
UNIVERSITÀ DEGLI STUDI DELLA CALABRIA

Facoltà di Scienze Matematiche Fisiche e Naturali
Dipartimento di Chimica

Dottorato di ricerca in Metodologie Chimiche Inorganiche
XXIII ciclo

AREA 03 –Scienze Chimiche, SSD-CHIM03/Chimica generale ed inorganica

***MECHANISM DETERMINATION OF
REACTIONS CATALYZED
BY METALLO-ENZYMES***

Supervisore
Prof.ssa Marirosa Toscano

Coordinatore
Ch.mo Prof. Nino Russo

Candidato
Orazio Amata

A.A. 2009-2010

*A mio figlio.
scienza e conoscenza sono effimere
senza una speranza per il futuro*

Gli enzimi portano a termine una gran quantità di funzioni all'interno degli organismi viventi: è proprio grazie a questi potenti catalizzatori biologici che le cellule riescono a compiere e a controllare le reazioni chimiche loro necessarie per ricavare energia e quindi perpetuarsi.

Gli aspetti più significativi riguardanti gli enzimi, che li rendono così indispensabili per la vita, risiedono nella loro capacità di funzionare in condizioni blande di pH e temperatura, di operare selettivamente su un substrato discriminando tra molecole molto simili tra loro e di essere regio- e stereo-specifici.

Il presente lavoro di tesi ha come oggetto di studio proprio gli enzimi e, più in particolare, la determinazione per via computazionale dei meccanismi di alcune reazioni catalizzate da metalloenzimi. Lo sviluppo dell'hardware, contemporaneo a quello di software sempre più potenti e performanti, ha infatti reso possibile la trattazione quantomeccanica di sistemi di dimensione notevole come quella che caratterizza in genere le molecole biologiche.

Il lavoro si prefigge di contribuire a chiarire i dettagli meccanicistici di processi catalitici per i quali non esistono certezze sperimentali a riguardo. Nello stesso tempo, si propone come scopo quello di fornire ulteriori esempi utili a dimostrare la validità e l'affidabilità della metodica utilizzata per gli studi rivolti a sistemi enzimatici.

La presente esposizione si compone di quattro capitoli.

I primi due hanno valore introduttivo e trattano delle caratteristiche generali degli enzimi e della teoria alla base della metodica utilizzata, rispettivamente.

Il capitolo 3 espone a grandi linee la necessità di dover impiegare dei modelli di sito catalitico dell'enzima ed il protocollo seguito nella loro costruzione, per concludersi con una breve trattazione sulla necessità di operare i calcoli in fase proteica.

L'ultimo capitolo, il capitolo 4, espone criticamente i risultati ottenuti negli studi effettuati che hanno riguardato in particolare:

- *L'Insulin-Degrading Enzyme (IDE) correlato a gravi patologie quali il Diabete Mellito e il Morbo di Alzheimer.*
- *La Mitochondrial Processing Peptidase (MPP), appartenente anch'essa come l'IDE alla famiglia delle inverzincine, è implicato nel corretto funzionamento dei mitocondri, le centrali energetiche delle cellule.*
- *Una Anidraasi Carbonica di recente scoperta, presente in organismi marini quali le diatomee, in grado di utilizzare al posto del consueto ione Zn^{2+} lo ione Cd^{2+} finora considerato tossico per ogni forma di vita.*

ELENCO DELLE PUBBLICAZIONI

- I. *Human insulin-degrading enzyme working mechanism.*
Orazio Amata, Tiziana Marino, Nino Russo, Marirosa Toscano
J. Am. Chem. Soc **2009**, *131*, 14804–14811

- II. *Catalytic activity of a ζ -class zinc and cadmium containing carbonic anhydrase. Compared work mechanisms.*
Orazio Amata, Tiziana Marino, Nino Russo, Marirosa Toscano
Phys. Chem. Chem. Phys., DOI:10.1039/C0CP01053G

INDICE

Prefazione	i
Elenco Pubblicazioni	ii
Indice	iii

Capitolo I: La Catalisi Cellulare	1
1.1. Gli enzimi	2
1.2. Cinetica enzimatica	7

Capitolo II: Aspetti Teorici	13
2.1. L'equazione di Schrödinger nell'approssimazione di Born-Hoppenheimer ...	14
2.2. Ipersuperfici	17
2.3. Risolvere l'equazione di Schrödinger elettronica: il determinante di Slater ..	20
2.4. Il metodo Hartree-Fock.....	23
2.5. L'equazione di Roothaan	25
2.6. La Teoria del funzionale della densità	28
2.7. Lo schema di Kohn-Sham.....	30

Capitolo III: L'implementazione pratica	39
3.1. La necessità di modelli	40
3.2. Il protocollo seguito	41
3.3. I calcoli in solvente.....	43

Capitolo IV: Gli studi effettuati	53
4.1. Le inverzincine	54
4.2. L'Insulin-Degrading Enzyme.....	55
4.2.1. Il meccanismo catalitico	56
4.2.2. I modelli studiati.....	58
4.2.3. I risultati ottenuti per l'IDE.....	60
4.3. La Mitochondrial Processing Peptidase: il suo ruolo nei mitocondri	63
4.3.1. Le caratteristiche strutturali della MPP	65
4.3.2. Il meccanismo ipotizzato	66
4.3.3. Il modello utilizzato	67
4.3.4. I risultati dello studio sulla MPP	70

4.4. Le Anidrasi Carboniche.....	73
4.4.1. La CDCA1.....	76
4.4.2. Il meccanismo catalitico delle CA	77
4.4.3. Il modello utilizzato per la CDCA1	78
4.4.4. I risultati ottenuti per la CDCA1.....	79
Appendice.....	89
Articolo I.....	91
Articolo II.....	103

CAPITOLO I: LA CATALISI CELLULARE

1.1 Gli enzimi

Ogni cellula vivente, a prescindere dalla sua origine, per poter espletare le sue funzioni e riprodursi necessita di un costante rifornimento di energia [1].

La fonte di questa energia può essere la radiazione solare o una fonte luminosa o, ancora, una qualche molecola ad alto contenuto energetico.

In ogni caso, questa energia deve essere trasformata per poter essere resa fruibile e questo avviene grazie ad una serie di reazioni chimiche, le quali, tuttavia, nelle condizioni di reazione compatibili con la vita delle cellule, quantunque favorite dal punto di vista termodinamico, possono non esserlo da quello cinetico.

È necessario, pertanto, che qualcosa renda tali reazioni fattibili nelle condizioni suddette ed in tempi compatibili con la stabilità chimica delle molecole coinvolte [2].

Questo compito è assolto, appunto, dagli *enzimi*. Tecnicamente gli enzimi possono essere definiti come proteine¹ ad attività catalitica [3].

L'importanza degli enzimi è tanto maggiore se si considera come da essi dipendano tutti i percorsi metabolici all'interno di una cellula come pure l'accrescimento e la duplicazione cellulare e funzioni come la motilità, la digestione o la trasduzione di segnali.

Ogni cellula contiene qualche migliaio di enzimi, ognuno dei quali catalizza uno specifico tipo di reazione di uno o più composti (substrati) aventi qualche caratteristica comune. La reazione risulta essere sia *regiospecifica* che *stereospecifica*: di tutti i possibili siti di reazione presenti sul substrato solo uno è quello chiamato in causa nella formazione di prodotti aventi una ben precisa configurazione stereochimica [2].

Detto in altri termini, perché la funzione degli enzimi abbia un senso e sia compatibile con le esigenze della cellula, ognuno di essi deve essere altamente efficiente e specifico.

Ogni enzima ha un nome sistematico, generalmente terminante con il suffisso *-asi*, che ne identifica la specificità sia a livello di substrato che di reazione catalizzata. Alcuni enzimi, tuttavia, possiedono diverse denominazioni, magari perché identificati in momenti storici differenti o tramite differenti metodiche: si rende così necessaria una catalogazione sistematica che identifichi il più precisamente possibile un enzima.

Tale catalogazione, nota come catalogazione EC, si basa su un gruppo di quattro numeri il primo dei quali indica la classe a cui appartiene l'enzima (vedi Tab 1).

¹ Nella realtà esistono anche acidi nucleici (RNA) ad attività catalitica: i Ribozimi [4]. Tuttavia la stragrande maggioranza degli enzimi noti sono di natura proteica e, pertanto, la trattazione che segue riguarda principalmente questi.

Numero EC	Denominazione	Tipo di reazione
1	Ossidoreduttasi	Trasferimento di elettroni
2	Transferasi	Trasferimenti di gruppi
3	Idrolasi	Idrolisi
4	Liasi	Addizioni di gruppi a doppi legami o rimozioni di gruppi con formazione di doppi legami
5	Isomerasi	Isomerizzazione
6	Ligasi	Condensazione

Tab. 1: Classificazione degli enzimi in base al tipo di reazione catalizzata

Il secondo ed il terzo numero, invece, specificano, rispettivamente, il tipo di substrato e il tipo di gruppo accettore o di gruppo rimosso, mentre l'ultimo è un numero progressivo indicante l'ordine di inclusione dell'enzima nella specifica sotto-sottoclasse.

In quanto proteine, gli enzimi [5] sono costituiti da un numero variabile (da qualche a diverse centinaia) di amminoacidi legati tra loro attraverso un legame ammidico (peptidico) ed in una ben precisa sequenza (struttura primaria), specifica di ognuno di essi.

La natura poliammidica degli enzimi riveste una grande importanza poiché consente loro di avere una struttura tridimensionale con il giusto compromesso tra rigidità e flessibilità.

Dipendentemente dalla sequenza primaria e dalla natura degli amminoacidi, ovvero se idrofilici o idrofobici, ogni catena polipeptidica origina delle strutture tridimensionali locali (struttura secondaria) come le eliche α o i foglietti β . Queste a loro volta interagiscono tra loro originando delle strutture via via più complesse (strutture supersecondarie) fino a determinare la struttura dell'intero polipeptide (struttura terziaria).

Le diverse catene polipeptidiche, infine, possono interagire tra loro formando il complesso enzimatico² vero e proprio (struttura quaternaria). In questo caso ogni singola catena polipeptidica costituisce una sub-unità enzimatica.

Talvolta, per esplicare la loro azione, gli enzimi richiedono anche la presenza di cofattori e/o gruppi prostetici, ovvero di molecole non proteiche o piccoli ioni metallici (ad esempio nel caso dei metallo-enzimi).

Tutte queste strutture sono tenute assieme da legami cosiddetti deboli quali ponti disolfuro³, legami a idrogeno, interazioni ioniche e idrofobiche [6]. È importante

² Esistono, naturalmente, anche numerosi esempi di proteine costituite da una singola catena polipeptidiche e quindi prive di una struttura quaternaria.

realizzare che l'alterazione della struttura dell'enzima nativo, provocata da agenti esterni o anche per modifica di uno solo degli amminoacidi che lo costituiscono, può comportare la perdita completa della sua funzionalità [7].

In quanto catalizzatori gli enzimi agiscono sulla cinetica della reazione abbassandone l'energia di attivazione o promuovendo un differente meccanismo ad energia minore rispetto la reazione non catalizzata (Fig. 1), senza tuttavia partecipare formalmente ad essa.

La loro azione, pertanto, si esplica solo a livello di velocità di reazione che risulta accelerata, mentre non si ha alcun effetto sull'equilibrio termodinamico.

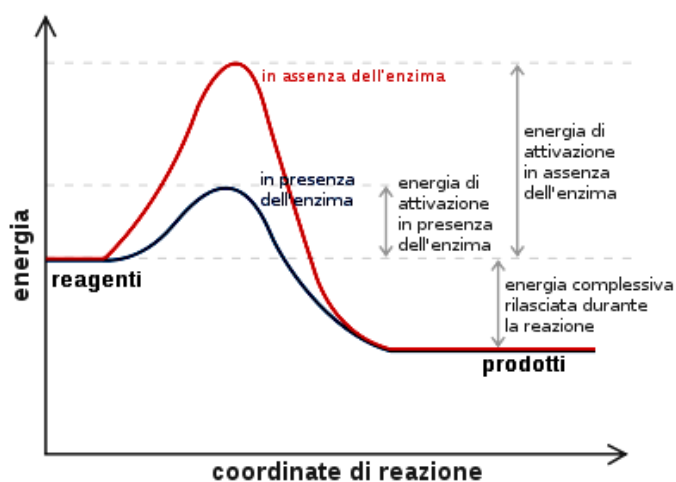


Fig. 1: Profili energetici di una reazione in presenza ed in assenza di un enzima

Questo è reso possibile in quanto l'azione enzimatica si esplica principalmente a livello di entropia di reazione [5]: tutti i gruppi chimici necessari e responsabili della reazione, appartenenti tanto all'enzima quanto al substrato, sono posizionati al punto giusto, nella corretta orientazione e alle giuste distanze perché essa abbia luogo.

