

**Università degli Studi Federico II di Napoli**

Dottorato di Ricerca in Biologia Computazionale e  
Bionformatica XXIV ciclo

**Determinazione *ab initio* di chemical shift  
per la validazione strutturale del sito di  
coordinazione del metallo in superfamiglie  
di metalloproteine**

R.G. Viglione

Tutor: Prof. Orlando Crescenzi  
Co-tutor: Prof. Giovanni Paoletta



## INDICE

CAPITOLO 1 .....	6
INTRODUZIONE .....	6
1.1 Scopo del lavoro .....	6
1.2 Metodi per la determinazione della struttura delle proteine .....	6
1.3 NMR e struttura delle proteine .....	8
1.4 Chemical shift .....	9
1.5 Significato di chemical shift e chemical shielding .....	10
1.7 Che cosa è il modelling molecolare? .....	12
1.8 Superfici di energia potenziale .....	13
CAPITOLO 2 .....	15
TEORIA DEL FUNZIONALE DELLA DENSITA' ELETTRONICA .....	15
2.1 Introduzione alla chimica quantistica .....	15
2.2 Equazione di Shrödinger indipendente dal tempo .....	16
2.3 Approssimazione di Born-Oppheneimer .....	17
2.4 Metodo di Hartree-Fock .....	18
2.5 Correlazione Elettronica .....	20
2.6 Teoria del funzionale della densità elettronica .....	20
2.7 Densità elettronica e matrici densità .....	21
2.8 Teoremi di Hohenberg e Kohn .....	22
2.9 Teorema di Kohn e Sham .....	24
2.10 Il modello SCF generalizzato .....	27
2.11 Local Density Approximation .....	29
2.12 Local Spin Density Approximation .....	29
2.13 Generalised Gradient Approximation .....	30
2.14 Funzionali meta-GGA .....	31
2.15 La connessione adiabatica .....	32
2.16 I vantaggi dei metodi DFT .....	33
CAPITOLO 3 .....	35
Meccanica Molecolare .....	35
3.1 Un semplice campo di forza in meccanica molecolare .....	36
3.2 Caratteristiche generali di un campo di forza .....	37
3.3 Stretching di legame .....	38
3.4 Bending di legame .....	38
3.5 Potenziale torsionale .....	39
3.6 Bending out of plane .....	40
3.7 Termini incrociati .....	41
3.8 Interazioni di non legame .....	42
3.9 Interazioni elettrostatiche .....	43
3.10 Cariche puntiformi .....	44
3.11 Cariche derivate dal potenziale elettrostatico molecolare .....	45
3.12 Interazioni di van der Waals .....	47
3.13 Legame a idrogeno .....	47
3.14 Parametrizzazione di un campo di forza .....	48
3.15 Limiti e punti di forza della meccanica molecolare .....	49
3.16 Performance e qualità dei campi di forza .....	51
3.17 Differenze tra i campi di forza .....	52

3.18 Esecuzione di un calcolo MM .....	52
CAPITOLO 4 .....	54
DINAMICA MOLECOLARE .....	54
4.1 Basi della dinamica molecolare .....	55
4.2 Approssimazione adiabatica .....	57
4.3 Integrazione delle equazioni del moto .....	57
<i>Predictor corrector</i> .....	60
4.4 Potenziali .....	61
<i>Limiti dei Potenziali Empirici</i> .....	61
4.5 Inclusione del Solvente in una Simulazione MD .....	63
4.6 Set-up e running di una Simulazione MD .....	65
4.7 Proprietà dipendenti dal tempo .....	67
4.8 Dinamica di un sistema sottoposto a vincoli: SHAKE .....	68
CAPITOLO 5 .....	69
METODI IBRIDI QUANTISTICI/CLASSICI .....	69
5.1 Introduzione .....	69
5.2 Terminologia .....	70
5.3 Energia QM/MM .....	70
<i>Schemi sottrattivi</i> .....	71
<i>Schemi additivi</i> .....	71
5.4 Interazioni QM-MM bonded .....	73
5.5 Interazioni non-bonded tra gli strati .....	73
5.6 Interazione elettrostatica .....	74
<i>Embedding meccanico</i> .....	74
<i>Embedding elettrostatico</i> .....	74
<i>Embedding polarizzato</i> .....	75
5.7 Trattazione del legame all'interfaccia .....	76
5.8 Link atom .....	77
5.9 Forze sui link atoms .....	78
5.10 Metodo ONIOM .....	79
CAPITOLO 6 .....	82
Risultati e Discussione .....	82
6.1 Elettrostatica delle proteine .....	82
6.2 Calcolo dei pKa nella Calbindina .....	84
6.3 Calcolo dei Chemical Shieldings .....	86
<i>Scelta del metodo computazionale</i> .....	86
<i>Electrostatic embedding</i> .....	88
<i>Messa a punto della partizione QM/MM</i> .....	89
6.4 Verso calcoli realistici .....	95
<i>Dinamica molecolare della CAB</i> .....	95
<i>Calcoli di chemical shift</i> .....	98
6.5 Conclusioni .....	104
APPENDICE A .....	106
Approssimazione LCAO e set di funzioni di base .....	106
APPENDICE B .....	109
Densità di corrente elettronica e proprietà magnetiche molecolari .....	109
Il problema del gauge .....	111
APPENDICE C .....	113

Proteine calcium-binding.....	113
APPENDICE D .....	115
Legami a idrogeno della CAB.....	115
Schermi magnetici della Apo-CAB.....	116
Schermi magnetici della Holo-CAB.....	123
BIBLIOGRAFIA .....	129

# CAPITOLO 1

## INTRODUZIONE

### 1.1 Scopo del lavoro

Le metalloproteine mostrano numerose e diversificate attività biologiche, come il trasferimento elettronico e la catalisi. La risonanza magnetica nucleare (NMR) è un eccellente strumento per indagare sulla struttura tridimensionale delle proteine. Tuttavia la determinazione strutturale nelle metalloproteine richiede una conoscenza *a priori* del sito di coordinazione del metallo a causa di una mancanza di *restraints* ione metallico-proteina, derivanti dagli esperimenti [1].

I chemical shift mostrano una notevole sensibilità alla coordinazione del metallo, tuttavia la loro dipendenza da fattori stereoelettronici è complessa e l'informazione strutturale insita è difficile da estrarre.

Questo progetto ha lo scopo di introdurre il calcolo *ab initio* dei chemical shift come uno strumento per la validazione strutturale dei siti di binding delle metallo-proteine. A tale proposito verranno prese in considerazione le *EF-hand* calcium-binding proteine per applicare il protocollo che si intende sviluppare, poiché per questa famiglia di metallo-proteine esistono accurati database sia strutturali che di proprietà NMR. Lo scopo del protocollo che si intende implementare è ampio, con possibilità di applicazioni per una sistematica esplorazione di altri metalli oltre allo ione calcio come anche di differenti sistemi proteici.

### 1.2 Metodi per la determinazione della struttura delle proteine

La conoscenza strutturale delle macromolecole biologiche, come le proteine, è cruciale per una comprensione piena delle loro funzioni oltre che per numerose applicazioni pratiche,

come l'ingegneria delle proteine ed il *drug design*. La cristallografia a raggi X e la spettroscopia di Risonanza Magnetica Nucleare (NMR) sono le due tecniche principali usate per ottenere questa informazione. Fondati su principi biofisici questi metodi sperimentali dovrebbero, entro certi limiti di errore, condurre alla medesima struttura proteica, inoltre essi non forniscono direttamente un'immagine tridimensionale della macromolecola, ma piuttosto la struttura viene determinata attraverso complicate procedure di interpretazione dei dati grezzi.

In un esperimento ai raggi X può generalmente essere conseguita una più elevata qualità strutturale, rispetto agli esperimenti NMR; inoltre quando sia disponibile un cristallo adatto, la procedura per la determinazione della struttura ai raggi X è più rapida; infine con la diffrazione possono essere studiati anche sistemi biomolecolari molto grandi. A dispetto della elevata risoluzione strutturale, il principale svantaggio di questa tecnica è la crescita di cristalli di qualità appropriata; tra l'altro la presenza di molecole di solvente e di effetti di impaccamento possono alterare la struttura finale.

In un esperimento NMR invece la proteina è studiata in soluzione (in micelle, membrane...) ovvero in condizioni vicine a quelle biologiche; a differenza dei raggi X la spettroscopia NMR permette di osservare gli atomi di idrogeno; sono assenti distorsioni strutturali dovute a contatti nel cristallo, e questo è un aspetto importante soprattutto per molecole flessibili come i peptidi; è possibile studiare anche stati *unfolded* o parzialmente avvolti; l'NMR è una tecnica sensibile ai processi dinamici e le proteine, essendo flessibili, esibiscono anisotropia nel loro moto ed anzi esistono come un insieme dinamico di conformazioni. Gli studi sperimentali NMR forniscono informazioni sulla struttura proteica proprio a partire dai dati ottenuti per tale insieme di "conformer".

La prima determinazione della struttura 3D di una proteina, conseguita usando tecniche NMR fu accolta inizialmente con un certo scetticismo [2]. Da allora, però, numerose indagini sperimentali e diversi sviluppi tecnologici (ad esempio l'introduzione di più potenti magneti o della spettroscopia multidimensionale) hanno consentito all'NMR di affermarsi come uno strumento standard per la determinazione della struttura tridimensionale delle proteine ed in genere delle macromolecole biologiche, ad una risoluzione atomica, spesso risultando una valida alternativa alla cristallografia a raggi X. Oggi è possibile determinare le strutture di biomolecole fino 900 kDa sulla base di esperimenti NMR ed infatti risulta che, attualmente, oltre il 15% delle strutture proteiche,

depositate nel Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)) siano state ottenute attraverso tecniche NMR.

Un'altra tecnica disponibile per la determinazione strutturale è il *modeling* molecolare, che negli anni recenti ha conosciuto enormi progressi [3], infatti lo sviluppo di metodi di solvatazione, della dinamica molecolare e la diffusione di calcolatori sempre più potenti hanno consentito simulazioni e calcoli sempre più accurati [4].

Il *focus* del prossimo paragrafo è sull'utilizzo della spettroscopia NMR nella determinazione strutturale.

### 1.3 NMR e struttura delle proteine

Le principali fonti di informazioni per la determinazione strutturale via NMR sono il Nuclear Overhauser Effect (NOE), le costanti di accoppiamento J, l'accoppiamento dipolare [5].

In particolare il successo dell'applicazione della spettroscopia NMR alla determinazione della struttura delle proteine è largamente basata sui NOE, considerati come una fonte primaria di informazione strutturale. I segnali NOE forniscono stime delle distanze tra protoni dalle quali è possibile ricostruire le strutture complete delle biomolecole:

$$NOE_{ij} \propto \frac{1}{r_{ij}^6}$$

dove  $NOE_{ij}$  è l'aumento di intensità osservato tra i nuclei  $i$  e  $j$  mentre  $r_{ij}$  è la distanza tra i due nuclei. Questa informazione è davvero notevole nella determinazione della struttura, poiché protoni che sono distanti nel *framework* covalente possono essere vicini nello spazio.

Un fonte addizionale di informazione strutturale è rappresentata dalle costanti di accoppiamento J, che sono legate allo splitting delle linee in un multipletto di uno spettro NMR. Quando i protoni sono separati da tre legami, J dipende dall'angolo di torsione in accordo all'equazione di Karplus [6]. Le costanti J possono dunque essere usate per il *refinement* degli angoli di torsione nella determinazione strutturale.

Anche le costanti di accoppiamento dipolare forniscono informazioni che insieme con i NOE possono essere di grande aiuto nel determinare la struttura delle proteine. Questo accoppiamento deriva dall'interazione magnetica dei momenti nucleari [7], per cui risulta non nullo solo in soluzioni anisotrope.

#### **1.4 Chemical shift**

Recentemente, diversi articoli comparsi in letteratura hanno suggerito che strutture di proteine ragionevolmente accurate possano essere determinate direttamente dai chemical shift [8], ed è stato illustrato da vari gruppi di ricerca come effettivamente i chemical shift possano essere usati per definire le strutture degli stati nativi di proteine ad alta risoluzione, senza misure sperimentali aggiuntive. A differenza dei NOE, i chemical shift sono facili da misurare e non sono particolarmente sensibili ai vincoli sulle dimensioni della proteina.

Nuclei non equivalenti sperimentano nella stessa molecola un differente campo magnetico. Questa differenza dà origine a differenti risonanze (o segnali) nello spettro. Nel caso di un protone, lo schermo diamagnetico del suo unico elettrone è relativamente piccolo e dunque l'effetto dei gruppi di atomi vicini influenza in maniera significativa il chemical shift.

Fin dalla prima dimostrazione riportata da Arnold nel 1951 dei tre picchi separati nello spettro protonico dell'etanolo, è stato riconosciuto che i chemical shift sono sensibili alla struttura chimica ed infatti l'NMR è divenuto uno degli strumenti più importanti per la determinazione delle strutture covalenti delle molecole. Ma i chemical shift sono sensibili anche all'ambiente intorno al nucleo in esame e dunque possono risultare utili nella determinazione della struttura non-covalente e supramolecolare [9]. Quando analizzati in maniera opportuna i chemical shift possono essere usati per desumere strutture secondarie di proteine, flessibilità, angoli diedri, orientazione di catene laterali, interazioni ioniche o elettrostatiche, legami a idrogeno. Il chemical shift associato ad uno specifico atomo è una sommatoria di numerosi fattori, che possono essere influenzati da contatti tra residui, ad esempio prossimità ad anelli aromatici che si trovano in posizioni lontane nella sequenza proteica. L'interpretazione di tali effetti potrebbe in linea di principio consentire la caratterizzazione dettagliata dell'intorno di ciascun atomo nella proteina. Anche in assenza di una definizione completa della struttura, variazioni nei chemical shift potrebbero fornire

informazioni importanti, come il *mapping* dei siti di binding, ad esempio delle metallo-proteine.

### 1.5 Significato di chemical shift e chemical shielding

Il tensore di schermo chimico (chemical shielding)  $\sigma$  mette in relazione il campo magnetico effettivo  $H^{eff}$ , sperimentato dal nucleo probe, con il campo applicato dello spettrometro ( $H_0$ ):

$$H^{eff} = (1 - \sigma) H_0 \quad (1.1)$$

Il tensore di chemical shift collega quindi lo schermo assoluto con quello di un composto di riferimento. L'energia di interazione con il dipolo magnetico nucleare  $\mu$  è

$$E = -\mu \cdot H^{eff}$$

E' proprio questo momento magnetico, che dipende dallo spin nucleare  $I$ , che permette l'osservazione degli spettri di assorbimento NMR, causati da transizioni tra livelli di spin nucleare. L'atomo di calcio è un esempio di nuclide che non presenta risposta NMR.

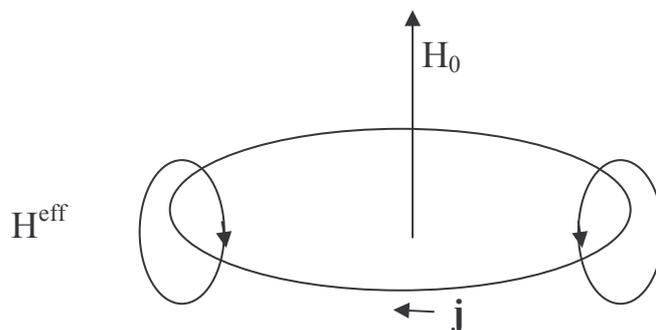
Il tensore di schermo può essere scritto in termini di derivata seconda dell'energia rispetto al campo esterno e al dipolo magnetico nucleare:

$$1 - \sigma = \left( \frac{\partial^2 E}{\partial \mu \partial H_0} \right)$$

cioè, espresso nel linguaggio della teoria delle perturbazioni, lo schermo chimico rappresenta una proprietà di risposta al secondo ordine.

Un modello fisico utile ad illustrare i fondamenti teorici del chemical shift è riportato schematicamente in figura 1.1.

Il campo esterno (dello spettrometro) induce correnti elettroniche  $j$ , che dipendono dall'orientazione e dalla natura chimica della molecola. Queste correnti a loro volta danno origine a campi locali, che interagiscono con i momenti magnetici nucleari.



**Figura 1.1**

Se è nota la distribuzione di corrente  $j$ , che è funzione della posizione, il campo magnetico associato può essere determinato dalla legge di Biot-Savart:

$$H^{eff} = -\frac{1}{c} \int \frac{r \times j(r)}{r^3} dr \quad (1.2)$$

Da questo punto di vista, dunque il compito di calcolare teoricamente i chemical shift può essere diviso in due parti: trovare la densità di corrente elettronica  $j$  e calcolare i campi attraverso l'equazione (1.2). Questo può essere realizzato o sulla base di dettagliati calcoli di struttura elettronica o mediante semplici forme funzionali, con parametri empirici ottenuti da dati sperimentali.

La parte isotropa di  $\sigma$ , ovvero 1/3 della somma dei suoi elementi diagonali, è il parametro che viene comunemente determinato nelle misure allo stato liquido.

## 1.6 Calcoli teorici di chemical shift in proteine

Lo schermo chimico riflette l'intorno elettronico dei nuclei osservati, quindi contiene importanti informazioni circa la struttura molecolare e la dinamica conformazionale. Una dettagliata comprensione delle fonti dei vari contributi allo schermo chimico nelle proteine non è critica solo nella predizione dei chemical shift e dunque nell'assegnazione dei segnali NMR ma è anche potenzialmente importante per migliorare la caratterizzazione strutturale delle proteine e l'analisi dinamica di questi biopolimeri. La spettroscopia NMR comunque richiede una interpretazione dei chemical shift, oltre che delle costanti di accoppiamento spin-spin e dei tempi di rilassamento. A tale scopo possono risultare utili i

metodi teorici e computazionali. Negli ultimi decenni del resto ci sono stati notevoli progressi nello sviluppo di schemi teorici e computazionali per il calcolo sempre più accurato dei chemical shift di peptidi e proteine. I metodi di calcolo variano da quelli basati sulla superficie di schermo empirica [10] alla omologia di sequenza a quelli che adottano calcoli quantistici *ab initio*. I calcoli da principi primi sono molto accurati, anzi rappresentano l'approccio più accurato, tuttavia essi richiedono anche significativi sforzi computazionali, in quanto dipendono in maniera non favorevole dal numero di elettroni della molecola, ciò che pone molti sistemi di interesse biologico al di fuori del loro campo di applicazione. Tuttavia è possibile ridurre notevolmente lo sforzo computazionale nella determinazione *ab initio* dei parametri NMR, ricorrendo ad esempio a sistemi modello "troncati" e adottando metodi ibridi meccanica quantistica/ meccanica molecolare. Lo sviluppo degli approcci misti è guidato dall'idea generale che sistemi chimici di grandi dimensioni possono essere suddivisi in una regione importante dal punto di vista elettronico, che richiede un trattamento quantomeccanico ed il resto del sistema che agisce solo in maniera "perturbativa" e quindi ammette anche una descrizione classica. Il risultato è che in molti casi è possibile ottenere risultati accurati con una frazione soltanto del costo computazionale dei metodi convenzionali della chimica quantistica. Questi approcci verranno illustrati in dettaglio nel Capitolo 5.

### **1.7 Che cosa è il modelling molecolare?**

Per modello si intende una descrizione semplificata o idealizzata, spesso in termini matematici, di un sistema o di un processo, atta a facilitare calcoli e predizioni. Il *molecular modelling* fa riferimento dunque alla simulazione del comportamento di molecole e sistemi molecolari [11]. Naturalmente è possibile realizzare semplici studi anche con modelli meccanici ma oggi il *molecular modelling* è invariabilmente associato al *modelling* al calcolatore, infatti le tecniche computazionali hanno rivoluzionato questa branca delle scienze a tal punto che la maggior parte dei calcoli non potrebbero essere eseguiti senza l'uso di un computer. Questo non implica che un modello più sofisticato sia necessariamente migliore di uno più semplice, tuttavia i computer hanno senz'altro

ampliato il *range* di modelli che possono essere proposti e di sistemi ai quali tali modelli possono essere applicati.

I modelli che molti chimici hanno incontrato per primi sono i modelli “sticks” proposti da Dreiding o i modelli “space fillings” di Corey, Pauling e Koltun, ovvero modelli connessi essenzialmente alle rappresentazioni tridimensionali delle strutture molecolari. Tuttavia il *modelling* molecolare riguarda anche modelli più astratti, molti dei quali hanno una storia distinta. Un esempio è la meccanica quantistica i cui principi hanno anticipato di molti decenni la costruzione dei primi computer.

Può accadere che vengano confusi termini come chimica teorica, chimica computazionale e *modelling* molecolare. Spesso la chimica teorica viene considerata come un sinonimo di chimica quantistica mentre la chimica computazionale generalmente concerne non solo la meccanica quantistica ma anche la meccanica molecolare, procedure di minimizzazione, analisi conformazionale, simulazioni ed altri metodi basati sull’uso del calcolatore per la predizione e la comprensione del comportamento dei sistemi molecolari. Qualunque tecnica teorica o computazionale che ambisca a raggiungere un siffatto scopo può essere, comunque, considerata un esempio di *modelling* molecolare. Se una distinzione deve essere fatta, riguarda piuttosto l’enfasi che il *modelling* molecolare ripone nella rappresentazione e manipolazione delle strutture molecolari e nelle proprietà che da tali strutture dipendono.

Un soggetto che è strettamente legato al *modelling* è l’informatica chimica, che a sua volta abbraccia due discipline la cheminformatica e la bioinformatica, sviluppatasi anche come conseguenza dell’introduzione di tecniche sperimentali come la sintesi combinatoriale e le *automated sequencing machines*.

## **1.8 Superfici di energia potenziale**

Nel *modelling molecolare*, si assume invariabilmente la validità dell’approssimazione di Born-Oppeneimer, che come verrà descritto nei prossimi Capitoli, permette di separare i moti nucleari ed elettronici in una molecola. Gli elettroni infatti molto più leggeri e veloci dei nuclei possono rapidamente adattarsi a qualunque variazione delle posizioni nucleari, di conseguenza l’energia della molecola nel suo stato fondamentale può essere considerata

funzione delle sole coordinate nucleari. Dunque se soltanto alcuni o tutti i nuclei si muovono l'energia cambia e queste variazioni possono essere considerate come movimenti su una superficie multidimensionale chiamata superficie di energia potenziale. Particolarmente interessanti sono i punti stazionari di tale superficie, ovvero i punti in cui le forze su tutti gli atomi sono nulle, dove le derivate prime dell'energia rispetto alle posizioni nucleari sono pari a zero. I punti di minimo sono punti stazionari corrispondenti a strutture stabili. Esistono due modi comunemente adottati per specificare le posizioni di atomi e molecole: specificare le coordinate cartesiane  $(x,y,z)$  di tutti gli atomi presenti o, alternativamente, usare coordinate interne, che descrivono la posizione di ciascun atomo rispetto agli altri atomi del sistema. Ad eccezione del primo, ogni atomo ha tre coordinate interne: la distanza dall'atomo precedentemente definito, l'angolo formato con altri due atomi definiti in precedenza ed infine l'angolo diedro formato dall'atomo in questione e da altri tre atomi già definiti. E' sempre possibile convertire coordinate interne in cartesiane e viceversa. Generalmente le coordinate interne sono utili per descrivere la relazione tra gli atomi in una molecola singola, mentre per un insieme di molecole discrete è più appropriato ricorrere alle coordinate cartesiane.

Nei prossimi capitoli verranno presi in considerazione tre dei modelli teorici per il calcolo delle struttura e dell'energetica dei sistemi molecolari maggiormente diffusi: i modelli della chimica quantistica, la meccanica molecolare e la dinamica molecolare.

## CAPITOLO 2

### TEORIA DEL FUNZIONALE DELLA DENSITA' ELETTRONICA

#### 2.1 Introduzione alla chimica quantistica

La chimica quantistica è una disciplina che applica i metodi della meccanica quantistica ai sistemi atomici e molecolari per affrontare problemi in ambito chimico.

Negli anni venti Schrödinger introdusse il formalismo matematico per descrivere il moto di un insieme di elettroni e di atomi sotto forma di una equazione d'onda. Questa equazione considerata la chiave di volta della chimica quantistica, sfortunatamente non è risolvibile esattamente per sistemi atomici o molecolari, che abbiano più di un solo elettrone. L'obiettivo della chimica quantistica non relativistica è quella di ottenere una soluzione dell'equazione di Schrödinger che sia il più possibile vicina alla soluzione del sistema fisico reale.

In tal senso un primo approccio è stato sviluppato nel 1927 da Hartree, che propose un metodo per calcolare le funzioni d'onda e le energie approssimate di atomi e ioni: era nato il *metodo self consistent field* (campo auto-coerente). In seguito Slater rese questo metodo direttamente applicabile proponendo la decomposizione della funzione d'onda in prodotto di funzioni monoelettroniche. Nel 1930 Slater e Fock introdussero nel metodo di Hartree il principio di antisimmetria della funzione d'onda (dovuto a Pauli). Il principio di esclusione di Pauli è dunque rispettato usando un determinante di Slater nel calcolo del campo auto-coerente.

Il metodo Hartree-Fock (HF) è stato però realmente usato solo a partire dagli anni cinquanta con l'invenzione dei calcolatori, che hanno permesso di ampliare notevolmente le possibilità di applicazione. Tuttavia l'approssimazione HF è insufficiente per consentire un calcolo accurato delle proprietà chimiche. La creazione di legami tra atomi per formare molecole è dovuta alla condivisione degli elettroni più esterni degli atomi, che prendono il nome di elettroni di valenza. La descrizione dei legami chimici è corretta solo prendendo in considerazione la correlazione simultanea tra questi elettroni. Questo fenomeno prende il nome di correlazione elettronica. L'approccio HF tuttavia, essendo un metodo di campo

medio – si considera che un elettrone subisca l’influenza media degli altri elettroni – non consente il trattamento della correlazione tra elettroni. E’ stato necessario dunque andare al di là dell’approssimazione HF. Tutti i metodi *ab initio*, così denominati perché stabiliti a partire solo dai principi primi della meccanica quantistica, che prendono in considerazione la correlazione elettronica si chiamano metodi *post Hartree-Fock*.

Parallelamente allo sviluppo di tali metodi negli anni sessanta è stata proposta la teoria della Funzione della Densità Elettronica (DFT). Essa ripone sulla nozione di densità elettronica, una quantità dipendente da tre variabili, che rappresenta la probabilità di presenza di un elettrone, conoscendo la posizione di tutti gli altri. In questo modo, il problema della funzione d’onda elettronica a  $3N$  variabili (con  $N$  numero degli elettroni considerati) viene ridotta ad un problema a 3 variabili. Il DFT è basato sui teoremi dimostrati da Hohenberg e Kohn ed è divenuto molto diffuso grazie al formalismo introdotto da Kohn e Sham, che utilizza un gioco di equazioni simili a quelle di Hartree-Fock e che introduce i funzionali di scambio-correlazione per la modellizzazione della correlazione degli elettroni. Oggi i metodi basati sul DFT sono quelli maggiormente usati in chimica teorica e computazionale per il calcolo delle proprietà molecolari, in quanto permettono di includere, in qualche maniera, la correlazione elettronica anche per sistemi costituiti da un numero importante di elettroni ad un costo computazionale confrontabile con quello di un calcolo HF.

## 2.2 Equazione di Schrödinger indipendente dal tempo

Qualsiasi problema riguardante la struttura elettronica della materia può essere trattato, almeno in linea di principio, risolvendo l’equazione di Schrödinger dipendente dal tempo. Nella maggior parte dei casi tuttavia si è interessati a sistemi atomici o molecolari non soggetti ad interazioni dipendenti dal tempo. Lo studio di tali sistemi può essere affrontato per mezzo dell’equazione di Schrödinger indipendente dal tempo.

Consideriamo un sistema costituito da  $N_a$  atomi e  $N$  elettroni. Per ottenere l’energia  $E$  o la funzione d’onda  $\Phi$ , che descrive il comportamento del sistema occorre risolvere l’equazione di Schrödinger indipendente dal tempo [12,13]

$$H|\Phi\rangle = E|\Phi\rangle \quad (2.1)$$

dove  $H$  è l'Hamiltoniano non relativistico, non magnetico, così definito

$$H = -\frac{1}{2} \sum_i^N \nabla_i^2 - \sum_A^{N_a} \frac{1}{2M_A} \nabla_A^2 - \sum_i^N \sum_A^{N_a} \frac{Z_A}{r_{iA}} + \sum_i^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_i^N \sum_{B>A}^{N_a} \frac{Z_A Z_B}{R_{AB}}$$

L'Hamiltoniano contiene i seguenti contributi: l'energia cinetica degli elettroni, quella degli atomi, il potenziale di interazione nucleo-elettrone, il potenziale di interazione elettrone-elettrone ed infine il potenziale di interazione nucleo-nucleo.

### 2.3 Approssimazione di Born-Oppeneimer

L'Hamiltoniano definito nell'equazione ad autovalori (2.1) è estremamente complesso. Di fondamentale importanza in chimica quantistica è dunque la possibilità di separazione tra le coordinate nucleari ed elettroniche. Fisicamente, la massa dei nuclei, molto maggiore rispetto a quella degli elettroni, consente di affermare che il moto dei primi avviene su una scala dei tempi molto più lenta rispetto al moto dei secondi. Ciò equivale al fatto che, con buona approssimazione, le proprietà degli elettroni possono essere definite assumendo che lo 'scheletro' della molecola, formato dai nuclei, sia immobile. Fondamentalmente, questa è l'approssimazione di Born-Oppenheimer (BO) [12].

Il moto elettronico può allora essere separato da quello nucleare: gli elettroni si muovono su una superficie di energia potenziale nel campo dei nuclei. Dal punto di vista elettronico l'energia cinetica nucleare è trascurabile e il termine di repulsione tra i nuclei può essere considerato costante. L'Hamiltoniano elettronico è quindi definito come

$$H_{el} = \sum_i^N h_i + \sum_i^N \sum_{j>i}^N \frac{1}{r_{ij}}, \quad (2.2a)$$

dove

$$h_i = -\frac{1}{2} \nabla_i^2 - \sum_A^{N_a} \frac{Z_A}{r_{iA}} \quad (2.2b)$$

Allora

$$H_{el} |\Phi_{el}\rangle = E_{el} |\Phi_{el}\rangle$$

L'autovalore dell'Hamiltoniano elettronico è l'energia elettronica. Essa dipende parametricamente dalle coordinate dei nuclei  $E_{el} = E_{el}(\{R_A\})$ . Quanto alla funzione d'onda elettronica, essa dipende esplicitamente dalle coordinate elettroniche e parametricamente da quelle atomiche:

$$\Phi_{el} = \Phi_{el}(\{\mathbf{r}_i\}; \{R_A\}).$$

L'energia totale per nuclei fissi è dunque data da

$$E_{tot} = E_{el} + \sum_A \sum_{B>A} \frac{Z_A Z_B}{R_{AB}}.$$

Una volta che le equazioni elettroniche sono risolte, è possibile risolvere il problema nucleare. Poiché gli elettroni si spostano molto più velocemente dei nuclei, le coordinate elettroniche vengono sostituite con il loro valore medio. L'Hamiltoniano nucleare per il moto degli elettroni nel campo medio degli elettroni è espresso come

$$H_{nucl} = -\sum_A \frac{1}{2M_A} \nabla_A^2 + E_{tot}(\{R_A\})$$

le cui soluzioni descrivono le vibrazioni, rotazioni e traslazioni di una molecola.

## 2.4 Metodo di Hartree-Fock

L'approssimazione di Hartree-Fock è alla base del primo e forse più usato metodo di calcolo *ab initio* per la determinazione di strutture elettroniche molecolari. Deve perciò essere chiaro che si tratta di una approssimazione e non di un metodo esatto di risoluzione dell'equazione di Schrödinger elettronica, ed ha per fine la determinazione dello stato fondamentale di un sistema di elettroni. L'Hamiltoniano elettronico di una molecola di  $N$  elettroni è dato dalle relazioni (2.2) in cui  $h_i$ , l'Hamiltoniano monoelettronico di *core* di ciascun elettrone è relativo all'energia cinetica e di interazione con i nuclei.

L'approssimazione di Hartree consiste nell'identificare la funzione d'onda multielettronica (autofunzione dell'Hamiltoniano elettronico) come il prodotto di Hartree di



## 2.5 Correlazione Elettronica

La maggior parte delle quantità chimiche di interesse sono collegate a differenze di energia e non a valori assoluti. Poiché una grande parte dell'errore nel modello di Hartree-Fock è costante, in molti casi le differenze di energia possono essere calcolate con una ragionevole precisione.

Sebbene il metodo di Hartree-Fock possa dare buoni risultati per molti problemi, in genere esso fallisce ogni volta che ci sono riarrangiamenti sostanziali nella distribuzione elettronica, per esempio quando si verifica la rottura e la formazione di legami chimici. In tali situazioni la cancellazione dell'errore viene a mancare e il metodo di Hartree-Fock conseguentemente produce risultati mediocri o incorretti. Diviene allora importante descrivere l'effetto dinamico in virtù del quale gli elettroni si evitano l'un l'altro. In un quadro classico ciò riflette il fatto che i movimenti degli elettroni devono essere correlati, per cui se uno di essi si trova in un lato della molecola, un altro preferisce stare nell'altro lato per minimizzare la repulsione. La cosiddetta 'energia di correlazione' è definita come la differenza tra l'energia esatta dello stato fondamentale e quella determinata nell'approssimazione di Hartree-Fock:

$$E^{corr} = E^{esatta} - E^{HF}.$$

Un approccio per recuperare la correlazione elettronica e che rappresenta la base di partenza di metodi *postHF* più complessi è il metodo dell'Interazione di Configurazione o CI, che consiste nello sviluppare la funzione d'onda come una somma di determinanti di Slater, che rappresentano le eccitazioni elettroniche.

Una metodologia sostanzialmente equivalente per il calcolo di  $E^{corr}$  è basata su un trattamento perturbativo e va sotto il nome di correzione di Moller-Plesset (o metodo MPPT, *Moller-Plesset perturbation theory*).

## 2.6 Teoria del funzionale della densità elettronica

Una alternativa al trattamento della correlazione elettronica è rappresentata dalla teoria del funzionale della densità elettronica, secondo la quale l'energia di un sistema

multielettronico può essere espressa attraverso la densità elettronica, che è il quadrato della funzione d'onda, integrata su  $N-1$  coordinate. Dunque mentre la funzione d'onda di un sistema a  $N$  elettroni dipende da  $3N$  coordinate spaziali e da  $N$  coordinate di spin, per cui la sua complessità aumenta con il numero di variabili, la densità elettronica dipende sempre dalle 3 coordinate spaziali e dalla coordinata di spin, e per questo resta indipendente dalla grandezza del sistema. Lo scopo dei metodi basati sulla teoria del funzionale della densità elettronica (DFT) è quello di trovare un funzionale (ovvero una funzione dipendente da una altra funzione) che permetta di collegare la densità con l'energia [13].

## 2.7 Densità elettronica e matrici densità

Il punto di partenza del DFT è l'espressione della densità elettronica. Per un sistema ad  $N$  elettroni descritto dalla funzione d'onda  $\psi$ , la probabilità di trovare l'elettrone che etichettiamo come 1 nella porzione di spazio compresa tra  $x_I$  e  $x_I + dx_I$  è [14]

$$\rho(x_I)dx_I = N \int \psi(x_1, x_2, \dots, x_N) \psi^*(x_1, x_2, \dots, x_N) dx_2 \dots dx_N$$

con  $x_I$  coordinata di spin-orbita.  $\rho(r_I)$  è la funzione associata all'elettrone 1.

L'integrazione sulla coordinata di spin  $\sigma_I$  permette di definire la densità elettronica come

$$\rho(r_I) = \int \rho(x_I) d\sigma_I$$

L'integrazione della densità rispetto a  $r_I$  (la coordinata spaziale dell'elettrone 1) consente di ritrovare il numero di elettroni,

$$\int \rho(r_I) dr_I = N$$

Allo stesso modo si definisce la densità di ordine 2, che rappresenta la densità di probabilità di trovare un elettrone simultaneamente in  $x_I$  e  $x_2$ :

$$\gamma_2(x_1, x_2) = N(N-1) \int \psi(x_1, x_2, \dots, x_N) \psi^*(x_1, x_2, \dots, x_N) dx_3 \dots dx_N$$

L'integrazione rispetto alle variabili di spin  $\sigma_I$  e  $\sigma_2$  permette di ottenere la densità di probabilità bielettronica  $\gamma_2(r_1, r_2)$

$$\gamma_2(r_1, r_2) = \int \gamma_2(x_1, x_2) d\sigma_1 d\sigma_2$$

Un'altra quantità molto utile è la matrice densità ridotta di ordine 1

$$\rho_1(r_1, r_1') = \int \rho_1(x, x') d\sigma_1 d\sigma_2$$

Si può dimostrare che

$$E = \langle \psi | H | \psi \rangle =$$

$$\langle \psi | \hat{T} + \hat{V} + \hat{V}_{ee} | \psi \rangle = T + V + V_{ee} = \underbrace{\int dr_1 \left( -\frac{1}{2} \nabla^2(r_1) \right) \rho_1(r_1, r_1')}_{T[\rho]} + \underbrace{\int v_{ext}(r_1) \rho(r_1, r_1) dr_1}_{v_{ext}[\rho]} + \underbrace{\frac{1}{2} \iint \frac{\gamma_2(r_1, r_2)}{r_{12}} dr_1 dr_2}_{V_{ee}[\rho]}$$

dove  $T$  è l'energia cinetica, che richiede la matrice densità ad un corpo,  $V = v_{ext}$  è l'energia di attrazione nucleare (potenziale esterno) che richiede la funzione densità, mentre  $V_{ee}$  è l'energia di repulsione inter-elettronica che richiede la funzione densità a due corpi. La densità bieletronica, che rappresenta la densità di probabilità di trovare simultaneamente un elettrone in  $r_1$  e  $r_2$ , a sua volta, si può riscrivere utilizzando le densità di probabilità correlate e decorrelate

$$\gamma_2(r_1, r_2) = \rho(r_1)\rho(r_2) + \rho(r_1)\rho_{xc}(r_1, r_2)$$

in cui  $\rho_{xc}(r_1, r_2)$  rappresenta la densità di una buca di scambio-correlazione.

