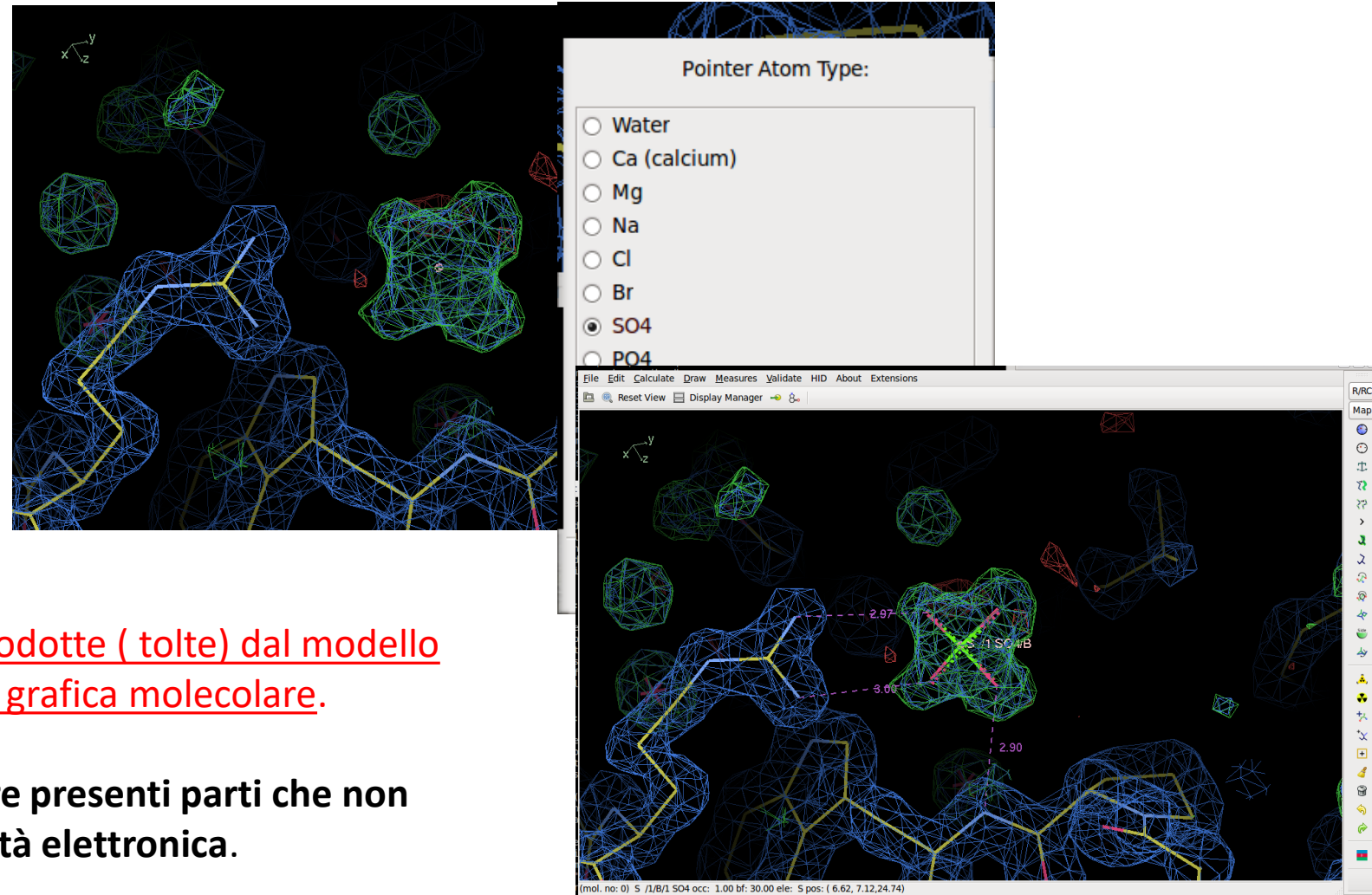
Completamento del Modello

Nel corso della procedura di affinamento possono presentarsi parti aggiuntive più o meno aspettate:

- Molecole o Ioni presenti nel liquido madre
- Molecole di precipitante
- Molecole di crioprotettore
- Modifiche della proteina (ossidazioni, acilazioni etc...)

Tutte queste parti devono essere introdotte (tolte) dal modello molecolare nel corso delle sessioni di grafica molecolare.

Nel modello finale non devono essere presenti parti che non siano coerenti con le mappe di densità elettronica.