Si ha, inoltre, un effetto anche sulla entalpia della reazione. L'interazione tra l'enzima ed il suo substrato, infatti, rende quest'ultimo più attivo e può distorcerne la conformazione.

È importante, in ogni caso, che le interazioni tra enzima e substrato non risultino troppo forti, sia per quanto concerne il complesso iniziale che per tutte le altre tappe del meccanismo catalitico: in caso contrario, uno o più stadi del meccanismo sarebbero stabilizzati troppo e l'enzima perderebbe di efficacia.

³ Anche se questo è un esempio di legame covalente, il ponte disolfuro è piuttosto labile a causa della facilità con cui è possibile ridurre lo zolfo; anche shock termici possono provocare la rottura di tale legame.

La catalisi ha luogo in una specifica regione dell'enzima denominata *cavità catalitica* o centro attivo, mentre gli amminoacidi specificatamente coinvolti nell'interazione con il substrato e nel processo catalitico costituiscono il *sito attivo*.

Gli enzimi, infatti, risultano in genere molto più grandi dei loro substrati per poter soddisfare la grande richiesta entropica legata alla costituzione del centro attivo e la sua conservazione [2,8,9]. Inoltre, essi necessitano di regioni di riconoscimento per il substrato e per l'interazione con altre biomolecole presenti nella cellula [10].

In particolare, l'attività enzimatica [5] può essere modulata dall'interazione con tali biomolecole (effettori allosterici) che possono fungere anche da promotori o da inibitori. Il meccanismo più semplice con cui queste biomolecole agiscono consiste in un semplice cambiamento conformazionale che comporta una riorientazione di residui coinvolti nella catalisi.

Sono possibili, tuttavia, specie per quel che concerne l'inibizione, anche altri meccanismi come la competizione per il sito di legame del substrato (inibizione competitiva) oppure la formazione di una supramolecola con il complesso enzima-substrato (inibizione acompetitiva) che risulta inattivo. Può anche accadere che il prodotto finale di una via metabolica inibisca un qualche enzima a monte (inibizione da feedback negativo) mentre un eccesso di reagente la promuova.

La catalisi può avvenire attraverso diversi meccanismi [2]: se promossa da ioni H^+ o OH^- prende il nome, rispettivamente, di *catalisi acida specifica* e *catalisi basica specifica*, mentre se promossa da un acido generico o una base generica prende il nome di *catalisi acida generale* o *catalisi basica generale*; quando invece è operata da un nucleofilo con formazione di un intermedio transiente in cui l'enzima è legato in maniera covalente al substrato o con cessione di densità elettronica ci si riferisce ad essa con il termine di *catalisi nucleofilica* o covalente; se, infine, il meccanismo prevede la sottrazione di densità elettronica, come ad esempio in presenza di metalli che stabilizzino cariche elettriche negative o acquistino elettroni, si parla di *catalisi elettrofilica*.

Naturalmente nel caso degli enzimi possono benissimo operare più meccanismi in sinergia tra loro durante il ciclo catalitico.

Per stabilire come opera un enzima è dunque necessario non solo conoscere la struttura nativa dell'enzima, ma anche se esso si avvale di cofattori e di quali.

Nel caso dei metalloenzimi, poi, anche altre questioni come la localizzazione del sito catalitico, l'identità del metallo e il suo stato di ossidazione, la stechiometria, la natura dei leganti e la loro identità assumono particolare rilievo.

Inoltre, risulta ugualmente determinante conoscere le strutture degli enzimi con i suoi substrati e tutti gli intermedi della reazione e i prodotti.

Determinare queste strutture [9], tuttavia, non è cosa né semplice né banale.

La determinazione della struttura dell'enzima nativo ed i suoi particolari strutturali, ad esempio, possono essere ricavati da esperimenti di diffrazione ai raggi x [11], dopo però che questo sia stato isolato, purificarlo al 100 % in quantità apprezzabile, precipitato in modo da ottenere i cristalli per l'esperimento curando di non denaturare l'enzima!!!

Alcuni dettagli del meccanismo possono essere dedotti da misure della cinetica di reazione in differenti condizioni: ad esempio variando il pH o la forza ionica del solvente o impiegando tecniche isotopiche o, ancora, ricorrendo all'impiego di inibitori.

Spesso i substrati delle reazioni enzimatiche sono altre proteine o biomolecole prive di una qualche caratteristica chimico-fisica facilmente misurabile.

Questo comporta una duplice difficoltà nel condurre esperimenti di cinetica enzimatica giacché il substrato naturale, se molto complesso, è difficile da sintetizzare in laboratorio. Per poter essere impiegato, quindi, sono richiesti il suo isolamento e la sua purificazione da fonti naturali con conseguente aumento dei costi economici dell'esperimento. In secondo luogo, è necessario poter seguire nel tempo il decorso della reazione e questo può essere fatto solo monitorando una qualche proprietà chimico-fisica, ad esempio spettroscopica, correlata alla reazione. In aggiunta, l'alta efficienza degli enzimi costringe a dover ricorrere a tecniche analitiche che siano in grado di fornire una risposta in tempi adeguati.

Queste difficoltà possono essere aggirate ricorrendo a substrati modello per effettuare le misure cinetiche. Questi presentano anche un vantaggio aggiuntivo perché permettono di variare i parametri chimico-strutturali del substrato, traendone importanti indizi sul meccanismo enzimatico.

Anche l'utilizzo di forme enzimatiche mutate risulta di grande aiuto nella determinazione del meccanismo in quanto consente di individuare gli amminoacidi importanti per l'azione catalitica attraverso, ad esempio, la conduzione di esperimenti di cinetica enzimatica. In più, grazie a questa metodica, si può sperare di determinare la struttura del complesso enzima-substrato naturale o di qualche altro intermedio del ciclo catalitico.

Anche in questi casi, tuttavia, è facile ottenere falsi indizi in quanto l'importanza di un particolare amminoacido può dipendere da un suo ruolo strutturale e non da un diretto coinvolgimento nel meccanismo. Inoltre, anche se la mutazione interessasse realmente un amminoacido importante per la catalisi, può darsi che questa si ripercuota in maniera massiccia sull'intera geometria del sito catalitico: la struttura così ottenuta può risultare non più conforme all'originale.

In ogni caso, i dati ottenuti sperimentalmente con le varie metodiche, combinati insieme, conducono all'elaborazione di un'ipotesi per il meccanismo di reazione seguito dall'enzima. Spesso, tuttavia, non c'è modo di verificare tale ipotesi.

A tal proposito, però, l'utilizzo di tecniche informatiche e più specificatamente di codici di calcolo basati sulla teoria quantomeccanica consente di superare questo ostacolo [12-21].

La simulazione al computer permette, infatti, contrariamente all'esperimento, non solo la determinazione degli intermedi stabili di reazione ma anche quella degli stati di transizione consentendo così la definizione dell'intero meccanismo catalitico.

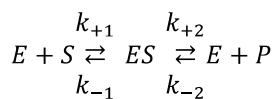
La conoscenza del meccanismo catalitico degli enzimi riveste particolare importanza per le possibili ricadute non solo limitate alle applicazioni industriali ma anche in campo farmacologico, medico, ambientale, economico e sociologico.

Infatti, sta assumendo un'importanza sempre crescente la possibilità di produrre catalizzatori con le proprietà volute attraverso la modifica di enzimi già esistenti o creando sistemi chimici che li imitano. In tal senso l'applicazione di modelli teorici può risultare determinante soprattutto nella fase predittiva indirizzando la ricerca ad opportuni esperimenti e limitando, di conseguenza, le spese per apparecchiature costose.

1.2 Cinetica enzimatica

La cinetica enzimatica [5] si basa sulla ipotesi di L. Michaelis e M. Menten ed è valida per enzimi non allosterici (enzimi la cui attività non dipende da regolatori) operanti su un singolo substrato.

In pratica, secondo tale ipotesi, la reazione catalizzata procede attraverso due semitappe, la prima delle quali, veloce, consiste nella formazione del complesso enzima-substrato (ES), mentre la seconda, lenta e cinematicamente limitante, nella conversione di ES nei prodotti ($E+P$):



Nell'equazione k_{+1} è la costante per la formazione di ES , k_{-1} la costante di dissociazione di ES , k_{+2} la costante per la conversione di ES nei prodotti e k_{-2} la costante per la riconversione dei prodotti in ES . Spesso, k_{+2} è indicata anche come k_{cat} o numero di turn-

over, rappresentando il numero di molecole di substrato convertite nell'unità di tempo per molecola di enzima.

La velocità della reazione è così determinata dalla velocità di scomparsa del complesso ES :

$$V = -\frac{d ES}{dt}$$

Inoltre, poiché la concentrazione del substrato varia con il procedere della reazione, si effettuano misure della velocità iniziale V_0 .

Questo consente anche di trascurare k_{-2} , in quanto la concentrazione dei prodotti è piccola, e di esprimere V_0 come:

$$V_0 = k_{+2} ES$$

Introducendo l'ipotesi di stato stazionario per ES , si ottiene

$$k_{+1} E S = k_{-1} ES + k_{+2} ES$$

Posto che $E = E_{tot} - ES$, dove E_{tot} è la quantità totale di enzima e risolvendo in funzione di ES , si ricava

$$ES = \frac{E_{tot} S}{S + k_M}$$

dove

$$k_M = \frac{k_{-1} + k_{+2}}{k_{+1}}$$

è la costante di Michaelis-Menten ed è una misura dell'affinità dell'enzima per il substrato: più basso è il suo valore maggiore è l'affinità.

La velocità della reazione può così risciversi

$$V_0 = k_{+2} \frac{E_{tot} S}{S + k_M}$$

Per bassi valori di S , l'enzima libero E è in eccesso e la sua concentrazione può considerarsi costante: la velocità di reazione mostra così un andamento lineare (Fig. 2). Tuttavia, all'aumentare della concentrazione del substrato, la velocità di reazione aumenta fino ad un valore massimo V_{max} , a cui tende asintoticamente, perdendo di linearità.

Questo dipende dal fatto che, mano a mano che aumenta S , sempre più molecole di enzima formano il complesso ES finché tutto l'enzima libero è consumato e la velocità risulta indipendente da S . L'enzima è così detto saturo.

I due più importanti parametri cinetici di un enzima sono quanto rapidamente raggiunge la saturazione (ovvero il valore di k_M) e quale è la V_{max} (legata invece al valore di k_{+2}) che può raggiungere.

Il rapporto tra k_{+2} e k_M definisce una nuova costante, la *costante di specificità* e rappresenta una misura dell'efficienza catalitica di un enzima. Il massimo teorico di questa costante è chiamato *limite di diffusione* e vale 10^8 - 10^9 $M^{-1} s^{-1}$: in questo caso ogni collisione tra enzima e substrato ha come effetto la formazione dei prodotti. Enzimi che possiedono una tale efficienza sono detti *cataliticamente perfetti*.

In condizioni di saturazione la velocità può essere espressa come:

$$V_{max} = k_{+2} E_{tot}$$

e la velocità come

$$V_0 = \frac{V_{max} S}{S + k_M}$$

In particolare, quando $k_M = S$ risulta

$$V_0 = \frac{1}{2} V_{max}$$

La k_M , dunque, è pari a quel valore di S per il quale la velocità della reazione è pari a metà della velocità massima.

I valori di k_M e V_{max} possono, quindi, essere ricavati sperimentalmente dal grafico della velocità. Tuttavia, poiché la velocità tende asintoticamente al suo massimo (Fig. 2), la determinazione di V_{max} è soggetta ad imprecisioni.