## 2.8 Teoremi di Hohenberg e Kohn

Tra il 1920 ed il 1930 Fermi e Thomas intuirono l'importanza della densità elettronica di una molecola, come una quantità in qualche modo sufficiente per poter conoscere tutte le altre proprietà del sistema, tra cui l'energia. Tuttavia gli sforzi per ottenere buone energie dalla sola  $\rho$  risultarono vani. La situazione cambiò radicalmente negli anni sessanta, quando furono enunciati due teoremi [14], in base ai quali il modello proposto da Fermi e Thomas, ma anche le successive modificazioni possono essere visti come approssimazioni di una teoria esatta: la Teoria del Funzionale Densità.

Il primo teorema di Hohenberg-Kohn afferma che *la densità  $\rho$  determina univocamente il potenziale esterno  $v_{ext}$  a meno di una costante.*

Dunque qualsiasi osservabile di uno stato fondamentale stazionario, non degenere può essere calcolato esattamente, almeno in teoria, a partire dalla densità elettronica di questo stato. Infatti l'Hamiltoniano di un sistema dipende dal numero di elettroni e dal potenziale esterno, che rappresenta l'interazione tra elettroni e nuclei. Alla conoscenza della densità corrisponde quella del potenziale, che a sua volta determina quella dell'Hamiltoniano e dunque dei valori medi associati.

Questo teorema è valido quando la densità è  $N$ -rappresentabile, ovvero nel caso in cui l'integrazione della densità sulla coordinata spaziale permetta di ritrovare il numero di elettroni  $N$  e la densità sia non negativa e quando la densità è anche  $V$ -rappresentabile, cioè quando esista un potenziale esterno da cui essa possa essere derivata. Nelle condizioni di validità di questo teorema, il funzionale dell'energia si esprime come

$$E[\rho] = T[\rho] + V_{ee}[\rho] + v_{ext}[\rho]$$

in cui  $T[\rho]$  è il funzionale dell'energia cinetica,  $V_{ee}[\rho]$  è il potenziale di interazione e  $v_{ext}[\rho]$  è il funzionale del potenziale esterno, ovvero l'attrazione nucleo-elettrone. Le prime due quantità vengono generalmente raccolti in un funzionale universale  $F_{HK}$ , indipendente dal potenziale esterno

$$E[\rho] = F_{HK}[\rho] + \int \rho(r_1) v_{ext}(r_1) dr_1$$

Il secondo teorema di Hohenberg e Kohn può essere enunciato dicendo che *l'energia di uno stato fondamentale non degenere può essere determinato, esattamente in linea di principio, attraverso la densità che minimizza l'energia dello stato fondamentale.*

Questo significa che per la densità  $\rho$ ,  $N$ -rappresentabile e  $V$ -rappresentabile, si ha

$$E_0 \leq E[\rho]$$

Allora per ricavare l'energia dello stato fondamentale bisogna minimizzare il funzionale energia

$$\left[ \frac{\delta E[\rho]}{\delta \rho} \right] = 0$$

La minimizzazione dell'energia passa attraverso il formalismo di Lagrange, ovvero definendo una Lagrangiana  $L$  avente come restrizione la  $N$ -rappresentabilità della densità

$$L[\rho] = E[\rho] - \mu \left[ \int \rho(r) dr - N \right]$$

dove  $\mu$  è il moltiplicatore di Lagrange incognito. La minimizzazione di questa Lagrangiana implica che

$$\frac{\delta}{\delta \rho} \left\{ E[\rho] - \mu \left[ \int \rho(r) dr - N \right] \right\} = 0. \quad (2.3)$$

Poiché il differenziale di un funzionale è definito come

$$\delta F = \int \frac{\delta F}{\delta f(x)} \delta f(x) dx,$$

è possibile riscrivere la (2.3) come

$$\int \left\{ \frac{\delta E[\rho]}{\delta \rho(r)} - \mu \right\} \delta \rho(r) dr = 0. \quad (2.4)$$

Questa ultima relazione è denominata *equazione fondamentale del DFT* e implica che

$$\mu = \frac{\delta E[\rho]}{\delta \rho(r)} = v_{\text{ext}}(r) + \frac{\delta F_{HK}}{\delta \rho(r)} \quad (2.5)$$

## 2.9 Teorema di Kohn e Sham

L'equazione (2.5) ha una forma semplice, tuttavia il funzionale  $F_{HK}$  non è noto. Più precisamente non è conosciuto con molta precisione il funzionale dell'energia cinetica  $T[\rho]$ . D'altro canto  $T[\rho]$  è molto ben definito se è nota la funzione d'onda. Nel 1965 Kohn

e Sham (KS) proposero un formalismo per calcolare l'energia  $E[\rho]$ , basandosi su una funzione d'onda costituita da determinanti di Slater [13,14]. Il punto di partenza è un Hamiltoniano di riferimento con elettroni non interagenti tra loro e sottoposti solo all'azione del potenziale esterno  $v_s(r)$ . Questo Hamiltoniano è tale che la densità, che si ottiene dalla risoluzione delle equazioni ad esso associate, sia uguale alla densità reale del sistema. In questa approssimazione un siffatto Hamiltoniano si scrive:

$$H_{KS} = \sum_{i=1}^N h_i = \sum_{i=1}^N -\frac{1}{2} \nabla^2(i) + \sum_{i=1}^N v_s(i).$$

Questo Hamiltoniano ha come soluzione un determinante di Slater  $\psi_S$ . Gli orbitali molecolari di questo determinante sono ottenuti risolvendo l'equazione

$$\left[ -\frac{1}{2} \nabla^2 + v_s(r) \right] \phi_i = \varepsilon_i \phi_i$$

la cui risoluzione permette di determinare il valore del funzionale dell'energia cinetica senza interazione  $T_s$

$$T_s[\rho] = 2 \sum_{i=1}^{N_{occ}/2} \langle \phi_i | -\frac{1}{2} \nabla^2 | \phi_i \rangle$$

e la densità

$$\rho = 2 \sum_{i=1}^{N_{occ}/2} |\phi_i|^2$$

Per questo sistema senza interazione l'equazione fondamentale del DFT diventa

$$\mu = \frac{\delta T_s[\rho]}{\delta \rho[r]} + v_s(r). \quad (2.6)$$

Essendo note  $T_s[\rho]$  e  $\Delta T[\rho] = T[\rho] - T_s[\rho]$ , cioè la differenza tra l'energia cinetica del sistema con e senza interazione, l'obiettivo è scrivere l'energia del sistema  $E[\rho]$  in funzione di  $T_s[\rho]$ , che è nota in quanto calcolata con l'aiuto degli orbitali molecolari. Si può dimostrare infatti che

$$E[\rho] = T_s[\rho] + \int v_{ext}(r) \rho(r) dr + J[\rho] + E_{xc}[\rho]$$

dove

$$J[\rho] = \frac{1}{2} \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2,$$

$$E_{XC} = \Delta T[\rho] + V_{ee}[\rho] - J[\rho].$$

$J[\rho]$  è il potenziale di Coulomb mentre  $E_{XC}[\rho]$  è il funzionale di scambio-correlazione, in cui vengono inclusi tutti i termini che non sono noti esattamente, in particolare tutte le interazioni a molte particelle.

E' importante notare che il formalismo di Kohn-Sham non è basato univocamente sull'uso della densità elettronica in senso stretto, dacché il calcolo del termine  $T_S[\rho]$  viene eseguito a partire dagli orbitali molecolari. Tuttavia questa approssimazione è la maniera più efficace di applicare il DFT. L'applicazione dell'equazione fondamentale del DFT, ovvero la (2.5) permette, dopo alcuni riarrangiamenti di esprimere  $\mu$  nella forma seguente

$$\mu = \frac{\delta T_S[\rho]}{\delta \rho(r)} + v_{eff}(r) \quad (2.7)$$

dove  $v_{eff}(r) = v_{XC}(r) + v_{ext}(r) + \int \frac{\rho(r_2)}{r_{12}} dr_2$ , con  $v_{XC} = \frac{\delta E_{XC}[\rho]}{\delta \rho(r)}$ .

Confrontando la (2.7) con la (2.6) si constata che l'equazione che bisogna risolvere per un sistema interagente è simile a quella di un sistema senza interazioni, l'unica differenza è nell'uso di potenziali differenti. E' necessario allora risolvere

$$h_{KS}\phi_i = \varepsilon_i\phi_i \quad (2.8)$$

con l'Hamiltoniano di KS definito come

$$h_{KS} = -\frac{1}{2}\nabla^2 + v_{eff}(r)$$

Questa equazione essendo di tipo HF può essere risolta nel modo seguente. Innanzitutto viene adottato un insieme di orbitali molecolari di prova allo scopo di determinare la densità. Questa ultima permette il calcolo del potenziale effettivo  $v_{eff}(r)$ , che a sua volta consente di determinare le soluzioni della (2.8), autovalori e autovettori. Questa procedura deve essere ripetuta fino alla convergenza. Nell'espressione dell'Hamiltoniano di Kohn e Sham la sola incognita è il potenziale di scambio-correlazione  $v_{XC}$ . Più la conoscenza di questa quantità sarà precisa più la densità sarà conosciuta con precisione e dunque più l'energia risulterà vicina a quella esatta. Questo implica che è necessario trovare una

espressione per il funzionale di scambio-correlazione, che si avvicini il più possibile all'espressione esatta. A tale proposito sono state suggerite diverse approssimazioni. La più semplice consiste nell'approccio *Local Density Approximation*, che consiste nel considerare la densità assimilabile a quella di un gas uniforme di elettroni, che varia lentamente con la posizione. Invece la *Local Spin Density Approximation* è un modello LDA che introduce anche lo spin. Altri approcci sono i metodi *Generalised Gradient Approximation* GGA, i meta-GGA ed infine i funzionali ibridi.

## 2.10 Il modello SCF generalizzato

La soluzione del problema KS, espresso dall'equazione (2.8), è del tutto analoga alla procedura utilizzata per il modello HF, entrambi i metodi infatti si basano su una risoluzione di tipo self consistent field (SCF) per la determinazione dell'energia totale. Le equazioni SCF sono costruite espandendo le funzioni  $\phi_i$  in termini di un set di funzioni di base monoelettroniche (vedi Appendice A):

$$\phi_i(r) = \sum_{\mu=1}^m C_{i\mu} \chi_{\mu}(r)$$

in modo tale che la densità elettronica possa essere descritta in termini di queste funzioni e della corrispondente matrice densità P

$$\rho(r) = \sum_{\mu\nu} P_{\mu\nu} \chi_{\mu}(r) \chi_{\nu}(r)$$

L'energia totale SCF è :

$$E_{SCF} = \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu} P_{\lambda\sigma} (\mu\nu|\lambda\sigma) + E_{xc}$$

dove  $H_{\mu\nu}$  sono gli elementi di matrice dell'operatore monoelettronico

$$H^{core} = -\frac{1}{2}\nabla^2 + V_N$$

In un formalismo di spin non ristretto

$$\begin{aligned}\rho(r) &= \rho^\alpha(r) + \rho^\beta(r) \\ P &= P^\alpha + P^\beta\end{aligned}$$

e l'equazione di Kohn-Sham (2.8) è trasformata in due set di equazioni secolari, una per ogni funzione di spin, dalle quali sono ottenuti gli autovalori e le autofunzioni

$$F^\sigma C^\sigma = S C^\sigma \varepsilon^\sigma$$

con le condizioni di ortonormalità

$$(C^\sigma)^\dagger S C^\sigma = I$$

dove  $\sigma$  è riferito allo spin  $\alpha$  o  $\beta$ . Le matrici di Fock  $F^\sigma$  hanno elementi

$$F_{\mu\nu}^\sigma = \frac{\partial E_{SCF}}{\partial P_{\mu\nu}^\sigma}$$

$C^\sigma$  sono le matrici dei coefficienti degli orbitali molecolari,  $S$  è la matrice di sovrapposizione atomica e  $\varepsilon^\sigma$  sono le matrici diagonali degli autovalori KS. Ora la differenza sostanziale tra gli approcci HF e KS risiede nel termine  $E_{XC}$ , il quale è

$$E_{xc}^{HF} = -\frac{1}{2} \left( \sum_{\mu\nu\lambda\tau} P_{\mu\nu}^\alpha P_{\lambda\tau}^\alpha + P_{\mu\nu}^\beta P_{\lambda\tau}^\beta \right) (\mu\lambda|\nu\tau) \quad (2.9)$$

nella teoria HF, e

$$E_{xc}^{KS} = \int f(\rho^\alpha \rho^\beta) dr \quad (2.10)$$

nella teoria KS.

In molti casi l'integrale (2.10) non può essere calcolato analiticamente e quindi deve essere ottenuto mediante quadratura numerica.

## 2.11 Local Density Approximation

L'approssimazione LDA è la più semplice e consiste nel considerare la densità del sistema equivalente a quella di un gas di elettroni uniforme. Essa varia dunque lentamente in funzione della posizione  $r_1$ . Differenti valori per l'energia di scambio possono essere usati. Ad esempio l'energia di scambio formulata da Dirac ha la seguente forma

$$E_X^{LDA}[\rho] = -C_X \int \rho^{4/3}(r) dr ,$$

dove  $C_X = -1.5(3/8\pi)^{1/3}$ .

Il metodo  $X_\alpha$  proposto da Slater [15] invece suggerisce per il funzionale di scambio la formula seguente:

$$E_X^{LDA}[\rho] = -\frac{3}{2}\alpha C_X \int \rho^{4/3}(r) dr$$

dove  $\alpha$  è una variabile che quando assume il valore 2/3 consente di ritrovare l'espressione di Dirac. E' stato trovato che generalmente il valore 0.75 permette di raggiungere buoni risultati per atomi e molecole. L'energia di correlazione  $E_c[\rho]$  è nulla nel metodo elaborato da Slater.

Un funzionale di correlazione  $E_c$ , da utilizzare in connessione con quello di scambio definito da Dirac, è stato suggerito da Vosko, Wilk, Nusair [16] e da Perdew e Wang [17]. I parametri ottenuti per questo funzionale sono stati determinati sulla base dell'interpolazione di un insieme di valori di energia di correlazione ottenuti da calcoli Monte Carlo quantistici.

## 2.12 Local Spin Density Approximation

Per i sistemi a guscio chiuso le densità di spin  $\rho^\alpha$  e  $\rho^\beta$  sono identiche. Non è questo il caso di sistemi con orbitali parzialmente occupati. In tal caso bisogna risolvere le equazioni per

densità  $\alpha$  e  $\beta$  separatamente. Questa maniera di procedere è equivalente al metodo *unrestricted* Hartree-Fock ed è dunque denominato *Unrestricted Kohn-Sham* (UKS).

### 2.13 Generalised Gradient Approximation

Questo approccio introduce il gradiente della densità nella descrizione degli effetti di scambio e correlazione. A differenza del metodo LDA, che è locale, in questo caso vengono incorporati degli elementi non locali. In effetti, il gradiente della densità permette di prendere in considerazione la variazione della densità in prossimità di ciascun punto. Si scrive dunque

$$E_{xc}^{GGA}[\rho, \nabla \rho] = \int f(\rho, \nabla \rho) dr.$$

Si può dimostrare che lo scambio esatto presenta, per distanze  $r$  tra i due elettroni molto grandi un valore limite pari a  $-0.5/r$ , mentre nel caso dello scambio LDA tale limite tende esponenzialmente a zero. Anche se è difficile sviluppare potenziali GGA, sono stati proposti funzionali che consentono di ottenere i limiti corretti. Ad esempio, l'energia di scambio-correlazione nell'approccio non empirico di Perdew, Burke e Ernzerhof [18] è

$$\begin{aligned} E_X^{PBE}[\rho] &= \int \rho E_X^{LDA} F_X(s) dr, \\ F_X(s) &= 1 + \kappa - \frac{\kappa}{1 + bs^2 / \kappa}, \quad (2.11) \\ s &= \frac{|\nabla \rho|}{2\kappa_F \rho} \end{aligned}$$

mentre il funzionale di correlazione si scrive

$$E_C^{PBE} = \int dr \rho [E_C^{LDA} + H(t)],$$

$$\begin{aligned}
H(t) &= \gamma \ln \left\{ 1 + \frac{\beta t^2}{\gamma} \left[ \frac{1 + At^2}{1 + At^2 + A^2 t^4} \right] \right\}, \\
A &= \frac{\beta}{\gamma (\exp(-E_C^{LDA} / \gamma) - 1)}, \\
t &= \frac{|\nabla \rho|}{2k_S \rho}
\end{aligned} \tag{2.12}$$

I differenti parametri,  $k_S$ ,  $\alpha$ ,  $\gamma$ ,  $a$ ,  $b$ ,  $k_F$ ,  $\kappa$  che compaiono nelle equazioni (2.11) e (2.12) non vengono calibrati sulla base di dati sperimentali ma sono determinati dalle varie condizioni. Ad esempio quando la densità diventa costante, deve essere ritrovato il gas di elettroni. Dunque devono essere rispettati i comportamenti asintotici delle energie e dei potenziali.

## 2.14 Funzionali meta-GGA

In questo caso i funzionali di scambio-correlazione dipendono non solo dalla densità e dalla sua derivata prima, ma anche dalle sue derivate di ordine superiore. Si introduce a tale proposito la densità di energia cinetica  $\tau(r)$

$$\tau(r) = \frac{1}{2} \sum_{i=1}^{N_{occ}} |\nabla \phi_i(r)|^2$$

Questa dipende esplicitamente dagli orbitali occupati e ottenuti a seguito di un calcolo Kohn-Sham. Questi orbitali sono funzionali della densità, per cui

$$E_{XC}^{matGGA}[\rho] = \int f(\rho, \nabla \rho, \tau, \Delta \rho) dr,$$

dove  $\Delta \rho$  è il Laplaciano della densità.

## 2.15 La connessione adiabatca

Studi recenti hanno mostrato che grazie ad una parziale inclusione dello scambio HF (il che conduce ai cosiddetti metodi ibridi) i risultati ottenuti sono migliori di quelli che vengono dall'utilizzo delle GGA.

La recente introduzione dei funzionali ibridi da parte di Becke [19] è basata sulla formula di connessione adiabatca[20]. Questa formula è usualmente espressa nella forma :

$$E_{xc} = \int_0^1 U_{xc,\lambda} d\lambda$$

con

$$U_{xc,\lambda} = \langle \Psi_\lambda | V_{ee} | \Psi_\lambda \rangle - \frac{e^2}{2} \iint \frac{\rho(r)\rho(r')}{|r-r'|} dr dr' \quad (2.13)$$

dove  $\lambda$  è un parametro di accoppiamento elettronico, che modula la repulsione di Coulomb tra elettroni, e  $U_{XC,\lambda}$  è la corrispondente energia potenziale di scambio-correlazione per l'interazione elettrone-elettrone ad una forza intermedia di accoppiamento  $\lambda$ . Come evidenziato da Becke [21] questa formula connette il sistema di riferimento non interagente di Kohn-Sham (definito da  $\lambda = 0$ ) con il sistema reale pienamente interagente ( $\lambda = 1$ ), attraverso un continuo di sistemi parzialmente interagenti ( $0 \leq \lambda \leq 1$ ), i quali hanno tutti la stessa densità elettronica  $\rho$ , cioè la densità del sistema reale. Becke ha dimostrato che la più semplice approssimazione alla equazione (2.13) può essere espressa mediante la formula :

$$E_{xc}^{ACM} = \frac{1}{2} (E_{xc,\lambda=0} + E_{xc,\lambda=1}) \quad (2.14)$$

Un'applicazione della equazione (2.14) è l'approccio ACM definito *half-and-half* (h&h), in cui il termine di scambio è composto per metà di un contributo HF e per metà di un contributo DFT; in particolare si utilizza lo scambio HF per il sistema con  $\lambda = 0$ , mentre per il sistema pienamente interagente si da una stima dell'energia di scambio-correlazione ricorrendo all'approssimazione LSD.

Un più valido approccio è stato successivamente ottenuto dallo stesso Becke che ha proposto per il funzionale  $E_{XC}$  la seguente equazione a tre parametri :

$$E_{xc}^{ACM} = a_{x0} E_x^{LSD} + (1 - a_{x0}) E_x^{HF} + a_{x1} \Delta E_x^{GGA} + E_c^{LSD} + a_c \Delta E_c^{GGA}$$

dove  $\Delta E_x^{GGA}$  e  $\Delta E_c^{GGA}$  sono le correzioni di gradiente generalizzato per i contributi di scambio e di correlazione, mentre  $E_x^{LSD}$  e  $E_c^{LSD}$  sono i rispettivi contributi LSD. I tre parametri semiempirici  $a_{x0}$ ,  $a_{x1}$  e  $a_c$  sono stati determinati fittando i calori di formazione di un set standard di molecole. Nel suo lavoro originario Becke utilizzò il funzionale di correlazione di Perdew-Wang ottenendo la seguente approssimazione per il funzionale di scambio-correlazione :

$$E_{XC}^{B3PW91} = a_{x0}E_X^{LSD} + (1-a_{x0})E_X^{HF} + a_{x1}\Delta E_X^B + E_C^{PW91} + \Delta E_C^{PW91}$$

Se invece si utilizza il funzionale di scambio di Becke [22] insieme con il funzionale di correlazione LYP, con gli stessi parametri del B3PW91, si ottiene un altro funzionale ibrido, molto diffuso e popolare, che prende il nome di B3LYP.

In particolare i valori delle costanti sono:

$$a_{x0} = 0.20$$

$$a_{x1} = 0.72$$

$$a_c = 0.80$$

L'introduzione dello scambio esatto permette di migliorare i risultati anche se la frazione di scambio da usare dipende dal sistema studiato. Generalmente  $a_{x0} = 0.2$  rappresenta un valore corretto per numerosi sistemi.

## 2.16 I vantaggi dei metodi DFT

Il vantaggio evidente nell'utilizzo dei funzionali della densità consiste nella formulazione di un problema quantomeccanico nelle tre coordinate spaziali  $r$  rispetto alle  $3N$  variabili del problema, espresso invece in termini della funzione d'onda  $\Psi$  : il costo computazionale è notevolmente ridotto a tutto vantaggio della trattazione di sistemi chimicamente interessanti, solitamente inaccessibili per le dimensioni medio-alte e quindi per il numero elevato di elettroni attivi. A tal proposito occorre ricordare che negli approcci *post*-HF la quantità di calcoli necessaria alla risoluzione della equazione di Schrodinger cresce con la 7<sup>a</sup>-8<sup>a</sup> potenza del numero degli elettroni attivi, mentre nel caso degli approcci DFT tale aumento è molto meno impressionante (3<sup>a</sup>-4<sup>a</sup> potenza).

Un altro grande pregio di tali metodi consiste nel fatto che essi trattano con la stessa accuratezza sia sistemi a guscio chiuso che sistemi a guscio aperto, per tutti gli elementi della tavola periodica. Inoltre i metodi DF sono molto interessanti, poiché essi sono capaci di includere gran parte dell'energia di correlazione in un formalismo che richiede essenzialmente le stesse risorse del metodo Hartree-Fock; in particolare l'approccio di Kohn-Sham ha condotto ad implementazioni DF che sono molto simili ai corrispondenti algoritmi HF; d'altro canto, da un punto di vista chimico, tutti i metodi Kohn-Sham mantengono la semplicità interpretativa dei metodi a singolo determinante.

Dunque un ragionevole compromesso tra accuratezza e costi computazionali è quello di ricorrere ai metodi che derivano dalla Teoria del Funzionale della Densità elettronica. Questo è un aspetto notevole di cui bisogna tener conto se si vogliono studiare sistemi molecolari complessi con un certo grado di accuratezza e senza eccessivi costi computazionali.

Dall'inizio degli anni '90 del secolo scorso, grazie a questi numerosi aspetti vantaggiosi, la teoria del funzionale della densità, che sino ad allora occupava una posizione periferica in ambito chimico, essendo stata sviluppata per lo studio di solidi cristallini, grazie anche all'introduzione degli orbitali di Kohn e Sham, ha svolto un ruolo da protagonista nell'ambito della modellistica chimica. Oggi, la quasi totalità delle simulazioni da principi primi, fatte in ambito chimico, si basano sul metodo DFT, mentre solo 10-15 anni fa non più del 10% degli studi si basavano su questo approccio.

## CAPITOLO 3

### Meccanica Molecolare

Molti problemi che vengono affrontati con il *modelling* molecolare riguardano sistemi fisici troppo grandi per i metodi quantomeccanici anche quando vengono ignorati una parte degli elettroni, come negli schemi semiempirici. Viceversa i metodi *force field*, altrimenti noti come meccanica molecolare (MM), ignorano il moto degli elettroni e calcolano l'energia del sistema come funzione delle sole coordinate nucleari. Dunque la meccanica molecolare è invariabilmente usata per eseguire calcoli su sistemi molecolari con significativi numeri di atomi. In alcuni casi i campi di forza possono fornire risposte accurate come calcoli quantistici di buon livello in una frazione dei tempi macchina richiesti dai metodi più sofisticati. Tuttavia la meccanica molecolare non può fornire proprietà che dipendono dalla distribuzione elettronica nello spazio molecolare.

La meccanica molecolare ovviamente si basa sulla validità di diverse assunzioni. La prima di queste è l'approssimazione di Born-Oppheneimer senza la quale sarebbe impossibile scrivere l'energia come funzione di tutte le coordinate nucleari. La meccanica molecolare inoltre si fonda su un modello semplice per descrivere le interazioni di un sistema, con contributi derivanti da processi come lo *stretching* dei legami, il *bending*, la rotazione attorno ai legami semplici. Tuttavia anche quando si usino semplici funzioni, applicando ad esempio la legge di Hooke, per descrivere questi contributi, i campi di forza possono fornire risultati accettabili. Infine un attributo di fondamentale importanza di un campo di forza è la trasferibilità, che consente di applicare un set di parametri, sviluppati e testati su un relativamente piccolo numero di casi, ad un più ampio ventaglio di problemi di interesse fisico e chimico. Quindi i parametri, sviluppati a partire da dati su piccole molecole, possono essere usati per studiare molecole più grandi come i polimeri.

### 3.1 Un semplice campo di forza in meccanica molecolare

Molti campi di forza usati in meccanica molecolare possono essere interpretati in termini di una semplice descrizione a quattro componenti delle forze intra- e inter- molecolari all'interno del sistema. Funzioni tipo *penalties energetiche* vengono associate alle deviazioni degli angoli e dei legami dai loro valori di *riferimento* o di *equilibrio*, un'altra funzione, descrive come l'energia varia con la rotazione dei legami ed infine il campo di forza contiene anche termini che descrivono le interazioni tra parti del sistema non legate tra loro. Campi di forza più sofisticati possono includere dei termini aggiuntivi ma invariabilmente contengono quattro componenti di questo tipo. Una caratteristica di una siffatta rappresentazione è che i vari termini possono essere collegati alle variazioni di specifiche coordinate interne come lunghezze di legame, angoli, rotazioni attorno ai legami e movimenti di atomi rispetto ad altri atomi. Tale caratteristica permette di comprendere come le variazioni nei parametri del campo di forza influenzino le sue *performance* ed aiuta anche nel processo di parametrizzazione. La forma funzionale che permette la modellizzazione di singole molecole o insiemi di atomi e/o molecole è [11]

$$V(r^N) = \sum_{\text{legami}} \frac{k_l}{2} (l_i - l_{i,0})^2 + \sum_{\text{angoli}} \frac{k_a}{2} (\vartheta_i - \vartheta_{i,0})^2 + \sum_{\text{torsioni}} V_n \frac{(1 + \cos(n\omega - \gamma))}{2} + \sum_i^N \sum_{j=i+1}^N \left[ \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) 4\varepsilon_{ij} + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right] \quad (3.1)$$

$V(r^N)$  denota l'energia potenziale, che è una funzione delle posizioni  $r$  di  $N$  particelle (generalmente atomi). Il primo termine in questa equazione rappresenta l'interazione tra coppie di atomi legati ed attraverso un potenziale armonico fornisce l'incremento di energia che si ha quando la lunghezza di un legame  $l_i$  devia dal valore all'equilibrio  $l_{i,0}$ . Il secondo termine è una somma su tutti gli angoli di valenza della molecola, di nuovo rappresentato da un potenziale armonico. Il terzo termine è un potenziale torsionale, mentre il quarto contributo deriva da un termine *non-bonded* cioè di non legame, calcolato tra tutte le coppie di atomi  $i$  e  $j$  che sono in molecole differenti o nella medesima molecola, ma separati da almeno tre legami. In un campo di forza questo ultimo termine è

generalmente rappresentato da un potenziale Coulombiano, per le interazioni elettrostatiche, e da un potenziale di Lennard-Jones per le interazioni di van der Waals.

### 3.2 Caratteristiche generali di un campo di forza

Per definire un campo di forza è necessario specificare non solo la forma funzionale dell'energia potenziale ma anche i parametri, ad esempio le varie costanti  $k_{ij}$ ,  $V_n$  e  $\sigma_{ij}$  nell'equazione (3.1). La trasferibilità sia della forma funzionale che dei parametri è un'importante caratteristica di un campo di forza, che debba essere usato per eseguire calcoli a carattere predittivo. Trasferibilità, come già accennato, significa che lo stesso set di parametri può essere usato per descrivere una serie di molecole tra loro correlate, piuttosto che definire un nuovo set di parametri per ogni singola molecola. Un punto importante da tenere in considerazione è che i campi di forza sono empirici; non c'è una forma corretta per un campo di forza. Se una forma funzionale dimostra di poter conseguire performance migliori di un'altra è probabile che venga favorita. Le forme funzionali usate in meccanica molecolare sono generalmente un buon compromesso tra accuratezza ed efficienza computazionale, d'altronde aumentando le performance dei computer diventa progressivamente possibile incorporare modelli sempre più sofisticati, per una sempre più realistica descrizione delle interazioni presenti in un dato sistema [24]. Un concetto comune alla maggior parte dei campi di forza è quello di *tipo atomico*. Quando si prepara un *input* per un calcolo quantomeccanico è generalmente necessario specificare i numeri atomici dei nuclei presenti, insieme con la geometria del sistema, la carica complessiva e la molteplicità di spin. Per un campo di forza carica e molteplicità non sono esplicitamente richieste ma è necessario assegnare un tipo di atomo a ciascun atomo presente nel sistema. Il tipo di atomo è molto più del numero atomico, in quanto esso fornisce anche informazioni circa lo stato di ibridizzazione e talvolta anche sull'*environment* chimico locale. Nel campo di forza AMBER [23] all'atomo di carbonio tra gli anelli a 5 e 6 termini (ad esempio nell'amminoacido triptofano) è assegnato un tipo atomico diverso da quello assegnato all'atomo di carbonio in un anello isolato a cinque atomi (ad esempio nell'istidina) che a sua volta è differente dal tipo atomico di un atomo di carbonio nel benzene. In effetti i campi di forza sviluppati per il *modelling* di specifiche

classi di molecole (soprattutto proteine e acidi nucleici nel caso di AMBER) utilizzano un maggior numero di specifici tipi atomici rispetto ai campi di forza *general-purpose*.

### 3.3 Stretching di legame

Per descrivere la curva di energia potenziale di un tipico legame chimico sono state proposte varie forme funzionali tra cui il potenziale di Morse. Tuttavia in meccanica molecolare raramente si effettuano calcoli in cui i legami sono lontani dai rispettivi valori di equilibrio, per cui l'approccio più elementare consiste nell'usare la formula della legge di Hooke, in cui l'energia varia con il quadrato dello spostamento dalla lunghezza di legame di riferimento

$$v(l) = \frac{k}{2}(l - l_0)^2$$

In particolare l'energia potenziale di stretching è la somma delle energie potenziali di stretching di ogni legame della molecola. Campi di forza come AMBER [23] usano solo una funzione quadratica, mentre altri introducono per il potenziale di stretching anche termini cubici o quartici.

Un particolare campo di forza conterrà i valori per i parametri  $k$  e  $l_0$  per tutti i possibili legami formati dai tipi atomici di quel campo di forza.

### 3.4 Bending di legame

Anche l'energia potenziale di bending è presa come la somma di energie potenziali per il bending di ciascun angolo di legame della molecola e la deviazione degli angoli dai loro valori di riferimento è frequentemente descritta adottando un potenziale armonico

$$v(\vartheta) = \frac{k}{2}(\vartheta - \vartheta_0)^2$$

Il contributo di ogni angolo è caratterizzato da una costante di forza e da un valore di riferimento. Ovviamente è necessaria meno energia per conseguire la distorsione di un

angolo dall'equilibrio piuttosto che per "stirare" o comprimere un legame, dunque le costanti di forza sono proporzionalmente più piccole che nel caso dello stretching.

Come per i termini di legame, l'accuratezza del campo di forza può essere migliorata incorporando termini di ordine superiore a quello quadratico.

### 3.5 Potenziale torsionale

Lo stretching di legame ed il bending degli angoli di valenza vengono definiti gradi di libertà *hard*, in quanto sono necessarie quantità sostanziali di energia per causare apprezzabili deformazioni dai rispettivi valori di riferimento. Tuttavia una parte significativa delle variazioni strutturali ed energetiche di un certo sistema è dovuto ad un delicato equilibrio tra contributi torsionali e di non legame.

L'esistenza di barriere rotazionali intorno ai legami chimici è fondamentale per comprendere le proprietà strutturali delle molecole e la relativa analisi conformazionale. I tre minimi di energia delle conformazioni *staggered* ed i tre minimi di energia delle strutture *eclipsed* dell'etano sono un classico esempio di come l'energia cambi con la rotazione intorno al legame chimico. Calcoli quantomeccanici suggeriscono che questa barriera di rotazione può essere causata dalle interazioni antilegame tra gli atomi di idrogeno sulle estremità opposte della molecola e che tali interazioni vengono minimizzate quando le conformazioni sono *staggered* mentre sono su un massimo dell'energia quando la conformazione è *eclipsed*. Affinché un campo di forza possa simulare la variazioni conformazionali dovute alle rotazioni attorno ai legami è essenziale che esso rappresenti propriamente i profili di energia associati a tali variazioni. La maggior parte dei campi di forza (anche se non tutti) che sono stati sviluppati per molecole organiche usa in maniera esplicita potenziali torsionali con un contributo da ciascuno dei quartetti di atomi legati A-B-C-D nel sistema. Allora, ad esempio, ci saranno 9 termini torsionali per l'etano e 24 per il benzene. I potenziali torsionali sono spesso espressi come un'espansione in serie di funzioni coseno, dunque una possibile forma funzionale è

$$v(\omega) = \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \quad (3.2)$$

dove  $\omega$  è l'angolo di torsione mentre  $V_n$  è denominato altezza di barriera, anche se questa terminologia è da considerarsi ovviamente fuorviante quando sia presente nell'espansione più di un termine; d'altro canto però,  $V_n$  fornisce una indicazione qualitativa delle barriere di rotazione relative; per esempio esso sarà maggiore per un legame ammidico piuttosto che per un legame semplice tra due atomi di carbonio  $sp^3$ . Nella formula (3.2)  $n$  di  $V_n$  è la molteplicità, fornisce perciò il numero di punti di minimo nella funzione al variare dell'angolo di rotazione da  $0^\circ$  a  $360^\circ$ ;  $\gamma$  è il fattore di fase che determina dove l'angolo torsionale passa per il suo valore di minimo. Ad esempio un doppio legame tra due atomi di carbonio  $sp^2$  avrà  $n=2$  e  $\gamma=180^\circ$ , con minimi a  $0^\circ$  e  $180^\circ$ .

Molti termini torsionali nel campo di forza AMBER contengono solo un contributo dall'espansione in serie dei coseni, ma per alcuni legami è necessario includere più di un termine. Un'altra caratteristica di AMBER è il suo uso dei parametri torsionali: il profilo di energia della rotazione attorno ad un legame dipende soltanto dal tipo di atomo dei due atomi centrali e non dal tipo di atomo degli atomi terminali. Per esempio a tutti gli angoli diedri nei quali il legame centrale è tra due atomi di carbonio ibridizzati  $sp^3$  sono assegnati gli stessi parametri torsionali tranne la torsione in un caso speciale come O-C-C-O. Nel suo trattamento del potenziale torsionale AMBER assume una posizione intermedia tra campi di forza che usano un solo termine nell'espansione di  $v(\omega)$  e quelli che sistematicamente usano più termini per tutti gli angoli diedri. Con una attenta parametrizzazione un campo di forza che utilizza più di un termine nell'espansione di  $v(\omega)$  come il campo di forza MM2 [25] dovrebbe conseguire risultati più accurati rispetto a quello che utilizza un solo termine, con uno svantaggio legato però alla necessità di un gran numero di parametri anche per simulare un modesto novero di molecole e sistemi.