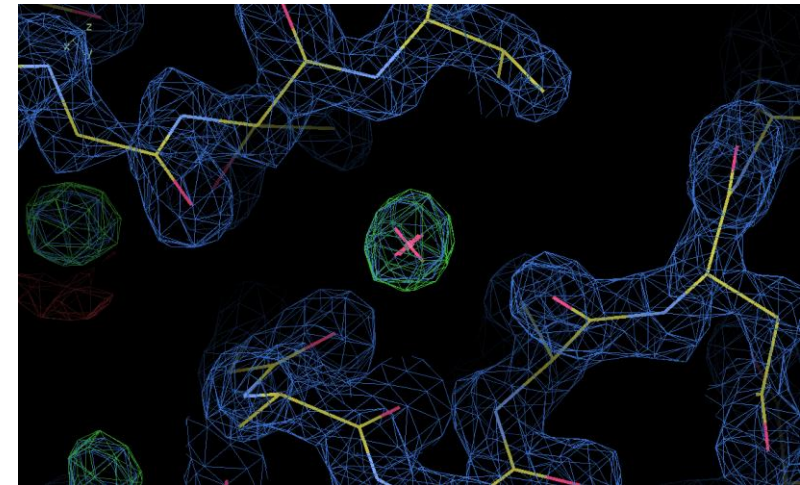
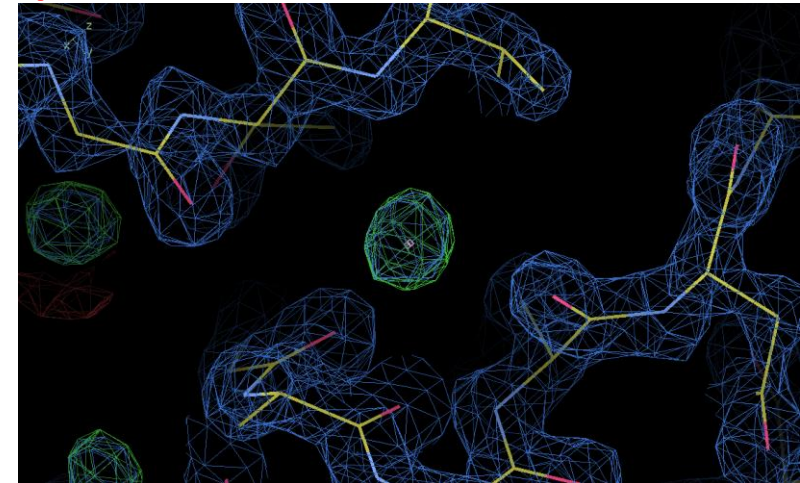
Molecole di Acqua

Le macromolecole biologiche presentano sempre delle molecole di acqua coordinate con i residui superficiali della macromolecola (**primo e secondo shell di idratazione**)

La 'visibilità' delle molecole di acqua dipende dalla risoluzione. A bassa risoluzione è probabile che non siano visibili. **Ad alta risoluzione sono visibili in modo molto chiaro.**

Le molecole di acqua presentano delle geometrie di interazione abbastanza definite.

Devono essere introdotte nel modello, se possibile, poiché talvolta hanno un ruolo funzionale preciso.



Costruzione Automatica (*Arp-Warp*)

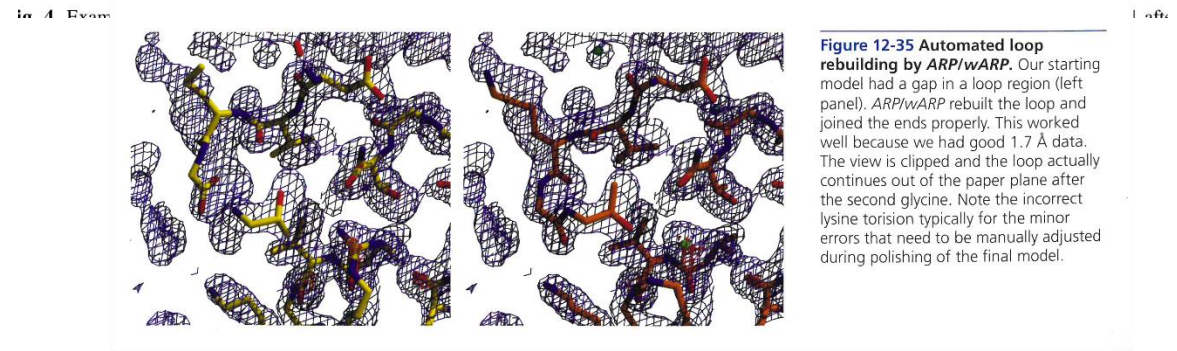
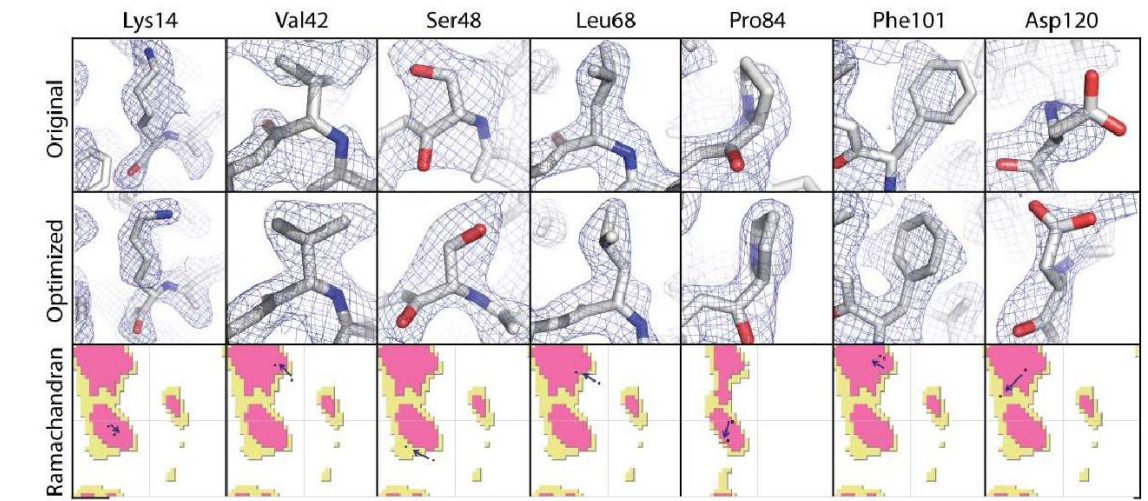
Come per la costruzione del modello iniziale, anche per l'ottimizzazione e il completamento del modello si può far ricorso a procedure automatiche (Model re-building).

I motivi sono fondamentalmente due:

- Minore intervento umano
- Eliminare la soggettività nel processo di ottimizzazione

La procedura può essere piuttosto lunga, ma viene svolta dal computer.

- Non è in grado di risolvere problematiche non-aspettate
- In genere non dà risultati migliori, ma è comunque un risparmio di tempo per il cristallografo



Protocollo generale