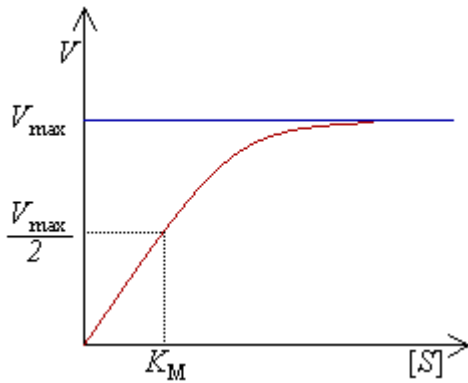


Fig. 2: Andamento della velocità di reazione enzimatica in funzione della concentrazione di substrato S .

Questa difficoltà può essere eliminata linearizzando l'equazione della velocità:

$$\frac{1}{V_0} = \frac{1}{S} \frac{k_M}{V_{max}} + \frac{1}{V_{max}}$$

L'equazione così riscritta rappresenta una retta in un grafico in cui in ascisse sono riportati i valori di $1/S$ e in ordinate quelli di $1/V_0$. Per tale motivo tale grafico è noto come grafico dei doppi reciproci o di Lineweaver-Burk.

L'intercetta delle ordinate fornisce il valore di V_{max} mentre quella delle ascisse il valore di k_M (Fig. 3).

I valori di V_{max} e k_M , naturalmente, sono ottenuti per estrapolazione poiché $1/S$ non può essere né nullo né tantomeno assumere valori negativi.

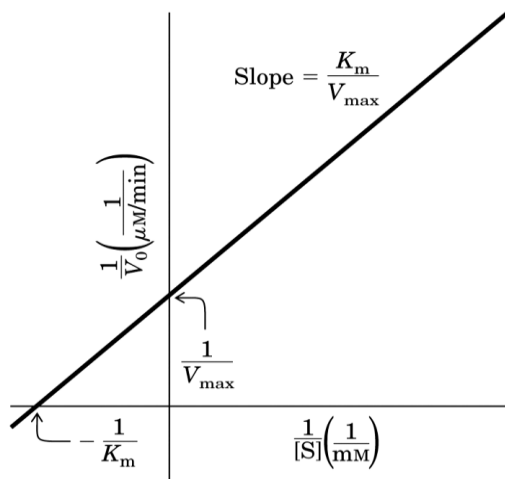


Fig. 3: Diagramma di Lineweaver-Burk o dei doppi reciproci

Bibliografia

- [1] H. Firket – *La Cellula Vivente* – 1994, Newton Compton ed., Roma
- [2] H. Dugas – *Biorganic Chemistry. A chemical approach to enzyme action* – 3rd Ed. – 1996, Springer-Verlag, New York
- [3] J. Daintith – *A concise dictionary of Chemistry* – 1991, Oxford University Press, New York
- [4] K. Kruger, P.J. Grabowski, A.J. Zaugg, J. Sands, D.E. Gottschling e T.R. Cech – *Cell* – 1982, 31, 147
- [5] D.L Nelson, M.M. Cox – *I principi di biochimica di Lehninger* – 4^a Ed. – 2006, Zanichelli, Bologna
- [6] G.E.Schultz, R.H.Schimer – *Principles of protein structure* – 1979, Springer-Verlag, New York
- [7] C.B. Anfinsen – *Science* – 1973, 181, 223
- [8] A. Ferscht – *Enzyme Structure and Mechanism* – 1977, Freeman, San Francisco
- [9] D.A. Kraut, K.S. Carroll, D. Herschlag – *Annu. Rev. Biochem.* – 2003, 72, 517
- [10] P. A. Srere – *Trends Biochem. Sci.* – 1984, 9, 387
- [11] M. Sommerhalter, R.L. Lieberman, A.C. Rosenzweig – *Inorg. Chem.* – 2005, 44, 770
- [12] D.J. Tantillo, J. Chen, K.N. Houk – *Curr. Opinion Chem. Biol.* – 1998, 2, 743
- [13] R.A. Friesner, M.D. Beachy – *Curr. Opinion. Struct. Biol.* – 1998, 8, 257
- [14] T.C. Bruice, K. Kahn - *Curr. Opinion Chem. Biol.* – 2000, 4, 540
- [15] V. Gogonea, D. Suarez, A. van der Vaart, K.M. Merz jr. – *Curr. Opinion. Struct. Biol.* – 2001, 11, 217
- [16] M.J. Field – *J. Comput. Chem.* – 2002, 23, 48
- [17] A. Warshel – *Annu. Rev. Biophys. Biomol. Struct.* – 2003, 32, 425
- [18] T. Lovell, F. Himo, W. Han, L. Noodleman – *Coord. Chem. Rev.* – 2003, 238-239, 211
- [19] A.J. Mulholland – *Drug Discoveri Today: Biosilico* – 2005, 10, 1393
- [20] M.J. Ramos, P.A. Fernandes – *Acc. Chem. Res.* – 2008, 41, 689
- [21] K.H. Hopmann, F. Himo – *Quantum Chemical Modeling of Enzymatic Reactions – Applications to Epoxide-Transforming Enzymes in Comprehensive Natural Products II. Chemistry and Biology* – 2010, L. Mander & H.-W. Liu Eds, Elsevier, Oxford

CAPITOLO II: ASPETTI TEORICI

2.1 L'equazione di Schrödinger nell'approssimazione di Born-Oppenheimer

La descrizione di un sistema microscopico può essere adeguatamente effettuata ricorrendo ad equazioni quantomeccaniche agli autovalori [1-4].

Nel caso di una molecola contenente N elettroni ed M nuclei, ad esempio, per la descrizione di fenomeni non relativistici ed indipendenti dal tempo, è possibile scrivere, facendo uso delle unità atomiche⁴, la seguente equazione per l'energia del sistema (equazione di Schrödinger) [5]:

$$\begin{aligned}
 H\Psi(r, R) = & -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{\substack{j=1 \\ j>i}}^N \frac{1}{r_{ij}} + \sum_{A=1}^N \sum_{\substack{B=1 \\ B>A}}^M \frac{Z_A Z_B}{R_{AB}} \Psi(r, R) \\
 = & E\Psi(r, R)
 \end{aligned}$$

dove: \hat{H} è l'operatore Hamiltoniano per l'energia

m_A la massa del nucleo A

$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ è, in coordinate cartesiane, l'operatore derivata seconda

Z_A e Z_B , rispettivamente, le cariche del nucleo A e del nucleo B

r_{iA} , r_{ij} e R_{AB} , rispettivamente, le distanze dell'elettrone i dal nucleo A, tra gli elettroni i e j e tra i nuclei A e B

E è autovalore dell'autofunzione, ovvero l'energia del sistema in esame

Il simbolo $\Psi(r, R)$ è utilizzato per designare l'autofunzione soluzione dell'equazione e descrittiva del sistema; essa è funzione delle coordinate degli elettroni r_i e delle coordinate nucleari R_A .

L'espressione dell'operatore Hamiltoniano risulta essere formata da più termini i primi due dei quali rappresentano, rispettivamente, l'energia cinetica degli elettroni e l'energia cinetica dei nuclei, mentre i restanti tre termini energetici, rispettivamente, l'attrazione nuclei-elettroni, la repulsione interelettronica e la repulsione internucleare.

L'equazione di Schrödinger può essere semplificata [6,7] se si considera che il rapporto tra la massa del nucleo e dell'elettrone è circa 1800: i nuclei, dunque, possono essere considerati

⁴ Secondo queste unità ogni costante fisica è espressa come multiplo o combinazione di costanti fondamentali quali m_e , \hbar , $4\pi\epsilon_0$, $|e|$ che assumono tutte il valore 1. L'energia, pertanto, risulta essere espressa in E_h (Hartree; $1 E_h = 27,211 \text{ eV}$ o $627,51 \text{ kcal/mol}$).

fermi rispetto agli elettroni. Il termine relativo all'energia cinetica dei nuclei nell'Hamiltoniano è, pertanto, trascurabile, mentre il termine relativo alla repulsione internucleare si traduce in un termine costante. Si ottiene, così, il cosiddetto Hamiltoniano elettronico:

$$H_{el} \Psi(r, R)_{el} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{\substack{j=1 \\ j>i}}^M \frac{1}{r_{ij}} \Psi(r, R)_{el} = E(R)_{el} \Psi(r, R)_{el}$$

L'autofunzione e l'autovalore di questa equazione dipendono entrambe in maniera parametrica dalle coordinate nucleari: cambiando queste cambiano sia l'espressione di $\Psi(r, R)_{el}$ che il valore di $E(R)_{el}$.

Nella realtà, la soluzione di questo problema agli autovalori, così come quello coinvolgente l'Hamiltoniano esatto, non produce un'unica soluzione bensì un insieme non limitato di autofunzioni, tutte ortogonali tra loro e linearmente indipendenti.

Le autofunzioni $\Psi_i(r, R)$ dell'Hamiltoniano esatto \hat{H} possono essere espresse, quindi, come combinazione lineare delle $\Psi_i(r, R)_{el}$:

$$\Psi_j(r, R) = \sum_i \Psi_i(r, R)_{el} \phi_{ij}(R)$$

dove $\phi_{ij}(R)$ sono i coefficienti dell'espansione dipendenti dalle coordinate nucleari.

Introducendo l'ipotesi che il moto nucleare non accoppi differenti stati elettronici (approssimazione adiabatica)⁵ è possibile limitare l'espansione al solo primo termine:

$$\Psi_i(r, R) \approx \Psi_i(r, R)_{el} \phi_i(R)$$

Detto in altri termini, introdurre l'approssimazione adiabatica equivale ad affermare che i moti elettronici e nucleari possono essere separati e le autofunzioni dell'Hamiltoniano esatto espresse come prodotto di autofunzioni elettroniche e nucleari.

L'equazione esatta del sistema può, pertanto, essere riscritta come:

⁵ Tale ipotesi è tanto più valida quanto più gli stati elettronici sono separati in energia e dunque, in special modo, per lo stato fondamentale.

$$\begin{aligned}
H\Psi_i(r, R)_{el}\phi_i(R) &= -\frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \nabla_A^2 + H_{el} + \sum_{\substack{A=1 \\ B>A}}^N \sum_{B=1}^M \frac{Z_A Z_B}{R_{AB}} \Psi_i(r, R)_{el}\phi_i(R) \\
&= -\frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \nabla_A^2 + E_i(R)_{tot} \Psi_i(r, R)_{el}\phi_i(R) \approx E_i \Psi_i(r, R)_{el}\phi_i(R)
\end{aligned}$$

dove si è posto

$$E_i(R) + \sum_{\substack{A=1 \\ B>A}}^N \sum_{B=1}^M \frac{Z_A Z_B}{R_{AB}} = E_i(R)_{tot}$$

Moltiplicando la precedente espressione per $\Psi_i^*(r, R)_{el}$ ed integrando rispetto alle coordinate elettroniche, si ottiene:

$$-\frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \int \Psi_i^*(r, R)_{el} \nabla_A^2 \Psi_i(r, R)_{el} dr + E_i(R)_{tot} \int \phi_i(R) \approx E_i \int \phi_i(R)$$

che è un'equazione agli autovalori con autofunzioni $\phi_i(R)$.

Questa equazione contiene un operatore differenziale la cui espressione cambia al variare di R a causa della dipendenza di $\Psi_i^*(r, R)_{el}$ dalle coordinate nucleari.

Tuttavia, poiché i nuclei sono molto più pesanti degli elettroni, è ragionevole assumere che $\Psi_i^*(r, R)_{el}$ cambi molto poco in funzione di R (approssimazione di Born-Oppenheimer) e che pertanto le sue derivate prima e seconda siano trascurabili.

Si può così scrivere:

$$\int \Psi_i^*(r, R)_{el} \nabla_A^2 \Psi_i(r, R)_{el} dr \phi_i(R) \approx \int \Psi_i^*(r, R)_{el} \Psi_i(r, R)_{el} dr \nabla_A^2 \phi_i(R) = \nabla_A^2 \phi_i(R)$$

L'equazione di Schrödinger per il sistema completo può così essere espressa come:

$$-\frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \nabla_A^2 \phi_i(R) + E_i \phi_i(R) \approx E_i \phi_i(R)$$

Questa equazione equivale a dire che le $\phi_i R$ possono essere interpretate come funzioni d'onda descrittive i moti molecolari che avvengono su una superficie di energia potenziale $E_i R_{tot}$. I nuclei, pertanto, nell'approssimazione di Born-Oppenheimer si muovono lungo una superficie di energia potenziale ottenuta risolvendo l'Hamiltoniano elettronico.