### 3.6 Bending out of plane

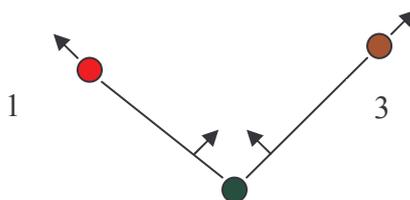
La molecola di ciclobutanone  $C_3H_6C=O$  presenta un atomo di ossigeno legato attraverso un doppio legame ad uno degli atomi di carbonio di un anello a 4 membri. Il carbonio carbonilico ed i tre atomi ad esso legati giacciono nello stesso piano. L'angolo di legame CCC al carbonio carbonilico è vicino a  $90^\circ$  e gli angoli di legame CCO sono prossimi a

135°. Poiché questi angoli deviano considerevolmente dagli angoli di riferimento di circa 120° al carbonio carbonilico, i termini di bending determineranno lo spostamento della posizione di equilibrio dell'ossigeno al di sopra del piano dell'anello di atomi di carbonio, per cui gli angoli CCO assumeranno un valore molto più vicino a 120°. Onde assicurare la planarità dei tre atomi legati al carbonio carbonilico (favorita dal legame  $\pi$  tra C e O) viene incluso nel potenziale anche un termine di bending *out-of-plane* (*oop*) ad ogni carbonio carbonilico. Questo termine può avere una forma del tipo  $\frac{1}{2}k_{oop}\chi_{oop}^2$ , dove  $\chi_{oop}$  è l'angolo tra il legame CO ed il piano formato dal carbonio del carbonile e dai due atomi ad esso legati. Simili termini *out-of-plane* vengono utilizzati per rinforzare la planarità all'atomo di azoto nelle ammidi [26]. Un termine *out-of-plane* viene spesso usato anche per il carbonio legato con un doppio legame ad un altro carbonio, dacché questo permette di riprodurre meglio le frequenze vibrazionali.

Al posto di termini di bending *oop* alcuni campi di forza usano termini torsionali impropri per ottenere il medesimo risultato. Un angolo di torsione si dice improprio quando i quattro atomi non sono legati nella sequenza 1-2-3-4. È interessante osservare che non sempre è necessario introdurre questi termini *oop*, che possono anche avere effetti deleteri sulle performance del campo di forza.

### 3.7 Termini incrociati

La presenza di termini incrociati in un campo di forza riflette l'accoppiamento tra coordinate interne. Ad esempio, se un angolo di valenza diminuisce, i legami adiacenti si allungano per ridurre le interazioni tra gli atomi 1,3 come illustrato nella figura riportata sotto.



I termini incrociati sono risultati essere importanti nei campi di forza progettati per predire gli spettri vibrazionali e che sono stati anche i precursori degli attuali campi di forza, per cui non sorprende che tali termini siano spesso inclusi nei campi di forza della meccanica molecolare per raggiungere performance ottimali. In principio dovrebbero essere inclusi termini incrociati tra tutti i contributi al campo di forza. Tuttavia solo pochi termini incrociati sono davvero necessari per riprodurre accuratamente le proprietà strutturali, mentre generalmente le interazioni, che coinvolgono moti che sono lontani in una molecola possono essere posti uguali a zero. Varie forme funzionali sono possibili per i termini incrociati, ad esempio per il contributo stretching-stretching da due legami denominati 1 e 2 si può scrivere

$$v(l_1, l_2) = \frac{k_{l_1, l_2}}{2} [(l_1 - l_{1,0})(l_2 - l_{2,0})]$$

Maple, Dinur e Hagler hanno usato calcoli quantomeccanici per valutare l'importanza dei termini incrociati trovando in ultima analisi che quelli più importanti sono stretching-stretching, stretching-bending, bending-bending. [27].

### 3.8 Interazioni di non legame

Molecole indipendenti ed atomi interagiscono mediante forze *non-bonded*, le quali giocano un ruolo importante anche nel determinare la struttura di specie molecolari individuali. Tali forze hanno scarsa importanza nel definire la struttura "locale" ma sono i principali fattori nel determinare la struttura "globale" delle grandi molecole (si pensi ad esempio al ripiegamento delle proteine)

Tali interazioni non dipendono da una relazione di *bonding* tra atomi ma sono interazioni "attraverso lo spazio", solitamente modellizzate come una funzione di qualche potenza inversa della distanza. I termini *non-bonded* in un campo di forza vengono solitamente classificati in due gruppi: interazioni elettrostatiche e interazioni di van der Waals.

### 3.9 Interazioni elettrostatiche

La distribuzione di carica in un sistema molecolare può essere rappresentata in vari modi, un approccio semplice ed abbastanza comune consiste nell'assegnare considerate cariche puntiformi a ciascun atomo, capaci di riprodurre le proprietà elettrostatiche della molecola (cariche parziali atomiche).

Il contributo elettrostatico  $V_{es}$  al potenziale è comunemente calcolato come la somma delle interazioni elettrostatiche coinvolgenti tutte le coppie di atomi eccetto le coppie 1,2, e 1,3, ovvero  $V_{es} = \sum_{1 \geq 4} V_{es,ij}$  in cui gli atomi  $i$  e  $j$  hanno almeno una relazione 1,4 e ogni termine  $V_{es,ij}$  viene calcolato assegnando cariche parziali  $Q_i$  ed applicando la legge di Coulomb

$$V_{es,ij} = \frac{Q_i Q_j}{\epsilon_r r_{ij}}$$

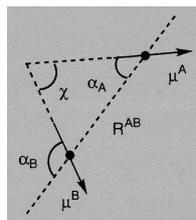
Vari metodi vengono usati per assegnare le cariche parziali atomiche. In Amber le cariche vengono trovate in modo da riprodurre potenziali elettrostatici calcolati a livello Hartree-Fock/6-31G\* di analoghe molecole più piccole. Ad esempio per le cariche sugli atomi del residuo amminoacidico  $-\text{NHCH}(\text{R})\text{C}(\text{O})-$  di una proteina si usano le cariche degli atomi corrispondenti della molecola  $\text{CH}_3\text{C}(\text{O})\text{NHCHRC}(\text{O})\text{NH}_2$ . In maniera analoga si procede per le cariche parziali di altri biopolimeri come DNA e RNA.

In OPLS, le cariche parziali sono assegnate agli atomi delle proteine sulla base del tipo atomico. Per esempio a tutti gli atomi di ossigeno del gruppo ammidico viene assegnata carica parziale -0.50 mentre ai carboni del medesimo gruppo una carica pari a +0.50. Diversi altri schemi basati ad esempio sull'elettronegatività sono stati proposti per assegnare le cariche parziali [11,26].

Alcuni campi di forza riducono  $V_{es,ij}$  di ogni coppia di atomi 1,4 moltiplicando per un fattore di scaling (0.75 nel campo di forza MMFF94 e 0.5 in AMBER).

In alternativa, il calcolo dell'energia elettrostatica può essere eseguito assegnando a ciascun legame un momento di dipolo. L'interazione tra due dipoli illustrata nella figura, riportata di seguito, è data da

$$V_{es}(\mathbf{R}) = (\mu_A \mu_B / \epsilon R^3) (\cos \chi - 3 \cos \alpha_A \cos \alpha_B)$$



Se parametrizzate in maniera corretta, non esistono sostanziali differenze nelle prestazioni di queste due rappresentazioni del potenziale elettrostatico  $V_{es}$ ; anzi nel caso di distanze di interazione molto grandi i due metodi sono perfettamente equivalenti.

La costante dielettrica  $\epsilon$  nelle due descrizioni può essere modificata in modo da rappresentare l'effetto delle molecole di solvente o il fatto che alcune interazioni tra siti molto distanti possono in realtà avvenire "attraverso" parti della stessa molecola (effetti di polarizzazione). Un valore di 1 per  $\epsilon$  corrisponde al vuoto, mentre un valore elevato riduce l'importanza delle interazioni tra cariche a lunga distanza. In genere si usa un valore compreso tra 1 e 4. In alcune applicazioni la costante dielettrica è resa dipendente dalla distanza  $\epsilon = \epsilon_0 R$ . Non esistono giustificazioni teoriche per questa pratica, ma l'efficienza del calcolo aumenta considerevolmente (in quanto si evita un'operazione di estrazione di radice quadrata) e i risultati ottenuti sembrano ragionevoli.

### 3.10 Cariche puntiformi

Ritorniamo alle cariche puntiformi in maniera più dettagliata anche perché questa rappresentazione del potenziale elettrostatico, come vedremo, rappresenta un aspetto rilevante del protocollo computazionale proposto in questo lavoro di Tesi.

Dato l'uso diffuso del modello delle cariche atomiche parziali è importante considerare come sono ottenute. Per specie chimiche semplici le cariche atomiche richieste per riprodurre i momenti elettrici possono essere calcolate esattamente se la geometria è nota. Ad esempio il momento di dipolo di HF può essere riprodotto ponendo cariche opposte pari a  $0.413e$  su i due nuclei ed assumendo una lunghezza di legame pari a  $0.917 \text{ \AA}$ . Analogamente la geometria tetraedrica del metano  $\text{CH}_4$  implica che ogni atomo di idrogeno abbia carica uguale ad un quarto di quella del carbonio. La molecola elettricamente neutra presenta momenti di dipolo e quadrupolo pari a zero ed un momento di ottupolo non nullo,

che può essere riprodotto usando un carica sull'idrogeno approssimativamente uguale a  $0.14e$ .

In alcuni casi le cariche atomiche sono scelte in modo da riprodurre proprietà termodinamiche calcolate con simulazioni Monte Carlo oppure di dinamica molecolare: viene eseguita una serie di simulazioni ed il modello di carica viene modificato fino ad ottenere un soddisfacente accordo con i dati sperimentali. Questo approccio viene considerato il solo realmente praticabile per piccole molecole o semplici modelli.

Le proprietà elettrostatiche di una molecola sono una conseguenza della distribuzione degli elettroni e dei nuclei e dunque sarebbe ragionevole assumere che un set di cariche parziali atomiche possano essere ottenute dai calcoli quantomeccanici. Tuttavia le cariche parziali non sono quantità sperimentalmente osservabili e non possono essere calcolate in maniera non ambigua dalla funzione d'onda. Questo spiega perché siano stati proposti numerosi metodi per determinare le cariche parziali. Questi set di cariche ad ogni modo essendo dipendenti dalla costituzione della molecola, da come gli atomi sono legati tra loro, vengono considerati inappropriati per rappresentare accuratamente le interazioni tra molecole e riprodurre quindi proprietà come il potenziale elettrostatico. L'importanza del potenziale elettrostatico nelle interazioni intermolecolari ha ispirato la messa a punto di schemi che calcolano le cariche in maniera consistente con questa proprietà.

### **3.11 Cariche derivate dal potenziale elettrostatico molecolare**

Il potenziale elettrostatico in un punto dello spazio è la forza che agisce su una carica positiva posizionata in quel punto. Esso è una quantità osservabile che può essere ricavata dalla funzione d'onda. Inoltre il potenziale elettrostatico è una proprietà continua e non è facilmente rappresentato da una funzione analitica, conseguentemente è necessario derivare una rappresentazione discreta. L'obiettivo è ottenere un set di cariche atomiche parziali che meglio riproduca il potenziale elettrostatico quantomeccanico in una serie di punti intorno alla molecola. Una soluzione a questo problema è stata suggerita da Cox e Williams [28]. Il potenziale elettrostatico in ciascuno dei punti scelti è calcolato dalla funzione d'onda e una procedura di minimi quadrati viene usata per determinare il set di cariche atomiche parziali, che meglio riproducono il potenziale elettrostatico in quei punti,

con il vincolo che la somma delle cariche sia uguale alla carica netta del sistema. La procedura di *fitting* minimizza la somma dei quadrati delle differenze nel potenziale elettrostatico. Allora se  $\phi_i^0$  è il potenziale elettrostatico in un dato punto e se  $\phi_i^{calc}$  è il valore ottenuto dal modello di carica, l'obiettivo è minimizzare la seguente funzione

$$R = \sum_{i=1}^{N_{punti}} w_i (\phi_i^0 - \phi_i^{calc})^2$$

dove  $N_{punti}$  è il numero di punti e  $w_i$  un fattore che pesa punti diversi con differenti gradi di "importanza" nel processo di *fitting*.

I punti  $i$  in cui il potenziale viene fittato possono essere scelti in vari modi ma ovviamente è importante prendere punti nella regione spaziale dove è più importante modellizzare correttamente le interazioni intermolecolari. Questa regione è appena oltre i raggi di van der Waals degli atomi coinvolti. Oltre a quello di Cox e Williams sono state proposti altri metodi come la procedura CHELP, che utilizza i moltiplicatori di Lagrange invece dei minimi quadrati e adotta una griglia cubica di punti da cui vengono scartati tutti i punti che giacciono all'interno del raggio di van der Waals e quelli distanti più di 2.8 Å da ogni atomo.

L'algoritmo di Singh e Kollman inizialmente proposto per derivare le cariche nel campo di forza Amber utilizzava punti su una serie di superfici molecolari, costruite usando raggi di van der Waals crescenti per gli atomi [29]. Successivamente è stata usata una versione modificata di questo metodo, che ha preso il nome di *Restrained Electrostatic Potential Fit* o RESP [30]. L'algoritmo RESP usa dei vincoli sugli atomi più pesanti dell'idrogeno H, che hanno l'effetto di ridurre le cariche su alcuni atomi. Le cariche RESP variano poco con la conformazione molecolare.

Generalmente la MM è usata per il *modelling* di sistemi costituiti da migliaia di atomi come i polimeri. Come vengono derivate le cariche per tali specie? Chiaramente non è possibile usare di routine calcoli quantistici su una molecola con così tanti atomi, per cui il sistema deve essere suddiviso in frammenti di opportune dimensioni. In molti casi i sistemi polimerici sono costruiti legando insieme chimicamente le unità monometriche. Le cariche per ogni monomero potrebbero allora essere ottenute da calcoli realizzati su opportuni frammenti che ricreino la situazione chimica locale del frammento all'interno della macromolecola. Per esempio le cariche atomiche parziali per gli amminoacidi sono spesso

ottenute da calcoli su frammenti dipeptidici, che meglio riflette l'ambiente dell'amminoacido nella proteina, rispetto all'amminoacido isolato.

I set di carica ottenuti dal *fitting* del potenziale elettrostatico possono dipendere dal set di base usato per derivare la funzione d'onda. Tuttavia è stato trovato che non sempre le cariche migliorano adottando set di base più grandi e che il set di base 6-31G\* fornisce, in genere, risultati ragionevoli.

### 3.12 Interazioni di van der Waals

Il contributo di van der Waals è generalmente la somma delle interazioni che coinvolgono tutte le possibili interazioni 1,4, 1,5, 1,6 ... di coppie di atomi, mentre le interazioni 1,2, e 1,3 sono considerate implicitamente incluse nei parametri di *stretching* e *bending*. Ogni termine di interazione di van der Waals è la somma di un'attrazione dovuta alle forze di dispersione di London e di una repulsione principalmente dovuta alla repulsione di Pauli. AMBER ed altri campi di forza adottano un potenziale di Lennard-Jones 12-6 mentre altri campi di forza come MM2 usano una funzione con un termine attrattivo proporzionale a  $r^{-6}$  ed un termine repulsivo proporzionale a  $e^{-ar}$ , dove  $a$  è un parametro.

Le interazioni *non bonded* consumano gran parte del tempo necessario a calcolare il potenziale di un sistema molecolare molto grande. Per rendere più veloci i calcoli di meccanica molecolare sulle macromolecole molti programmi usano un *cutoff* in modo che molti contributi di van der Waals vengano omessi per coppie di atomi distanti tra loro più di una certa distanza fissata. Le interazioni di van der Waals sono proporzionali a grandi distanze a  $1/r^6$  (*short range*) e dunque tipicamente viene adottata una distanza di *cutoff* di 8 Å.

### 3.13 Legame a idrogeno

Il legame idrogeno richiede un'attenzione speciale. Questi legami sono formati tra gli atomi di idrogeno legati ad atomi elettronegativi e lone pairs. Alcuni campi di forza sostituiscono il potenziale 6-12 di Lennard-Jones con un termine esplicito per il legame a idrogeno, spesso adottando un potenziale 10-12 di Lennard-Jones, che ha una forma

funzionale di questo tipo  $A/r^{12} - C/r^{10}$ . Questa funzione viene adottata per rappresentare l'interazione tra l'atomo di idrogeno del donatore e l'eteroatomo accettore ed il suo uso è inteso a migliorare l'accuratezza con la quale è predetta la geometria del sistema del legame a ponte di idrogeno. A volte viene modificato anche il funzionale, di modo che tenga conto anche di un termine direzionale. Altri campi di forza incorporano funzioni più complicate mentre altri ancora non contengono termini espliciti per il legame a idrogeno, che viene riprodotto attraverso la combinazione di interazioni elettrostatiche e di van der Waals.

### 3.14 Parametrizzazione di un campo di forza

La procedura generale per trovare i valori dei parametri di un campo di forza consiste nell'usare dati sperimentali o risultati teorici *ab initio*, onde scegliere un set iniziale di valori dei parametri e dunque variare questi parametri in modo da minimizzare le deviazioni delle proprietà molecolari predette dal campo di forza rispetto alle medesime proprietà misurate sperimentalmente o calcolate a livello quantomeccanico, di un set di molecole opportunamente scelto, che prende il nome di set di riferimento o *training set*. Le proprietà molecolari utilizzate nella parametrizzazione includono strutture molecolari, differenze di energie conformazionali, frequenze vibrazionali, barriere di rotazione, momenti di dipolo ed interazioni intermolecolari. I primi campi di forza si basavano esclusivamente sui dati sperimentali ma col tempo i calcoli *ab initio*, capaci di trattare accuratamente molecole sempre più grandi, sono stati usati con una frequenza crescente nelle procedure di parametrizzazione [31].

Come stime iniziali delle lunghezze e degli angoli di riferimento possono essere presi in considerazione tipici valori sperimentali o valori calcolati con metodi *ab initio*, valori che possono essere migliorati in modo da minimizzare gli errori nelle lunghezze ed angoli di legame delle molecole del *training set*. Le costanti di forza di bending e di stretching nelle molecole poliatomiche non sono quantità sperimentalmente osservabili, poiché le frequenze vibrazionali sono collegate in maniera complicata a tali costanti di forza. Tuttavia è possibile scrivere un'espressione approssimata per l'energia potenziale vibrazionale come somma di termini quadratici di stretching di legame, di angoli di

bending indipendenti, di torsioni che coinvolgano angoli diedri indipendenti, più vari termini incrociati. Un tale campo di forza detto di valenza non è un campo di forza di meccanica molecolare, che, invece, include anche interazioni *non-bonded* ed in cui gli angoli di bending e gli angoli diedri non sono tutti indipendenti tra loro. Però usando un siffatto campo di forza di valenza, combinato con dati di frequenze vibrazionali di un certo numero di composti organici, Schachtschneider e Snyder [24] hanno ottenuto costanti di forza di *stretching* e di *bending* che sono approssimativamente trasferibili e che possono essere usate come stime iniziali di questi parametri. Un modo alternativo per ottenere stime iniziali delle costanti di forza è attraverso le derivate seconde analitiche di energie elettroniche molecolari *ab initio*. Le costanti di forza possono poi essere migliorate minimizzando gli errori sulle frequenze vibrazionali predette dal campo di forza.

Per i parametri di van der Waals si assumono invece regole e formule empiriche più o meno elaborate che fanno uso di dati di energie di sublimazione di composti molecolari, le quali dipendono proprio dalle interazioni di van der Waals. Alcuni campi di forza come Amber94 e OPLS regolano i parametri di van der Waals in modo da conseguire buoni risultati per densità ed entalpie di vaporizzazione trovate con simulazioni Monte Carlo di composti liquidi.

Poiché tutti i parametri in un campo di forza influenzano tutte le proprietà molecolari calcolate, la parametrizzazione è una procedura iterativa. Per adottare un approccio di meccanica molecolare c'è bisogno di molti dati per parametrizzare il campo di forza, infatti questo metodo è di difficile applicazione a composti nuovi, anche se i calcoli *ab initio* possono aiutare molto nel processo di parametrizzazione. Un campo di forza empirico può contenere centinaia o addirittura migliaia di parametri: costanti di forza, lunghezze e angoli di legame di riferimento, parametri torsionali, di van der Waals, cariche elettrostatiche...

### **3.15 Limiti e punti di forza della meccanica molecolare**

Il principale vantaggio dei modelli della meccanica molecolare rispetto a modelli più sofisticati come quelli quantomeccanici è la loro semplicità. Eccetto che per sistemi molto piccoli il costo computazionale è quasi completamente dominato dalla valutazione dei

termini Coulombiani e di van der Waals, il cui numero dipende dal quadrato del numero di atomi nel sistema. Tuttavia questi termini vanno a zero rapidamente all'aumentare della distanza interatomica ed in pratica il costo computazionale scala linearmente con la dimensione molecolare per molecole sufficientemente grandi. Dunque i calcoli della meccanica molecolare possono essere eseguiti, senza particolari problemi, su macromolecole costituite da migliaia di atomi. Inoltre i calcoli sono sufficientemente rapidi da permettere ricerche conformazionali in sistemi contenenti fino a 100-200 atomi. L'analisi conformazionale è forse l'applicazione più importante della meccanica molecolare. Il fatto che i modelli della meccanica molecolare classica siano parametrizzati può rappresentare anche un vantaggio rispetto ai modelli della chimica quantistica. Infatti è possibile, almeno in linea di principio, costruire modelli di meccanica molecolare che riproducono accuratamente i dati sperimentali di sistemi noti e magari anticipare dati ed informazioni su composti "incogniti" ma strettamente relazionati a quelli che già sono stati studiati.

I modelli di meccanica molecolare presentano tuttavia anche importanti limiti. Innanzitutto essi sono limitati essenzialmente alla descrizione di geometrie e conformazioni di equilibrio. Inoltre i calcoli non forniscono ovviamente alcuna informazione sul *bonding* e sulla distribuzione elettronica molecolare, che è un aspetto chiave nella modellizzazione della reattività e selettività chimica. I campi di forza disponibili non sono stati parametrizzati per trattare forme di non-equilibrio, come stati di transizione di reazione, anche se non c'è nessuna fondamentale ragione affinché questo non possa essere eseguito, magari usando i risultati dei calcoli quantomeccanici come base per la parametrizzazione. Infine è bene sottolineare che la meccanica molecolare è essenzialmente uno schema di interpolazione, il successo del quale dipende non solo da buoni parametri ma anche da relazioni sistematiche tra sistemi molecolari. La meccanica molecolare dunque non è capace generalmente di conseguire particolari successi nel caso di nuove molecole, al di fuori del *range* di parametrizzazione.

Per concludere, bisogna osservare che la MM è un metodo che non permette di stabilire l'affidabilità del risultato ottenuto all'interno della teoria stessa (come invece accade per molti metodi quantomeccanici). La qualità di un risultato può essere valutata solo mediante un confronto con i dati sperimentali.

### 3.16 Performance e qualità dei campi di forza

Quando un campo di forza parametrizzato con cura è applicato a composti simili a quelli adottati per la parametrizzazione si possono ottenere risultati estremamente accurati. Per composti organici monofunzionali un campo di forza come MM3 commette errori sul calcolo delle variazioni di entalpia confrontabili con l'errore sperimentale, mentre fornisce lunghezze di legame con un errore di 0.01 Å ed angoli di legame con errori al di sotto di 2°, rispetto ai corrispondenti valori determinati attraverso gli esperimenti. Naturalmente non è lecito attendersi altrettanto buone performance quando la meccanica molecolare è applicata a classi di composti molto ampie. Per un campo di forza con 60 tipi di atomi il numero di possibili angoli diedri è circa  $3 \times 10^6$ . Nessun campo di forza disponibile avrà, ad esempio, i parametri necessari per tutti i possibili angoli diedri e frequentemente ci si trova di fronte al problema della mancanza di parametri, anche se la maggior parte dei programmi di meccanica molecolare utilizza una qualche procedura per stimare un valore da associare al parametro mancante. Quando, però, vengono usati tali stime per i parametri, la attendibilità del risultato potrebbe in alcuni casi diminuire. Da questo punto di vista i campi di forza sviluppati per le proteine ed attualmente disponibili presentano carenze talvolta significative.

L'introduzione di interfacce grafiche *user friendly* già a partire dagli anni novanta ha condotto ad una diffusione notevole dei programmi di meccanica molecolare. In generale un buon programma informa sempre l'utilizzatore della qualità dei parametri usati, in quanto questa informazione è utile per valutare la qualità dei risultati. Alcuni programmi però utilizzano il necessario numero di parametri "generici" per effettuare il calcolo e, in casi estremi, è possibile che per una particolare molecola non esista nessun buon parametro e quindi si ottengano risultati privi di senso. La possibilità di effettuare un calcolo non dà nessuna garanzia sulla qualità dei risultati ottenuti.

La qualità dei parametri di un campo di forza è essenziale per giudicare quanto siano credibili i risultati. Se la molecola in questione utilizza parametri basati su un gran numero di dati sperimentali di buona qualità i risultati del calcolo possono essere affidabili almeno quanto un risultato sperimentale. D'altra parte se vengono impiegati solo pochi dati sperimentali e molti parametri "generici" i risultati dovrebbero essere considerati con cautela. Infatti se questi parametri di scarsa qualità vengono usati per descrivere parti poco

importanti della molecola la situazione non presenta problemi (in molti casi queste parti della molecola possono essere eliminate o sostituite da gruppi più semplici). Al contrario se i parametri di scarsa qualità influenzano direttamente la proprietà di interesse i risultati possono essere potenzialmente fuorvianti.

### 3.17 Differenze tra i campi di forza

Attualmente vengono usati molti campi di forza differenti, e altri ne verranno sviluppati. Tutti differiscono essenzialmente per tre aspetti fondamentali:

- 1) La forma della funzione impiegata per ciascun termine energetico
- 2) Il numero di termini incrociati introdotti
- 3) Il tipo di dati usati durante la parametrizzazione

Esistono due tendenze generali. Se l'obiettivo è quello di descrivere sistemi molto grandi (macromolecole o biopolimeri) la forma del funzionale viene mantenuta la più semplice possibile. Questo significa che per  $V_{\text{str}}$  ed  $V_{\text{bend}}$  si assume la forma armonica (o addirittura queste coordinate vengono bloccate al loro valore naturale), si tralasciano i termini misti ed  $V_{\text{vdW}}$  assume la forma di Lennard-Jones. Un campo di forza di questo tipo è spesso definito "armonico" o di "classe I". Al contrario, se si vogliono descrivere molecole medio-piccole con un elevato grado di accuratezza verranno impiegati molti termini misti,  $V_{\text{str}}$  ed  $V_{\text{bend}}$  saranno espanse fino al terzo o quarto ordine e  $V_{\text{vdW}}$  assume una formulazione di tipo esponenziale. Questo tipo di campo di forza, che si propone non solo di riprodurre la geometria e l'energia relativa, ma anche altre proprietà come le frequenze vibrazionali, viene spesso denominato di "classe II".

### 3.18 Esecuzione di un calcolo MM

Dovrebbe essere chiaro che i metodi basati su un campo di forza sono dei modelli del sistema quantomeccanico "reale". Tralasciare completamente gli elettroni fa sì che chi usa un metodo MM deve definire anticipatamente quali e che tipo di legami sono presenti nella molecola in esame. Per effettuare un calcolo bisogna fornire al programma il tipo di atomi presenti, la loro connettività, una geometria iniziale. Le prime due informazioni

definiscono univocamente il funzionale (forma e parametri) tranne nel caso in cui alcuni parametri vengano assegnati a partire dalla geometria attuale (sistemi  $\pi$  coniugati, alcuni composti di coordinazione, ecc.). Tutte queste informazioni vengono fornite al programma tramite un file di input. Tutte le informazioni necessarie possono essere ricavate in modo automatico da una rappresentazione tridimensionale anche molto approssimativa della molecola [31]. Perciò i programmi più moderni consentono di disegnare la molecola con una interfaccia grafica più o meno evoluta ed assegnano poi automaticamente i tipi atomici sulla base del simbolo e della connettività, convertendo anche la rappresentazione grafica in coordinate cartesiane. Queste informazioni vengono quindi trasferite alla componente del programma che calcola il potenziale del campo di forza (o energia sterica) ed eventualmente al programma di ottimizzazione della geometria.

## CAPITOLO 4

### DINAMICA MOLECOLARE

Un approccio computazionale molto usato sia nella scienza dei materiali che nello studio di biomolecole è la dinamica molecolare (MD). Con questo metodo possono essere seguiti passo per passo e simulati nel dettaglio il moto e le proprietà di un sistema di particelle.

Di conseguenza, le simulazioni di MD permettono di indagare le proprietà dinamico-strutturali anche di grossi sistemi biologici, in condizioni molto simili a quelle fisiologiche e per ampie finestre di tempo, per cui possono dare informazioni molto dettagliate su fenomeni quali cambiamenti conformazionali di proteine e acidi nucleici, difficili da estrarre con altri approcci come ad esempio, il metodo Monte Carlo. Il prezzo da pagare può essere elevato in termini di tempi di calcolo: simulazioni MD di sistemi complessi possono essere assai dispendiose. Tuttavia l'incremento di prestazioni dei computer da una parte, il raffinamento del metodo e degli algoritmi dall'altro, hanno reso possibili calcoli impensabili fino a pochi anni fa.

Con la MD possono essere studiati in particolare:

- stabilità di proteine;
- cambi conformazionali
- protein folding (di piccole proteine!)
- riconoscimento molecolare: interazioni tra proteine, DNA, membrane, complessi
- trasporto ionico in sistemi biologici

Tali caratteristiche stanno portando all'utilizzo diffuso di questo metodo per interpretare i dati sperimentali e per progettare nuovi esperimenti. Questa tecnica di simulazione fornisce, ad esempio un valido aiuto nel design di nuovi farmaci e nella determinazione di strutture di molecole complesse (in combinazione con tecniche sperimentali quali diffrazione di raggi X e NMR).

Il metodo fu introdotto inizialmente da Alder e Wainwright negli anni '50 per lo studio di sistemi composti da sfere rigide e portò rapidamente ottimi risultati anche nello studio di liquidi semplici [32]. La prima simulazione di dinamica molecolare di un sistema realistico, l'acqua in fase liquida, risale al 1974 [32]. Tre anni dopo apparve la prima

simulazione di proteine (*bovine pancreatic trypsin inhibitor*, BPTI) che fu realizzata nel 1977 da McCammon e altri [33]. Questo lavoro, benché effettuato con la proteina nel vuoto, utilizzando un “rozzo” potenziale di meccanica molecolare e per una durata di soli 9.2 ps, servì a cambiare l’opinione corrente secondo cui le proteine erano strutture rigide e portò alla concezione di proteine come sistemi dinamici, i cui moti interni hanno un importante ruolo funzionale.

Due anni dopo, con la prima simulazione della BPTI si comprese che i fattori di temperatura  $B$ , misurati dai cristallografi in esperimenti di diffrazione di raggi X di cristalli di proteina, potevano essere utilizzati per avere informazioni sui moti interni delle proteine e furono utilizzati per una comparazione con le fluttuazioni quadratiche medie ottenute dalle simulazioni di dinamica molecolare.

La simulazione della BPTI aprì la strada alla simulazione di dinamica molecolare classica di molecole proteiche; da allora infatti sono apparsi numerosi articoli tra i quali va almeno ricordato il primo lavoro in cui una proteina è stata immersa in un box, nel quale le molecole di acqua erano considerate in maniera esplicita [34]. La possibilità di simulare una molecola proteica in acqua esplicita rese più concreta l’opportunità di comparare dati simulati e dati sperimentali, in particolare con dati spettroscopici quali la depolarizzazione di fluorescenza, i parametri di ordine misurati per NMR, l’effetto del solvente e della temperatura sulla struttura e sulla dinamica delle proteine.

Le simulazioni MD di proteine in acqua oggi raggiungono scale temporali fino ai nanosecondi; tuttavia, in assenza di solvente, sono riportati anche casi di simulazioni fino ai millisecondi.

Più lentamente l’approccio simulativo si estese anche ad altri sistemi biologici. Nella seconda metà degli anni ottanta si effettuarono le prime simulazioni di DNA e successivamente di RNA, mentre nella seconda metà degli anni novanta fu eseguita la simulazione delle prime proteine di membrana immerse in un doppio strato lipidico.

#### **4.1 Basi della dinamica molecolare**

La MD si basa sulla seconda legge di Newton. L’evoluzione temporale di un sistema di molti atomi è data dall’equazione

$$F_i = m_i a_i \quad i = 1, 2, \dots, N$$

Integrando le equazioni del moto si può ricavare la traiettoria, ovvero posizioni  $r_i(t)$  in funzione del tempo e velocità  $v_i(t)$ . Il metodo è deterministico e la traiettoria è in linea di principio esattamente determinata dalle condizioni iniziali.

Assunzione fondamentale è che la forza possa essere espressa come il gradiente di un'energia potenziale  $U$

$$F_i = -\nabla_i U(r_1, \dots, r_N)$$

che combinata con la precedente relazione fornisce

$$-\nabla_i U = m_i dr_i/dt^2$$

Per calcolare le traiettorie degli atomi sono necessarie, oltre alla funzione energia potenziale  $U$ , anche delle posizioni iniziali (diciamo al tempo  $t=0$ ) e delle velocità iniziali. Le posizioni iniziali vengono ricavate da dati sperimentali provenienti, per esempio dalla cristallografia a raggi X; le velocità iniziali  $v_i(t=0)$  sono spesso prese da una distribuzione Gaussiana adatta alla temperatura di partenza, facendo in modo che la quantità di moto complessiva del sistema sia nulla, ovvero  $P = \sum_i m_i v_i = 0$ .

In assenza di accorgimenti, l'energia totale in una simulazione MD è conservata. Si parla allora di dinamica  $NVE$ , per indicare che numero di particelle, volume ed energia sono conservate. Un'ovvia conseguenza è che la dinamica  $NVE$  esplora l'insieme micro canonico: per sistemi ergodici la media temporale calcolata sulla traiettoria della dinamica molecolare equivale alla media statistica sull'insieme microcanonico. La temperatura del sistema varia durante la simulazione ed è determinata dall'energia cinetica istantanea

$$\frac{3}{2} N k_b T(t) = \sum_i \frac{1}{2} m_i v_i^2(t)$$

Sono possibili comunque altri tipi di dinamica molecolare. Per esempio, l'insieme canonico può essere esplorato con la dinamica  $NVT$ , in cui si aggiunge un opportuno termostato che mantiene il sistema ad una certa temperatura data. E' anche possibile realizzare una dinamica  $NPT$ , in cui il sistema è tenuto a temperatura e pressione dati.

## 4.2 Approssimazione adiabatica

E' lecito che il moto degli atomi sia determinato solo da una energia potenziale che dipende soltanto dalle posizioni dei nuclei? In realtà gli elettroni sono molto più leggeri dei nuclei (almeno 1800 volte nel caso più sfavorevole) e di conseguenza il loro moto è notevolmente più veloce di quello dei nuclei. E' possibile allora assumere che gli elettroni in ogni istante restino nel loro stato fondamentale, corrispondente alla posizione istantanea dei nuclei (approssimazione adiabatica o di Born-Oppenheimer, Capitolo 2). Si tratta di una eccellente approssimazione con un ampio intervallo di validità. Essa garantisce l'esistenza di un potenziale interatomico  $U$  e ne dà anche la formulazione in termini dell'energia di un sistema di elettroni sotto il campo dei nuclei. Nella MD *ab initio* il potenziale  $U$  è esplicitamente calcolato a partire dalle proprietà elettroniche. I risultati sono molto accurati ma i costi numerici e computazionali sono molto elevati rispetto alla MD classica in cui si usano forme semplici del potenziale, come verrà illustrato nelle prossime sezioni.

## 4.3 Integrazione delle equazioni del moto

Il potenziale  $U$  è funzione delle posizioni di tutti gli atomi del sistema; ovviamente, a causa della complessità di tale funzione le equazioni del moto devono essere risolte numericamente [11]. Il modo più ovvio è quello di adottare un *metodo alle differenze finite*, infatti è necessario discretizzare le equazioni, ovvero passare ad una descrizione in cui il tempo diventa una variabile discreta. Si sceglie allora un intervallo di tempo (time step)  $\delta t$ , piccolo (ma non più del necessario: il più grande possibile che ci permetta di descrivere il moto che ci aspettiamo).

Questo modo di procedere è soggetto ad errori. Gli errori sono di due tipi : 1) errori dovuti a troncamento e 2) errori dovuti ad arrotondamento. Gli errori dovuti a troncamento sono quelli che derivano da un'espansione della funzione esatta e dal fermarsi ad un certo termine. Per esempio considerare la funzione  $x(t)$ , che descriva la semplice cinematica di un corpo materiale, come uno sviluppo in serie e fermarsi ad un certo termine

$$x(t+\delta t) = x(t) + \dot{x}(t) \delta t + 1/2 \ddot{x}(t) \delta t^2 + \dots$$

Se ci si ferma al secondo termine si considera il moto con velocità costante, mentre se ci si ferma al terzo termine si considera il moto con accelerazione costante. In entrambi i casi si fa un'approssimazione e si commette un errore dovuto al troncamento dello sviluppo in serie. L'errore di troncamento è misurato dal primo termine, diverso da zero, che è omissso dalla serie. Un metodo il cui errore varia come  $(\delta t)^{n+1}$  è detto di ordine  $n$ . L'errore di arrotondamento dipende dal fatto tecnico che nei calcoli non possiamo considerare tutte le cifre significative, dalle approssimazioni nel calcolo di radici quadrate, esponenziali ...