Le soluzioni dell'equazione di Schrödinger nucleare descrivono vibrazioni, traslazioni e rotazioni di una molecola che si trova nel particolare stato elettronico (solitamente lo stato fondamentale) considerato.

2.2 Ipersuperfici

La rappresentazione grafica della variazione di $E_i R_{tot}$ in funzione delle coordinate nucleari genera una cosiddetta ipersuperficie, in virtù della dipendenza dai numerosi parametri geometrici⁶.

Ad ogni stato elettronico può essere associata una differente ipersuperficie [7], ognuna delle quali presenterà punti di massimo e minimo relativi in corrispondenza di particolari valori dei parametri geometrici che rappresentano i valori di equilibrio [4].

In corrispondenza di tali punti deve risultare $\frac{\partial E_i R_{tot}}{\partial R_i} = 0$, per ognuna delle coordinate nucleari R_i : tutte le componenti del vettore gradiente, calcolate in siffatti punti, devono essere nulle.

Ora, la derivata prima dell'energia potenziale è pari alla forza applicata cambiata di segno: dire, cioè, che la derivata prima è nulla in quel punto significa dire che anche le forze che li agiscono sono nulle. Il punto è detto così *stazionario*.

Tuttavia, per poter distinguere tra punti di minimo e di massimo non è sufficiente considerare solo il vettore gradiente ma è necessario ricorrere alle derivate seconde dell'energia potenziale [8]: se tali derivate sono tutte positive allora il punto in questione è un minimo, viceversa se qualcuna di queste derivate è negativa, allora si avrà un punto di massimo.

Le derivate seconde possono essere organizzate in una matrice nota come *matrice Hessiana*.

L'espressione analitica esatta di $E_i R_{tot}$ non è però nota: essa può essere, tuttavia, approssimata mediante un'espansione in serie di Taylor intorno ad un punto stazionario [4].

Limitandosi al termine quadratico (approssimazione armonica):

⁶ L'ipersuperficie per una generica molecola contenente M nuclei può essere costruita riferendosi alle 3M coordinate cartesiane. Tuttavia, risulta più conveniente utilizzare un sistema di 3M-6 coordinate interne, separando i moti vibrazionali da quelli traslazionali e rotazionali (nell'approssimazione di rotore rigido) [9].

$$E_i R_{tot} \cong E_i R_{0 tot} + R - R_0 + \frac{\partial E_i R_{tot}}{\partial R} \Big|_{R=R_0} + \frac{1}{2} R - R_0 + \frac{\partial^2 E_i R_{tot}}{\partial R^2} \Big|_{R=R_0} R - R_0$$

$$E_i R_{tot} \cong E_i R_{0 tot} + \frac{1}{2} R - R_0 + \frac{\partial^2 E_i R_{tot}}{\partial R^2} \Big|_{R=R_0} R - R_0$$

L'equazione di Schrödinger nucleare diventa allora

$$-\frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \nabla_A^2 R + E_i R_{0 tot} + \frac{1}{2} R - R_0 + \frac{\partial^2 E_i R_{tot}}{\partial R^2} \Big|_{R=R_0} R - R_0 \quad \phi_i R = E_i \phi_i R$$

$$-\frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \nabla_A^2 R + \frac{1}{2} R - R_0 + \frac{\partial^2 E_i R_{tot}}{\partial R^2} \Big|_{R=R_0} R - R_0 \quad \phi_i R = E_i - E_i R_{0 tot} \phi_i R$$

$$-\frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \nabla_A^2 R + \frac{1}{2} R - R_0 + \frac{\partial^2 E_i R_{tot}}{\partial R^2} \Big|_{R=R_0} R - R_0 \quad \phi_i R = E' \phi_i R$$

Introducendo le coordinate pesate sulla massa $Y_i = \overline{m_i} \Delta R_i$ definite in termini di spostamento dal punto di equilibrio invece che in termini di posizione assoluta, si ottiene

$$-\frac{1}{2} \sum_{A=1}^{3M} \frac{\partial^2}{\partial Y^2} + \frac{1}{2} Y^\dagger H Y \quad \phi_i Y = E' \phi_i Y$$

dove H è la matrice Hessiana e la sommatoria al primo termine si estende ora su tutte le $3M$ coordinate cartesiani nucleari.

Se a questo punto si opera una trasformazione unitaria che diagonalizzi H , trasformando gli spostamenti nucleari in coordinate normali Q

$$U^\dagger H U = k \quad U^\dagger U = U U^\dagger = I$$

$$Q = U^\dagger Y \Leftrightarrow Q_k = \sum_j U_{kj}^\dagger Y_j$$

si ottiene

$$\frac{\partial}{\partial Y_i} = \sum_k \frac{\partial Q_k}{\partial Q_k} \frac{\partial Q_k}{\partial Y_i} = \sum_k \frac{\partial}{\partial Q_k} U_{ki}^\dagger$$

e

$$Y^\dagger H Y = Y^\dagger U U^\dagger H U U^\dagger Y = Q^\dagger k Q$$

che sostituiti nell'equazione di Schrödinger nucleare danno

$$\sum_{A=1}^{3M} -\frac{1}{2} \frac{\partial^2}{\partial Q_A^2} + \frac{1}{2} k_A Q_A^2 \phi_i Q = E' \phi_i Q$$

Quest'ultima equazione mostra come l'Hamiltoniano possa essere espresso come somma di 3M Hamiltoniani indipendenti, ognuno dei quali conduce a un'equazione agli autovalori indipendente⁷.

Nel contempo la funzione d'onda può essere espressa come prodotto di funzioni d'onda, ognuna dipendente da una singola coordinata normale.

$$-\frac{1}{2} \frac{\partial^2}{\partial Q_A^2} + \frac{1}{2} k_A Q_A^2 \phi_A Q_A = \epsilon_A \phi_A Q_A$$

$$E' = \epsilon_A + \epsilon_B + \epsilon_C + \dots + \epsilon_{3M}$$

$$\phi_i Q = \phi_A Q_A \phi_B Q_B \phi_C Q_C \dots \phi_{3M} Q_{3M}$$

Consideriamo, ora, il caso di una generica reazione chimica: alcuni legami si rompono ed altri si formano. Tutto ciò significa spostarsi da un punto di minimo ad un altro punto di minimo nell'ipersuperficie passando per un punto di massima energia o *stato di transizione*. (Fig. 3)

I modi di connessione di tali minimi, tuttavia, risultano essere molteplici rendendo necessaria così una più rigorosa definizione per lo stato di transizione: esso è il massimo lungo il cammino a minima energia. In altre parole, lo stato di transizione è un punto di sella di ordine uno.

Tradotto in termini matematici, gli autovalori k_i della matrice Hessiana devono risultare tutti positivi eccetto uno.

⁷ Ciascuna di queste 3M equazioni è in realtà un'equazione di Schroedinger per un oscillatore armonico (una molla) con costante di forza k (ovvero gli autovalori della matrice Hessiana) e frequenza $\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}$.

Ma, poiché tali autovalori corrispondono alle costanti di forza dei modi normali di vibrazione della molecola, dire che un autovalore assume segno negativo equivale a dire che quel modo normale di vibrazione ha una frequenza immaginaria.

Lo stato di transizione, dunque, può essere identificato da una frequenza di vibrazione immaginaria.

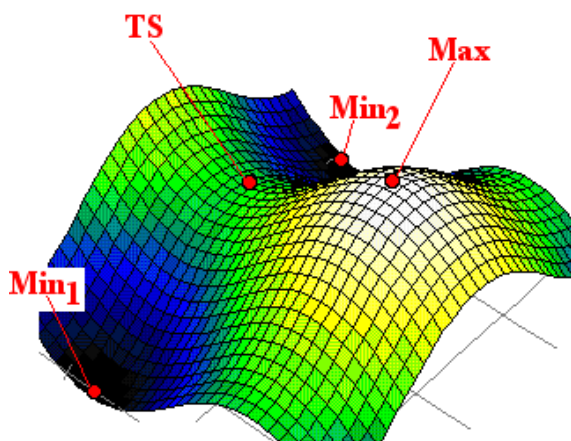


Fig. 3: Rappresentazione grafica di una ipersuperficie con evidenziati i punti di massimo, di minimo e di sella.

Visualizzare il profilo energetico di una reazione, o più esattamente del meccanismo preso in considerazione, risulta pressoché impossibile utilizzando l'ipersuperficie a causa della sua elevata dimensionalità; tuttavia, ritagliandola lungo il cammino di reazione ed appiattendola in un piano è possibile ottenere la canonica rappresentazione bidimensionale della variazione di energia in funzione della *coordinata di reazione*, definendo quest'ultima come l'insieme dei parametri geometrici che variano nel corso della reazione.

2.3 Risolvere l'equazione di Schrödinger elettronica: il determinante di Slater

L'importanza di risolvere l'equazione di Schrödinger per gli elettroni è argomento che discende direttamente da quanto finora esposto.

Risolvere tale equazione, però, non è cosa né semplice né banale, poiché richiede la formulazione esplicita sia dell'Hamiltoniano corrispondente che delle sue autofunzioni.

Nel caso di sistemi monoelettronici, tuttavia, si può pervenire facilmente ad una risoluzione esatta del problema, giacché l'Hamiltoniano risulta essere semplificato dall'assenza di interazioni tra elettroni, ovvero il termine contenente l'operatore $\frac{1}{r_{ij}}$ è assente.

Nel caso più frequente di sistemi a più elettroni, però, l'impossibilità di determinare le posizioni relative degli elettroni e la loro evoluzione nel tempo rende impossibile valutare esattamente tale termine: è necessario, dunque, introdurre delle approssimazioni.

La più semplice delle approssimazioni consiste nel considerare il sistema plurielettronico come costituito da elettroni non-interagenti [2].

Una prima conseguenza di tale approssimazione è quella che si riflette nell'espressione delle autofunzioni che possono così essere fattorizzate come prodotto di funzioni, ognuna delle quali dipendenti da un singolo elettrone (*prodotto di Hartree*) [10-12]

$$\psi r_1 \dots r_i \dots r_j \dots r_N = \prod_{i=1}^N \varphi_i r_i = \varphi_1 r_1 \varphi_2 r_2 \varphi_3 r_3 \dots$$

Ognuna di delle $\varphi_i(r_i)$ è in realtà formata dal prodotto di una funzione dipendente dalle coordinate spaziali dell'elettrone (*spatial orbitals*⁸) e di una funzione dello spin⁹. Tale prodotto è indicato come *spin orbitals*¹⁰.

$$\varphi_i r_i = \chi_i x, y, z \alpha \omega_1$$

$$\varphi_i r_i = \chi_i x, y, z \beta \omega_2$$

Per comodità di calcolo gli *spin orbitals* sono assunti essere ortonormali, ovvero

$$\int \varphi_i^* r_i \varphi_j r_j dr = \delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Il prodotto di Hartree, però, soffre di un problema di fondo: viola l'indistinguibilità degli elettroni poiché assegna uno specifico elettrone ad una specifica funzione. Inoltre, per gli elettroni vale il principio di antisimmetria per cui la funzione d'onda deve cambiare di segno per effetto dell'interscambio delle coordinate di una qualsiasi coppia di elettroni [13].

$$\psi r_1 \dots r_i \dots r_j \dots r_N = -\psi r_1 \dots r_j \dots r_i \dots r_N$$

Il prodotto di Hartree viola anche tale principio. Esso, pertanto, risulta non essere una buona funzione d'onda per il sistema. Scegliendo però un'opportuna combinazione lineare dei prodotti di Hartree ottenuti tramite tutti i possibili interscambi delle r_i , è possibile ottenere una nuova funzione d'onda che rispetti il principio di antisimmetria [14].

⁸ Un *Orbitale spaziale* è una funzione che descrive la distribuzione spaziale di un elettrone tale che $\chi_i x, y, z \int d\tau$ rappresenta la probabilità di trovare l'elettrone nel volume di spazio $d\tau$.

⁹ Nella trattazione non-relativistica l'Hamiltoniano non dipende dalle coordinate di spin, cosa questa che consente di separare le variabili spaziali da quelle di spin.

¹⁰ Per sistemi plurieltronici monoatomici gli spin orbitals sono approssimati con gli orbitali atomici idrogenoidi, mentre per sistemi plurieltronici pluriatomici sono generalmente formati da combinazioni lineari di orbitali atomici idrogenoidi centrati sui diversi atomi. Gli spin orbitals che differiscono solo per la funzione di spin sono indicati come *restricted*, mentre se differiscono anche per la funzione spaziale sono indicati come *unrestricted*.

La funzione d'onda così ottenuta può essere rappresentata in termini di un determinante noto come *determinante di Slater* [15,16]

$$\psi_{r_1 \dots r_i \dots r_j \dots r_N} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1 r_1 & \varphi_2 r_1 & \dots & \varphi_N r_1 \\ \varphi_1 r_2 & \varphi_2 r_2 & \dots & \varphi_N r_2 \\ \dots & \dots & \dots & \dots \\ \varphi_1 r_N & \varphi_2 r_N & \dots & \varphi_N r_N \end{vmatrix}$$

dove il termine $\frac{1}{\sqrt{N!}}$ è un fattore di normalizzazione.

Il determinante di Slater può essere rappresentato anche con una notazione più compatta, come

$$\psi_{r_1 \dots r_i \dots r_j \dots r_N} = |\varphi_1 r_1 \varphi_2 r_2 \dots \varphi_N r_N| = |\varphi_1 \varphi_2 \dots \varphi_N|$$

In questo caso il fattore di normalizzazione è sottinteso, mentre il ket contiene gli elementi diagonali del determinante. Poiché lo scambio di due colonne nel determinante provoca un cambio di segno della funzione, l'ordine in cui compaiono le funzioni riveste grande importanza.

La bontà della scelta fatta utilizzando una particolare funzione d'onda, approssimazione dell'esatta funzione d'onda dello stato fondamentale del sistema, può essere testata attraverso il principio variazionale per il quale il valore di aspettazione dell'Hamiltoniano è un limite superiore dell'energia esatta del sistema, purché la funzione di prova sia normalizzata ed obbedisca alle appropriate condizioni al contorno.

$$E \leq \varepsilon_0 = \psi_0 H \psi_0$$

La migliore funzione d'onda possibile per la descrizione dello stato fondamentale di un sistema, dunque, è quella che minimizza l'energia.

Nel caso di un determinante di Slater la flessibilità variazionale è data solo dalla scelta degli spin orbitals.

2.4 Il metodo Hartree-Fock

I risultati ottenuti con l'approssimazione di elettroni non-interagenti sono piuttosto distanti dal ragionevole errore (1-2 kcal): se valutare esattamente il termine di repulsione interelettronica risulta impossibile, trascurarlo del tutto non conduce a risultati migliori [1,3,17,18].

Ciononostante il determinante di Slater risulta essere un buon punto di partenza per ottenere approssimazioni migliori.

Ricordando l'espressione dell'Hamiltoniano completo

$$H = -\frac{1}{2} \sum_i \nabla_i^2 r_i - \sum_{i, A=1}^M \frac{Z_A}{r_{iA}} + \frac{1}{2} \sum_{i, j \neq i} \frac{1}{r_{ij}} = \sum_i h_i + \frac{1}{2} \sum_{i, j \neq i} \frac{1}{r_{ij}}$$

dove $h_i = -\frac{1}{2} \nabla_i^2 r_i - \sum_{A=1}^M \frac{Z_A}{r_{iA}}$ è il cosiddetto *Hamiltoniano di core* e descrive l'energia cinetica e l'energia potenziale dell' i-esimo elettrone nel campo dei nuclei (il "core" appunto): esso è pertanto un operatore monoelettronico.

Il valore di aspettazione E_0 per l'energia dello stato fondamentale risulta essere

$$E_0 = \psi_0 H \psi_0 = \sum_i \varphi_i h_i \varphi_i + \frac{1}{2} \sum_{i, j \neq i} \int dr_i dr_j \varphi_i^* r_i \varphi_j^* r_j r_{ij}^{-1} \varphi_i r_i \varphi_j r_j + \\ - \int dr_i dr_j \varphi_i^* r_i \varphi_j^* r_j r_{ij}^{-1} \varphi_j r_i \varphi_i r_j$$

La minimizzazione di E_0 sotto il vincolo che le $\varphi_i r_i$ siano ortonormali conduce alle ben note equazioni di Hartree-Fock

$$h_i \varphi_i r_i + \sum_{j \neq i} \int dr_j \varphi_j r_j r_{ij}^{-1} \varphi_i r_i - \sum_{j \neq i} \int dr_j \varphi_j^* r_j \varphi_i^* r_j r_{ij}^{-1} \varphi_j r_i = \varepsilon_i \varphi_i r_i$$

dove

$$J_j r_i = \int dr_j \varphi_j r_j r_{ij}^{-1}$$

è il termine di Coulomb, ovvero il potenziale monoelettronico ottenuto mediando sulle coordinate spaziali e di spin le repulsioni tra gli elettroni i e j , pesato sulla probabilità che l'elettrone j occupi l'elemento di volume dr_j intorno a r_j . Sommando su $i \neq j$ si ottiene il potenziale medio che agisce su un elettrone in $\varphi_i r_i$ derivante dagli altri $N-1$ elettroni¹¹. Poiché il risultato dell'applicazione di tale potenziale sullo spin orbital $\varphi_i r_i$ dipende solamente dal valore che assume φ_i nel punto r_i , $J_j r_i$ è detto *locale*.

Il secondo termine di sommatoria non ha, invece, un corrispettivo "classico" e può essere definito solo dall'effetto che opera sulla $\varphi_i r_i$.

$$K_j r_i \varphi_i r_i = \int dr_j \varphi_j^* r_j r_{ij}^{-1} \varphi_i r_j \varphi_j r_i$$

Esso è indicato come operatore di scambio in quanto porta allo scambio delle variabili negli spin orbitals φ_i e φ_j e deriva dalla natura antisimmetrica del determinante e dal fatto che il moto di elettroni con spin uguale risulta essere correlato. Poiché il risultato dell'applicazione di $K_j r_i$ su $\varphi_i r_i$ dipende dal valore di $\varphi_i r_j$ nello spazio, tale termine è indicato anche come potenziale *non-locale*.

Definendo l'operatore di *Fock* come somma dell'operatore Hamiltoniano di core $h r_i$ e di un effettivo potenziale monoelettronico (*potenziale di Hartree-Fock*) $v^{HF} r_i$, ovvero

$$f r_i = h r_i + v^{HF} r_i = h r_i + \sum_j J_j r_i - \sum_j K_j r_i$$

essendo

$$v^{HF} r_i = \sum_j J_j r_i - \sum_j K_j r_i$$

è possibile riscrivere l'equazione di Hartree-Fock come un'equazione agli autovalori

$$f r_i \varphi_i r_i = \varepsilon_i \varphi_i r_i$$

¹¹ Considerare nella sommatoria anche il termine $i=j$ sembrerebbe un non senso dal punto di vista fisico poiché tale termine rappresenterebbe la repulsione di un elettrone per sé stesso. Tuttavia questo stesso termine risulta essere uguale, a meno del segno, a quello derivante dalla sommatoria su K_i con il risultato finale che l'introduzione di tali termini non altera la definizione del potenziale di Fock.

con gli autovalori ε_i rappresentanti le energie degli spin orbitals¹².

Nella realtà l'equazione di Hartree-Fock è una pseudo-equazione agli autovalori in quanto l'operatore di Fock, attraverso l'espressione di v^{HF} , dipende dalle sue soluzioni φ_i r_i .

Detto in altri termini, l'equazione di Hartree-fock non è un'equazione lineare e per essere risolta richiede una procedura iterativa nota come *Self-consistent field (SCF)*. Ciononostante, il metodo Hartree-Fock ha il merito di sostituire il complicato problema a più elettroni con un problema monoelettronico in cui si tiene conto della repulsione interelettronica in maniera mediata.

L'impiego di un determinante di Slater come approssimazione della funzione d'onda esatta, infatti, consente di catturare molta parte della fisica dei sistemi plurieltronici, risultando tuttavia in un'energia calcolata per il sistema in esame sempre superiore alla vera energia.

La differenza tra l'esatto valore di energia e quello calcolato rappresenta la cosiddetta energia di correlazione, la quale risulta essere sempre negativa.

Questa differenza tra valore calcolato e valore reale può essere attribuita principalmente alla repulsione istantanea degli elettroni, non considerata nell'espressione del v^{HF} r_i . In pratica nello schema HF gli elettroni sono spesso troppo vicini l'un l'altro con conseguente repulsione troppo elevata e sovrastima dell'energia cinetica, mentre l'attrazione nucleo elettrone risulta o troppo piccola o troppo grande.

Questo tipo di effetto va sotto il nome di correlazione dinamica ed è un effetto a corto raggio.

Un'altra possibile fonte di errore può essere ricondotta alla presenza di altri determinanti di Slater ad energia paragonabile a quella del determinante considerato. Questo tipo di errore è indicato come correlazione non-dinamica o statica ed è un effetto a lungo raggio.

La correlazione elettronica è stata oggetto di numerose trattazioni teoriche che hanno portato a sviluppare implementazioni del metodo HF, implementazioni che si basano su metodi perturbativi come l'MP2 o l'MP4 (Møller-Plesset al 2 e al 4 ordine) oppure su una migliore approssimazione della funzione d'onda come i metodi CI (interazione di configurazione), QCI (interazione di configurazione quadratica), CC (Cluster accoppiati) e loro derivati.

2.5 L'equazione di Roothaan

La risoluzione delle equazioni di Hartree-Fock può essere facilmente ottenuta convertendo tali equazioni in un insieme di equazioni algebriche risolvibili con la normale algebra

¹² L'energia dello stato fondamentale, tuttavia, non è pari alla somma delle energie di singoli spin orbital giacché ogni spin orbital considera le interazioni dell'i-esimo elettrone con i restanti N-1 elettroni e sommando su tutti gli N elettroni tali interazioni sarebbero contate due volte.

matriciale [1,17], semplicemente esprimendo gli orbitali spaziali come combinazione lineare di un set noto di k funzioni di base Γ_v ¹³

$$\chi_i = \sum_{v=1}^k C_{vi} \Gamma_v$$

Risolvere le equazioni di Hartree-Fock, pertanto, significa determinare i coefficienti C_{vi} di tali combinazioni lineari.

Riscrivendo l'equazione di Hartree-Fock in termini di orbitali spaziali e nell'ipotesi di molecole a guscio chiuso descritte da un singolo determinante di Slater, si ottiene

$$f' r_i \chi_i r_i = \varepsilon_i \chi_i r_i$$

dove

$$f' r_i = h r_i + v^{HF} r_i = h r_i + \sum_j^{N/2} 2J_j r_i - \sum_j^{N/2} K_j r_i$$

Moltiplicando a sinistra per Γ_μ ed integrando si ottiene la cosiddetta equazione di Roothaan:

$$\sum_v C_{vi} \int \Gamma_\mu^* r_i f' r_i \Gamma_v r_i = \varepsilon_i \sum_v C_{vi} \int \Gamma_\mu^* r_i \Gamma_v r_i$$

la quale, posto

¹³ Come funzioni di base si possono usare qualunque tipo di funzioni anche se generalmente si preferiscono funzioni di Slater che riproducono meglio gli orbitali atomici. Inoltre, ogni loro combinazione lineare richiede un minore numero di termini per riprodurre l'orbitale molecolare spaziale. Dal punto di vista del calcolo, tuttavia, le funzioni gaussiane risultano più vantaggiose, pur riproducendo peggio gli orbitali atomici e richiedendo un maggior numero di termini nell'espansione. In pratica si cerca di coniugare i pregi di questi due tipi di funzioni espandendo le funzioni di Slater in termini di una combinazione lineare fissa (contrazione) di funzioni (*primitive*) gaussiane (ad es. il set STO-NG: una funzione di Slater è formata da una combinazione lineare di N funzioni gaussiane). Il set di funzioni di base può essere migliorato raddoppiando il numero delle funzioni di base così che ognuna abbia un esponente leggermente più piccolo o leggermente più grande del valore utilizzato per il set STO-NG (set doppia zeta). Ad es. il set 6-31G utilizza una singola funzione di base combinazione lineare di 6 gaussiane per descrivere gli orbitali interni dell'atomo considerato, mentre gli orbitali di valenza sono descritti da due differenti funzioni, una combinazione lineare di 3 gaussiane e l'altra da una singola gaussiana. Ulteriori miglioramenti al set di base possono essere raggiunti includendo funzioni di polarizzazione ovvero orbitali d per gli elementi della prima riga (ad es. 6-31G*) e orbitali di tipo p per l'atomo di idrogeno (ad es. 6-31G**). Per gli atomi più pesanti, invece, per la descrizione degli elettroni interni si utilizza un potenziale di core effettivo o *pseudopotenziale* [19].

$$dr_i \Gamma_\mu^* r_i f' r_i \Gamma_\nu r_i = F_{\mu\nu}$$

e

$$dr_i \Gamma_\mu^* r_i \Gamma_\nu r_i = S_{\mu\nu}$$

può essere riscritta come

$$F_{\mu\nu} C_{\nu i} = \varepsilon_i S_{\mu\nu} C_{\nu i}$$

o in forma matriciale come

$$FC = SC\varepsilon$$

La matrice F di elementi $F_{\mu\nu}$ è la matrice di Fock, rappresentazione dell'operatore di Fock nel set di funzioni di base adoperato. Come tale essa è una matrice Hermitiana di ordine $k \times k$.