Entrambi i tipi di errore possono essere suddivisi in errori *locali* e *globali*. I primi sono gli errori che sono fatti in ogni passo e che non si amplificano nei calcoli successivi. I secondi sono gli errori che si accumulano nell'intero calcolo.

Gli errori di troncamento si riducono al diminuire del  $\delta t$ . Gli errori di arrotondamento si riducono utilizzando alta precisione nei calcoli e riducendo i calcoli stessi, cioè scrivendo codici non ridondanti ed efficienti.

Una grande varietà di algoritmi di integrazione è disponibile per risolvere le equazioni del moto. Poiché il calcolo delle forze ad ogni passo è molto costoso è preferibile usare schemi di integrazione, che autorizzino dei passi abbastanza ampi e nel contempo permettano di mantenere una precisione ed una stabilità sufficienti. E' noto che due traiettorie nello spazio delle fasi con una differenza arbitrariamente piccola delle condizioni iniziali divergono esponenzialmente dopo un tempo sufficientemente lungo. Così, nessun algoritmo fornirà una soluzione esatta con tempi molto lunghi. Un buon algoritmo di integrazione deve perciò essere capace di fornire una buona approssimazione delle equazioni differenziali su scale di tempi corti e assicurare, su scale di tempi lunghi, la conservazione delle costanti del moto, come l'energia totale ed il momento del sistema. Un algoritmo molto semplice che soddisfa la condizione di stabilità su tempi lunghi è l'algoritmo di Verlet delle velocità [11], che funziona nella maniera seguente. Date le posizioni atomiche  $r_i(t)$ , le velocità  $v_i(t)$  e le accelerazioni  $a_i(t)$  ad un certo tempo  $t$  si determinano le velocità al tempo  $t+\delta t$

$$v_i(t + \frac{\delta t}{2}) = v_i(t) + a_i(t) \frac{\delta t}{2}$$

Le posizioni atomiche al tempo  $t+\delta t$  sono allora calcolate come

$$r_i(t + \delta t) = r_i(t) + v_i(t + \frac{\delta t}{2})\delta t$$

Le forze  $f_i(t + dt)$  e le accelerazioni  $a_i(t + dt)$  al tempo  $t + \delta t$  possono dunque essere calcolate a partire dai nuovi valori delle posizioni  $r_i(t)$ . Infine le velocità al tempo  $t + \delta t$  sono dedotte

$$v_i(t + \delta t) = v_i(t + \frac{\delta t}{2}) + a_i(t + \delta t)\frac{\delta t}{2}$$

La risoluzione delle equazioni del moto è dunque realizzata effettuando questo schema ad ogni passo temporale.

Combinando diversamente gli sviluppi in serie di Taylor si possono ricavare altri algoritmi di integrazione come l'algoritmo di Verlet semplice. Per ogni particella si può scrivere infatti

$$r_i(t + \delta t) = r_i(t) + v_i(t)\delta t + \frac{1}{2}a_i(t)\delta t^2 + \frac{1}{6}a_i(t)\delta t^3 + O(\delta t^4)$$

$$r_i(t - \delta t) = r_i(t) - v_i(t)\delta t + \frac{1}{2}a_i(t)\delta t^2 - \frac{1}{6}a_i(t)\delta t^3 + O(\delta t^4)$$

Sommando e sottraendo le due relazioni e ricordando che  $F = ma$  si ottiene l'*algoritmo di Verlet*

$$r_i(t + \delta t) = r_i(t - \delta t) + 2r(t) + \frac{F(t)}{m}\delta^2 + O(\delta t^4) \quad (4.1a)$$

$$v_i(t) = \frac{r_i(t + \delta t) - r_i(t - \delta t)}{2\delta t} + O(\delta t^3) \quad (4.1b)$$

Da notare che le velocità non sono di fatto usate. L'algoritmo di Verlet è il prototipo di tutti gli algoritmi di integrazione ed è tuttora usato. E' efficiente perché richiede un solo calcolo di forze (la parte più costosa dal punto di vista computazionale) ad ogni passo. E' numericamente stabile ed ha una proprietà assai importante per una buona dinamica molecolare: conserva la simmetria per inversione temporale delle equazioni del moto. In altri termini, se si inverte la velocità e la direzione del tempo ad un certo istante, la traiettoria ritraccia all'inverso quella seguita fino ad allora. Si può dimostrare che Verlet

normale e *velocity Verlet* producono le stesse traiettorie. Un altro algoritmo molto usato è il *leap-frog* (a salto di rana)

$$r_i(t + \delta t) = r_i(t) + v_i(t + \frac{\delta t}{2})\delta t$$

$$v_i(t + \frac{\delta t}{2}) = v_i(t - \frac{\delta t}{2}) + \frac{F(t)}{m}\delta t$$

così chiamato perché velocità e posizione non sono calcolati allo stesso tempo ma “saltellano” fra tempi sfalsati di  $\delta t/2$ .

### *Predictor corrector*

L'integrazione delle equazioni del moto può essere alternativamente realizzata attraverso l'uso di un'altra famiglia di algoritmi, denominati *predictor-corrector*. Sono metodi che si basano sulla predizione approssimata delle posizioni al tempo  $t + \delta t$ , sul calcolo della forza per queste posizioni e da questa dell'accelerazione. Si correggono quindi le precedenti equazioni in qualche modo.

Ci sono moltissimi metodi *predictor-corrector* che sono stati proposti. Uno dei primi in ordine temporale è il seguente

$$r^p(t + \delta t) = r(t) + v(t)\delta t \quad (i)$$

$$\text{calcolo dell'accelerazione corrispondente a } r^p \text{ ovvero } a(t + \delta t) \quad (ii)$$

$$v(t + \delta t) = v(t) + \delta t[a(t + \delta t) + a(t)]/2 \quad (iii)$$

$$r^c(t + \delta t) = r(t) + \delta t[v(t + \delta t) + v(t)]/2 \quad (iv)$$

nella (iv) si considera in pratica la velocità media (tra quella al tempo  $t$  e quella predetta al tempo successivo) nell'intervallo. Si itera dal punto 2 fino a che la differenza tra le due posizioni calcolate in due iterazioni successive,  $r^{c''}(t + \delta t)$  e  $r^{c'}(t + \delta t)$ , non sia minore di un valore prefissato. Questi metodi in genere richiedono il calcolo delle forze più volte per ogni passo di integrazione e quindi richiedono molto tempo di calcolo.

#### 4.4 Potenziali

L'utilità di effettuare dei calcoli di dinamica molecolare ripone in gran parte sulla qualità dei potenziali di interazione che vengono scelti. In generale i potenziali più semplici, i potenziali di coppia, sono adeguati per sistemi monoatomici. Dunque è necessario andare al di là di questa semplice approssimazione dell'interazione a due corpi, ciò che è particolarmente vero per i sistemi fortemente covalenti. In tutti i casi i parametri dei potenziali sono ottenuti riproducendo proprietà del sistema. [11].

Lo studio teorico dei sistemi biologici permette di ricavare le relazioni tra struttura, funzione e dinamica a livello atomico. Molti dei problemi di interesse coinvolgono molti atomi e dunque per tali situazioni non è ancora possibile adottare una trattazione quantistica. Tuttavia, i problemi diventano molto più trattabili utilizzando dei potenziali empirici, che sono molto meno dispendiosi dal punto di vista computazionale, anche se ciò porta ad una serie di approssimazioni e limitazioni. Le funzioni di energia potenziale (campi di forza) utilizzate portano ad un buon compromesso tra l'accuratezza fisica e l'efficienza computazionale. In effetti, tali funzioni sono in genere calibrate su dati sperimentali e calcoli quantistici attraverso cui vengono ricavati i parametri dai quali tali funzioni dipendono. Tra gli insiemi di funzioni di energia potenziale più usati, citiamo quelli adottati dai più noti pacchetti di software per simulazioni MD: AMBER, CHARMM, GROMOS.

Una seria limitazione di tali approcci è la quasi impossibilità di descrivere drastici cambiamenti nella struttura elettronica del sistema: eventi come la formazione o la rottura di un legame non sono permessi, o più esattamente, sono male descritti. Una migliore descrizione richiederebbe approcci misti in cui parte delle interazioni sono trattate quantisticamente. Se tuttavia si è interessati, come nel caso presente, a seguire la dinamica di molecole biologiche (la calbindina D9k) in assenza di reazioni chimiche significative, tale limitazione non costituisce un problema.

##### *Limiti dei Potenziali Empirici*

I potenziali empirici sono funzioni semplici e differenziabili delle coordinate atomiche, dalle quali è facile ricavare analiticamente la forza (vettoriale) che agisce su ogni atomo.

Ciò permette di fare la simulazione della dinamica in modo efficiente. Le limitazioni imposte dai potenziali introdotti sono essenzialmente di due tipi:

- I parametri che definiscono le funzioni di interazione fra ogni atomo del sistema e gli altri atomi vengono assegnate all'inizio e rimangono fissi per tutta la simulazione. Questo non permette di tenere conto del fatto che il comportamento di certi atomi (per esempio, per l'ossigeno e l'azoto) è influenzato dagli atomi vicini.
- I potenziali sono costruiti attraverso la somma di interazioni di coppia: non sono considerate interazioni a tre o più particelle. Questo non permette di tenere conto di effetti legati alla polarizzazione di certi gruppi e/o a ionizzazioni di certi residui amminoacidici, indotti dal campo elettrico di tutta la proteina.

Infine, le funzioni di energia potenziale non includono effetti entropici. Quindi un minimo di energia non corrisponde necessariamente ad uno stato di equilibrio. Di conseguenza le simulazioni MD sono fatte in modo da riprodurre condizioni analoghe a quelle in cui normalmente sono condotti gli esperimenti, e cioè a temperatura e/o pressione costante.

Il termine energetico più dispendioso in una simulazione MD è quello associato alle interazioni non-bonded (elettrostatiche e van der Waals). Tali interazioni decrescono lentamente con la distanza fra gli atomi, per cui tale termine dovrebbe essere valutato per ogni coppia di atomi del sistema. Ciò implica che il tempo necessario per la sua valutazione cresce con il quadrato del numero di atomi  $N$  del sistema ( $N^2$ ). Per velocizzare tale calcolo viene posto un *cut-off* alla distanza di interazione. Nel corso degli anni si sono sviluppati vari modi di attuazione del *cut-off*:

- Troncamento. L'energia d'interazione viene semplicemente posta a zero per distanze maggiori di  $r_{cut-off}$ . Dato che questo metodo (brutale) porta spesso a grosse e innaturali fluttuazioni nell'energia, è poco utilizzato.
- Shift. Viene modificata l'intera funzione potenziale in modo che a  $r_{cut-off}$  sia nulla. Tale metodo porta a distanze di equilibrio più piccole di quelle sperimentali.
- Switch. La funzione energia viene modificata solo in un piccolo intervallo di distanze. Tale metodo ha il problema di generare forze notevoli nella regione di switch, che possono perturbare l'equilibrio.

L'importanza delle interazioni elettrostatiche a lungo raggio nei sistemi biologici è dimostrato da un grande numero di studi sperimentali. In effetti, in una simulazione MD, l'inclusione di interazioni a raggio più elevato si ottiene semplicemente aumentando  $r_{cut-off}$ ,

con il risultato però di aumentare notevolmente il costo computazionale. A causa di ciò, spesso, tali interazioni vengono completamente ignorate, introducendo la severa approssimazione di non tenere conto delle proprietà dielettriche del mezzo. Più recentemente, sono stati introdotti vari modelli che permettono l'inclusione delle interazioni elettrostatiche a lungo range nelle simulazioni MD.

#### 4.5 Inclusione del Solvente in una Simulazione MD

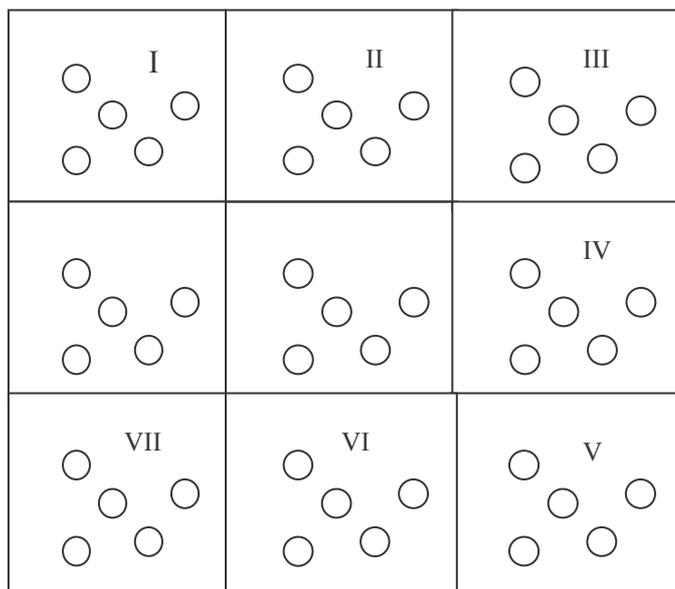
L'acqua ha un ruolo fondamentale su struttura, dinamica e termodinamica delle molecole biologiche. Uno dei più importanti effetti del solvente è quello di schermo per le interazioni elettrostatiche. L'inclusione del solvente nelle simulazioni può essere fatta in vari modi. Il modo implicito più semplice è quello di includere una costante elettrostatica nel termine elettrostatico dell'energia potenziale, usando se necessario una costante dielettrica effettiva dipendente dalla distanza  $\epsilon_{\text{eff}} = \epsilon(r_{ij})$  (a tale proposito vedi anche Cap. 3). Il modo esplicito consiste invece nell'includere molecole d'acqua nel sistema. Le molecole forniscono direttamente l'effetto di schermo (o almeno la parte legata al riorientamento delle stesse).

In questa trattazione più dettagliata bisogna utilizzare convenienti condizioni al contorno, sia per prevenire la diffusione delle molecole d'acqua in zone troppo lontane dalla proteina, che per permettere la simulazione e il calcolo delle proprietà macroscopiche, utilizzando un numero limitato di molecole di solvente. Le condizioni al contorno periodiche permettono di fare la simulazione con un numero relativamente piccolo di molecole di solvente e in modo che le particelle risentano di forze analoghe a quelle che risentirebbero nel bulk della soluzione. Qui sotto in Figura 4.1 è mostrata la situazione in due dimensioni.

Il quadrato centrale è circondato dagli otto quadrati vicini e le coordinate delle particelle immagine in tali quadrati sono ottenute mediante semplici traslazioni delle posizioni delle corrispondenti particelle nel quadrato principale. Le forze nel quadrato principale sono calcolate utilizzando tutte le particelle (anche quelle dei quadrati secondari).

Il raggio di *cutoff* è in genere scelto in modo che una particella del quadrato primario non veda la sua immagine nei quadrati secondari.

Ci sono casi in cui non è utile considerare le condizioni al contorno periodiche e in cui, anche sfruttando la notevole potenza dei computer, le simulazioni vengono fatte incorporando un grande numero di molecole di solvente.



**Figura 4.1**

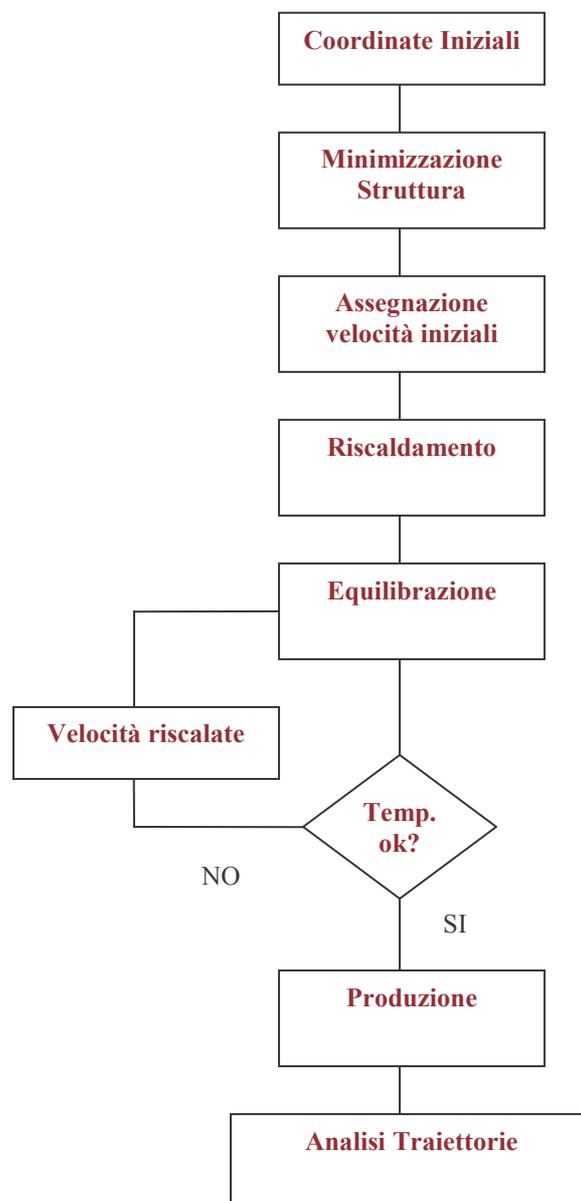
In questi casi, la proteina viene circondata da una sfera d'acqua, ottenuta utilizzando un potenziale al bordo che confina le molecole d'acqua in tale volume (vedi figura in basso a sinistra). In altri casi, l'acqua viene confinata solo nella parte interessante della simulazione (vedi parte destra della figura 4.2).



**Figura 4.2**

#### 4.6 Set-up e running di una Simulazione MD

Come per le simulazioni Monte Carlo, il risultato di una simulazione MD è una serie temporale di conformazioni corrispondenti alle traiettorie o cammini seguiti da ogni atomo. La maggior parte delle simulazioni MD sono effettuate a  $N$ ,  $V$  ed  $E$  costanti (insieme microcanonico), ma è possibile fare simulazioni anche a  $N$ ,  $T$  e  $P$  costanti che meglio riproducono le condizioni sperimentali. Qui di seguito è illustrato uno schematico diagramma a blocchi che riassume le varie tappe di una simulazione.



- **Inizializzazione.** Per cominciare si sceglie la configurazione iniziale del sistema a  $t = 0$ . Nella maggior parte dei casi, per la simulazione di biomolecole si parte dalla struttura (cristallografica o NMR) ottenuta dal Brookhaven Protein Databank (PDB). La scelta della configurazione iniziale deve essere fatta con attenzione perché essa può influenzare la qualità della simulazione. Prima di partire con la vera e propria simulazione è spesso bene procedere ad una minimizzazione dell'energia della struttura. A questo punto vengono addizionate le molecole d'acqua: spesso le molecole d'acqua sono prese da una precedente simulazione fatta in un box di dimensioni opportune; la proteina viene inserita in tale box eliminando le molecole d'acqua ad essa sovrapposte. A questo punto si effettua una nuova minimizzazione dell'energia per permettere all'acqua ed alla proteina di assestarsi nella nuova situazione.

- **Riscaldamento.** Vengono assegnate le velocità iniziali (in modo da avere una bassa temperatura) e si avvia la simulazione. Nel caso di solvente esplicito, prima si blocca la proteina e si lascia termalizzare le molecole d'acqua, poi si sblocca la proteina e si comincia l'evoluzione del sistema *proteina + acqua* insieme. Durante la fase di riscaldamento periodicamente le velocità vengono riscaldate in modo da via via aumentare la temperatura fino ad arrivare a quella prescritta.

- **Equilibratura.** Raggiunta la temperatura prescritta la simulazione continua in modo che possano essere monitorate quantità come pressione, temperatura ed energia. In questa fase, talvolta è necessario, per ovviare a innaturali fluttuazioni della temperatura, riscaldare le velocità opportunamente.

La fase continua fino a quando l'equilibratura dei parametri suddetti non viene raggiunta.

- **Fase di Produzione.** Nella fase finale, dopo l'equilibratura, lasciando andare la simulazione, si possono ricavare le medie dei parametri termodinamici di interesse. Tale fase può corrispondere a varie centinaia di picosecondi a nanosecondi.

Durante una simulazione MD, le coordinate e le velocità degli atomi vengono memorizzate e successivamente utilizzate per l'analisi. Si possono calcolare medie di quantità come l'energia (cinetica e/o potenziale):

$$\langle E \rangle = \frac{1}{M} \sum_i E_i$$

dove  $M$  è il numero di passi di MD,  $E_i = E_i(t)$  al tempo  $t_i = t_0 + (i-1)\delta t$

Si possono poi calcolare altri tipi di grandezze quali il raggio di girazione in funzione del tempo e la sua media; o le cosiddette funzioni di correlazioni di coppia,  $g_{AB}(R)$ , che danno la probabilità di trovare un atomo di tipo B a distanza R da un atomo di tipo A.

#### 4.7 Proprietà dipendenti dal tempo

La MD genera configurazioni del sistema nel tempo e quindi le simulazioni MD possono essere usate per calcolare proprietà dipendenti dal tempo. Questo è il principale vantaggio rispetto al metodo Monte Carlo. Le proprietà dipendenti dal tempo sono spesso calcolate come coefficienti di correlazione temporale [11].

Le funzioni di correlazione temporale calcolano come il valore di una grandezza dinamica  $A(t)$  è correlato al valore di un'altra grandezza  $B(t)$ . Alcune funzioni di correlazione temporale possono essere utilizzate per il calcolo dei coefficienti di trasporto come la viscosità, la conducibilità termica e il coefficiente di diffusione. Le funzioni di correlazione temporale vengono inoltre utilizzate nella spettroscopia per calcolare gli spettri.

Supponiamo di avere due quantità A e B misurate con una serie di misure. Per ognuna possiamo calcolare la varianza  $\sigma^2$  dove  $\sigma$  è lo scarto quadratico medio o deviazione standard. Il coefficiente di correlazione tra le due quantità è definito come

$$c_{AB} = \frac{\langle \partial A \partial B \rangle}{\sigma(A)\sigma(B)}$$

Con  $c_{AA}$  si indica il coefficiente di autocorrelazione. L'idea del coefficiente di correlazione può essere estesa considerando il caso in cui le quantità A e B vengano calcolate a due tempi diversi separati da un certo intervallo t. In tal caso il coefficiente di correlazione dipenderà da t e quindi  $c_{AB}(t)$  prende il nome di funzione di correlazione temporale

$$c_{AB}(t) = \frac{\langle \partial A(t_0) \partial B(t_0 + t) \rangle}{\sigma(A)\sigma(B)} \quad (4.2)$$

in cui la media è effettuata su tutti i valori di  $t_0$ . Se le quantità  $A(t)$  e  $B(t)$  sono rappresentate da funzioni continue nel tempo la (4.2) diventa

$$c_{AB}(t) = \lim_{\tau \rightarrow \infty} \int \frac{\partial A(t_0) \partial B(t_0 + t) dt_0}{\sigma(A) \sigma(B)}$$

Se le quantità A e B sono identiche si avrà allora  $c_{AA}(t)$  che prende il nome di funzione di autocorrelazione temporale. Il suo integrale tra 0 e infinito viene chiamato tempo di correlazione. Per tempi lunghi  $c_{AA}$  tende al limite  $\langle A \rangle^2$ .

#### 4.8 Dinamica di un sistema sottoposto a vincoli: SHAKE

Nelle simulazioni si usa spesso mantenere costanti le distanze di legame in una molecola, rimuovendo dal sistema quei gradi di libertà che hanno la frequenza più alta, assumendo, naturalmente, che la proprietà che si voglia studiare sia indipendente da tali gradi di libertà [11]. Ciò permette di utilizzare un tempo di integrazione maggiore, con un notevole risparmio di tempo di calcolo, senza inficiare la stabilità numerica. Infatti, se i legami atomo pesante-idrogeno sono vincolati ad avere una lunghezza costante, i moti con le frequenze più alte saranno le vibrazioni atomo pesante-atomo pesante che tipicamente sono un fattore 2-5 volte più piccole in intensità, rispetto a quelle coinvolgenti l'atomo H. Uno degli algoritmi comunemente usato, per eliminare questi gradi di libertà, è un metodo iterativo che va sotto il nome di SHAKE [11,13]. Nel contesto dell'algoritmo di Verlet si prende in considerazione dapprima uno step senza vincoli (eq. 4.1a), successivamente si introducono i vincoli con la seguente relazione

$$r_i(t + \delta t) = r_i^o(t + \delta t) + \delta r_i(t)$$

dove  $r_i^o(t + \delta t)$  è la posizione dopo lo step *unconstrained* e  $\delta r_i(t)$  è il vettore spostamento richiesto per soddisfare un set di equazioni di vincolo accoppiate, definite come

$$\delta r_i(t) = \frac{2\delta t^2}{m_i} \sum_j \lambda_{ij} r_{ij}(t)$$

dove i moltiplicatori di Lagrange vengono determinati iterativamente.

## CAPITOLO 5

### METODI IBRIDI QUANTISTICI/CLASSICI

#### 5.1 Introduzione

I metodi della chimica quantistica consentono di simulare reazioni chimiche ed altri processi complessi, ma a causa dei costi computazionali la loro applicazione è ancora limitata allo studio di molecole relativamente piccole (un centinaio di atomi con i metodi DFT).

I sistemi di interesse biochimico invece sono costituiti da un gran numero di atomi. Le simulazioni della meccanica molecolare (MM) permettono lo studio di siffatti sistemi quando non si voglia descrivere la loro struttura elettronica.

Una soluzione astuta per trattare questi casi consiste nell'adottare una descrizione quantistica (QM) per la porzione del sistema per la quale è importante tenere in conto la struttura elettronica ed una descrizione classica (MM) per il resto del sistema.

Questo approccio denominato correntemente metodo misto QM/MM è stato suggerito ed usato nel 1976 da Warshel e Levit [35] per lo studio della reazione enzimatica del lisozima. In questo lavoro, l'energia del sistema viene decomposta in due parti, una che include i termini classici della meccanica molecolare e l'altra che include un Hamiltoniano quantomeccanico semiempirico che prende in considerazione la polarizzazione della porzione quantistica attraverso cariche puntiformi della parte classica. I legami chimici tra parte QM e MM vengono "tagliati" ed orbitali ibridi vengono posti sugli atomi della frontiera. Dieci anni più tardi Singh e Kollman [36] sono stati i primi a pubblicare una ottimizzazione di geometria QM/MM con una parte quantistica *ab initio*.

Nel 1990, Field, Bash e Karplus [37] hanno descritto in dettaglio l'accoppiamento di un Hamiltoniano semiempirico con i campi di forza CHARMM.

In questo lavoro di tesi è stato adottato un semplice trattamento ibrido QM/MM per studiare sistematicamente la convergenza dei chemical shift NMR della *calcium-binding* Calbindina D9k, come funzione della dimensione della regione QM.

## 5.2 Terminologia

In un approccio QM/MM il sistema  $S$  è partizionato in una regione *inner* ( $I$ ) che deve essere trattata con la quantomeccanica e una regione *outer* ( $O$ ) descritta da un campo di forza (Figura 5.1)

Le regioni interna ( $I$ ) ed esterna ( $O$ ) sono frequentemente denominate anche QM ed MM, rispettivamente. Ogni atomo dell'intero sistema viene assegnato ad uno dei due sottosistemi. Il *boundary* designa la regione dove le procedure standard QM e MM vengono in qualche modo modificate.

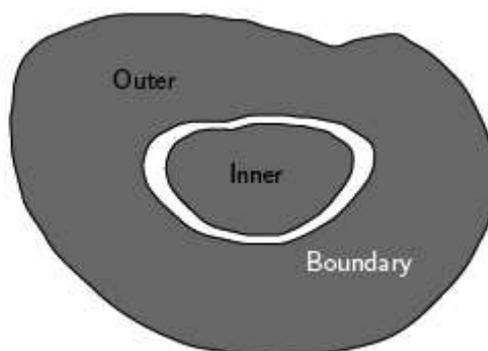


Figura 5.1: Divisione del sistema nei sottosistemi *Inner* e *Outer*

Ci sono tre tipi di interazioni in un modello QM/MM: interazione tra gli atomi nel sottosistema MM, interazioni tra nuclei ed elettroni esplicite nel sottosistema QM e le interazioni tra le due regioni  $I$  e  $O$ . Poiché questo ultimo tipo di interazioni sono generalmente forti, non è possibile scrivere l'energia di  $S$  semplicemente come la somma delle energie dei suoi sottosistemi.

## 5.3 Energia QM/MM

I metodi QM/MM che si classificano in base alla forma che assume l'espressione dell'energia del sistema, si dividono in due ampie categorie: schemi additivi e schemi sottrattivi.

### *Schemi sottrattivi*

Negli schemi QM/MM sottrattivi vengono eseguiti tre calcoli indipendenti: un calcolo sull'intero sistema  $S$ , un calcolo QM sul sottosistema *inner*  $I$ , un calcolo MM di  $I$ . L'energia totale del sistema viene ricavata sommando i primi due termini e sottraendo l'ultimo, onde evitare un conteggio doppio dei contributi. Da questo punto di vista l'espressione dell'energia può essere considerata come una correzione o estrapolazione. Se si utilizza l'implementazione dei *link atom* per saturare le valenze libere nello strato interno (vedi sezioni seguenti) l'energia totale assume la seguente forma:

$$E_{QM/MM}^{sottrattiva}(S) = E_{MM}(S) + E_{QM}(I + L) - E_{MM}(I + L) \quad 5.1$$

Il pedice denota il livello di calcolo, mentre in parentesi viene riportato il sistema sul quale tale calcolo viene eseguito.

Concettualmente, lo schema sottrattivo può essere visto come un approccio MM in cui una certa regione è stata rimossa e sostituita con un trattamento di livello più alto. Il suo pregio più importante è la semplicità, che rende abbastanza agevole anche l'implementazione del metodo. Non è necessario nessun termine esplicito di accoppiamento, evitando così modifiche alle procedure standard QM e MM. Il termine sottratto corregge per qualunque artefatto causato dai *link atom*, ammesso che i parametri del campo di forza MM riferiti ai *link atom* riproducano ragionevolmente il potenziale QM.

Uno svantaggio, in uno schema di questo tipo, è che l'accoppiamento tra i sottosistemi è trattato completamente a livello MM. Questo può risultare problematico soprattutto per le interazioni elettrostatiche, che sono descritte adottando cariche atomiche fissate, ad un livello MM. Di conseguenza, uno schema strettamente sottrattivo QM/MM implica un *embedding meccanico*, quindi la densità QM non è polarizzata dall'ambiente circostante (vedi sezioni successive).

### *Schemi additivi*

L'espressione dell'energia per uno schema additivo QM/MM è invece

$$E_{QM/MM}^{additiva}(S) = E_{MM}(O) + E_{QM}(I + L) - E_{QM-MM}(I, O) \quad (5.2)$$

In contrasto con lo schema sottrattivo, il calcolo MM è eseguito solo sul sistema più esterno. Il sistema interno viene trattato come nell'equazione (5.1), mentre l'ultimo termine

rappresenta l'accoppiamento tra le due regioni. Esso solitamente include le interazioni di van der Waals ed elettrostatiche tra i gli atomi QM e MM. Il termine di accoppiamento elettrostatico (sezione 5.6) è quello con il maggiore impatto oltre ad essere oneroso per quanto inerisce alla implementazione.

Se si assume un approccio *link atom* con *embedding* meccanico, è possibile derivare l'espressione additiva dell'energia dallo schema sottrattivo [38]. Per un campo di forze classico l'energia totale può essere decomposta in contributi provenienti da differenti set di atomi, ovvero

$$E_{MM}(S) = E_{MM}(O) + E_{MM}(I) + E_{MM}(I, O)$$

$$E_{MM}(I + L) = E_{MM}(I) + E_{MM}(L) + E_{MM}(I, L)$$

Sostituendo queste equazioni nella (5.1) si arriva alla seguente equazione per l'energia QM/MM

$$E_{QM/MM}^{additiva}(S) = E_{MM}(O) + E_{QM}(I + L) - E_{QM-MM}(I, O) - [E_{MM}(L) + E_{MM}(I + L)]$$

Poiché si è assunto un formalismo *embedding* meccanico,  $E_{MM}(I, O)$  rappresenta l'accoppiamento tra le regioni *inner* e *outer* ed è equivalente all'ultimo termine dell'equazione (5.2). Il termine tra parentesi quadre nell'espressione, sopra riportata, viene chiamata "correzione *link atom*" e contiene sia l'energia di interazione tra i *link atom* che quella tra *link atom* ed atomi dello strato interno ( $I$ ). Poiché queste interazioni sono anche incluse nel termine  $E_{QM}(I + L)$ , tale correzione è formalmente giustificata.

L'equazione di lavoro adottata dalla maggioranza degli schemi QM/MM è dunque l'equazione (5.2). Quello che definisce un particolare metodo QM/MM è la descrizione esatta del termine di accoppiamento QM-MM,  $E_{QM-MM}(I, O)$ .

L'interazione tra centri QM e MM può essere decomposta in accordo con le interazioni che vengono prese in considerazione nel campo di forza

$$E_{QM-MM}(I, O) = E_{QM-MM}^b + E_{QM-MM}^{vdW} + E_{QM-MM}^{elec}$$

Il primo addendo descrive le interazioni di legame QM-MM, localizzate in corrispondenza della frontiera, mentre gli ultimi due termini descrivono le interazioni QM-MM *non bonded*.

#### 5.4 Interazioni QM-MM bonded

Esiste una differenza operativa tra schemi sottrattivi e additivi, rispetto al trattamento dei *link atom*, che conduce a differenti termini nell'espressione finale dell'energia.

Uno schema sottrattivo rimuove le interazioni di legame QM-MM, cioè quelle che coinvolgono la regione *I+L* e conserva tutti i termini di legame misti QM-MM, quindi implicitamente corregge per i *link atom*.

In un approccio utilizzato da Eurenus *et al.* [39], allo scopo di evitare duplicati delle interazioni calcolate quantisticamente, vengono inclusi solamente i termini di legame, che coinvolgono un atomo QM e un atomo MM, tutti gli angoli ed i diedri con uno o due atomi centrali MM. Tutte le altre interazioni *bonded* sono descritte attraverso l'Hamiltoniano QM.

#### 5.5 Interazioni non-bonded tra gli strati

Il trattamento delle interazioni di van der Waals è generalmente descritto da un potenziale di Lennard-Jones e segue il modello adottato nel calcolo MM. Il problema è che questi parametri sono validi solo nell'ambito della parametrizzazione per la quale sono stati derivati, ovvero per le interazioni MM-MM. Fortunatamente la natura *short-range* dell'interazione di van der Waals rende questa restrizione meno rigorosa. Mentre ogni atomo nella regione più interna interagisce con tutti gli atomi dalla regione MM, soltanto quelli più vicini contribuiscono in maniera significativa, dunque i parametri non ottimizzati influenzano solo gli atomi QM vicini alla frontiera. Una soluzione potrebbe essere quella di spostare la regione di confine più lontano dagli atomi QM coinvolti.

## 5.6 Interazione elettrostatica

L'interazione elettrostatica tra la densità di carica QM ed il modello di carica usato per la meccanica molecolare -può essere trattata a differenti livelli di sofisticatezza. Sono stati definiti tre approcci [38]: *embedding meccanico*, *embedding elettrostatico* e *embedding polarizzato*.

### *Embedding meccanico*

Nell'*embedding* meccanico il calcolo QM viene realizzato in fase gas, senza inclusione delle cariche dalla regione *outer* nell'Hamiltoniano. Questo significa che la densità QM non viene polarizzata. In questo approccio l'interazione elettrostatica tra regioni QM e MM è omessa o al più si prende in considerazione usando il codice MM. Se si sceglie di trattare questa interazione, le cariche elettrostatiche MM per gli atomi nella regione *inner* devono essere derivate quando il sistema evolve lungo il cammino di reazione. Questo può funzionare se si prende una regione *inner* sufficientemente ampia, tuttavia si tratta di una procedura assolutamente non banale, che potrebbe fornire anche cariche non consistenti con il modello di carica presente nel campo di forza.

### *Embedding elettrostatico*

I principali difetti dell'*embedding* meccanico possono essere eliminati effettuando il calcolo in presenza della distribuzione di carica delle regioni circostanti. La polarizzazione dello strato QM viene realizzata incorporando le cariche puntiformi MM come termini monoelettronici nell'Hamiltoniano quantistico:

$$H_{QM-MM}^{el} = - \sum_i^{elettroni} \sum_{M \in O} \frac{q_M}{|r_i - r_M|} + \sum_{\alpha \in I} \sum_{M \in O} \frac{q_M Z_\alpha}{|r_\alpha - r_M|}$$

dove  $q_M$  sono le cariche puntuali MM e  $Z_\alpha$  le cariche nucleari degli atomi QM, gli indici  $i$ ,  $M$  e  $\alpha$  si riferiscono rispettivamente a tutti gli elettroni, alle cariche MM, ai nuclei QM.

Con questo approccio, la struttura elettronica della regione interna può adattarsi alla distribuzione dell'ambiente intorno. L'interazione elettrostatica QM-MM è incorporata nel calcolo QM, che è così più accurato, anche se aumenta il costo computazionale,

specialmente per calcoli della forza elettrostatica dovuti alla densità QM, che agisce su numerose cariche MM.