Dall'espressione dell'operatore di Fock è possibile notare come la sua matrice di rappresentazione F si componga in realtà di due parti, una monoelettronica invariante $H_{\mu\nu}^{core}$ contenente i termini di energia cinetica elettronica e del potenziale attrattivo elettrone – nuclei, l'altra $G_{\mu\nu}$ dipendente dalla *matrice densità*¹⁴ P, di elementi $P_{\lambda\sigma} = \sum_j^N 2C_{\lambda j} C_{\sigma j}^*$, e dagli integrali relativi ai termini di repulsione interelettronica e di scambio.

$$F_{\mu\nu} = H_{\mu\nu}^{core} + G_{\mu\nu}$$

dove

$$G_{\mu\nu} = \sum_j^N dr_i \Gamma_\mu^* r_i \left[2J_j r_i - K_j r_i \right] \Gamma_\nu r_i$$

¹⁴ La matrice densità è così chiamata poiché, dato un set di funzioni di base noto, essa specifica completamente la densità di carica: $\rho = 2 \sum_a^{N/2} \chi_a r_i^2 = \sum_{\mu\nu} 2 \sum_a^{N/2} C_{\mu a} C_{\nu a}^* \Gamma_\mu \Gamma_\nu^* = \sum_{\mu\nu} P_{\mu\nu} \Gamma_\mu \Gamma_\nu^*$

$$G_{\mu\nu} = \sum_{j, \lambda\sigma}^N C_{\lambda j} C_{\sigma j}^* \int dr_i dr_j \Gamma_{\mu}^* r_i \Gamma_{\nu} r_i r_{ij}^{-1} \Gamma_{\lambda}^* r_j \Gamma_{\sigma} r_j - \int dr_i dr_j \Gamma_{\mu}^* r_i \Gamma_{\lambda} r_i r_{ij}^{-1} \Gamma_{\sigma}^* r_j \Gamma_{\nu} r_j$$

$$G_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} \int dr_i dr_j \Gamma_{\mu}^* r_i \Gamma_{\nu} r_i r_{ij}^{-1} \Gamma_{\lambda}^* r_j \Gamma_{\sigma} r_j - \frac{1}{2} \int dr_i dr_j \Gamma_{\mu}^* r_i \Gamma_{\lambda} r_i r_{ij}^{-1} \Gamma_{\sigma}^* r_j \Gamma_{\nu} r_j$$

La matrice di Fock, dunque, dipende dalla matrice densità P e quindi dai coefficienti di espansione C_{vi} . L'equazione di Roothaan, pertanto, è un'equazione non lineare che richiede una procedura iterativa per poter essere risolta.

La matrice S, di elementi $S_{\mu\nu}$, invece, è la matrice di sovrapposizione delle funzioni del set di base. È anch'essa una matrice $k \times k$ Hermitiana ed i suoi elementi fuori diagonale costituiscono un'indicazione della dipendenza lineare delle funzioni di base: se uno di essi tende al valore di 1 (in valore assoluto) vuol dire che le funzioni di base ad esso associato tendono ad essere linearmente dipendenti ed i corrispondenti autovalori nulli. Se invece il set di funzioni di base è reso ortonormale¹⁵, ovvero se S è una matrice diagonale di elementi diagonali uguali ad 1, l'equazione di Roothaan può essere riscritta come $FC=C\varepsilon$ e gli autovettori C_i e gli autovalori ε_i possono essere trovati per diagonalizzazione della matrice F.

La matrice C, di elementi C_{vi} , infatti, è una matrice $k \times k$ formata dai coefficienti delle combinazioni lineari (autovettori), correlata, tramite la matrice densità $P_{\lambda\sigma}$, alla densità di carica, al potenziale $v^{HF} r_i$ e a $G_{\mu\nu}$.

La matrice ε , infine, è una matrice diagonale contenente le energie degli orbitali Hartree-Fock (autovalori).

2.6 La teoria del funzionale della densità

Le limitazioni dei metodi HF e post HF, soprattutto in termini di complessità e tempi di calcolo all'aumentare delle dimensioni del sistema considerato, possono essere ridimensionate adottando un differente approccio rispetto quello basato sulla funzione d'onda.

¹⁵ Operare direttamente sulle Γ_{ν} per renderle ortogonali non conviene poiché richiederebbe di ricalcolare gli integrali a due elettroni o comunque di trasformarli secondo il nuovo set di orbitali, consumando così molto tempo. Risulta più facile, invece, definire una nuova matrice dei coefficienti C' correlata alla matrice C tramite una matrice di trasformazione (non unitaria) X tale che $X^{\dagger}SX = 1$; l'equazione di Roothaan si riscrive così come $FXC' = SXC'\varepsilon$. Ora, moltiplicando a sinistra per X^{\dagger} si ottiene: $X^{\dagger}FXC' = X^{\dagger}SXC'\varepsilon$ ovvero $F'C' = C'\varepsilon$. Diagonalizzando F' si ricava C' e da questa la matrice C.

In altre parole, come già intuito nel 1927 da Thomas [20] e Fermi [21-23] e confermato dal primo teorema di Hohenber e Kohn (1964) [24], è possibile utilizzare la densità elettronica dello stato fondamentale ρ_0 per descrivere correttamente e completamente il sistema di interesse [3,18,25].

La densità elettronica, infatti, contiene in sé le informazioni relative al numero degli elettroni e alla posizione e carica dei nuclei. In aggiunta, essa può essere determinata sperimentalmente a differenza della funzione d'onda.

L'utilizzo della densità elettronica in luogo della funzione d'onda, in particolare, consente di passare dalla dipendenza dei calcoli dalle $4N$ coordinate delle $\varphi_i r_i$ alle sole 3 coordinate spaziali della ρ_0 .

Secondo questo nuovo approccio, denominato appunto *teoria del funzionale della densità* o DFT, è possibile esprimere l'energia dello stato fondamentale¹⁶ E_0 del sistema come funzionale¹⁷ di ρ_0 .

$$E_0 \rho_0 = T \rho_0 + E_{ee} \rho_0 + E_{Ne} \rho_0$$

dove

$$\rho_0 = N \int \varphi_i r_i^2 dr_2 \dots dr_N ds_i$$

L'integrale è su tutte le coordinate di spin e su tutte le coordinate spaziali tranne una.

L'ultimo dei tre termini che compongono l'espressione di $E_0 \rho_0$ rappresenta l'interazione tra nuclei ed elettroni ed è l'unico termine che dipende dal sistema in esame, mentre i primi due rappresentano rispettivamente l'energia cinetica elettronica e le interazioni interelettroniche e risultano indipendenti dal particolare sistema considerato, andando a formare il cosiddetto funzionale di Hohenberg-Kohn o funzionale universale $F \rho_0$.

$$F \rho_0 = T \rho_0 + E_{ee} \rho_0$$

¹⁶ La ρ_0 contiene anche le informazioni sugli stati elettronici eccitati del sistema ma in pratica non è nota l'espressione dei funzionali richiesti per ottenere queste informazioni.

¹⁷ Un funzionale è una regola matematica che assegna ad ogni funzione un numero: ad esempio l'integrale che definisce il perimetro o l'area di una curva chiusa. La sua dipendenza dalla funzione è indicata generalmente con le parentesi quadre.

Se tale funzionale fosse conosciuto esattamente allora l'equazione di Schrödinger sarebbe risolta esattamente. Purtroppo le forme esplicite di $T \rho_0$ e di $E_{ee} \rho_0$ sono completamente sconosciute.

L'unica parte di $F \rho_0$ nota è il contributo di Coulomb $J \rho_0$ al funzionale $E_{ee} \rho_0$.

$$E_{ee} \rho_0 = J \rho_0 + E_{ncl} \rho_0$$

dove

$$J \rho_0 = \frac{1}{2} \int \frac{\rho(r_i) \rho(r_j)}{r_{ij}} dr_i dr_j$$

Il termine $E_{ncl} \rho_0$ rappresenta tutte le interazioni non classiche tra gli elettroni, incluse la correzione per l'autointerazione¹⁸, l'effetto di scambio dovuto alla correlazione tra gli elettroni di uguale spin e l'effetto di correlazione elettronica legata alla carica degli elettroni e all'impossibilità di avvicinarsi troppo l'uno all'altro.

Per poter effettuare calcoli DFT, dunque, occorre fornire una qualche espressione approssimata per il funzionale $F \rho_0$. Tale richiesta equivale ad utilizzare un Hamiltoniano approssimato, facendo sì che il principio variazionale non sia più valido. In altre parole l'energia calcolata può stare benissimo al di sotto dell'energia reale.

2.7 Lo schema di Kohn-Sham

Il principale ostacolo da superare per ottenere accurati calcoli DFT è costituito dal modo in cui si determina l'energia cinetica elettronica. Tutti i tentativi di determinare tale quantità direttamente, infatti, si sono rivelati senza successo presentando errori ben superiori a quelli ottenuti con il metodo HF.

Basandosi su ciò, Kohn e Sham escogitarono uno schema indiretto basato su orbitali per il calcolo di tale energia [3,18,26].

In pratica, essi considerarono un sistema di riferimento costituito da elettroni non interagenti e, quindi, descrivibile tramite un determinante di Slater. Tale sistema sperimenta un potenziale locale $v_s(r)$ tale che la sua densità elettronica dello stato fondamentale ρ_s risulta uguale a quella del sistema reale ρ_0 .

¹⁸ Il contributo di Coulomb contiene anche un'interazione della densità elettronica con sé stessa, fisicamente priva di senso. Nel modello HF questo effetto era perfettamente corretto dall'integrale di scambio.

$$\rho_0 = \rho_s = \sum_i^N \varphi_i^2$$

Gli orbitali del sistema non-interagente φ_i , noti come orbitali di Kohn-sham, non hanno un reale significato fisico (come invece è per gli orbitali HF): semplicemente la somma dei loro quadrati è uguale alla densità elettronica.

Essi possono essere determinati in maniera analoga agli orbitali HF:

$$f^{\text{KS}} \varphi_i = -\frac{1}{2} \nabla^2 \varphi_i + v_s \varphi_i = \varepsilon_i \varphi_i$$

dove f^{KS} è l'operatore monoelettronico di Kohn-Sham.

Giova ricordare che, a dispetto della somiglianza, le equazioni di Kohn-Sham contengono un potenziale che è locale¹⁹, mentre nelle equazioni HF il potenziale è non-locale. Questo assicura anche un costo computazionale minore per lo schema KS.

L'energia cinetica degli elettroni non-interagenti T_s può essere calcolata come

$$T_s = -\frac{1}{2} \sum_i^N \varphi_i \nabla^2 \varphi_i$$

Essa naturalmente, non è uguale a quella del sistema reale, cionondimeno consente di calcolarne la maggior parte possibile con buona accuratezza.