Tuttavia restano alcune incongruenze. Il modello di cariche puntiformi non è stato derivato per una fedele rappresentazione della distribuzione di carica reale e la sua validità è collegata agli altri parametri del campo di forza. Così, in linea di principio, questo schema non è legittimo, ciò nonostante è divenuto oggi la procedura standard per questo tipo di calcoli, infatti l'esperienza mostra che esso generalmente fornisce risultati ragionevoli. Un'attenzione particolare dovrebbe essere prestata in corrispondenza del confine QM-MM, dove le cariche MM più vicine potrebbero creare un sovrappolarizzazione della densità elettronica.

Vari approcci con lo scopo di minimizzare la sovrappolarizzazione sono stati suggeriti in letteratura. Forse la maniera più semplice consiste nell'eliminare le cariche vicine alla regione QM. Differenti implementazioni di questa idea prevedono di eliminare solo la carica  $M_1$ , oppure le cariche  $M_1$  e  $M_2$  o ancora tre cariche  $M_1, M_2, M_3$  (per una definizione di  $M_1, M_2, M_3$  vedi figura 5.2 e sezione 5.8). Il pericolo di questi schemi è che la semplice eliminazione delle cariche in accordo con la connettività del sistema spesso creano cariche nette in prossimità della regione QM, che possono condurre a seri problemi. Un'alternativa all'eliminazione delle cariche è il *charge shifting*, in cui la carica  $M_1$  viene spostata sugli atomi  $M_2$  legati a  $M_1$ . Il dipolo risultante è compensato collocando una coppia di cariche puntiformi vicino a  $M_2$ .

### *Embedding polarizzato*

L'embedding elettrostatico fornisce una descrizione asimmetrica dell'interazione elettrostatica QM-MM. Mentre la polarizzazione della regione interna viene ammessa, quella della regione esterna no. In prima approssimazione però, le cariche fisse puntiformi MM possono essere polarizzate dalla regione QM, ma il campo risultante non agisce a sua volta sulla densità QM. Questa procedura può essere realizzata in maniera *self-consistent*, includendo i risultanti dipoli indotti nell'Hamiltoniano QM. Una varietà di modelli possono essere concepiti allo scopo di prendere in considerazione la polarizzazione classica: (i) il modello dei dipoli polarizzati, nel quale le polarizzabilità assegnate agli atomi interagiscono con il campo elettrico inducendo i dipoli; (ii) il modello degli oscillatori di Drude, che utilizza un molla per collegare un carica fissa ad una mobile; (iii)

il modello delle cariche “fluttuanti”, che prevede l’ottimizzazione delle cariche parziali, rispetto all’energia elettrostatica totale.

Questi schemi dell’*embedding* polarizzato hanno comunque una presenza davvero limitata nelle implementazioni dei campi forza, come del resto sono pochi i campi di forza MM polarizzati, proposti in letteratura. Come visto con l’*embedding* elettrostatico, l’incremento del livello di complessità determina l’aumento della complessità computazionale. In questo caso poi l’aggiunta di cicli iterativi *self-consistent* per la densità di carica QM ed il modello MM polarizzabile può condurre addirittura a problemi di convergenza.

### 5.7 Trattazione del legame all’interfaccia

Vari approcci sono stati proposti per trattare i legami covalenti che vengono tagliati nella regione della frontiera tra i *layer* QM e MM. Una soluzione a questo problema è quella di definire il sottosistema in maniera tale che non ci siano legami covalenti da tagliare nella zona di confine. Questo è triviale quando si effettui uno studio di solvatazione esplicita, in cui il soluto viene trattato a livello QM, mentre il solvente attraverso un campo di forza MM. Una circostanza del medesimo tipo si verifica nel caso di sistemi biomolecolari in cui i reagenti non siano legati covalentemente all’enzima ed il resto del residuo proteico non venga coinvolto direttamente nella trasformazione chimica. Tuttavia nella maggior parte dei casi è necessario effettuare il *cut* di diversi legami covalenti che attraversano l’interfaccia QM-MM. Ne derivano due problemi da affrontare: valenze non sature degli atomi QM coinvolti e la sovrappolarizzazione della densità QM a causa delle vicine cariche parziali MM.

I differenti modelli che forniscono una soluzione ai problemi dei legami covalenti alla frontiera QM-MM delle due regioni, o almeno la maggior parte di essi, possono essere classificati in due gruppi, quello degli *orbitali frozen* e quello dei *link atom*.

I metodi basati sugli orbitali localizzati “congelati” [40] ricorrono ad un set di orbitali ibridi per l’atomo QM per il quale il legame è stato spezzato. Questo approccio elegante, ha senso fisico, tuttavia presenta una formulazione ed una implementazione abbastanza laboriose e complesse.

I metodi basati sui *link atom* [36,37] introducono un extra atomo, spesso l'atomo di idrogeno, lungo il legame spezzato ad una distanza appropriata dagli atomi QM. Il *link atom* è trattato a livello quantistico mentre il legame tra gli atomi QM e MM è descritto dalla meccanica molecolare. Nel presente lavoro è stato adottato proprio l'approccio *link atom*, che viene discusso con maggior dettaglio nella prossima sezione.

## 5.8 Link atom

Anche se i metodi *link atom* hanno ricevuto numerose critiche, come la necessità di introdurre atomi extra e non fisici all'interno del sistema, la loro semplicità concettuale rimane un argomento significativo a sostegno del loro utilizzo.

Uno schema descrittivo di questo metodo è illustrato nella figura 5.2. Il link atom L è posizionato ad una certa distanza lungo il vettore del legame  $M_1-Q_1$ , anche se è completamente libero durante l'ottimizzazione della struttura. Questo può condurre ad incongruenze poiché  $Q_1-L$  e  $M_1-Q_1$  potrebbero deviare dall'allineamento, per cui un modo di evitare questo problema, è quello di introdurre un angolo artificiale  $L-Q_1-M_1$  con valore di equilibrio di  $0^\circ$ .

L'eccesso di gradi di libertà introdotti dal link atom può essere eliminato, costringendo l'atomo  $M_1$  a posizionarsi lungo il legame  $Q_1-L$ , come è stato suggerito da Maseras e Morokuma [41]. Questo approccio rimuove però quattro piuttosto che tre gradi di libertà, per cui la procedura è stata modificata in modo da far dipendere le coordinate del link atom dalle coordinate degli atomi  $Q_1$  e  $M_1$ :

$$r_L = r_{Q_1} + g(r_{M_1} - r_{Q_1})$$

dove  $r$  è un vettore posizione. Questa definizione posiziona L lungo il legame  $M_1-Q_1$  e mette in relazione la distanza  $Q_1-L$  con la distanza  $Q_1-M_1$  tramite il fattore di *scaling*  $g$ . E' possibile definire il fattore  $g$  come il rapporto tra le lunghezze di legame all'equilibrio  $Q_1-L$  e  $Q_1-M_1$ , riportate nel campo di forza adottato. Ad esempio se vogliamo tagliare un legame singolo C-C ed al suo posto usare un legame C-H, un ragionevole valore di  $g$  è dato dal valore standard delle lunghezze del legame C-H (1.08 Å) diviso per la lunghezza standard del legame C-C (1.528 Å), ovvero  $g = 0.709$ .

Un'altra possibilità è definire  $g$  come:

$$g = \frac{r_{Q_1-L}}{|r_{M_1} - r_{Q_1}|}$$

dove  $r_{Q_1-L}$  è una costante predefinita. Altre implementazioni tengono conto anche della rigidità dei legami  $Q_1-M_1$  e  $Q_1-L$ .

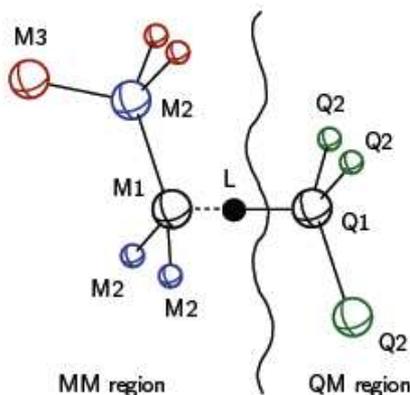


Figura 5.2. Illustrazione dello schema link atom. Il link atom L si trova lungo il legame  $Q_1-M_1$ . Gli atomi direttamente legati a  $M_1$  sono etichettati  $M_2$  se sono separati da  $M_1$  da due legami,  $M_3$  se vengono separati da  $M_1$  da tre legami... Stessa convenzione si applica agli atomi  $Q_M$ .

## 5.9 Forze sui link atoms

Se la posizione dell'atomo L è espressa come una funzione delle coordinate delle altre posizioni atomiche, le sue coordinate sono eliminate dal set che serve a descrivere l'intero sistema. In questa maniera l'atomo diventa "invisibile" alle ottimizzazioni di geometria e agli altri algoritmi, che debbono gestire tutto il set di coordinate. Tuttavia, poiché gli algoritmi dei programmi di chimica quantistica, trattano il *link atom* come gli altri atomi, le forze che agiscono su L sono non nulle.

Nello schema IMOMM [41] le forze sono "trasmesse" agli atomi vicini, lavorando in coordinate interne. Quando si lavora in coordinate cartesiane lo stesso effetto può essere ottenuto usando la *chain rule* per stabilire il contributo delle forze che agiscono su un

atomo K del sistema  $S$ , ovvero conoscere la derivata dell'energia rispetto ad una posizione atomica  $r_K$ :

$$\frac{\partial E}{\partial r_K} = \frac{\partial E}{\partial r_K} + \frac{\partial E}{\partial r_L} \frac{\partial r_L}{\partial r_K}$$

Il contributo del *link atom* alla forza scompare se l'atomo K non è coinvolto nella definizione del *link atom*.

## 5.10 Metodo ONIOM

Il metodo denominato ONIOM [42], termine che fa riferimento esplicito alla buccia della cipolla ma che è soprattutto l'acronimo di *our N-layered integrated molecular orbital + molecular mechanics*, è un approccio computazionale ibrido, appartenente alla categoria dei metodi misti sottrattivi, sviluppato da Morokuma e collaboratori. ONIOM usa i *link atom* per saturare i legami spezzati alla frontiera. La versione corrente è l'ultima di una serie di metodi misti che includono anche IMOMM (*integrated molecular orbital + molecular mechanics*) [41] e IMOMO (*integrated molecular orbital + molecular orbital*) [42]. IMOMM and IMOMO non dovrebbero essere considerati una sottocategoria del metodo ONIOM, poiché il posizionamento del *link atom* in ONIOM viene effettuato attraverso un fattore di scala, mentre in IMOMM e IMOMO il *link atom* si trova ad una distanza specificata dall'atomo a cui viene connesso. In ONIOM *two-layer* (QM:MM) l'energia totale del sistema è ottenuta da tre calcoli indipendenti

$$E^{ONIOM} = E^{real,MM} + E^{model,QM} - E^{model,MM} \quad (5.3)$$

Il sistema *real* contiene tutti gli atomi e viene trattato solo a livello MM. Il sistema modello contiene il sistema che è trattato a livello QM. Entrambi i calcoli QM e MM devono essere eseguiti per il sistema modello (Figura 5.3)

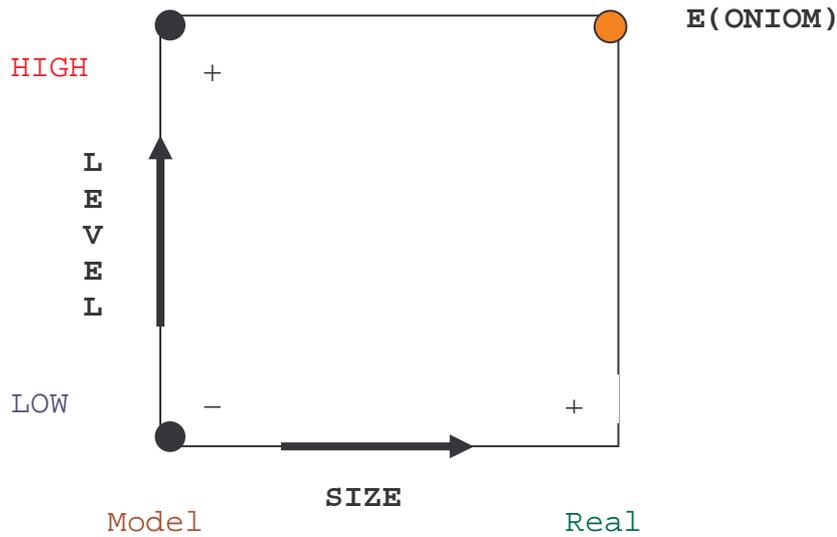


Figura 5.3: Schema ONIOM e terminologia

Nello schema ONIOM, la maggior parte dei termini MM nel sistema modello si trovano anche nel sistema reale e dunque si cancellano nella espressione dell'energia. La differenza tra i calcoli MM nel sistema reale e nel sistema modello, chiamata *S-value*, descrive il contributo che viene dalla regione MM, che a sua volta include sia l'energia della regione MM che l'interazione tra gli strati QM e MM:

$$S^{MM} = E^{real,MM} - E^{model,MM}$$

$S^{MM}$  gioca lo stesso ruolo di  $E_{MM}(O) + E_{QM-MM}(I,O)$  nel caso di uno schema additivo. Quando siano presenti interazioni non bonded tra le due regioni,  $S^{MM}$  è identico a  $E_{MM}(O) + E_{QM-MM}(I,O)$  e  $E^{ONIOM}$  diventa identica a  $E_{QM/MM}^{additivo}$ .

Come per qualunque schema sottrattivo, in ONIOM, la correzione per i *link atom* è inclusa nell'ultimo termine dell'equazione (5.3).

Come evidenziato, in precedenza, l'interazione tra il layer QM e quello MM è sempre inclusa nel livello computazionale più basso, attraverso il termine  $E^{real,MM}$ . Ne segue che ONIOM usa di default l'*embedding meccanico*. Recentemente però uno schema di *embedding elettrostatico* è stato implementato nel framework di ONIOM (ONIOM-EE) [43]:

$$E^{ONIOM-EE} = E^{real,MM} + E_v^{model,QM} - E_v^{model,MM} \quad (5.4)$$

in cui

$$E_v^{\text{model,MM}} = E^{\text{model,MM}} + \sum_J \sum_N \frac{s_{JN} q_J q_N}{r_{JN}}$$

e l'Hamiltoniano in fase gas corrispondente all'energia  $E_v^{\text{model,QM}}$  è

$$H^{\text{model}} = H_0^{\text{model}} - \sum_i \sum_J \frac{q_J}{r_{iJ}} + \sum_J \sum_N \frac{q_N Z_J}{r_{JN}}$$

dove i pedici nelle equazioni sopra riportate  $J, N, i$  si riferiscono rispettivamente agli atomi nel sistema reale, a quelli nel sistema modello e agli elettroni, mentre  $s_{JN}$  è un fattore di *scaling* della carica. Con questa implementazione l'interazione elettrostatica è valutata tre volte, una volta per ogni termine del membro di destra dell'equazione (5.4) (l'energia  $E^{\text{real,MM}}$  include anche l'interazione elettrostatica tra i due layer a livello MM). Poiché molti termini elettrostatici del primo e terzo termine si cancellano, il formalismo ONIOM-EE può essere riguardato come più o meno equivalente all'*embedding elettrostatico* di uno schema additivo QM/MM. Di nuovo, se interazioni non di legame sono presenti tra gli strati le due versioni dell'*embedding elettrostatico* diventano identiche. È importante notare che in ONIOM i *link atom* “vedono” le cariche circostanti. Quando i layer sono legati covalentemente, ONIOM adotta uno schema di *scaling*, onde superare i problemi connessi con la presenza di cariche puntiformi troppo vicine. Tipicamente le interazioni tra i centri, che sono separati da uno o due legami, vengono scalate a zero, quelle tra atomi separati da tre legami sono scalate di 0.5, mentre le interazioni tra cariche separate da oltre tre legami vengono pienamente incluse nel calcolo.

Il principale vantaggio del metodo ONIOM è senz'altro quello della sua flessibilità, permettendo la combinazione di differenti livelli di metodi QM con la MM in un singolo calcolo integrato dell'energia e di altre proprietà elettroniche. Questo aspetto ha contribuito notevolmente alla popolarità e diffusione di questo metodo ibrido, nella comunità scientifica.

## CAPITOLO 6

### Risultati e Discussione

#### 6.1 Elettrostatica delle proteine

E' ben noto che le interazioni Coulombiane, specialmente carica-carica, carica-dipolo e dipolo-dipolo svolgono un ruolo centrale sia nella determinazione della struttura che per la stabilità di una proteina, in dipendenza da grandezze macroscopiche come pH, forza ionica, temperatura. Questa nozione è a sua volta fondamentale per una più approfondita comprensione delle relazioni proprietà/struttura nelle biomolecole.

Il primo passo per la messa a punto del protocollo computazionale che si intende proporre è un calcolo elettrostatico sulla Calbindina D9k. A tale proposito sono state calcolate le costanti di dissociazione pKa dei residui acidi della proteina allo scopo di stabilire se i siti di binding del protone fossero protonati o deprotonati a pH fisiologico.

I calcoli di pKa inoltre permettono di capire se il valore della costante di dissociazione di un amminoacido titolabile nell'ambiente proteico differisca in maniera significativa dal valore del pKa che il medesimo amminoacido presenta in soluzione.

#### *Proteine e gruppi titolabili*

Una delle più frequenti reazioni chimiche che avvengono in una soluzione proteina-acqua è quella di protonazione/deprotonazione degli amminoacidi. Le parti di amminoacidi che legano o rilasciano protoni vengono riferite come gruppi titolabili della proteina. Allora la catena laterale di un acido aspartico contiene un singolo gruppo titolabile (il carbossile – COOH) mentre una lisina N-terminale contiene due gruppi titolabili, l'ammino gruppo N-terminale e l'ammino gruppo della catena laterale. I gruppi titolabili di una proteina si dividono in due categorie: acidi (Asp, Glu, Cys, Tyr e C-terminali) e basi (Hys, Lys, Arg, N-terminali). I gruppi titolabili basici sono carichi positivamente nel loro stato protonato e i gruppi titolabili acidi sono neutri nel loro stato protonato.

In generale, per gli acidi HA, la reazione di dissociazione è



La costante di dissociazione (o ionizzazione)  $\text{pK}_a = -\log(\text{K}_a)$  è definita dalla seguente equazione

$$\text{pK}_a = \text{pH} + \log\left(\frac{[\text{protonato}]}{[\text{deprotonato}]}\right)$$

dunque il  $\text{pK}_a$  di un sito può essere riguardato come il pH al quale le frazioni protonata e deprotonata sono uguali (infatti  $\text{pH} = -\log[\text{H}_3\text{O}^+]$ ). Se il pH è più alto del  $\text{pK}_a$  il sito è principalmente deprotonato, mentre se il valore di pH è più basso di  $\text{pK}_a$  allora il sito è quasi del tutto protonato. Dunque se è noto il valore di  $\text{pK}_a$  per un gruppo titolabile è possibile predire la sua carica elettrica se la proteina è in soluzione ad un dato pH.

E' possibile ottenere una stima dello stato di protonazione di una proteina semplicemente assumendo che i valori dei  $\text{pK}_a$  nella proteina *folded* siano gli stessi dei medesimi gruppi titolabili in acqua (composti modello). In alcune situazioni speciali, è tuttavia necessario ottenere una più dettagliata descrizione dei valori di  $\text{pK}_a$ , ad esempio quando si studia un meccanismo enzimatico o la stabilità del biopolimero; in questi casi le tecniche di calcolo di  $\text{pK}_a$  forniscono una maniera di calcolare l'effetto dell'intorno proteico sui valori di  $\text{pK}_a$  dei gruppi titolabili.

Quando una proteina si avvolge (*folding*) gli amminoacidi titolabili infatti passano da una soluzione ad un ambiente determinato dalla struttura tridimensionale della proteina. Un acido aspartico in una proteina *unfolded* è tipicamente in un ambiente che espone la catena laterale titolabile all'acqua. Nella proteina *folded* invece l'acido aspartico potrebbe trovarsi all'interno della macromolecola, senza esposizione al solvente, vicino ad altri gruppi titolabili ed interagirà presumibilmente con ioni (cariche permanenti) e dipoli presenti nella proteina. Tutti questi effetti alterano, ovviamente, il valore di  $\text{pK}_a$  della catena laterale dell'amminoacido.

Diversi pacchetti software sono disponibili per il calcolo dei valori di  $\text{pK}_a$  di una proteina. Alcuni metodi sono basati sulle soluzioni dell'equazione di Poisson-Boltzmann (PBE), che è una modifica dell'equazione di Poisson, che incorpora anche una descrizione degli effetti degli ioni del solvente sul campo elettrostatico intorno alla molecola

$$\nabla[\varepsilon(r)\nabla\phi(r)] = -4\pi(\rho_{ioni}(r) + \rho(r))$$

dove  $\rho_{ioni}$  è la densità di carica di tutti gli ioni, che dipende oltre che dalle concentrazioni e cariche delle specie ioniche presenti anche dalla temperatura T.

Questi metodi, spesso denominati anche come metodi FDPB (*Finite Difference Poisson-Boltzmann*), calcolano la variazione nel valore del pKa della catena laterale di un amminoacido quando questo passi da un ipotetico stato completamente solvatato alla sua posizione nella proteina. Per eseguire un siffatto calcolo è necessario disporre di strumenti teorici per calcolare l'effetto dell'interno di una proteina sul valore di pKa e, naturalmente, la conoscenza dei valori di pKa degli amminoacidi nel loro stato di solvatazione.

## 6.2 Calcolo dei pKa nella Calbindina

Il calcolo dei valori di pKa dei gruppi titolabili della Calbindina (CAB) a pH neutro è stato eseguito adottando il pacchetto MEAD (*Macroscopic Electrostatic with Atomic Detail*), sviluppato da Bashford e collaboratori [44]. Questo software genera energie elettrostatiche attraverso una soluzione alle differenze finite dell'equazione di Poisson-Boltzmann, che viene considerato un buon modello per descrivere macromolecole in soluzioni acquose. Nell'ambito di questo approccio la proteina viene modellizzata come un materiale poco dielettrico (con costante dielettrica generalmente tra 2.0 e 4.0) con cariche parziali immerse e circondate da un mezzo altamente dielettrico (il solvente infatti è considerato come un continuo con costante dielettrica pari a circa 80.0). I dettagli atomici della struttura proteica vengono usati per definire il confine ed il posizionamento delle cariche, mentre il metodo FDPB è usato per risolvere le risultanti equazioni differenziali parziali per il potenziale elettrostatico.

Per evitare valori di pKa non fisici è stato scelto di adottare una costante dielettrica  $\varepsilon = 40.0$  per schermare almeno parzialmente le interazioni tra gli atomi elettricamente carichi degli amminoacidi coinvolti nel binding dello ione calcio.

Nelle tabelle 1 e 2 viene riportato il confronto tra i valori di pKa sperimentali e quelli calcolati, rispettivamente per la forma *Apo* e per la forma *Holo* della CAB.

Tabella 1. Confronto tra i valori calcolati e sperimentali pKa della *Apo*-Calbindina

Residue	Calc.	Exp. <sup>a</sup>
LYS 1	11.2	10.6
GLU 4	3.6	3.8
GLU 5	3.3	3.4
LYS 7	11.3	11.4
GLU 11	4.1	4.7
LYS 12	11.4	11.1
LYS 16	13.7	10.1
GLU 17	4.4	3.6
GLU 26	4.3	4.1
LYS 29	10.9	11.3
LYS 41	11.2	10.9
ASP 47	3.4	3.0
GLU 48	4.5	4.6
GLU 64	4.3	3.8
LYS 71	11.1	10.7
LYS 72	11.6	11.3

(a) Valori sperimentali misurati da Kesvatera *et al.*[45]

Carica elettrica totale a pH= 7 -7

Table 2. Confronto tra i valori calcolati e sperimentali pKa della *Holo*-Calbindina

Residue	Calc.	Exp. <sup>a</sup>
LYS 1	11.6	10.6
GLU 4	4.0	3.8
GLU 5	3.3	3.4
LYS 7	11.9	11.4
GLU 11	4.4	4.7
LYS 12	11.7	11.1
LYS 16	11.9	10.1
GLU 17	4.0	3.6
GLU 26	4.1	4.1
LYS 29	11.2	11.3
LYS 41	10.7	10.9
ASP 47	2.4	3.0
GLU 48	4.1	4.6
GLU 64	4.1	3.8
LYS 71	10.7	10.7
LYS 72	11.6	11.3

(a) Valori sperimentali misurati da Kesvatera *et al.*[45]

Carica elettrica totale a pH= 7 -3

Come può essere osservato l'accordo è molto buono. Questo studio elettrostatico ci ha permesso di conoscere la cariche dei vari siti di protonazione della Calbindina D9k e ovviamente di avere una conferma sulla carica totale della proteina nelle forme *Apo* e *Holo* (Appendice C).

### 6.3 Calcolo dei Chemical Shieldings

#### *Scelta del metodo computazionale*

Svariati modelli empirici sono stati sviluppati e proposti in letteratura per calcolare i chemical shift delle proteine [10]. Tuttavia il successo dei metodi empirici richiede un *basis set* di chemical shift già noti per derivare un set di parametri di buon livello. Sicuramente non è semplice generalizzare un approccio empirico per trattare proteine con residui non standard, cofattori metallici (come nei metallo-enzimi) e complessi proteina-ligando.

Come già accennato nel Capitolo 1 la soluzione è rappresentata dai metodi di calcolo quantistici da principi primi, che tuttavia dipendono delle dimensioni del sistema molecolare. Tali approcci dipendono almeno dalla terza potenza del numero totale di elettroni (nel caso dei metodi HF e DFT). Dunque l'applicazione dei convenzionali metodi *ab initio* alle biomolecole, come le proteine, è troppo costosa e per alcuni versi una vera e propria sfida, per cui, è di fatto limitata solo a piccoli polipeptidi.

Tra i recenti sviluppi dei codici quantomeccanici si possono annoverare però diversi tentativi di simulare e descrivere questi sistemi macromolecolari a livello elettronico. Tra questi occupano un posizione prominente i metodi misti QM/MM, per i quali, come visto nel Capitolo 5, il trattamento QM è limitato alla parte del sistema per la quale si ritiene necessaria un dettagliata descrizione elettronica, mentre il resto viene trattato con la meccanica classica. Un'altra strategia è operativa invece nell'ambito dei metodi *linear scaling* che tendono a sviluppare astuti trucchi nell'implementazione del codice per fare aumentare il tempo di calcolo linearmente con il numero di atomi, costituenti la molecola. Infine, i ben noti metodi semiempirici, che oggi sono ancora più efficienti rispetto al

passato sia per quanto concerne le risorse di calcolo richieste che per l'attendibilità dei loro risultati e che, quindi, rappresentano un interessante compromesso tra i sofisticati metodi *ab initio* e quelli meramente classici, che fanno uso di un campo di forza empirico.

Nel caso dei calcoli NMR, il frammento importante è il cromoforo NMR ed i chemical shift degli atomi del cromoforo sono influenzati dalla presenza di legami idrogeno, dagli angoli diedri del *backbone* molecolare, dall'orientazione delle catene laterali degli amminoacidi prossimi nello spazio intorno ad esso. Poiché allora i chemical shift sono altamente dipendenti dall'intorno chimico locale e dagli effetti elettrostatici *long range*, il trattamento del cromoforo da solo è inadeguato. Per queste ragioni è stato scelto di usare un metodo misto QM/MM per i calcoli NMR per le proteine *calcium-binding*. Nell'ambito di un siffatto approccio, i metodi *ab initio* quantistici possono essere applicati per descrivere le principali interazioni locali *short range* che contribuiscono ai *chemical shift*. Le interazioni locali *short range* sono dovute in gran parte ai residui connessi tra loro in maniera sequenziale, ma anche a legami idrogeno, effetti di correnti di anello ed altre interazioni di van der Waals e elettrostatiche con i residui che sono vicini nello spazio anche se lontani nella sequenza. Invece il modello MM classico fornisce il potenziale elettrostatico *long range*.

D'altro canto è stato mostrato [46] che una buona accuratezza nel calcolo NMR può essere conseguita adottando i metodi ibridi QM/MM, poiché la distribuzione di densità elettronica locale è adeguata a descrivere gli effetti QM sui chemical shift NMR (che sono proprietà fisiche locali). Recentemente, in letteratura, sono stati riportati calcoli di chemical shift di proteine realizzati con un approccio ibrido QM/MM. Generalmente la parte QM è trattata con la Teoria del Funzionale della Densità elettronica, considerato uno dei metodi quantistici più appropriati per trattare le metalloproteine [47].

In questo lavoro, in particolare, è stato adottato il metodo ONIOM nell'ambito di un *framework* quantistico/classico, in cui il sistema viene suddiviso in due parti: la parte importante, che nel caso presente corrisponde alla regione sulla quale si effettuano i calcoli NMR, trattata con la quantomeccanica, al livello di teoria DFT, mentre il resto del sistema viene modellizzato tramite cariche parziali puntiformi dal campo di forze Amber. La frontiera tra i due *layers* è basata su uno schema *hydrogen link atom* e le interazioni elettrostatiche tra le regioni QM e MM sono incluse nel calcolo, adottando l'implementazione dell'*embedding elettronico* di ONIOM (ONIOM-EE). In tale maniera la

risultante descrizione quantistica include la polarizzazione del modello ridotto (QM) indotta dall'ambiente circostante.

Una particolare attenzione nei calcoli di proprietà magnetiche come tensori di schermo o suscettibilità magnetiche deve essere dedicata all'invarianza del gauge (descritta in Appendice B). Esistono vari metodi proposti in letteratura che permettono di affrontare questo problema.

Il metodo adottato in questo lavoro è il metodo GIAO (*Gauge Invariant Atomic Orbitals*) [48]. I GIAO sono funzioni di base atomiche in cui le funzioni convenzionali vengono moltiplicate per un fattore complesso, che contiene il potenziale vettore. Può essere mostrato che quando si usano questi orbitali gli elementi di matrice dell'Hamiltoniano dipendono solo da differenze tra i potenziali vettori, in tal modo ogni singolo integrale è indipendente dall'origine del potenziale vettore

$$\chi_A = \chi_A(0)e^{\left[\frac{-i}{2c}(\vec{B} \times \vec{R}_A) \cdot \vec{r}\right]}$$

dove  $\chi_A(0)$  è un orbitale atomico gaussiano,  $B$  è il campo magnetico esterno, mentre  $R_A$  e  $r$  sono i vettori spaziali di nucleo ed elettrone, rispettivamente.

La componente del tensore di schermo chimico è la derivata seconda dell'energia elettronica rispetto al campo magnetico esterno  $B$  e al momento magnetico del nucleo, calcolata in accordo con il formalismo della teoria delle perturbazioni (vedi Appendice B). Nel presente lavoro, quindi, le costanti di schermo (e in ultima analisi i chemical shifts) sono state ottenute, in accordo con il metodo GIAO, a livello DFT(B3LYP)-ONIOM-EE della teoria. Come funzionale della densità è stato scelto il funzionale ibrido B3LYP, la cui abilità a riprodurre gli schermi NMR anche con set di base piccoli è stata ben documentata [47]. Tutti i calcoli di *chemical shielding* sono stati realizzati con il pacchetto di programmi Gaussian 09 [49], in parte implementato anche nel gruppo di ricerca del Laboratorio di Struttura e Dinamica Molecolare.

### *Electrostatic embedding*

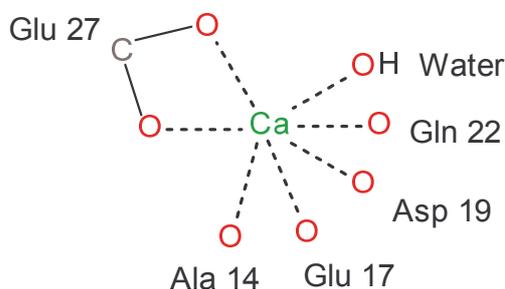
L'accoppiamento elettrostatico tra la densità di carica QM e il modello di carica usato nella regione MM può essere preso in considerazione a differenti livelli. Nello schema adottato, ovvero quello dell'embedding elettronico o elettrostatico, la struttura elettronica della

regione *inner* può adattarsi ai cambiamenti che avvengono nella distribuzione di carica dell'ambiente circostante ed è automaticamente polarizzata da questa. L'interazione elettrostatica QM-MM è trattata a livello QM e dunque fornisce una accurata descrizione ma aumenta anche se di poco la richiesta di risorse computazionali, specialmente per il calcolo delle forze di Coulomb che si esplicano tra la densità di carica QM e le numerose cariche puntiformi MM. Una cura particolare è richiesta per il *boundary* QM/MM, dove le cariche MM sono posizionate in prossimità della regione QM e della sua densità elettronica, ciò che può causare una sovrappolarizzazione. Questo problema è pronunciato specialmente quando il *boundary* attraversa un legame covalente. Per trattare la carica negli approcci *link atom* sono stati testati tre schemi di eliminazione della carica, in cui cariche puntiformi MM vengono eliminate in maniera selezionata. Se solo la carica  $M_1$  è posta uguale a zero, si ha lo schema Z1. Se vengono eliminate le cariche  $M_1$  e  $M_2$  lo schema si chiama Z2. Lo schema Z3 infine denota un trattamento delle cariche in cui tutte le cariche  $M_1$ ,  $M_2$ ,  $M_3$  vengono azzerate. Sulla base di alcuni calcoli effettuati su peptidi modello è stato trovato che lo schema che fornisce i risultati più vicini ad un calcolo QM è quello denominato Z1. Il significato dei simboli  $M_1$ ,  $M_2$ ,  $M_3$  è riportato nel Capitolo 5 dedicato alla descrizione dei metodi ibridi.

#### *Messa a punto della partizione QM/MM*

Un aspetto importante da considerare quando si adottino metodi misti QM/MM concerne la scelta di una regione QM sufficientemente ampia in modo che l'intorno locale sia ragionevolmente modellizzato. Lo scopo di questo step inerisce quindi, prima di tutto, alla definizione della regione QM (quella dei calcoli NMR) intorno ad ogni ione calcio. Questa regione deve essere definita in maniera tale da includere i diversi contributi allo schermo chimico degli atomi che costituiscono i residui amminoacidici legati al calcio: i contributi che derivano dalla struttura elettronica, ma anche quelli dovuti alle altre parti del sistema. E' stato deciso di partire con la struttura sperimentale determinata attraverso l'NMR della forma *Apo* della Calbindina D9k (file in formato pdb identificato come 1CLB dal database Protein Data Bank). In tal modo si evitano in questa fase preliminare i possibili problemi collegati alla presenza del metallo, lo ione Calcio. Con il metodo di calcolo scelto e descritto nella sezione precedente abbiamo calcolato i tensori di chemical shielding degli

amminoacidi che sono presunti come ligandi del calcio: Ala 14, Glu 17, Glu 27, Asp 19, Gln 22 (Figura 6.1).



**Figura 6.1** Sito di coordinazione del calcio nella Calbindina

Successivamente è stato ampliato il *layer* quantistico, includendo anche gli amminoacidi che sono legati in maniera covalente a ciascun ligando  $i$ , e che è possibile denominare, rispettivamente, come residui  $i-1$  e  $i+1$ . Indi sono stati inclusi nella costruzione della regione QM anche gli amminoacidi che formano legami a idrogeno con ciascuno dei residui  $i$ ,  $i-1$  e  $i+1$ . Infatti è noto che i chemical shift dei terminali N e C' di un residuo  $i$  sono influenzati anche dai legami a ponte di idrogeno che coinvolgono i gruppi adiacenti  $i-1$  e  $i+1$ .

Tutti i legami a H della proteina CAB sono stati verificati usando un software freeware, per la visualizzazione e l'analisi di strutture molecolari, denominato *CHIMERA an extensible molecular modeling system* e sviluppato dal Resource for Biocomputing, Visualization, and Informatics (del National Institutes of Health NIGMS P41-GM103311) [50].

Nella seguente tabella vengono riportati i legami a H con gli amminoacidi donatori e accettori coinvolti.

Allo scopo di ottenere una convergenza interna dei chemical shielding calcolati per ciascun atomo dei ligandi del calcio, è stato incluso nell'intorno di ciascun ligando  $i$  anche l'effetto delle correnti di anello, generalmente associato alla presenza di residui amminoacidici con un anello aromatico, ad esempio Phe. Questo effetto viene così chiamato perché descritto fisicamente come una circolazione di elettroni sugli anelli aromatici dei sistemi  $\pi$ , che determina un "deschermaggio" nel medesimo piano dell'anello, esternamente ad esso e viceversa uno "schermaggio" nella regione sopra e sotto il piano dell'anello.

Si è deciso sulla base dei risultati dei calcoli di prendere in considerazione anelli aromatici nell'intorno chimico di ogni ligando, entro una distanza di circa 5 Å da ogni atomo del residuo *i*.

I risultati dei calcoli sono riportati nelle tabelle 3-7 in cui abbiamo riportato nella prima colonna lo schermo chimico di ciascun ligando *i* (l'amminoacido che partecipa al binding dello ione calcio), mentre nelle altre colonne abbiamo riportato i valori ottenuti adottando una layer sempre più ampio.

Tabella 3. Schermi magnetici isotropi in p.p.m. di Ala 14

Atomo	Ligando <i>i</i>	<i>i</i> +1/ <i>i</i> -1	<i>H-bond</i>	<i>H-bond</i> ( <i>i</i> +1 e <i>i</i> -1)	Effetti Spaziali
N	133.1725	134.4993	135.0898	134.9298	134.3348
H	25.9479	25.8297	23.9141	23.7234	23.6026
CA	139.3672	137.7715	138.8636	138.7649	138.1599
HA	27.9987	28.0759	28.1751	28.2257	27.7792
CB	173.6702	173.6368	173.3814	172.4838	172.5243
HB1	31.1654	31.1396	32.3741	32.3928	32.1763
HB2	30.9573	30.9826	31.6755	31.3348	31.4194
HB3	29.4915	29.4502	30.4848	30.5831	30.3105
C'	29.4460	31.7665	30.4991	29.9880	28.2907

Confronto tra gli schermi magnetici degli atomi dell'Ala 14 ottenuti aggiungendo i residui legati covalentemente (*i*+1 Ala15 e *i*-1 Tyr13), i residui che formano legami H (Phe 10 che si lega all'N dell'Ala 14 e Lys 16 che si lega all'O dell'Ala 14) e residui che formano legami idrogeno con i primi vicini (Glu 11 con NH di Ala 15). Gli schermi calcolati a livello b3lyp/GIAO-6-31g(d,p): cariche amber (ONIOM-EE) in ppm.

Tabella 4. Schermi magnetici isotropi in p.p.m. Glu 17

Atomo	Ligando <i>i</i>	<i>i</i> +1/ <i>i</i> -1	Effetti spaziali
N	127.2057	129.8402	129.6660
H	26.3463	25.4754	25.4659
CA	140.2198	138.6810	138.3986
HA	27.1524	27.3911	27.4669
CB	151.7486	151.5271	151.6767
HB2	31.1702	31.2117	31.1404
HB3	29.7838	29.8979	29.8546
CG	154.3620	154.3553	154.3588
HG2	29.3431	29.3231	29.3456
HG3	30.7077	30.8167	30.7114
C'	31.6754	31.2842	31.3184

Confronto tra gli schermi magnetici degli atomi di Glu 17 ottenuti aggiungendo i residui legati covalentemente (Lys 16 e Gly 18) a livello b3lyp/GIAO-6-31g\*: cariche amber (ONIOM-EE). Gli schermi sono in p.p.m. I residui che formano legami H già sono inclusi.

Tabella 5. Schermi magnetici isotropi in p.p.m. di Asp 19

Atomo	Ligando <i>i</i>	<i>H-bond</i>	Effetti spaziali
N	134.1473	135.8451	141.0513
H	24.3136	24.2035	24.1891
CA	137.2222	137.0175	136.9167
HA	27.5028	27.4344	27.3099
CB	144.5609	144.1081	143.4302
HB2	29.2946	29.2206	28.8908
HB3	28.4220	28.4170	28.3007
C	31.7595	31.7144	31.2073
C'	28.4071	31.7144	22.7331

Confronto tra gli schermi magnetici degli atomi di Asp 19 ottenuti aggiungendo residui che formano legami H (Lys 16 che forma legame H con CO di Asp19) a livello b3lyp/GIAO-6-31g\*: cariche amber (ONIOM-EE). Gli schermi sono in ppm.

Tabella 6. Schermi magnetici isotropi in p.p.m. di Gln 22

Atomi	Ligando <i>i</i>	<i>i+1/i-1</i>	<i>H-bond</i>	<i>H-bond</i> ( <i>i+1/i-1</i> )	Phe 50	Phe10/ Effetti spaziali
N	133.9793	133.1815	133.9259	132.7687	132.7950	130.9113
H	25.8489	25.6114	25.5731	25.5445	25.5209	25.3153
CA	134.0737	131.2249	131.0023	132.1308	132.1134	131.3848
HA	27.8587	28.0457	27.9784	27.7436	27.7243	27.2963
CB	157.6345	158.5265	158.4422	158.7891	158.7473	158.6509
HB2	28.7311	29.0266	28.9950	29.0499	29.0132	28.9164
HB3	30.4171	29.9930	29.9534	30.0104	29.9793	29.8804
CG	154.8301	154.8287	154.8424	153.2612	153.2323	153.0405
HG2	30.7020	30.7772	30.7605	30.9331	30.9081	30.8293
HG3	28.0854	28.4918	28.4652	28.4388	28.4112	28.3010
CD	28.9946	29.2466	29.3898	28.7353	28.6268	28.7015
N	155.7915	155.8792	155.7384	154.9314	154.9400	154.6244
HE21	24.8497	24.8474	24.8404	24.7874	24.7456	24.6917
HE22	25.3843	25.2964	25.2669	25.3215	25.2955	25.231
C'	37.8883	35.5636	35.5077	35.5470	35.4888	34.9399

Confronto tra gli schermi magnetici degli atomi di Gln 22 ottenuti aggiungendo i residui legati covalentemente (Asn 21 e Leu 23); i residui che formano legami H (Lys 16); i residui che formano legami H con i primi vicini. Gli schermi calcolati a livello a livello b3lyp/GIAO-6-31g(d,p): cariche amber (ONIOM-EE) sono in p.p.m.

Tabella 7. Schermi magnetici isotropi in p.p.m. di Glu 27

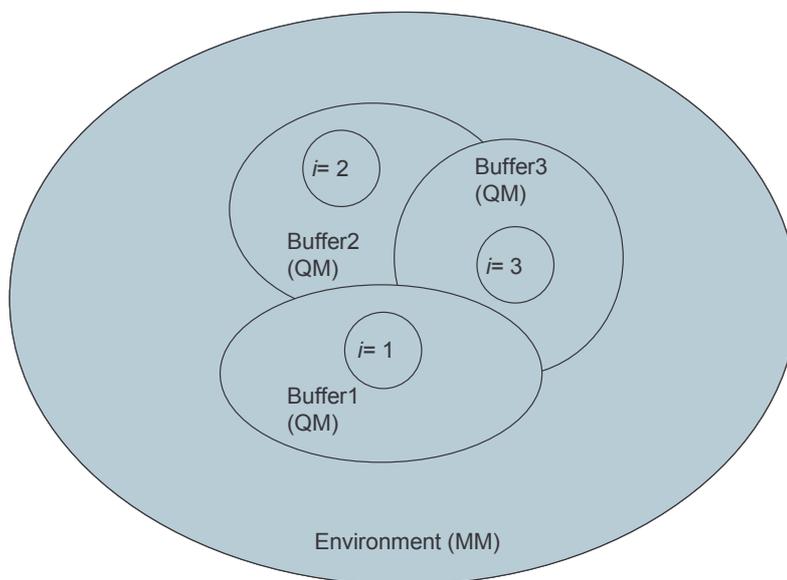
Atomi	Ligando $i$	$i+1/i-1$	$H-bond$	$H-bond (i+1/i-1)/$ Effetti spaziali
N	124.9811	134.9971	131.9575	128.3023
H	24.3482	24.4044	22.9669	23.1057
CA	131.6841	130.9038	130.3406	130.7429
HA	27.4695	27.7539	27.9572	27.5586
CB	161.2417	160.8505	162.6364	161.5899
HB2	30.5259	30.6664	29.8910	29.8711
HB3	30.4566	30.4798	30.3222	30.3898
CG	148.4537	148.7726	151.1574	148.8191
HG2	29.8123	29.8507	29.4696	29.8547
HG3	30.2123	30.1943	29.9881	29.9622
C'	31.4292	30.3818	27.2390	29.5935

Confronto tra gli schermi magnetici degli atomi di Glu27 ottenuti aggiungendo i residui legati covalentemente (Glu 26 e Leu 28) ; i residui che formano legami H (ser 24 con N di Glu 27 e Leu 31 con O di Glu 27) ; i residui che formano legami H con  $i-1$  e  $i+1$  (Leu 30 forma legame H con CO di Glu26). Gli schermi calcolati a livello b3lyp/GIAO-6-31g\*: cariche amber (ONIOM-EE) sono in ppm.

Nelle tabelle 3-7 sopra riportate con la notazione  $i+1/i-1$  si intende il ligando  $i$ , insieme con i residui ad esso legati covalentemente, con  $H-bond$  si intende un *layer* intorno al ligando che comprenda anche i residui che formano legami a ponte di idrogeno con  $i$ , mentre  $H-bond (i+1/i-1)$  indica un *layer* QM che contiene anche gli amminoacidi che hanno con  $i-1$  e  $i+1$  interazioni *hydrogen-bonding*. Gli effetti spaziali sono gli effetti eventualmente imputabili a residui presenti nell'intorno tridimensionale di  $i$ .

In questo approccio il sito di binding del calcio è virtualmente suddiviso in frammenti che comprendono innanzitutto i ligandi. I residui dentro un certo range da ciascun ligando vengono assegnati ad una regione *buffer* che in qualche modo circonda  $i$ . Lo scopo di questa zona buffer è quello di includere gli effetti locali quantistici sui chemical shift, in particolare interazioni di legame a idrogeno ed effetti di correnti d'anello. La regione frammento-centrica insieme con la zona di buffer sono trattate a livello QM/DFT, mentre il resto del sistema è descritto attraverso un modello di cariche puntiformi empirico. I residui alla frontiera che subiscono dei cut sono semplicemente completati con atomi di idrogeno per avere un frammento *closed-shell*. Ciascun calcolo QM/MM viene realizzato separatamente e da ciascun calcolo vengono estratti solo gli schermi magnetici degli atomi del frammento  $i$ . Una più dettagliata illustrazione dello schema di frammentazione messo a punto in questo lavoro di tesi è riportato in figura 6.2.

Lo scopo di questo approccio non è quello di ottenere l'energia totale della proteina, ma piuttosto calcolare una proprietà, più precisamente si intende sviluppare un metodo che permetta in ultima analisi di calcolare in maniera attendibile i chemical shift della macromolecola biologica.



**Figura 6.2** Schema di frammentazione della regione di binding del calcio

Adottando infatti un criterio generale e ragionevole per assegnare una zona di buffer a ciascun residuo  $i$  è possibile ridurre la dimensione della regione QM e raggiungere un compromesso tra l'accuratezza desiderata ed i costi computazionali.

La dimensione di ogni frammento (ligando+buffer) è indipendente dalla dimensione dell'intera proteina poiché ogni residuo ha soltanto un numero limitato di amminoacidi nelle proprie vicinanze, quindi il frammento più grande contiene circa 200-250 atomi soprattutto atomi di C,H, O, N, che rappresenta un calcolo assolutamente alla portata di un approccio DFT.

Il metodo che viene qui proposto è promettente e presenta una serie di caratteristiche interessanti. Innanzitutto la costruzione della matrice rappresentativa dell'Hamiltoniano dell'intero sistema molecolare viene evitata. Tutti i calcoli QM/MM sono parallelizzabili. Non c'è quindi necessità di diagonalizzare la matrice Hamiltoniana, operazione che rappresenta un vero e proprio collo di bottiglia nei calcoli *ab initio* di proprietà macromolecolari. Infine le richieste di memoria dipendono solo dalle dimensioni del frammento più grande in cui viene suddivisa l'intera regione QM e non aumentano con le

dimensioni dell'intero sistema. Questo approccio, in linea di principio, rende possibile il calcolo degli schermi NMR su proteine reali con migliaia di atomi.

#### 6.4 Verso calcoli realistici

Per migliorare l'attendibilità dei risultati si potrebbe adottare un set di base localmente più denso, ovvero un set di base più grande sui ligandi  $i$  e set di base più piccoli sulla restante regione QM.

Allo scopo di realizzare simulazioni di schermi magnetici più realistiche, che quindi includano sia gli effetti del solvente che gli effetti dei moti interni della proteina, delle fluttuazioni nelle posizioni atomiche, sui *chemical shift*, è possibile combinare la Dinamica Molecolare con gli approcci QM.

In tal modo è possibile, in linea di principio, ottenere valori numerici consistenti con gli esperimenti NMR.

Questo schema teorico consente di estrarre *snapshots* del nostro sistema, selezionati in un certo intervallo di tempo, quindi applica i calcoli quantistici degli schermi e infine media i risultati.

In particolare gli effetti solvente possono essere inclusi nei calcoli degli schermi magnetici, modelli di acqua esplicita nelle simulazioni MD.

##### *Dinamica molecolare della CAB*

Le simulazioni sono state eseguite utilizzando il campo di forza FF99SB implementato nella suite di programmi Amber10 [51]. Le strutture iniziali sono state scaricate da Protein Data Bank (PDB). La struttura della forma *Apo* della calbindina D9k (CAB) è stata presa dal primo modello di una struttura di risonanza magnetica nucleare (NMR) (file PDB ID 1CLB). Le simulazioni sono state effettuate adottando il modello di acqua TIP3P, in particolare è stato preso in considerazione un buffer di 8 angstrom di solvente intorno alla proteina in ogni direzione, il box risultante contiene ~3600 molecole di acqua.

La neutralizzazione della carica e la forza ionica fisiologica di  $\sim 0.1$  M sono state raggiunte aggiungendo appropriate quantità di ioni  $\text{Na}^+$  al box in cui viene fatta avvenire la simulazione.

Prima di far partire la dinamica, i *constraints* interni vengono fatti rilassare attraverso una minimizzazione dell'energia, allo scopo di rimuovere qualunque tipo di contatto non fisico tra residui, che possa condurre ad una dinamica molecolare instabile.

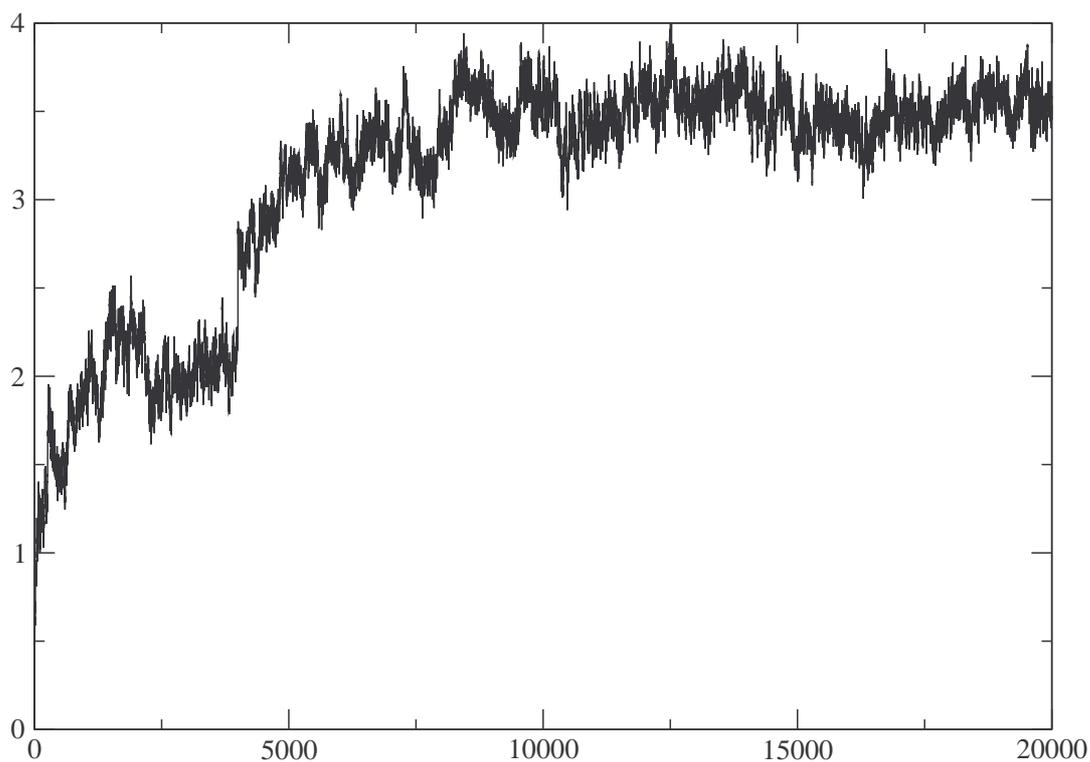
La procedura di minimizzazione adottata consiste in un approccio a due fasi. Nella prima (500 steps a soluto fissato) è stata tenuta ferma la macromolecola e sono state minimizzate solo le posizioni di acqua e ioni. Invece nella seconda fase (2500 steps) è stata effettuata la minimizzazione dell'intero sistema. Dopo la minimizzazione è stato eseguito anche un *run* di equilibratura per 20 ps sotto vincoli posizionali. In una prima è stato riscaldato il sistema da 0 K a 300 K. Allo scopo di assicurare che questo avvenga senza fluttuazioni esagerate nel soluto è stato usato per esso un debole *restraint*, come nella fase di minimizzazione. Lo scopo finale è eseguire una dinamica di produzione a temperatura e pressione costanti, cioè in condizioni da laboratorio. D'altro canto usare una pressione costante con vincoli potrebbe causare problemi, per cui inizialmente è stata eseguita una dinamica per 20 ps a volume costante, quando il sistema si è equilibrato, sono stati rimossi i vincoli e eseguito il *run* a pressione costante per altri 100 ps a  $T = 300\text{K}$ .

Per regolare la temperatura e mantenerla a 300 K è stato adottato un termostato di Langevin.

Per ridurre la complessità computazionale, per quanto possibile, abbiamo usato l'algoritmo di SHAKE in cui tutti i legami che coinvolgono l'idrogeno sono fissati. Rimuovere il moto degli idrogeni ha il fortunato effetto di rimuovere anche l'oscillazione a frequenza più alta, quella delle vibrazioni dell'idrogeno. Poiché l'oscillazione di frequenza più alta determina l'ampiezza dello *step* temporale, rimuovendo il moto degli atomi di H è possibile incrementare questo step temporale a 2 fs senza introdurre instabilità nella simulazione MD. Questa opportunità ha l'effetto di permettere di coprire una certa regione dello spazio delle fasi nella metà del tempo, poiché si ha bisogno di soltanto 50000 step per coprire 100 ps invece dei 100000 richiesti con uno step temporale di 1 fs.

Nella fase di produzione della presente simulazione, sono stati eseguiti ulteriori 20 ns di MD senza restraints sulla proteina, mantenendo la temperatura a 300 K con pressione costante ad 1 atm, condizioni periodiche, usando il termostato di Langevin, vincoli di

SHAKE sugli H ed uno step di tempo pari 2 fs. Poiché la simulazione prende un certo tempo in realtà vengono effettuati 100 run x 200 ps con ognuna delle simulazioni che continua da quelle precedenti. In questa maniera se un job va in crash per qualche motivo è possibile ripartire da dove si è arrivati. A seguito della fase di produzione sono stati analizzati i risultati. Innanzitutto si è notato che dopo un rilassamento iniziale gli andamenti delle proprietà principali del sistema temperatura, pressione, volume, densità ... restano quasi costanti. Questo indica che non ci sono stati problemi con la simulazione effettuata. Quindi sono state analizzate le strutture. Una utile misura da considerare è lo scarto quadratico medio (RMSd) dalla struttura di partenza. Per calcolare questa quantità in funzione del tempo è stato adottato uno strumento di Amber *ptraj*, considerando solo gli atomi principali del *backbone* della CAB. Questo valido *tool* aggiunge i files *mcdcd* prodotti dalla dinamica di Amber in una singola traiettoria. Esso consente anche, se è



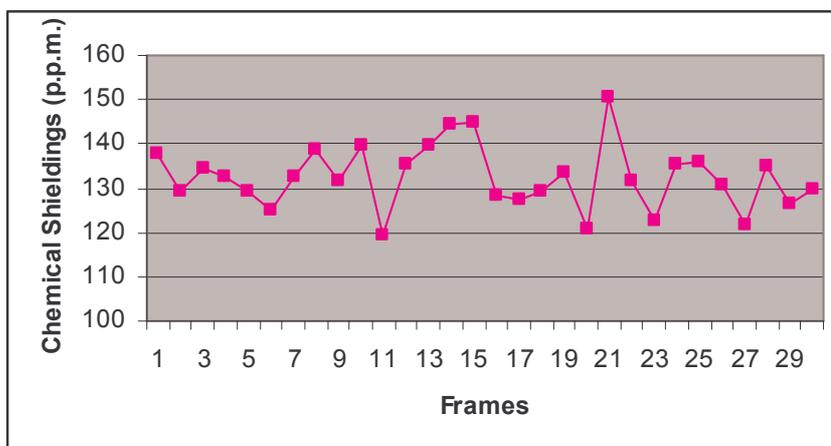
**RMSd degli atomi del backbone**

Anche nel loro studio di dinamica molecolare della Calbindina D9k nello stato *Apo*, Marchand and Roux hanno trovato che la differenza media rms per lo scheletro proteico (ad esclusione degli atomi di idrogeno) è 3.30 Å [52]. Questo comportamento è stato riportato in letteratura come una IPERFLESSIBILITA' della *Apo*-Calbindina. In questo lavoro è stato creato un semplice script (c-shell) che usando *ptraj* può estrarre *pdb frames* dalle traiettorie. Ogni *snapshot* in formato *pdb* viene trasformato in un file di coordinate cartesiane XYZ mediante uno script *python* ed un altro programma permette di modificare il file di coordinate in un formato utile per creare un input file Gaussian per il calcolo QM/MM.

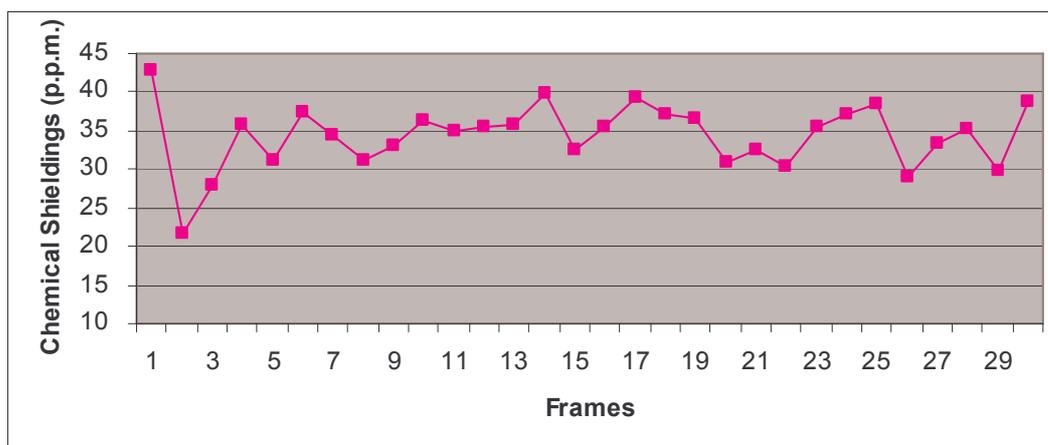
Per l'alanina-14 sono stati determinati 30 *frames* lungo la traiettoria su cui calcolare i chemical shift. I frames sono stati selezionati suddividendo in maniera regolare l'intervallo temporale in cui è stata effettuata la dinamica.

#### *Calcoli di chemical shift*

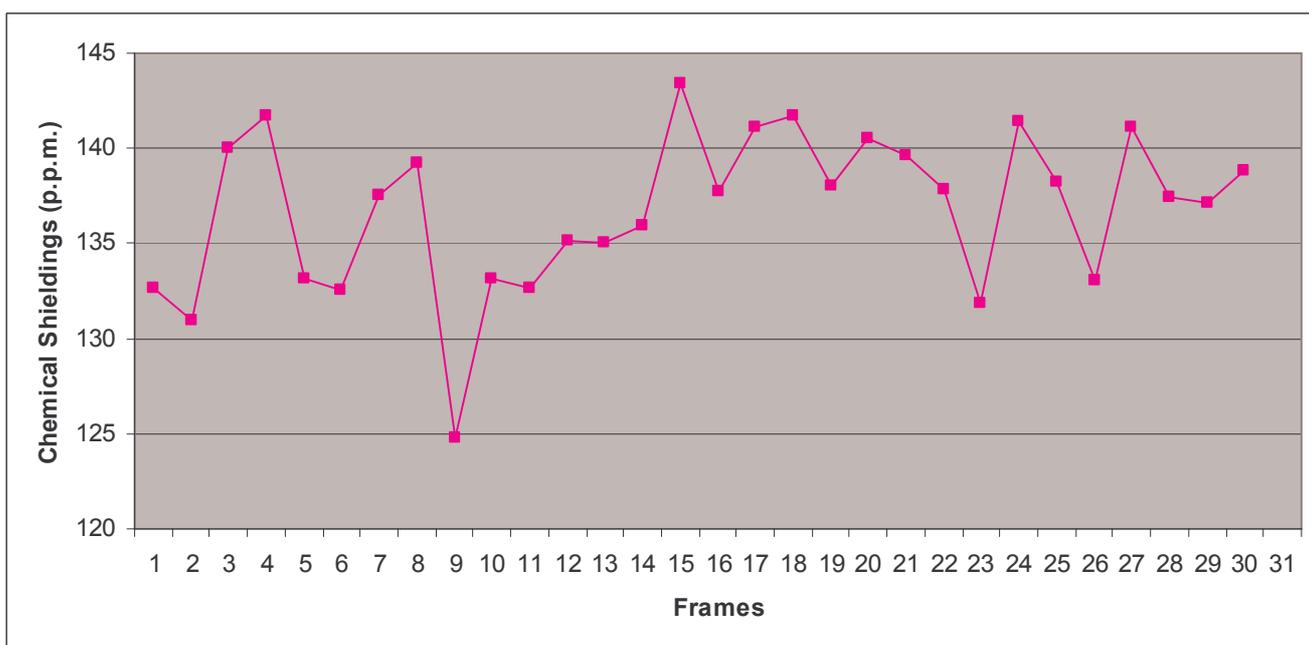
Le costanti di schermo  $\sigma$  in accordo con il metodo GIAO sono state ottenute a livello DFT-ONIOM-EE (B3LYP/6-31G(d,p):Amber) della teoria, adottando lo schema di partizione stabilito. I risultati per gli atomi N, C $^{\alpha}$ , C', H e H $^{\alpha}$  dell'Alanina-14 sono riportati nei grafici seguenti in cui viene illustrato l'andamento dello schermo magnetico in p.p.m. al variare dei *frame* selezionati dalla dinamica molecolare.



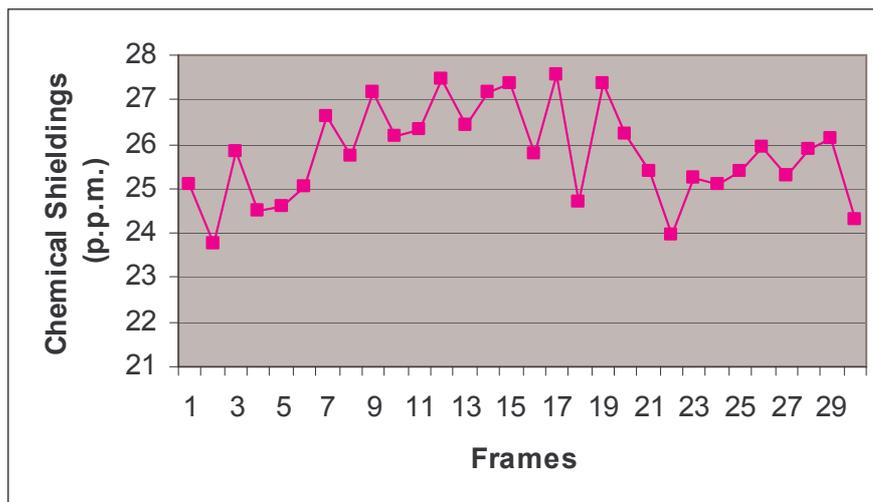
N



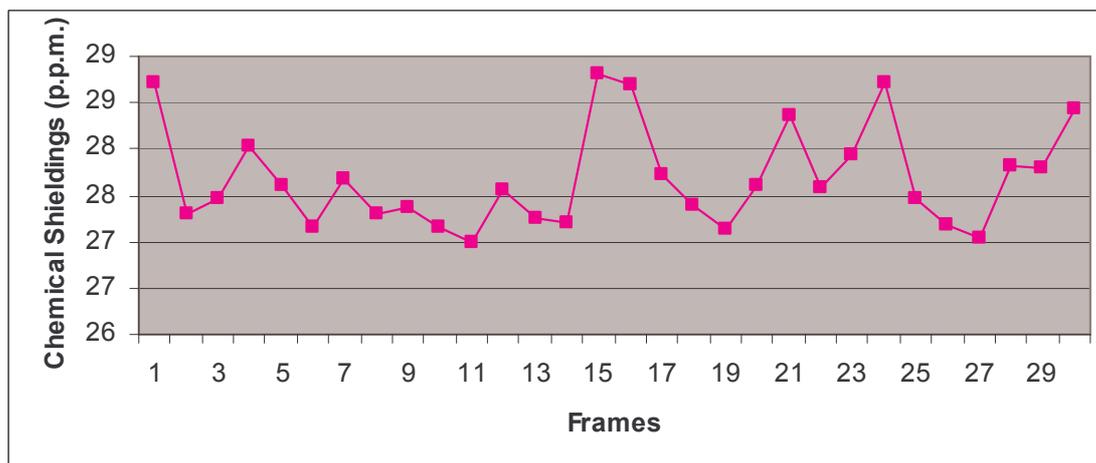
C'



C<sup>α</sup>



H<sup>N</sup>



H<sup>α</sup>

I medesimi grafici per Glu 17, Asp 19, Gln 22 e Glu 27 sono riportati nell'Appendice D. Probabilmente le grandi oscillazioni nei valori di spostamento chimico potrebbero essere spiegate tramite la *iperflessibilità* della forma *Apo* di CAB. In questo caso il confronto con i dati sperimentali è molto difficile.

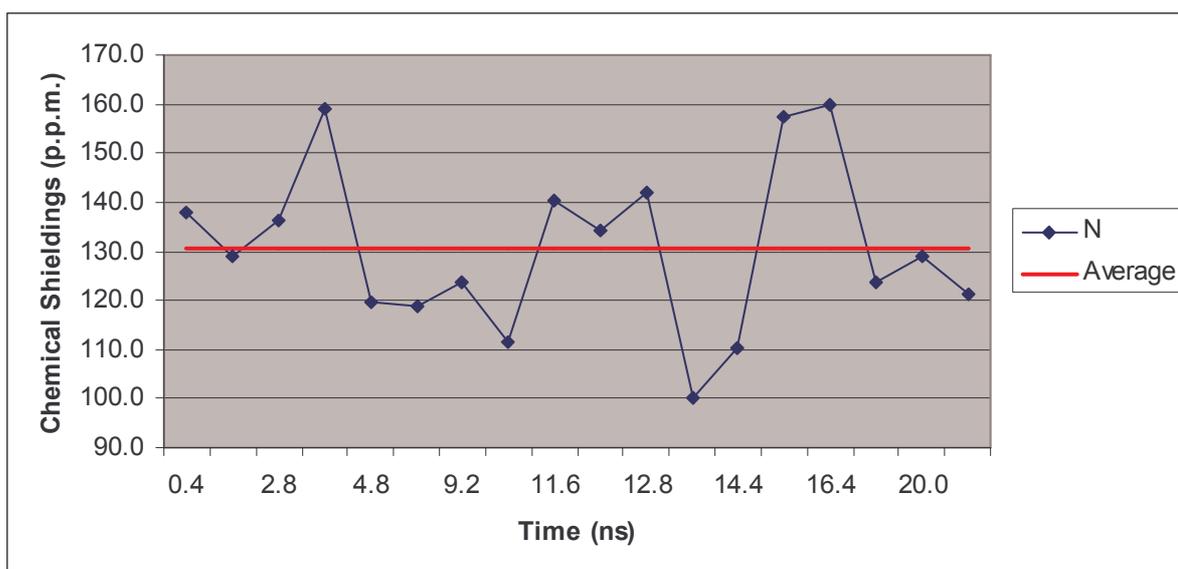
Lo stesso andamento è stato ottenuto anche adottando una base più grande 6-311+G(2d,p) sugli atomi del ligando e la base 6-31G(d,p) sugli altri atomi della regione QM.

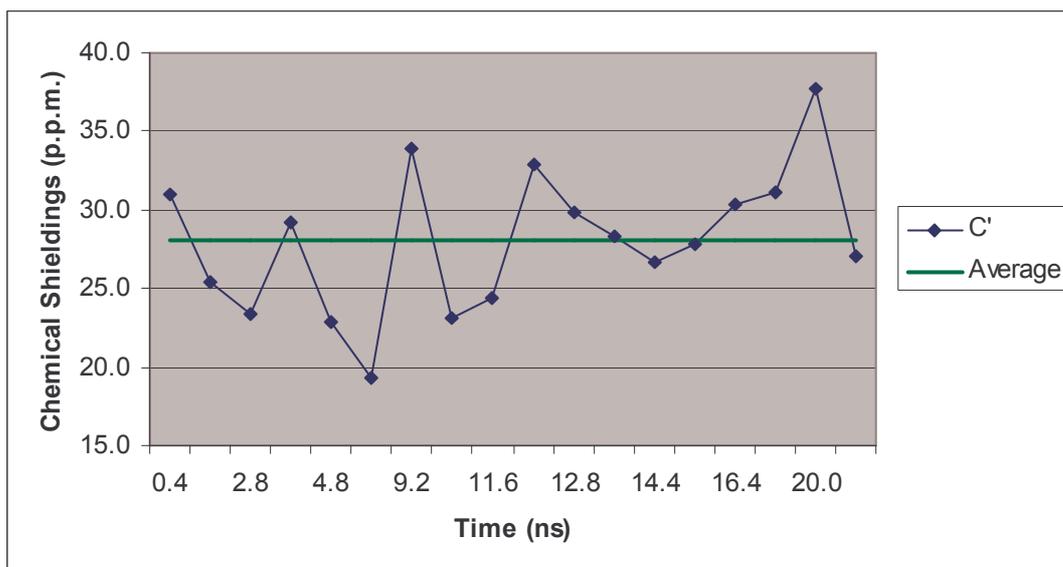
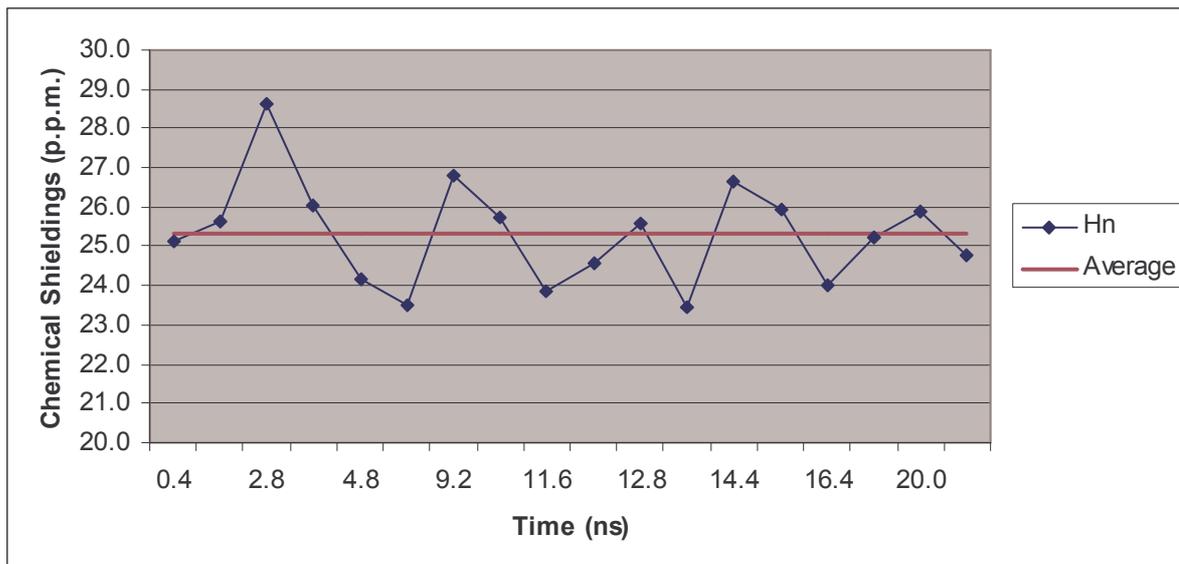
Tuttavia si può supporre che l'introduzione degli ioni metallici possa garantire un consolidamento della struttura e della geometria del sito di binding del calcio con una minore flessibilità della proteina in questa regione. Il passaggio successivo in questo lavoro

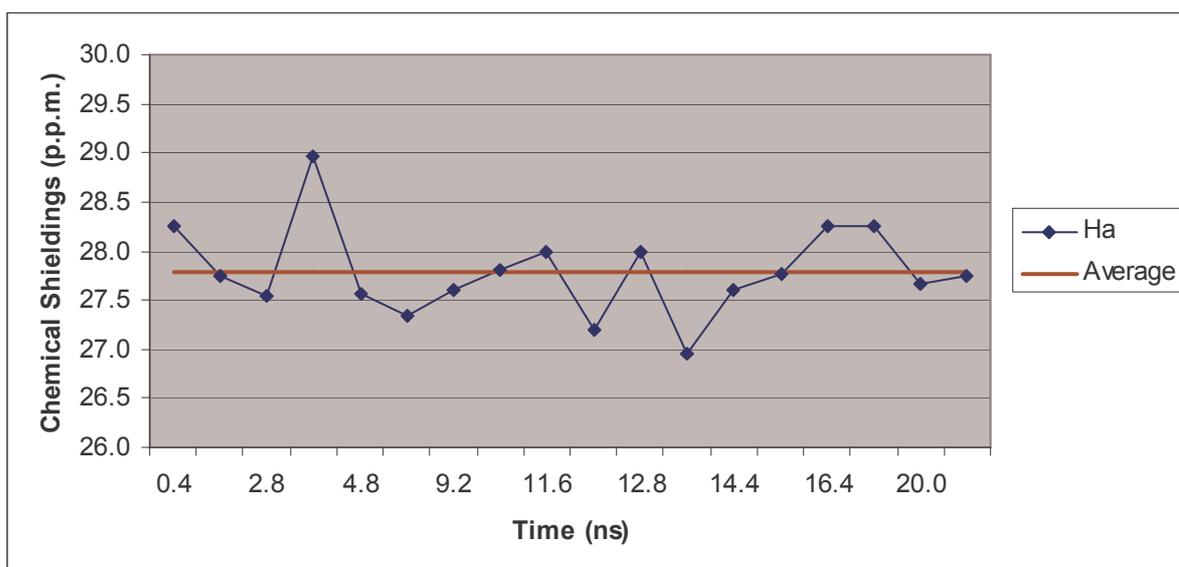
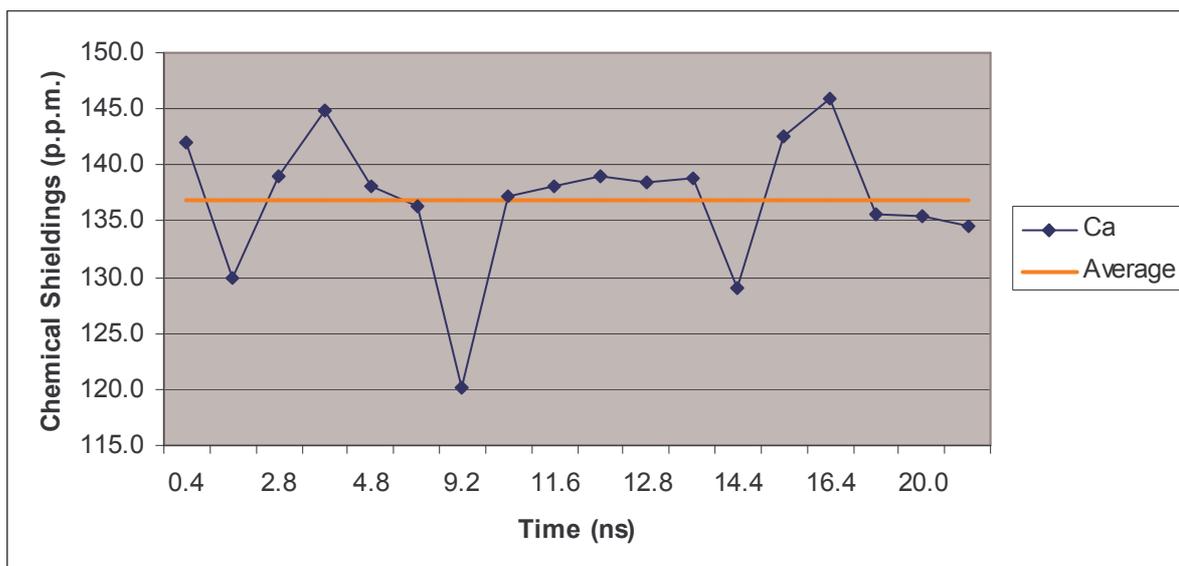
è stato allora quello di applicare il protocollo computazionale, finora sviluppato, anche alla forma *calcium-loaded* della Calbindina D9k.