La differenza tra valore reale e valore calcolato T_c è trattata separatamente e concorre a formare la cosiddetta energia di scambio e correlazione $E_{xc} \rho_0$ definita come

$$E_{xc} \rho_0 = T \rho_0 - T_s \rho_0 + E_{ee} \rho_0 - J \rho_0 = T_c \rho_0 + E_{ncl} \rho_0$$

In tal modo il funzionale universale può essere riscritto come

$$F \rho_0 = T_s \rho_0 + J \rho_0 + E_{xc} \rho_0$$

e quello per l'energia come

¹⁹ Esso dipende solo dal valore che assume in un punto r dello spazio, mentre è indipendente dal valore che assume in tutti gli altri punti. In termini di densità elettronica, invece, la sua dipendenza è in generale complessa e non-locale.

$$E_0 \rho_0 = F \rho_0 + E_{Ne} \rho_0 = T_s \rho_0 + J \rho_0 + E_{xc} \rho_0 + E_{Ne} \rho_0$$

Per poter ricavare gli orbitali φ_i e quindi l'energia cinetica T_s , però, è necessario fornire un'espressione per il potenziale $v_s r$. Ciò può essere fatto minimizzando il funzionale dell'energia per il sistema reale sotto il vincolo che gli orbitali φ_i rimangano ortonormali. Si ottiene così:

$$v_s r = \frac{\rho r_j}{r_{ij}} dr_j + V_{xc} r_i - \sum_A^M \frac{Z_A}{r_{iA}}$$

dove $V_{xc} r_i \equiv \frac{\delta E_{xc}}{\delta \rho}$ è il potenziale di scambio e correlazione ovvero la derivata dell'energia di scambio e correlazione rispetto la densità.

Anche in questo caso, come per HF, il potenziale dipende dalla densità e dunque dagli orbitali: anche le equazioni di Kohn-Sham, pertanto, devono essere risolte in maniera iterativa.

L'espressione del potenziale, inoltre, non contiene alcun riferimento allo spin elettronico²⁰: le soluzioni delle equazioni di Kohn-Sham sono così doppiamente degeneri ovvero per ogni ϵ_i ci sono due soluzioni indipendenti che condividono la stessa parte spaziale²¹.

Nella trattazione fin qui svolta non è stata introdotta alcuna approssimazione: lo schema di Kohn-Sham è dunque esatto in linea di principio. Tuttavia, non si conosce la forma esplicita per il funzionale $E_{xc} \rho_0$ né per il corrispondente potenziale $V_{xc} r_i$ ed è a questo punto che bisogna ricorrere a delle espressioni approssimate [27,28].

La bontà dell'approccio DFT, dunque, dipende solo dall'accuratezza scelta nell'approssimazione del funzionale di scambio e correlazione [29]. La ricerca di funzionali migliori si basa molto su intuizioni matematiche e fisiche, e si determinano procedendo di fatto per tentativi. Infatti, non è detto che approssimazioni che rispettino le condizioni fisiche al contorno del vero potenziale funzionino meglio di approssimazioni che non le rispettano:

²⁰ In linea di principio il formalismo KS è adatto per la descrizione di ogni tipo di molecola. Tuttavia, i funzionali approssimati utilizzati non offrono la flessibilità necessaria per la corretta descrizione della situazione open-shell: si preferisce, quindi, l'impiego di funzionali che dipendono esplicitamente dallo spin elettronico, capaci di fornire energie corrette qualitativamente ma densità errate. Tali funzionali sono indicati come *local spin-density functional* (LSDA).

²¹ La risoluzione delle equazioni KS parallela quella HF con la sola differenza consistente nel calcolo dell'integrale di scambio e correlazione. Nella pratica, dunque, si utilizza la stessa procedura e gli stessi set di base anche se questi sono stati concepiti ed ottimizzati per i metodi basati sulla funzione d'onda. I risultati dei calcoli DFT, infatti, specie per quanto concerne energie e geometrie di equilibrio, sono piuttosto insensibili riguardo al modo in cui i coefficienti e gli esponenti delle contrazioni sono state determinati; anche l'utilizzo degli pseudopotenziali comporta errori dello stesso ordine di grandezza di quelli introdotti nei calcoli HF.

il test più stringente per un nuovo funzionale si basa sull'accordo con accurati dati di riferimento sperimentali (energie di atomizzazione, di ionizzazione, di reazione, parametri strutturali, e così via).

La più semplice delle approssimazioni, nota come *local density approximation* (LDA), consiste nell'assumere che il funzionale di scambio e correlazione sia la somma integrale di contributi derivanti da ciascun punto nello spazio e che la densità di energia di scambio e correlazione ϵ_{xc} dipenda solo dalla densità elettronica²² e sia in ogni punto dello spazio uguale alla densità di energia di scambio e correlazione di un gas uniforme di elettroni di densità elettronica pari a quella del punto considerato, purché la densità vari in maniera sufficientemente lenta.

$$E_{xc}^{LDA} \rho_0 = \int \rho(r) \epsilon_{xc}(\rho(r)) dr$$

Questa assunzione consente di esprimere il funzionale E_{xc}^{LDA} con grandissima accuratezza poiché il gas uniforme di elettroni è l'unico sistema per cui sia nota la formulazione di tale funzionale.

In particolare la densità di energia di scambio e correlazione ϵ_{xc} può essere scomposta in un contributo legato allo scambio ϵ_x di cui si conosce l'esatta formula ed un contributo dovuto alla correlazione ϵ_c per la quale sono state proposte diverse espressioni analitiche.

In linea di principio è possibile combinare una qualsiasi formulazione del funzionale di scambio con una qualsiasi formulazione del funzionale di correlazione, ma solo poche combinazioni sono realmente utilizzate.

Il grande successo della LDA, legato principalmente al rispetto dei vincoli fisici richiesti dalla forma del potenziale reale, non è sufficiente per la maggior parte delle applicazioni in chimica. La LDA, infatti, sottostima il potenziale di scambio e sovrastima quello di correlazione portando a valori per le energie di legame che sono sovrastimati per di più di 30 kcal/mol.

È necessario, pertanto, ricorrere a delle approssimazioni più accurate, ammettendo una dipendenza della densità dell'energia di scambio ϵ_x e della densità di energia di correlazione non solo dalla densità elettronica ma anche dal suo gradiente [30-38]. Si ottiene così la cosiddetta *generalized gradient approximation* o GGA. Anche in questo caso, tuttavia,

²² Se si fa dipendere ϵ_{xc} non dalla densità elettronica totale ma dalle densità di spin si ottiene la cosiddetta *local spin-density approximation* o LSD.

l'accuratezza non è sufficiente per la corretta descrizione della chimica delle molecole, essendo l'errore medio di 5-7 kcal/mol.

La principale fonte di errore residua è legata all'espressione del funzionale di scambio e, in particolar modo, alla mancata correzione dell'autointerazione.

Nello schema HF questo problema è corretto dal termine di scambio che può essere calcolato esattamente. L'utilizzo del termine di scambio HF in luogo del funzionale di scambio DFT, tuttavia, non porta ai risultati sperati principalmente perché il funzionale di scambio DFT tiene conto anche della correlazione non-dinamica²³ che è invece trascurata nello schema HF. In altre parole, l'approssimazione usata nell'espressione del funzionale è locale, mentre lo scambio HF è non-locale e $E_{xc} \rho_0$ è relativamente localizzato.

Diventa naturale quindi mischiare $E_{xc}^{GGA} \rho_0$ con una certa quantità, espressa per mezzo di un parametro empirico, di scambio HF al fine di ottenere una migliore approssimazione al funzionale esatto [39-41]: si ottengono così i cosiddetti *funzionale ibridi* che consentono di ottenere misure accurate per la maggior parte delle applicazioni chimiche (2 kcal/mol).

La ricerca di nuove e sempre più adeguate approssimazioni ha portato, infine, alla formulazione di espressioni per i funzionali di scambio e di correlazione che tengano conto anche della densità di energia cinetica e del Laplaciano della densità. Questi costituiscono i cosiddetti funzionali *meta-GGA* [42-47] o *meta-ibridi GGA* (nel caso in cui contengono anche una certa quantità di scambio HF) [48].

Negli studi riportati nel presente lavoro di tesi, sono stati impiegati un funzionale ibrido, il B3LYP [41], ed un funzionale meta-ibrido, l'MPW1B1K [48].

L'espressione analitica per il funzionale B3LYP è:

$$E_{xc}^{B3LYP} = 1 - a E_x^{Slater-Dirac} + a E_x^{HF} + b E_x^{B88} + c E_c^{LYP} + (1 - c) E_c^{VWN}$$

dove: $E_x^{Slater-Dirac}$ è il funzionale di scambio per il gas elettronico uniforme [49,50]

E_x^{HF} è il termine di scambio cosiddetto "esatto", cioè calcolato con l'espressione HF utilizzando gli orbitali KS

E_x^{B88} è la correzione al gradiente per il funzionale di scambio [30]

E_c^{LYP} è il termine relativo al funzionale di correlazione derivato da Lee, Yang e Parr [51] a partire dall'espressione dell'energia di correlazione di Colle e Salvetti [52]

E_c^{VWN} è il funzionale di Vosko, Wilk e Nusair [53]. Poiché il funzionale di correlazione LYP contiene un termine *locale* ed uno *non-locale* di non facile

²³ La correlazione dinamica o a corto raggio è, invece, descritta da funzionale di correlazione.

separazione, il funzionale VWN è impiegato al solo fine di modificare il peso della componente locale

a, b e c sono tre parametri empirici fittati per riprodurre le energie di atomizzazione e ionizzazione e le affinità elettroniche del database G1 [54,55] e valgono, rispettivamente, 0.20, 0.72 e 0.81 [39].

Il B3LYP è uno dei funzionali più diffusi grazie all'accuratezza dimostrata con diversi database per quanto riguarda sia la termochimica che le geometrie molecolari [29,56].

Tuttavia, tale funzionale spesso tende a sottostimare le barriere energetiche per reazioni che presentano trasferimenti protonici o di atomi di idrogeno e a sovrastimare tutte le altre. In aggiunta, nel caso di complessi di metalli di transizione anche l'energetica è meno buona, presentando errori dell'ordine delle 3-5 kcal/mol [57,58].

Il funzionale MPWB1K è, invece, un funzionale piuttosto recente esplicitamente progettato per la cinetica di reazione [48].

La sua forma analitica è:

$$E_{xc}^{MPWB1K} = 1 - a E_x^{Slater-Dirac} + E_x^{mPW} + aE_x^{HF} + E_c^{B95}$$

dove: $E_x^{Slater-Dirac}$ e E_x^{HF} assumono lo stesso significato dato loro in precedenza

E_x^{mPW} è la correzione al gradiente per il funzionale di scambio PW91 modificato da Adamo e Barone [59]

E_c^{B95} è il funzionale di correlazione completo, includente cioè il termine locale, la correzione al gradiente e la dipendenza dalla densità di energia cinetica [47]

a è, infine, un parametro empirico che esprime la percentuale di scambio HF e fittato per riprodurre l'altezza delle barriere energetiche [48]; il suo valore è pari a 0.44

Il funzionale MPWB1K, in aggiunta alla capacità di determinare le barriere energetiche con maggiore precisione, sembra particolarmente adatto per la trattazione delle interazioni deboli [45], laddove il funzionale B3LYP mostra un'accuratezza molto minore [29,58].

Anche se il funzionale B3LYP è ormai superato per molte applicazioni dai funzionali di nuova generazione, esso resta ancora oggi uno dei più diffusi, soprattutto alla luce delle sue prestazioni generali e a differenza dei nuovi funzionali che operano bene solo per alcune proprietà.