Dopo aver effettuato una dinamica molecolare anche per la forma *calcium-loaded* della calbindina nelle stesse condizioni illustrate nella sezione (Dinamica Molecolare della Calbindina) sono stati selezionati un certo numero di frames su ciascuno dei quali è stato eseguito un calcolo di costanti di schermo a livello DFT-ONIOM-EE (B3LYP/6-31G(d,p):Amber) adottando ancora una volta orbitali *gauge-including*.

Di seguito vengono riportati i risultati per l'Alanina-14 nei seguenti grafici, in cui si può vedere come lo schermo magnetico di ciascun atomo degli snapshot selezionati dai risultati della dinamica molecolare con lo stesso criterio adottato per la forma *Apo* al variare del tempo (espresso in ns)







I corrispondenti risultati ottenuti per Glu 17, Asp 19, Gln 22 e Glu 27 sono riportati nell'Appendice D.

Come è possibile notare le oscillazioni nei valori di schermo magnetico per ciascun atomo sono ancora abbastanza evidenti.

Un risultato ovvio è che evidentemente i moti interni della proteina sono tali che la partizione fissa che è stata messa a punto ed adottata nei calcoli di *chemical shielding* non è del tutto adeguata ed andrebbe ampliata in modo da riuscire a tener conto del fatto che

L'intorno locale di ciascun atomo di cui si calcola il *chemical shielding* può variare al passare del tempo, cioè andando da un frame ad un altro del sistema in esame.

Tuttavia per quanto riguarda il protone del C<sup>α</sup> l'accordo con i dati sperimentali è buono, nei limiti dell'errore sperimentale. Il raffronto è stato fatto adottando come *riferimento* il DSS, su cui è stato fatto un calcolo di *chemical shielding* a livello B3LYP-6-31G(d,p) in solvente acquoso. Il solvente è stato simulato con il metodo PCM (Polarizable Continuum Model) come implementato nel pacchetto Gaussian 09.

Ricordando che il *chemical shift* per un atomo A è  $\delta^A = \sigma^{\text{rif}} - \sigma^A$ , si trova nel caso del *chemical shift* di H<sup>α</sup>

$$\text{Ala 14} \quad \delta^A = 3.912 \text{ vs } \delta^{\text{exp}} = 3.831$$

$$\text{Glu 17} \quad \delta^A = 4.308 \text{ vs } \delta^{\text{exp}} = 4.736$$

$$\text{Asp 19} \quad \delta^A = 4.770 \text{ vs } \delta^{\text{exp}} = 4.700$$

$$\text{Gln 22} \quad \delta^A = 4.526 \text{ vs } \delta^{\text{exp}} = 5.120$$

$$\text{Glu 27} \quad \delta^A = 4.018 \text{ vs } \delta^{\text{exp}} = 3.984$$

Questi risultati forniscono qualche conferma per la scelta metodo e della partizione QM/MM proposta. Risultati che vanno nella direzione del dato sperimentale forniscono anche una conferma per quanto riguarda la geometria adottata per il sito di *binding* del metallo.

## 6.5 Conclusioni

In questo lavoro di tesi è stato sviluppato un protocollo computazionale, che mira ad introdurre il calcolo *ab initio* dei *chemical shift* per la validazione della struttura del sito di coordinazione dello ione metallico nelle metallo proteine. A tale proposito è stato messo a punto uno schema di frammentazione del sito di *binding* in regioni QM centrate, approssimativamente, sui residui che si legano al catione. I frammenti importanti per il calcolo NMR vengono trattati a livello DFT, mentre il resto del sistema viene modellizzato

con cariche puntiformi MM. Il metodo proposto è parallelizzabile e non dipende dalle dimensioni del sistema in esame, dunque, almeno in linea di principio, può consentire la determinazione dei *chemical shift* di tutti i residui di una proteina (o metalloproteina) reale, anche costituita da migliaia di atomi. Questo approccio ibrido QM/MM è stato combinato con la dinamica molecolare, per tener conto degli effetti dei moti interni della proteina e del solvente sul *chemical shift*. Per i protoni  $H^\alpha$  il confronto del chemical shift ottenuto dai calcoli con i valori sperimentali è eccellente. Dunque il protocollo messo a punto in questo lavoro di Tesi può essere utile anche per la validazione strutturale dei siti di *binding* del metallo nelle metalloproteine.

In definitiva, si può ritenere che il protocollo sviluppato possa essere adottato per una esplorazione sistematica anche di altri metalli e classi strutturali.

## APPENDICE A

### Approssimazione LCAO e set di funzioni di base

L'approssimazione di HF conduce ad un set di equazioni differenziali accoppiate che dipendono dalle coordinate di un singolo elettrone. E' vantaggioso per la loro risoluzione introdurre un'altra approssimazione, che permette di trasformare le equazioni HF in un set di equazioni algebriche. E' ragionevole attendersi che le soluzioni monoelettroniche delle molecole assomiglino alle soluzioni per l'atomo di idrogeno. Dopotutto le molecole sono fatte da atomi e quindi gli orbitali molecolari possono essere espressi come combinazioni lineari di un set finito di funzioni noto come funzioni di base

$$\psi_i = \sum_{\mu} c_{\mu i} \chi_{\mu}$$

Dove  $c$  sono i coefficienti (incogniti) degli orbitali molecolari. Poiché i  $\chi$  sono generalmente centrati sugli atomi in corrispondenza delle posizioni nucleari, essi vengono denominati orbitali atomici e l'equazione sopra riportata prende il nome di *Linear Combination of Atomic Orbitals* o LCAO. Le approssimazioni di HF e LCAO applicate all'equazione di Schrödinger conducono alle equazioni di Roothan-Hall

$$Fc = \epsilon Sc$$

dove  $\epsilon$  sono le energie orbitali,  $S$  è la matrice di overlap (delle funzioni di base), mentre  $F$  è la matrice di Fock, che rappresenta un operatore analogo all'Hamiltoniano nell'equazione di Schrödinger.

I set di base usati in calcoli HF, DFT e metodi post-HF usano funzioni di tipo gaussiano. Le funzioni gaussiane sono funzioni, che hanno la forma delle soluzioni esatte dell'atomo di idrogeno (monoelettronico), ovvero comprendono una parte polinomiale nelle coordinate cartesiane (x, y, z) e una parte esponenziale in  $r^2$ .

La più semplice rappresentazione dell'orbitale atomico è chiamata set di base minimo. Questo comprende solo quelle funzioni richieste per accomodare tutti gli elettroni

dell'atomo, in pratica una sola funzione  $1s$  per idrogeno ed elio, un set di cinque funzioni ( $1s, 2s, 2px, 2py, 2pz$ ) per gli atomi dal litio al neon nella Tavola Periodica degli elementi ed un set di nove funzioni ( $1s, 2s, 2px, 2py, 2pz, 3s, 3px, 3py, 3pz$ ) per gli atomi dal sodio all'argon. Gli elementi di Terza e Quarta riga della Tavola Periodica degli Elementi vengono trattati in maniera simile. Nel caso degli elementi di transizione vi sono nove funzioni di *core* e nove funzioni che descrivono la valenza. In tutti i casi ciascuna funzione di base è espansa in termini di tre funzioni gaussiane, in cui gli esponenti e i coefficienti lineari sono stati determinati a partire da una procedura ai minimi quadrati per ottenere i *best fit* di funzioni esponenziali di Slater.

Il set di base minimo ha due principali difetti. Descrive molto meglio intorni molecolari sferici o quasi sferici che piuttosto intorni molecolari "asferici"; le funzioni sono centrate sugli atomi e quindi è abbastanza ristretta la loro flessibilità a descrivere distribuzioni elettroniche tra i nuclei, dunque legami chimici. Il primo difetto può essere in parte corretto, adottando due set di funzioni di base di valenza. Per esempio opportune combinazioni lineari determinate come soluzioni delle equazioni di Roothan-Hall permettono che un legame  $\sigma$  a partire dalla sovrapposizione di due orbitali  $p$  collineari sia più contratto di un legame  $\pi$  ottenuto da orbitali  $p$  tra loro paralleli.

Un basis set *split valence* rappresenta orbitali atomici di *core* mediante un set di funzioni e gli orbitali di valenza con due set di funzioni (funzioni di tipo  $s$  e  $p$  per gli elementi dei gruppi principali). Tra i più semplici set di base *split-valence* c'è quello denominato 6-31G in cui ogni orbitale di *core* atomico è espanso in termini di sei funzioni gaussiane, mentre le funzioni di base che rappresentano le componenti *inner* e *outer* degli orbitali atomici di valenza sono espanso in termini di tre e una funzioni gaussiane. Un set di base come 6-311G divide le funzioni di valenza in tre parti invece che due (3, 1, 1 gaussiane) garantendo una maggiore flessibilità. I coefficienti di espansione e gli esponenti delle gaussiane per la rappresentazione 6-31G vengono ottimizzati mediante minimizzazione dell'energia degli stati atomici fondamentali al livello HF, mentre nel caso del set di base 6-311G le minimizzazioni vengono eseguite con il metodo correlato MP2.

Il secondo difetto dei set di base minimo e *split valence* di avere funzioni centrate sugli atomi può essere corretto aggiungendo funzioni di tipo  $d$  sugli elementi dei gruppi principali (che hanno orbitali di valenza di tipo  $s$  e  $p$ ) e funzioni di tipo  $p$  sull'idrogeno, che ha orbitali di valenza di tipo  $s$ . Questo permette la "distorsione" delle distribuzioni

elettroniche lontano dalle posizioni nucleari. Tra i più semplici set di base di polarizzazione si possono annoverare i set di base 6-31G\* e 6-311G\*, che si costruiscono, rispettivamente, aggiungendo ai set di base 6-31G e 6-311G, funzioni di polarizzazione di tipo d, espresse come singole funzioni gaussiane per ogni atomo diverso dall'idrogeno. Gli esponenti delle gaussiane per le funzioni di polarizzazione vengono scelti in modo da dare le energie più basse per alcune molecole rappresentative. I set di base 6-31G\*\* e 6-311G\*\* aggiungono funzioni di polarizzazione di tipo *p* agli atomi di idrogeno. La polarizzazione degli orbitali *s* sugli atomi di idrogeno è necessaria per una accurata descrizione del legame chimico in molti sistemi, particolarmente quelli (le proteine ad esempio) in cui l'idrogeno funge da atomo ponte nel legame a ponte di idrogeno.

Infine i calcoli che coinvolgono specie anioniche spesso pongono speciali problemi. Questo perché gli *extra* elettroni possono solo essere debolmente associati a specifici atomi o coppie di atomi. In tali situazioni può essere necessario aggiungere ai set di base anche delle funzioni diffuse di tipo *s* e *p* designate con il simbolo “+” se vengono aggiunte solo agli atomi diversi dall'idrogeno, oppure “++” se le diffuse vengono aggiunte anche agli atomi di idrogeno, come ad esempio il set di base 6-311++G\*\*.

## APPENDICE B

### Densità di corrente elettronica e proprietà magnetiche molecolari

Un campo magnetico uniforme e statico induce circolazioni di carica nella nuvola elettronica di una molecola. L'entità di tali circolazioni dipenderà, oltre che dall'intensità del campo applicato, anche dalla forza che lega gli elettroni allo scheletro nucleare.

Per una molecola a guscio chiuso le principali proprietà di risposta allo stimolo magnetico sono la magnetizzabilità e lo schermo magnetico nucleare, le cui espressioni possono essere ricavate dalla teoria delle perturbazioni indipendente dal tempo, come derivate della energia della molecola rispetto al campo perturbante.

Ma queste stesse proprietà magnetiche possono essere formulate anche attraverso il vettore densità di corrente  $\mathbf{J}$ , che fornisce la densità di corrente elettronica indotta dal campo esterno  $\mathbf{B}$  in ogni punto dello spazio molecolare. Lo studio di questa funzione vettoriale fornisce una visualizzazione immediata dell'interazione tra la molecola e lo stimolo magnetico; la sua rappresentazione su basi quantitative oltre che qualitative può fornire, ad esempio, una interpretazione diretta dei fenomeni alla base della spettroscopia NMR, inoltre anche modelli interpretativi delle proprietà magnetiche molecolari, quale quello delle correnti di anello per i sistemi aromatici, possono trovare una verifica immediata.

Si supponga di avere una distribuzione di cariche con densità  $\rho(r)$  confinata in un volume  $V$ , sia  $S$  la superficie che racchiude  $V$  e  $dS = n dS$  il vettore diretto lungo la perpendicolare alla superficie. La variazione temporale della carica totale nel volume  $V$  è

$$\frac{\partial}{\partial t} \int_V \rho(r) d^3r = - \int_S \rho(r) v \cdot dS = - \int_V \nabla \cdot \mathbf{J}(r) d^3r. \quad \text{B.1}$$

L'ultima identità sorge dal teorema della divergenza di Gauss, con  $\mathbf{J} = \rho \mathbf{v}$ .

L'eq. (B.1) è chiamata equazione di continuità in forma integrale. Poiché essa vale per qualsiasi  $V$ , deve essere soddisfatta la condizione differenziale

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0$$

In accordo con tale condizione, la carica risulta essere localmente conservata; infatti un processo in cui una carica venga creata in certo punto e annichilita in altro rappresenta una situazione fisicamente proibita.

In meccanica quantistica, la matrice densità di probabilità per uno stato quantistico ad  $n$  elettroni  $|\psi\rangle$  è definita dalla relazione

$$P_1(r; r') = N \int \psi(x_1, x_2, \dots, x_N) \psi^*(x_1, x_2, \dots, x_N) dx_2 \dots dx_N$$

Gli elementi diagonali  $\gamma(r) = P_1(r; r)$  della matrice densità di probabilità danno la distribuzione di carica elettronica dello stato di riferimento attraverso la relazione  $\rho(\mathbf{r}) = -e\gamma(\mathbf{r})$ .

Dalla densità di probabilità è possibile ricavare anche la densità di probabilità di corrente elettronica

$$j(r) = \frac{1}{m_e} \Re[\pi P_1(r; r')]_{r'=r}$$

dove  $\pi$ , l'operatore per il momento meccanico elettronico è definito dalla relazione

$$\pi = p + \frac{e}{c} A.$$

La densità di corrente elettronica è definita dalla espressione  $\mathbf{J}(\mathbf{r}) = -e \mathbf{j}(\mathbf{r})$ .

La corrente  $\mathbf{J}$  può essere scritta anche come  $\mathbf{J} = \rho \mathbf{v}$  infatti  $\mathbf{j}(\mathbf{r})$  dipende esplicitamente dalla velocità locale.

Onde calcolare le proprietà di risposta magnetiche bisogna dunque conoscere la funzione densità di corrente  $\mathbf{J}$ , la quale può essere valutata a partire dalla funzione d'onda elettronica  $\Psi$ , che in presenza di campi magnetici indipendenti dal tempo può, a sua volta, essere espressa secondo uno sviluppo perturbativo

$$\Psi = \Psi^{(0)} + \Psi^B \cdot B + \dots$$

La suscettività magnetica e lo schermo magnetico nucleare al nucleo  $I$  possono essere ricavati come integrali del tensore di densità di corrente  $J_\alpha^{B\beta}$

$$\chi_{\alpha\delta} = -\left(\frac{\partial^2 W}{\partial B_\alpha \partial B_\delta}\right)_{B_\delta \rightarrow 0} \quad \text{e} \quad \sigma_{\alpha\delta}^I = -\left(\frac{\partial^2 W}{\partial \mu_{\alpha I} \partial B_\delta}\right)_{\mu_{\alpha I}, B_\delta \rightarrow 0}$$

dove le energie di interazione sono date da

$$W^{BB} = -\frac{1}{2c} \int d^3 r A^B \cdot J^B,$$

$$W^{\mu^I B} = -\frac{1}{c} \int d^3 r A^{\mu_I} \cdot J^B,$$

e  $\mu_I$  è il vettore di dipolo magnetico sull' $I$ -simo nucleo.

## Il problema del gauge

La densità di corrente totale e le proprietà magnetiche devono essere invarianti rispetto ad una trasformazione di gauge del potenziale vettore:

$$A^{B'} \rightarrow A^{B''} = A^{B'} + \nabla f,$$

$$A^{B'} = \frac{1}{2} B \times (r - r'),$$

dove  $f = f(r)$  è una funzione arbitraria, che tende a 0 per  $r$  che tende all'infinito. Una traslazione dell'origine del sistema di coordinate,

$$r' \rightarrow r'' = r' + d$$

è equivalente ad una trasformazione di gauge in cui  $f = (r'' - r') \cdot A^{B'}$ , pertanto, deve lasciare inalterate le proprietà magnetiche. Si può dimostrare che l'invarianza della energia di interazione in una traslazione del gauge del potenziale vettore implica anche la conservazione della carica e della densità di corrente. Questo risultato potrebbe essere ricavato utilizzando ad esempio il teorema di Nöther, in base al quale ogni simmetria continua, che caratterizza un qualunque sistema meccanico (nel presente caso lo shift

dell'origine del campo magnetico) comporta un corrispondente vincolo e quindi una legge di conservazione per quel sistema.

Il problema è che quando si calcolano le proprietà magnetiche e gli altri osservabili fisici con i convenzionali metodi approssimati della chimica quantistica si utilizzano set di autofunzioni dell'Hamiltoniano molecolare imperturbato, che sono troncati, o ciò che è lo stesso, basi di funzioni atomiche finite. Una conseguenza di ciò è che le grandezze calcolate presentano la fastidiosa caratteristica di dipendere dall'origine del potenziale vettore, quindi, solo utilizzando set di base molto estesi è possibile ottenere una (quasi) gauge-invarianza.

Diversi metodi sono stati introdotti in letteratura per la determinazione *ab initio* di proprietà magnetiche indipendenti dall'origine del gauge:

GIAO, IGLO, LORG, CSDGT. Tutti questi approcci mediante una distribuzione dell'origine del potenziale vettore conducono alle medesime conclusioni.

## APPENDICE C

### Proteine calcium-binding

Circa un terzo di tutte le proteine sono metalloproteine, complessi di proteine con ioni di metalli come ferro, calcio, rame, zinco.

Gli ioni metallici hanno una importanza critica nel determinare la funzione, la struttura o la stabilità della proteina. Infatti numerose funzioni biologiche essenziali richiedono ioni metallici e la maggior parte di siffatte funzioni coinvolgono metallo proteine. Queste proteine presentano svariate funzioni e supportano numerose attività biologiche, come il trasferimento elettronico, la catalisi, i processi di trasduzione dei segnali. Le metalloproteine svolgono dunque ruoli di primo piano per la vita sulla Terra, per cui l'abilità a comprendere il *binding* e l'attività del/i sito/i di coordinazione del metallo è di grande importanza biologica ma anche di notevole interesse per le applicazioni mediche.

Tra i metalli più importanti per la Vita vi è il Calcio che controlla numerosi processi, dalla mineralizzazione delle ossa al *signalling* cellulare. Lo ione  $\text{Ca}^{2+}$  interagisce con un gran numero di proteine, anche se la varietà di configurazioni dei siti di binding del calcio è piuttosto limitata. La più comune tra tali configurazioni è il motivo *EF-hand*. Questo dominio strutturale contiene un elemento topologico del tipo elica-*loop*-elica, in cui gli ioni calcio sono coordinati dai ligandi all'interno del *loop*, molto simile ad una mano in cui il pollice e l'indice siano alzati e separati tra loro,. Tale dominio consiste di due  $\alpha$ -eliche posizionate quasi perpendicolarmente l'una rispetto all'altra e collegate da una regione di *loop* (circa 12 amminoacidi) che generalmente lega gli ioni metallici. Il motivo prende il proprio nome dalla tradizionale nomenclatura usata per descrivere la parvalbumina. Il nome descrive non solo il *folding* polipeptidico, ma indica anche il potenziale moto che il *binding* del calcio può indurre. Lo ione calcio è coordinato in una configurazione di bpiramide pentagonale. Esso viene legato sia attraverso gli atomi del *backbone* proteico che mediante le catene laterali di residui amminoacidici come aspartato e glutammato, che essendo carichi negativamente interagiscono con la carica positiva del catione metallico. Asp o, in alternativa, Glu partecipano al *binding* con due atomi di ossigeno (ligando bidentato).

Le calcium-binding proteine svolgono un ruolo fondamentale nella trasduzione del segnale, contribuendo così a tutti gli aspetti del funzionamento cellulare, dall'omeostasi ai processi di apprendimento. La calmodulina ad esempio ha un effetto eccitatorio sui neuroni ed interagisce con altre proteine che controllano lo stato di *firing* dei neuroni.

Quando avviene il binding il dominio *EF-hand* può subire cambiamenti conformazionali che attivano le funzioni regolatorie della metalloproteina.

La Calbindina D9k è una proteina costituita da circa 1200 atomi (con massa molecolare pari a 9kDa), presente nelle cellule epiteliali dell'intestino dei mammiferi. Essa media il trasporto di calcio attraverso gli enterociti.

La forma calcium-loaded prende il nome di *Holo-Calbindina* (vedi figura C.1), mentre il termine *Apo* si riferisce alla forma senza il metallo.

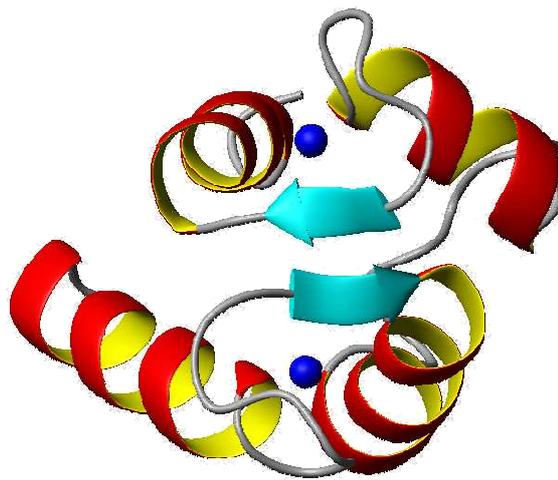


Figura C.1. Calbindina D9K e siti di *binding* del calcio Ca<sup>2+</sup>

## APPENDICE D

### Legami a idrogeno della CAB

Legami a idrogeno della Calbindina determinati con il software freeware *CHIMERA an extensible molecular modeling system*.

H-bonds(donor , acceptor):

SER 2.A OG	GLU 4.A OE1
LEU 6.A N	SER 2.A O
LYS 7.A N	PRO 3.A O
GLY 8.A N	GLU 4.A O
ILE 9.A N	GLU 5.A O
PHE 10.A N	LEU 6.A O
GLU 11.A N	LYS 7.A O
LYS 12.A N	GLY 8.A O
TYR 13.A N	ILE 9.A O
TYR 13.A N	PHE 10.A O
ALA 14.A N	PHE 10.A O
ALA 15.A N	GLU 11.A O
LYS 16.A N	ALA 14.A O
LYS 16.A NZ	ASP 19.A O
LYS 16.A NZ	GLN 22.A O
GLY 18.A N	GLU 17.A OE1
ASP 19.A N	GLU 17.A O
ASN 21.A N	ASP 19.A OD1
ASN 21.A ND2	GLU 11.A OE2
ASN 21.A ND2	PRO 20.A O
GLN 22.A NE2	GLY 57.A O
GLN 22.A NE2	GLU 60.A OE1
LEU 23.A N	VAL 61.A O
SER 24.A OG	GLU 27.A OE1
LYS 25.A NZ	GLY 59.A O
GLU 27.A N	SER 24.A OG
LEU 28.A N	SER 24.A O
LYS 29.A N	LYS 25.A O
LEU 30.A N	GLU 26.A O
LEU 31.A N	GLU 27.A O
LEU 32.A N	LEU 28.A O
GLN 33.A N	LYS 29.A O
THR 34.A N	LEU 30.A O
THR 34.A OG1	LEU 30.A O
GLU 35.A N	LEU 31.A O
PHE 36.A N	LEU 32.A O
LEU 39.A N	PHE 36.A O
LYS 41.A N	LEU 39.A O
GLY 42.A N	LEU 40.A O
GLY 43.A N	THR 45.A O
THR 45.A OG1	ASP 47.A OD2
LEU 46.A N	THR 45.A OG1
ASP 47.A N	THR 45.A OG1

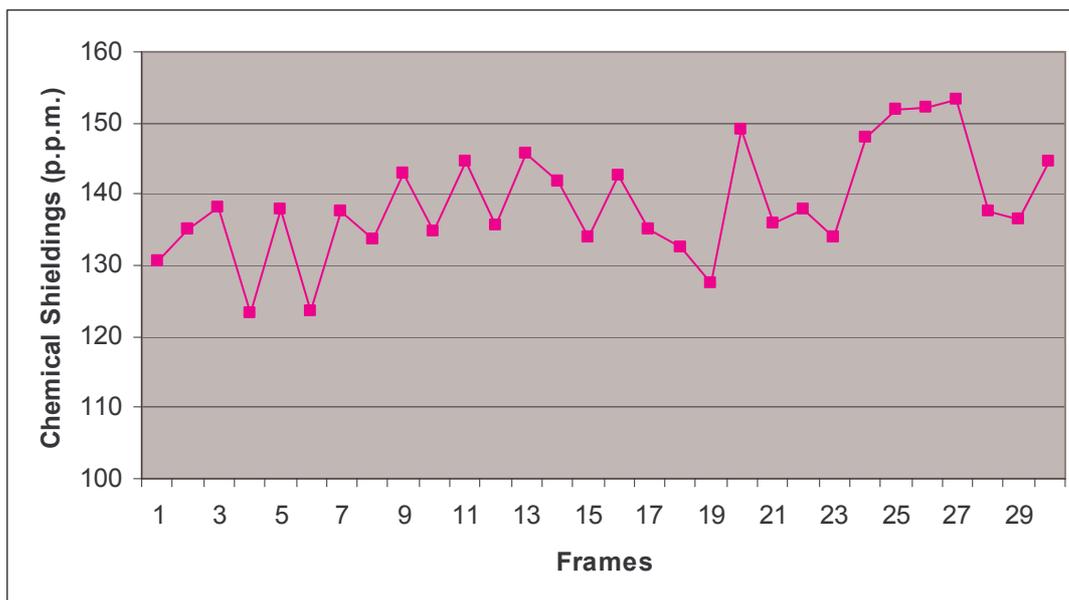
PHE 50.A N	LEU 46.A O
GLU 51.A N	ASP 47.A O
GLU 52.A N	GLU 48.A O
LEU 53.A N	LEU 49.A O
ASP 54.A N	PHE 50.A O
LYS 55.A N	LEU 53.A O
LYS 55.A NZ	GLU 52.A O
ASN 56.A ND2	SER 62.A OG
ASN 56.A ND2	GLU 65.A OE1
GLY 57.A N	GLU 65.A OE1
ASP 58.A N	ASN 56.A O
GLY 59.A N	ASP 54.A OD1
GLY 59.A N	ASP 54.A OD2
GLU 60.A N	GLY 57.A O
VAL 61.A N	LEU 23.A O
SER 62.A OG	GLN 22.A OE1
PHE 66.A N	SER 62.A O
GLN 67.A N	PHE 63.A O
VAL 68.A N	GLU 64.A O
LEU 69.A N	GLU 65.A O
VAL 70.A N	PHE 66.A O
LYS 71.A N	GLN 67.A O
LYS 71.A NZ	GLN 67.A OE1
LYS 72.A N	VAL 68.A O
ILE 73.A N	LEU 69.A O
SER 74.A N	VAL 70.A O
SER 74.A OG	VAL 70.A O
GLN 75.A N	LYS 71.A O
GLN 75.A NE2	LYS 71.A O

### Schermi magnetici della *Apo-CAB*

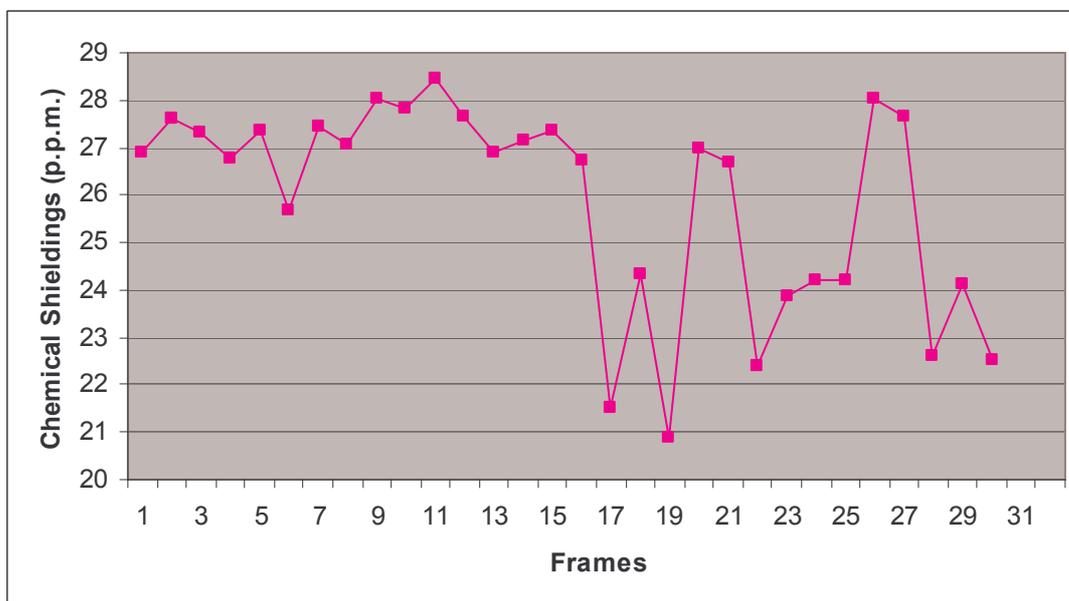
Costanti di schermo  $\sigma$ , in accordo con il metodo GIAO, calcolate a livello DFT-ONIOM-EE (b3lyp/6-31g(d,p):Amber) della teoria, adottando la schema di partizione descritto nel Capitolo 6.

Di seguito viene illustrato l'andamento dello schermo magnetico in p.p.m. al variare dei *frame* selezionati dalla simulazione di dinamica molecolare nell'intervallo 0-20 ns (Capitolo 6).

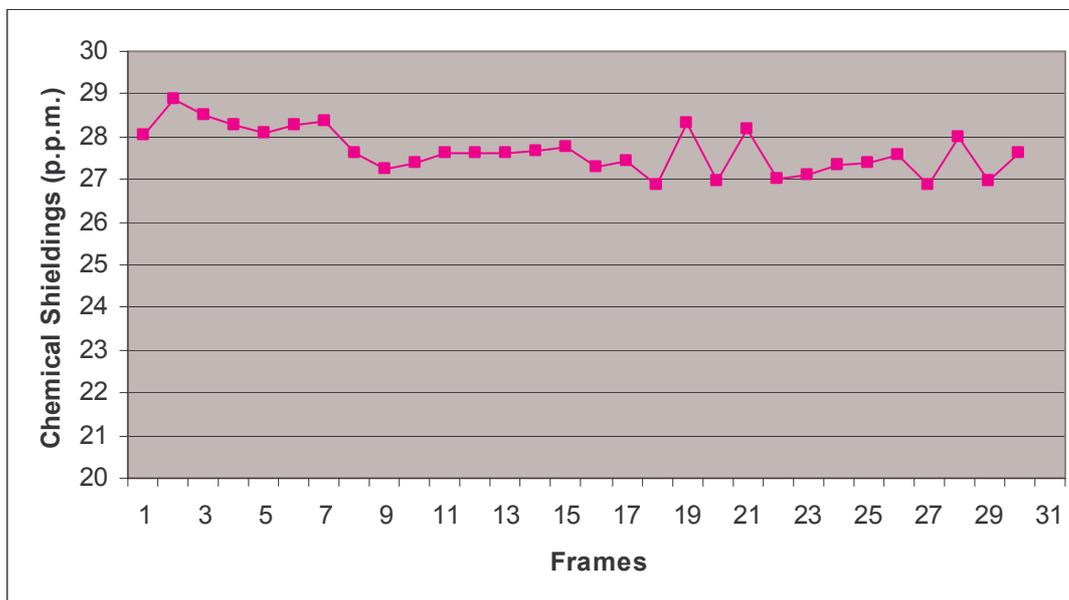
N Glu 17



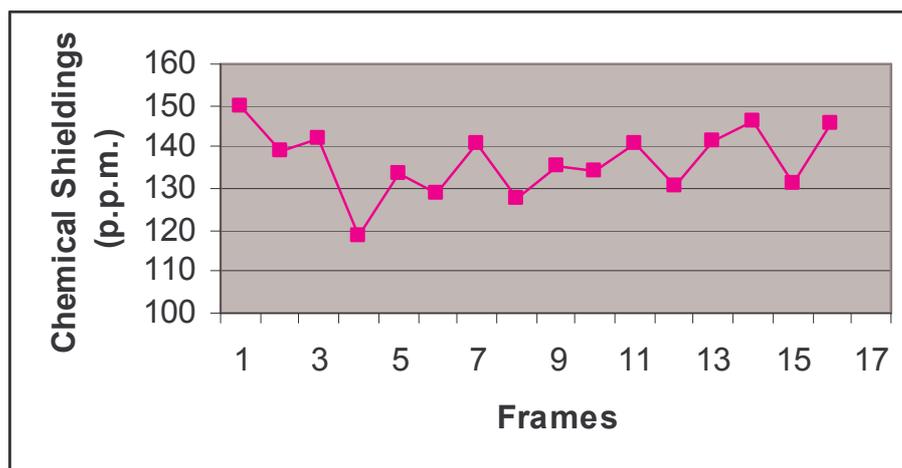
H<sup>N</sup> Glu 17



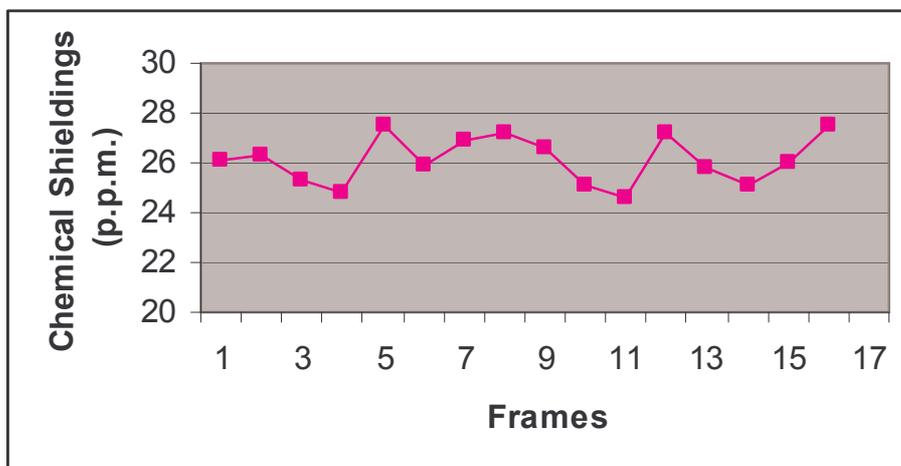
H<sup>α</sup> Glu 17



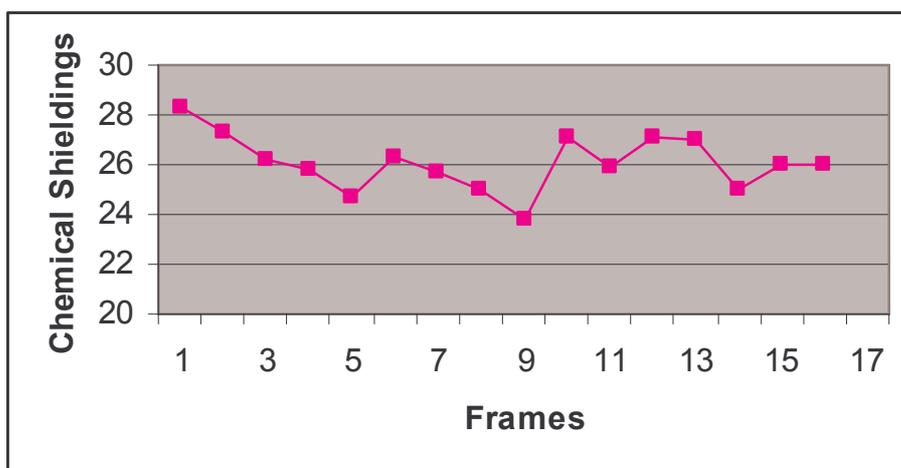
N Asp 19



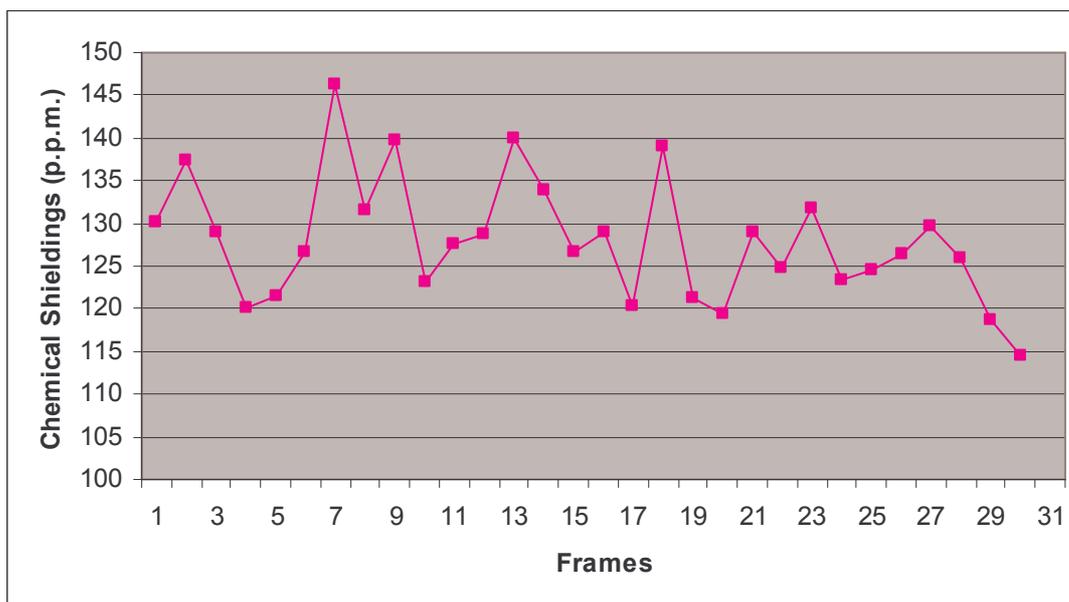
H<sup>N</sup> Asp 19



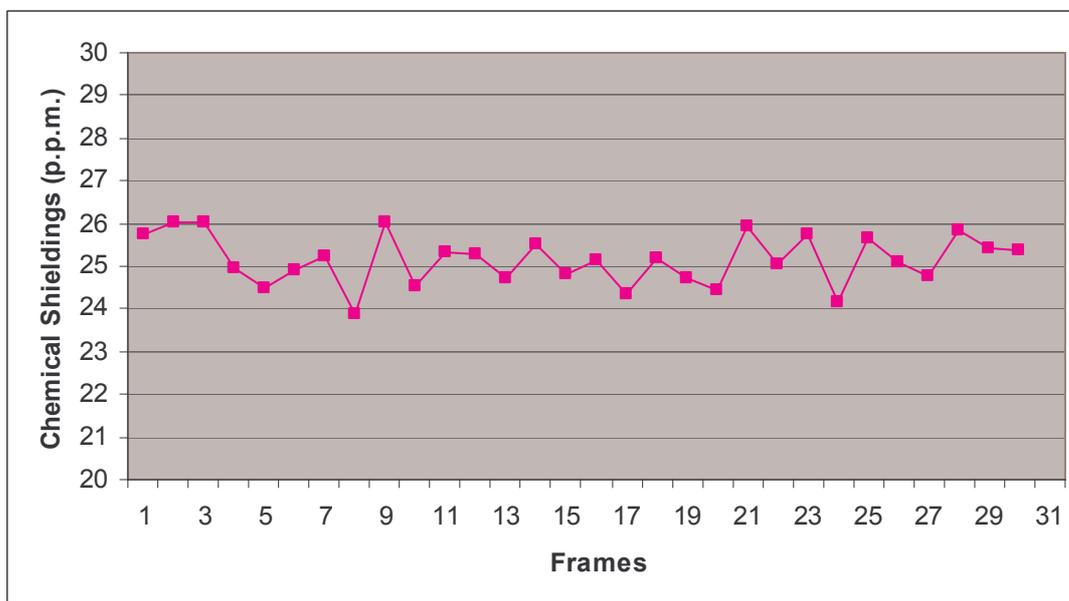
H<sup>α</sup> Asp 19



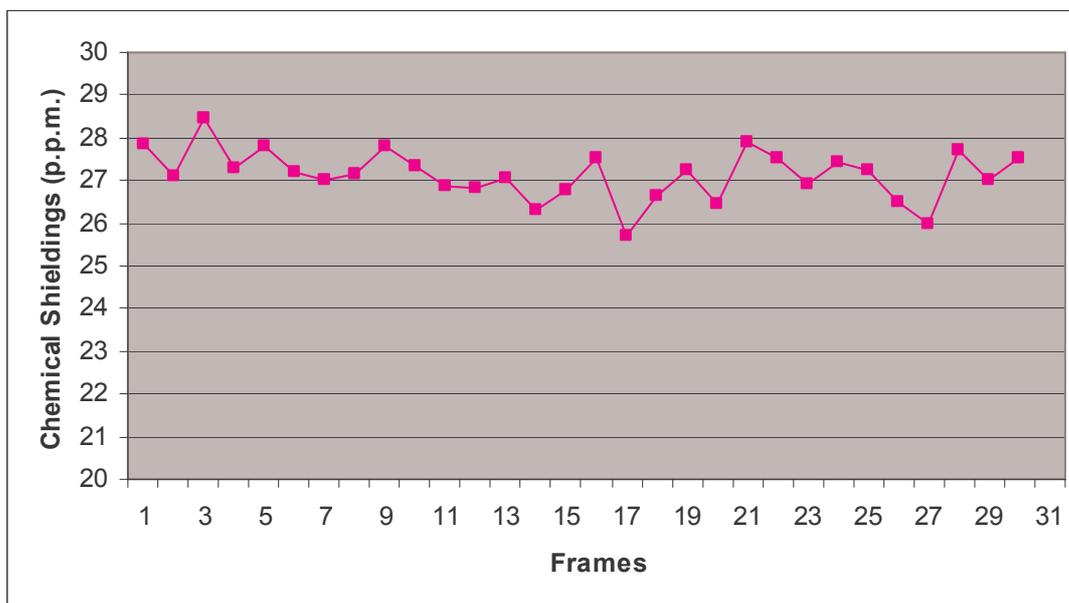
N Gln 22



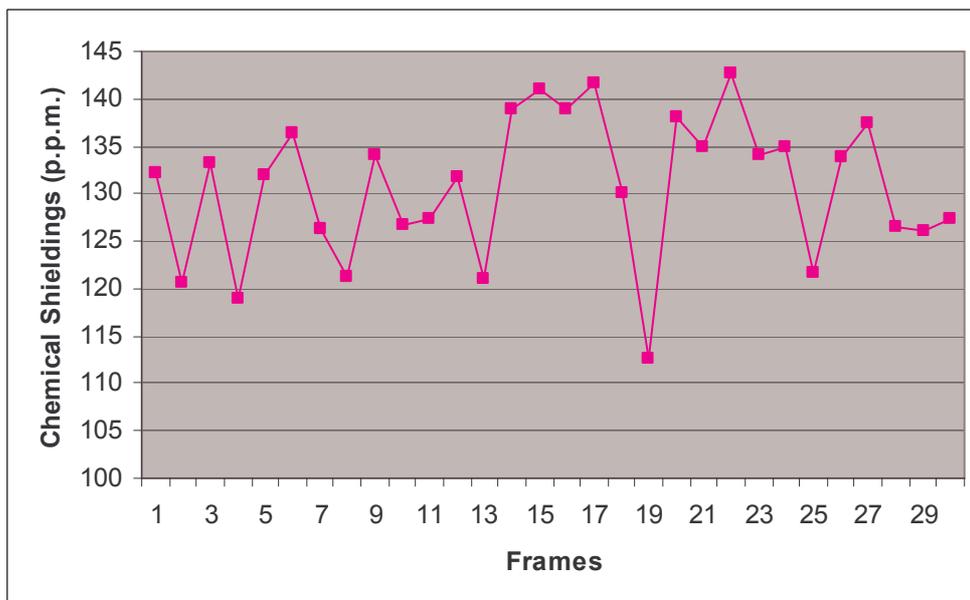
H<sup>N</sup> Gln 22



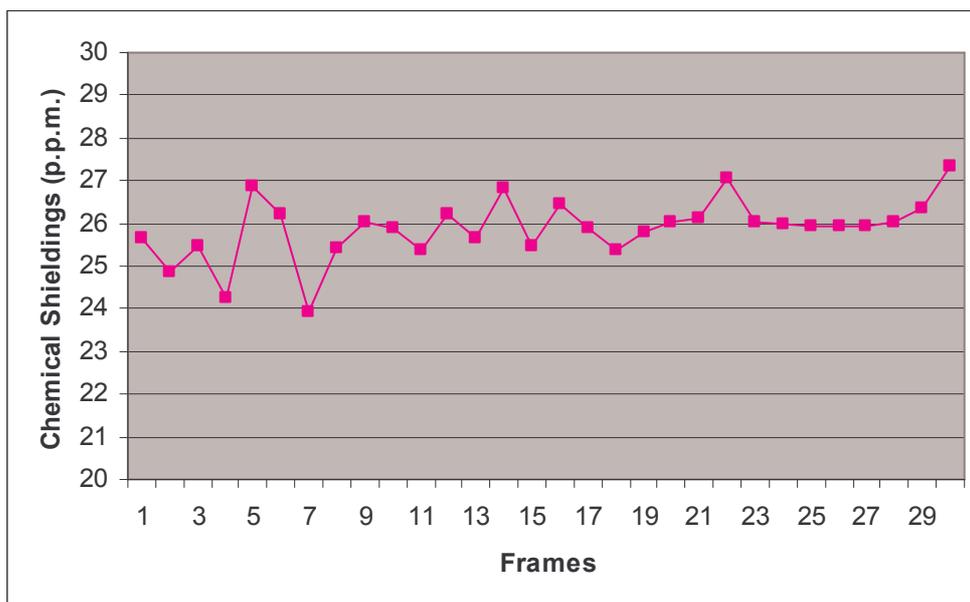
H<sup>α</sup> Gln 22



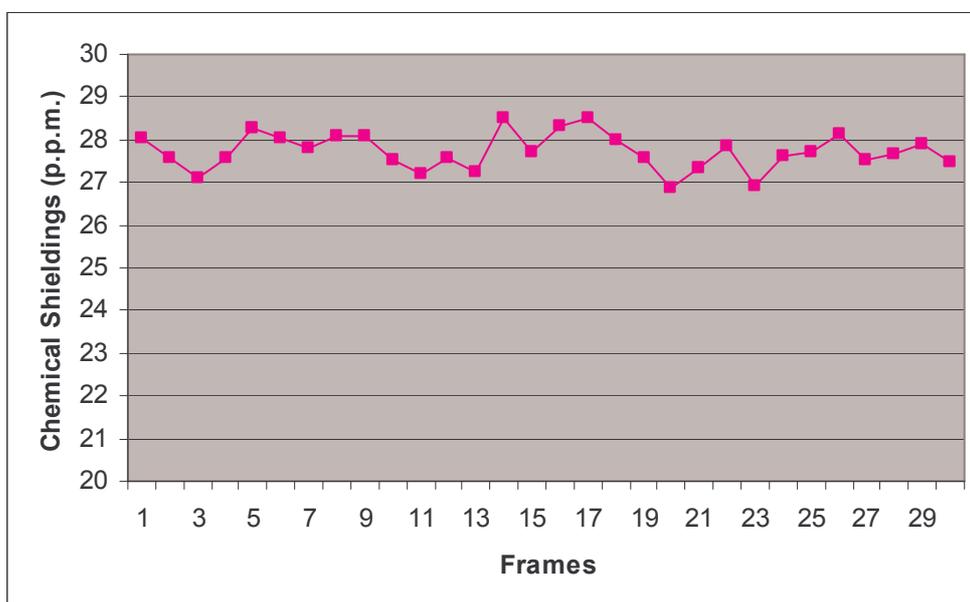
N Glu 27



$H^N$  Glu 27



$H^\alpha$  Glu 27

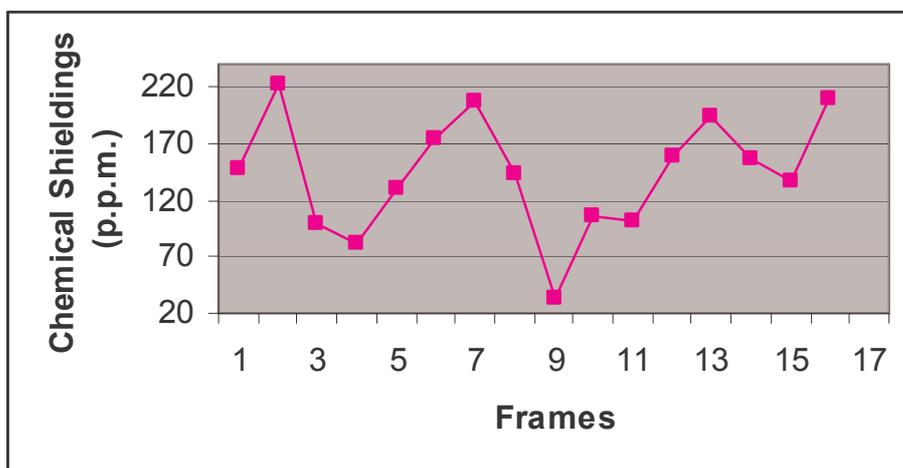


## Schermi magnetici della *Holo-CAB*

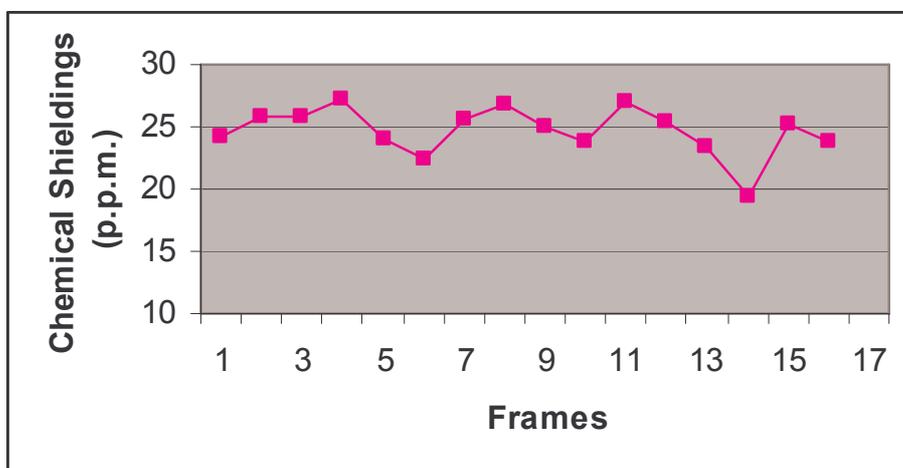
Costanti di schermo  $\sigma$ , calcolate a livello DFT-ONIOM-EE (b3lyp/6-31g(d,p):Amber) della teoria, adottando orbitali *gauge-including* e lo schema di partizione descritto nel Capitolo 6.

Di seguito vengono riportati i risultati per Glu 17, Asp 19, Gln 22 e Glu 27 nei seguenti grafici, in cui si può vedere come varia lo schermo magnetico di ciascun atomo degli *snapshot* selezionati dai risultati della dinamica molecolare con lo stesso criterio adottato per la forma *Apo*.

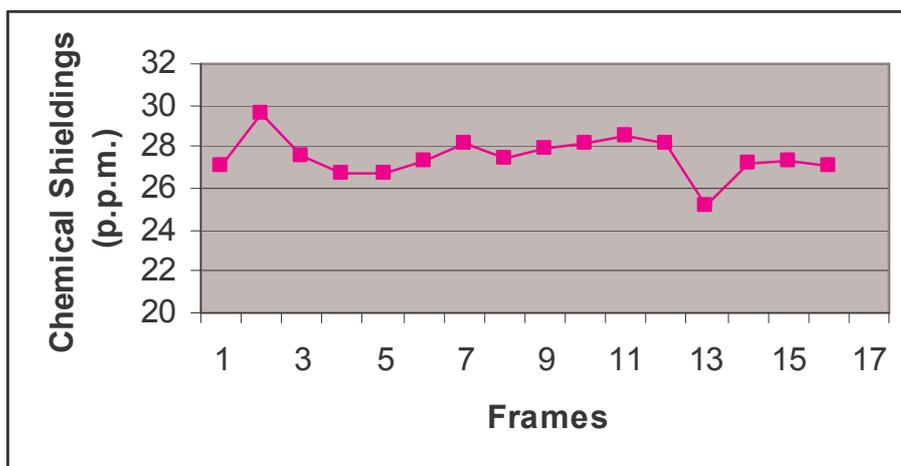
N Glu 17



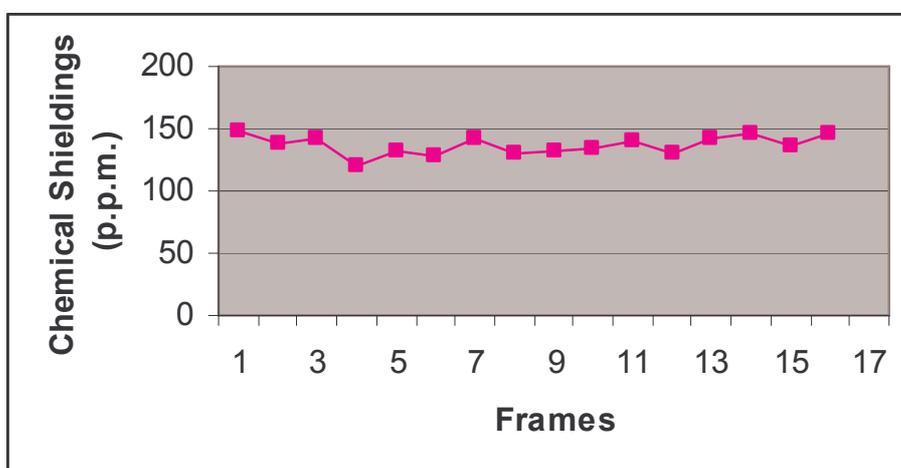
H<sup>N</sup> Glu 17



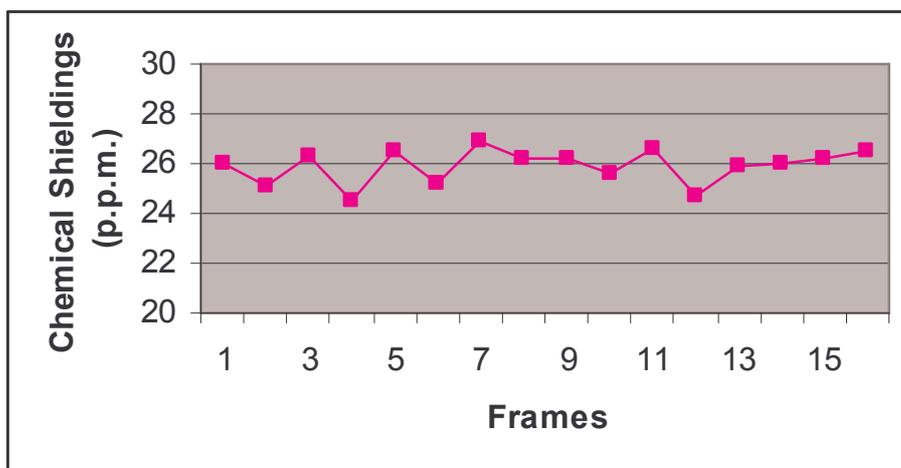
H<sup>α</sup> Glu 17



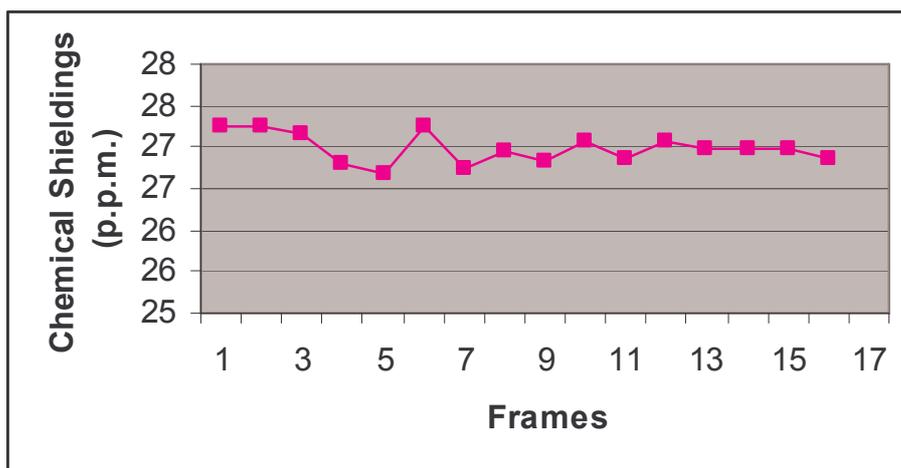
N Asp 19



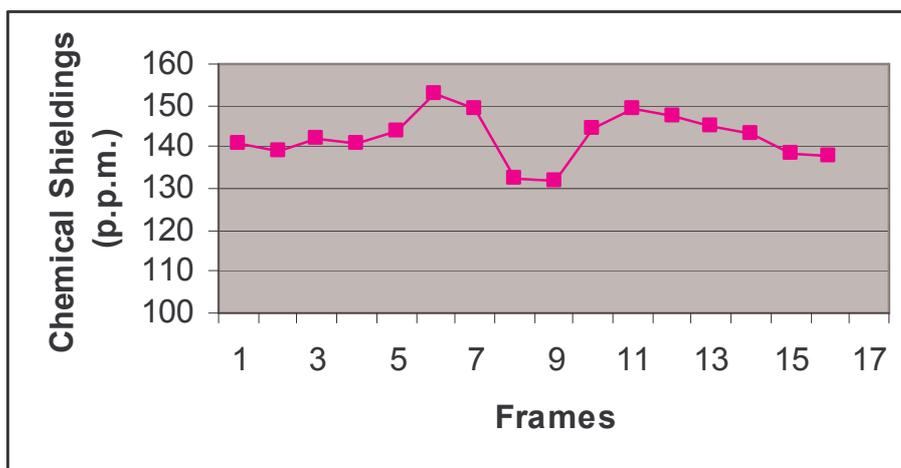
$H^N$  Asp 19



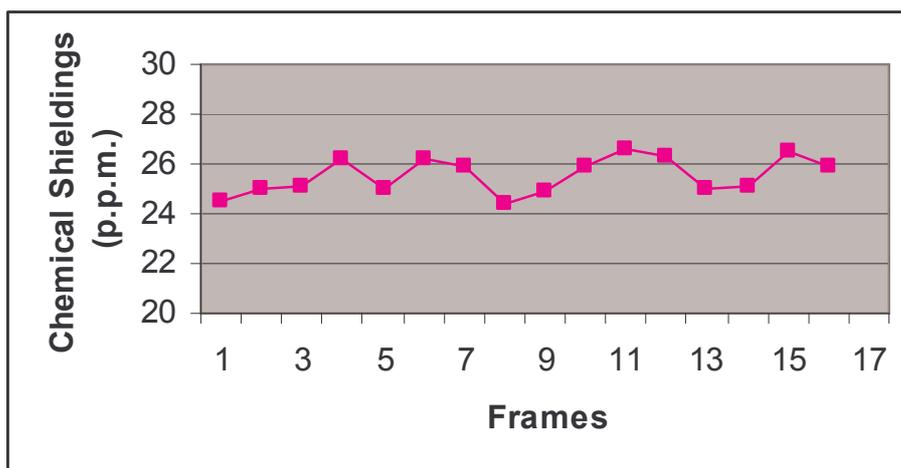
$H^\alpha$  Asp 19



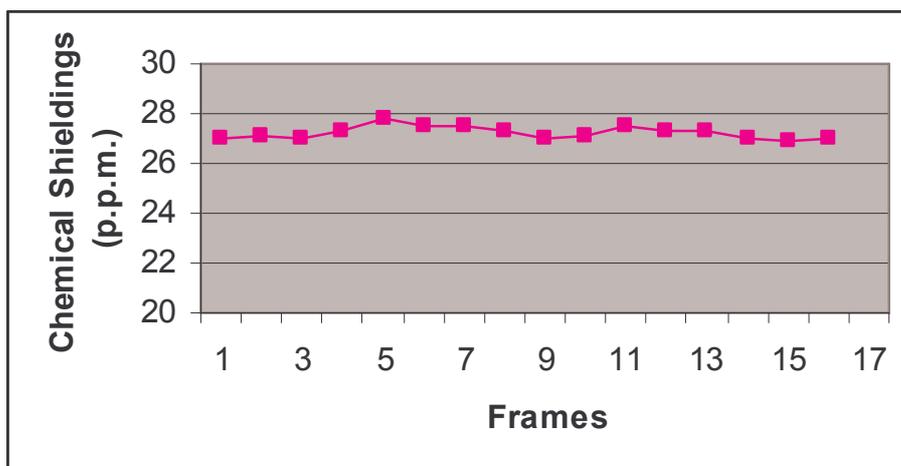
N Gln 22



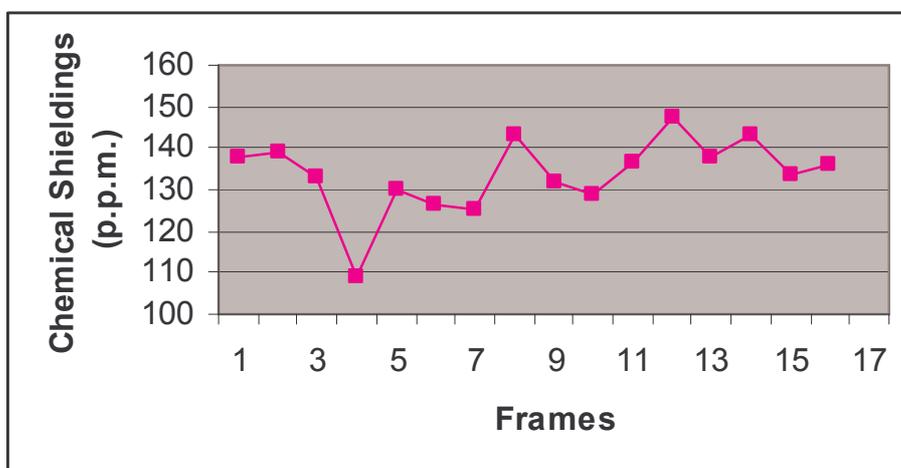
H<sup>N</sup> Gln 22



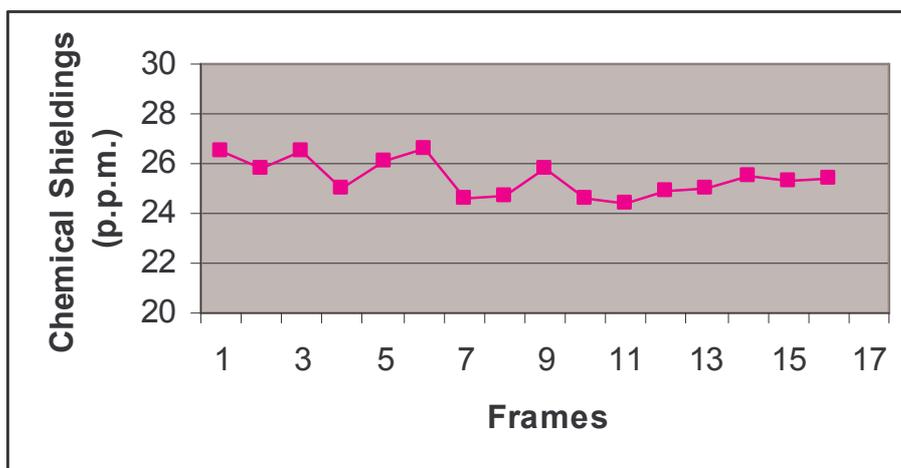
H<sup>α</sup> Glu 22



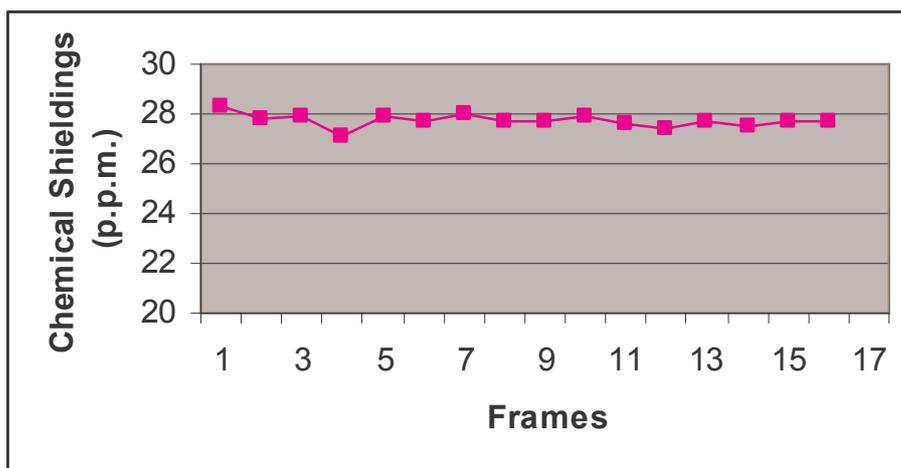
N Glu 27



$H^N$  Glu 27



$H^\alpha$  Glu 27



## BIBLIOGRAFIA

- [1] J. R. Calhoun, W. Liu, K. Spiegel, M. Dal Peraro, M. L. Klein, K. G. Valentine, A. J. Wand, W. F. DeGrado, *Structure* **16**, 210 (2008).
- [2] K. Wüthrich, *Nat. Struct. Biol.* **8**, 923 (2001).
- [3] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Chetham, *Acc. Chem. Res.* **33**, 889 (2000).
- [4] C. Hardin, T.V. Pogorelov, Z. Luthey-Schulten, *Curr. Opin. Struct. Biol.* **12**, 176 (2002).
- [5] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
- [6] M. Karplus, *J. Chem. Phys.* **30**, 11 (1959) ; M. Karplus, *J. Am. Chem. Soc.* **85**, 2870 (1963) .
- [7] H. Günther, *NMR spectroscopy: Basic Principles, Concepts and Applications in Chemistry*, 2nd Edition, Wiley, Chichester, 1995.
- [8] H. Gong, Y. Shen, G.D. Rose, *Protein Sci.*, **16**, 1515 (2007); A. Cavalli, X. Salvatella, C.M. Dobson, M. Vendruscolo, *Proc. Natl. Acad. Sci. USA*, **104**, 9615 (2007); Shen, Y.; Lange, O.; F. Delaglio, P. Rossi et al. *Proc. Natl. Acad. Sci. USA* **105**, 4685 (2008).
- [9] C. A. Hunter, M. J. Packer, C. Zonta, *Progress in Nuclear Magnetic Resonance Spectroscopy* **47**, 27 (2005).
- [10] S. Neal, A. M. Nip, H. Y. Zhang, H. Y.; D. S. Wishart, *J. Biomol. NMR*, **26**, 215 (2003); X. P. Xu, D. A. Case, *J. Biomol. NMR*, **21**, 321 (2001).
- [11] A. Leach, *Molecular Modelling: Principles and Applications*, 2nd Edition, Pearson, Edinburgh (2001).
- [12] I. N. Levine, *Quantum Chemistry*, 5th edition, Prentice Hall, Upper Saddle River (2000).
- [13] C. J. Cramer, *Essentials of Computational Chemistry*, 2nd edition, Wiley, Chichester (2004).
- [14] R. G. Parr, Y. Weitao, *Density-Functional Theory of Atoms and Molecule*, Oxford Science Publ. (1994).
- [15] J. C. Slater, *Phys. Rev.*, **81**, 385 (1951)
- [16] S. J. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).

- [17] J. P. Perdew and Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).
- [18] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [19] A.D. Becke, *J. Chem. Phys.*, **98**, 1372 (1993).
- [20] J. Harris, *Phys. Rev.*, **29**, 1648 (1984).
- [21] A. St-Amant, DR. Salahub, *Chem. Phys. Lett.*, **169**, 387 (1990).
- [22] A.D Becke, *J. Chem. Phys.*, **88**, 1053 (1988).
- [23] D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods, *J. Computat. Chem.* **26**, 1668 (2005).
- [24] U. Burkert and N.L. Allinger, *Molecular Mechanics*, ACS Monograph no. 177, American Chemical Society, Washington D.C. (1982).
- [25] N. L. Allinger, L. Norman L, *J. Am. Chem. Soc.* **99**, 8127 (1977).
- [26] T.A. Halgren, *J. Comput. Chem.* **17**, 533 (1996); G. Forgarasi, P.G: Szalay, *J. Phys. Chem. A*, **101**, 1400 (1997).
- [27] U. Dinur, A.T. Hagler, *New Approaches to Empirical Force Fields. Review of Computational Chemistry*, Chapter 4, 2, Lipkowitz, K.B.; Boyd, D.B., eds: VCH; New York, (1991).
- [28] S. R. Cox, D. E. Williams, *J. Comp. Chem.*, **2**, 304 (1981).
- [29] U. C. Singh, P.A. Kollman, *J. Comput. Chem.*, **5**, 129 (1984).
- [30] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, *J. Phys. Chem.*, **97**, 10269 (1993).
- [31] A.K. Rappe', C.J. Casewit, *Molecular Mechanics Across Chemistry*, University Science Books, Sausalito, CA, (1997).
- [32] A. Rahman, *Phys. Rev. A*, **136**, 405 (1964); F.H. Stillinger, A. Rahman, *J. Chem. Phys.*, **60**, 1545 (1974).
- [33] J.A. McCammon, B.R. Gelin, M. Karplus, *Nature* **267**, 585 (1977).
- [34] M. Levitt, R. Sharon, *Proc. Natl. Acad. Sci. USA* **85**, 7557 (1988).
- [35] A. Warshel, M. Levitt, *J. Mol. Biol.*, **103**, 227 (1976).
- [36] U. C. Singh, P. A. Kollman, *J. Comput. Chem.*, **7**, 718 (1986).
- [37] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.*, **11**, 700 (1990).
- [38] D. Bakowies and W. Thiel., *J. Phys. Chem.*, **100**, 10580 (1996).
- [39] K. P. Eurenium, D. C. Chatfield, B. R. Brooks, M. Hodosecek., *Int. J. Quantum Chem.*, **60**, 1189 (1996).

- [40] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, G. G. Ferenczy, *J. Comput. Chem.*, **15**, 269 (1994).
- [41] F. Maseras, K. Morokuma, *J. Comput. Chem.*, **16**, 1170 (1995).
- [42] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, K. Morokuma, *J. Phys. Chem.*, **100**, 19357 (1996).
- [43] T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, K. Morokuma, M. J. Frisch, *J. Chem. Theory Comput.*, **2**, 815 (2006).
- [44] D. Bashford, *Lecture Notes in Computer Science* **1343**, 233 (1997).
- [45] T. Kesvatera, B. Jönsson, E. Thulin, S. Linse, *Proteins*, **45**(2), 129 (2001).
- [46] Q. Cui, M.J. Karplus, *Phys. Chem. B*, **104**, 3721 (2000); A.C. Dedios, E. Oldfield, *Chem. Phys. Lett.*, **108**, 205 (1993),
- [47] U. Ryde, *Dalton Trans.*, 607 (2007).
- [48] R. Ditchfield, *Mol. Phys.*, **27**, 789 (1974).
- [49] Gaussian 09, Revision A.1, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
- [50] UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* 25(13), 1605 (2004).
- [51] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H.

Mathews, M.G. Seetin, C. Sagui, V. Babin and P.A. Kollman (2008), AMBER 10, University of California, San Francisco.

[52] S. Marchand , B. Roux, *Proteins*, **33**(2), 265 (1998).