Bibliografia

- [1] I.N. Levine – *Quantum Chemistry* – 3rd Ed. – 1983, Allyn and Bacon Boston
- [2] A. Szabo, N.S. Ostlund – *Modern Quantum Chemistry* – 1996, Dover Publications Inc., Mineola, New York
- [3] W. Koch, M.C. Holtausen – *A Chemist's Guide to Density Functional Theory* – 2nd Ed. – 2001, Wiley-VCH Verlag GmbH, Weinheim (DFR)
- [4] C.K. Skylaris – *Quantum Chemistry Techniques for Simulation of Molecular Properties* – School of Chemistry, University of Southampton:
<http://www.soton.ac.uk/~compchem/other/CHEM3023>
- [5] E. Schrödinger – *Ann. Physik.* – 1926, 79, 361
- [6] M. Born, J. Oppenheimer – *Ann. Physik.* – 1927, 84, 457
- [7] D. Dill – *Molecular Structure: Separating Electronic and Nuclear Motion. Notes on Quantum Mechanics* –
<http://quantum.bu.edu/notes/QuantumMechanics/MolecularStructure.pdf>
- [8] V. Pelmenchikov – *Theoretical Modeling of Enzyme Catalysis with focus on radical chemistry* – 2005, Universitetservice US AB, Stockholm
- [9] N.J.B. Green – *Quantum Mechanics 1: Foundations* – 2002, Oxford University Press, New York
- [10] D. Hartree – *Proc. Cam. Phil. Soc.* – 1928, 24, 89
- [11] D. Hartree – *Proc. Cam. Phil. Soc.* – 1928, 24, 111
- [12] D. Hartree – *Proc. Cam. Phil. Soc.* – 1928, 24, 426
- [13] W. Pauli – *Z. Physik.* – 1925, 31, 965
- [14] V. Fock – *Z. Physik.* – 1930, 61, 126
- [15] J. Slater – *Phys. Rev.* – 1929, 34, 1293
- [16] J. Slater – *Phys. Rev.* – 1930, 35, 509
- [17] C.C.J. Roothaan – *Rev. Mod. Phys.* – 1951, 23, 69
- [18] R.G. Parr, W. Yang – *Density Functional Theory of Atoms and Molecules* – 1989, Oxford University Press, New York
- [19] M. Dolg – *Effective core potentials in Modern Methods and Algorithms of Quantum Chemistry* – 2nd Ed. – 2000, J. Grotendorst ed.
- [20] L. Thomas – *Proc. Cam. Phil. Soc.* – 1927, 23, 542
- [21] E. Fermi – *Rend. Accad. Lincei* – 1927, 23, 542
- [22] E. Fermi – *Z. Physik.* – 1928, 48, 73
- [23] E. Fermi – *Rend. Accad. Lincei* – 1928, 7, 342

- [24] P. Hohenberg, W. Kohn – *Phys. Rev. B* – 1964, 136, 864
- [25] K. Burke – *The ABC of DFT* – 2007,
<http://chem.ps.uci.edu/~kieron/dft/book>
- [26] W. Kohn, L.J. Sham – *Phys. Rev. A* – 1965, 136, 1133
- [27] J.P. Perdew, A. Ruzsinszky, L.A. Constantin, J. Sun, G.I. Csonka – *J. Chem. Theory Comput.* – 2009, 5, 902
- [28] W. Kohn, A.D. Becke, R.G. Parr – *J. Phys. Chem.* – 1996, 100, 12974
- [29] S.F. Sousa, P.A. Fernandes, M.J. Ramos – *J. Phys. Chem. A* – 2007, 111, 1043
- [30] A.D. Becke – *Phys. Rev. A* – 1988, 38, 3098
- [31] A.D. Becke – *J. Chem. Phys.* – 1992, 96, 2155
- [32] A.D. Becke – *J. Chem. Phys.* – 1992, 97, 9173
- [33] J.P. Perdew, Y. Wang – *Phys. Rev. B* – 1992, 45, 13244
- [34] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiolhais – *Phys. Rev. B* – 1992, 46, 6671
- [35] J.P. Perdew, Y. Wang – *Phys. Rev. B* – 1986, 33, 8800
- [36] J.P. Perdew, K. Burke, Y. Wang – *Phys. Rev. B.* – 1996, 54, 16543
- [37] J.P. Perdew, K. Burke, M. Ernzerhof – *Phys. Rev. Lett.* – 1996, 77, 3865
- [38] J.P. Perdew, K. Burke, M. Ernzerhof – *Phys. Rev. Lett.* – 1997, 78, 1396
- [39] A.D. Becke – *J. Chem. Phys.* – 1993, 98, 5648
- [40] J.P. Perdew, M. Ernzerhof, K. Burke – *J. Chem. Phys.* – 1996, 105, 9982
- [41] P.J. Stephens, J.F. Devlin, C.F. Chabalowski, M.J. Frisch – *J. Phys. Chem.* – 1994, 98, 11623
- [42] J.P. Perdew, S. Kurth, A. Zupan, P. Blaha – *Phys. Rev. Lett.* – 1999, 82, 2544
- [43] J.P. Perdew, S. Kurth, A. Zupan, P. Blaha – *Phys. Rev. Lett.* – 1999, 82, 5179
- [44] J. Tao, J.P. Perdew, V.N. Storerov, G.E. Scuseria – *Phys. Rev. Lett.* – 2003, 91, 146401-1
- [45] J.B. Krieger, J. Chen, G.J. Iafrate, A. Savin – *Electron Correl. Mater. Prop.* – 1999, 463
- [46] T. Van Voorhis, G.E. Scuseria – *J. Chem. Phys.* – 1998, 109, 400
- [47] A.D. Becke – *J. Chem. Phys.* – 1996, 104, 1040
- [48] Y. Zhao, D.G. Truhlar – *J. Phys. Chem. A* – 2004, 108, 6908
- [49] P. Dirac – *Proc. Cam. Phil. Soc.* – 1930, 26, 376
- [50] J.C. Slater – *The Self-Consistent Field for Molecular and Solids: Quantum Theory of Molecular and Solids* – Vol.4 – 1974, McGraw-Hill, New York
- [51] C. Lee, W. Yang, R.G. Parr – *Phys. Rev. B* – 1988, 37, 785

- [52] R. Colle, D. Salvetti – *Theor. Chim. Acta* – 1975, 37, 329
- [53] S. Vosko, L. Wilk, M. Nusair – *Can. J. Phys.* – 1980, 58, 1200
- [54] J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari, L.A. Curtiss – *J. Chem. Phys.* – 1989, 90, 5622
- [55] L.A. Curtiss, C. Jones, G.W. Trucks, K. Raghavachari, J.A. Pople – *J. Chem. Phys.* – 1990, 93, 2537
- [56] L.A. Curtiss, P.C. Redfern, K. Raghavachari, J.A. Pople – *J. Chem. Phys.* – 1998, 109, 42
- [57] J.L. Durant – *Chem. Phys. Lett.* – 1996, 256, 595
- [58] P.E.M. Siegbahn – *J. Biol. Inorg. Chem.* – 2006, 11, 695
- [59] C. Adamo, V. Barone – *J. Chem. Phys.* – 1998, 108, 664

CAPITOLO III: L'IMPLEMENTAZIONE PRATICA

3.1 La necessità di modelli

Il grande sviluppo dei sistemi informatici, sia in termini di potenza che di incremento di nuovi e più veloci algoritmi per il calcolo, ha consentito di trattare in maniera quantomeccanica sistemi molecolari sempre più grandi e complessi con crescente affidabilità e precisione [1-12].

I sistemi biologici, tuttavia, risultano ancora non ancora pienamente trattabili a livello quantomeccanico a causa della loro elevata dimensionalità.

Le proteine, e quindi anche gli enzimi, ad esempio, sono composti da un numero di atomi che varia da qualche migliaio a qualche decina di migliaia: decisamente troppi anche per il più potente calcolatore fino ad oggi costruito²⁴. Pertanto, per poter trattare questi sistemi è necessario ricorrere a modelli approssimati.

Una scelta bilanciata tra il livello di teoria e le dimensioni del modello è una condizione imprescindibile per poter ottenere una riproduzione accurata dei dati sperimentali e validare così i risultati ottenuti [4].

Diversi riscontri indicano che non sempre la scelta di un modello molto grande (come quelli usati ad esempio in una trattazione QM/MM²⁵), risulta strategicamente migliore rispetto alla scelta di un modello ridotto (gas phase cluster o approccio supramolecolare) che rappresenti il più significativamente possibile, in relazione alle risorse di calcolo disponibili, il sistema di interesse [10]. Spesso, infatti, i risultati ottenuti nei due casi, nel limite di incertezza del metodo, sono paragonabili tra loro. L'utilizzo di modelli ridotti, tuttavia, consente di effettuare calcoli in tempi più contenuti e quindi di testare con maggiore attenzione differenti meccanismi. In aggiunta, i modelli ridotti permettono di monitorare più facilmente i cambi strutturali che intervengono durante la reazione e di escludere, di conseguenza, eventuali equilibri conformazionali tra minimi locali non correlati alla coordinata di reazione che condurrebbero ad artefatti [10-12].

La trattazione quantomeccanica di modelli ridotti, d'altra parte, presenta il limite di trascurare le interazioni a medio e lungo raggio tra i residui amminoacidici che compongono il modello ed il resto dell'enzima e tra questi ed il solvente. Bisogna però sottolineare che,

²⁴ Basti ricordare che la DFT scala come N^3 , HF come N^4 e i metodi post-HF anche come N^6 , dove N indica la dimensionalità del sistema. Bisogna, peraltro, registrare che molti sforzi sono stati fatti e si continuano a fare per raggiungere la linearità.

²⁵ La trattazione QM/MM consente di trattare il processo catalitico a livello quantomeccanico tenendo conto anche dell'influenza che l'intera struttura enzimatica (trattata invece a livello di meccanica molecolare) ha su questo. Questa, tuttavia, presenta alcuni problemi irrisolti relativi soprattutto alla trattazione dell'interfaccia QM-MM con conseguente possibilità di artefatti [4]. Un possibile stratagemma per ovviare a queste difficoltà consiste nel creare una regione intermedia tra quella QM e quella MM da trattare in maniera quantomeccanica ma ad un livello di teoria più basso [5].

eccezionati alcuni casi particolari come ad esempio reazioni con trasferimento di cariche, le maggiori variazioni energetiche sono associate alla rottura e alla formazione di legami chimici, solitamente confinati a livello di sito attivo dell'enzima.

La scelta di un modello piuttosto che un altro deve avere una sua giustificazione fondata su basi razionali [10]. Nella simulazione dei sistemi enzimatici in generale, le catene peptidiche e le connessioni tra queste ed i gruppi funzionali coinvolti nella catalisi possono essere eliminati; nel caso di metalloenzimi la prima sfera di coordinazione dello ione metallico deve essere naturalmente mantenuta ma anche i residui della seconda sfera di coordinazione, nel caso formino forti legami idrogeno con qualche gruppo coinvolto nella reazione o siano elettricamente carichi, devono essere considerati.

In ogni caso, la linea guida che detta la scelta di quali gruppi funzionali debbano essere compresi nel modello è costituita, laddove esistano, dai dati sperimentali biochimici, cinetici e strutturali o, comunque, dall'intuizione chimica.

3.2 Il protocollo seguito

Il punto di partenza di ogni procedura computazionale è rappresentato dalla definizione del modello da utilizzare nel protocollo di calcolo.

Tipicamente, la costituzione del modello si basa sulla struttura NMR o, molto più frequentemente, sulla struttura cristallografica²⁶ dell'enzima in esame [1,3,7-10,12].

In genere, tali strutture sono depositate nel *protein data bank* (PDB), una banca dati che raccoglie, appunto, le strutture tridimensionali di proteine, acidi nucleici e complessi di interesse biochimico determinati principalmente per cristallografia o spettroscopia NMR.

Esse possono essere scaricate liberamente dal sito web del Research Collaboratory for Structural Bioinformatics (RCSB): <http://www.pdb.org/pdb/home/home.do>

Le strutture scaricate possono essere visualizzate con numerosi software di interfaccia grafica, fra cui VMD (Visual Molecular Dynamics) [15] e GaussView [16].

Il modello da utilizzare per i calcoli può essere ottenuto modificando il file contenente le coordinate cristallografiche attraverso una procedura in due step.

Il file è prima aperto con l'applicazione WordPad di Windows che lo visualizza come file di testo. A questo punto le coordinate cristallografiche dei residui amminoacidici che non sono

²⁶ Partire dalle coordinate cristallografiche significa assumere implicitamente che la conformazione dell'enzima non risulti alterata dal processo di ottenimento dei cristalli o che sia comunque molto simile a quella in soluzione, cosa generalmente vera ma non vera in assoluto [9]. In alcuni casi, ad esempio, la simulazione quantomeccanica ha portato a reinterpretare i dati cristallografici [13,14].

coinvolti nella reazione sono semplicemente cancellate dal file: si ottiene, così, un file molto più leggero, più facile da visualizzare e da modificare.

Il file “leggero” ottenuto può essere visualizzato con GaussView che consente di raffinare il modello eliminando gli atomi superstiti non necessari e di saturare con atomi di idrogeno quelli che sono rimasti a valenza aperta ovvero che erano legati agli atomi eliminati.

Gaussview, infatti, consente di costruire strutture ex novo e di modificare i parametri geometrici quali distanze e angoli fra gli atomi o gruppi di atomi.

Questa possibilità di modifica della geometria della struttura del modello è di fondamentale importanza non solo nell’ottenimento del modello stesso da sottoporre ad ottimizzazione, quanto soprattutto perché consente di predisporre delle ipotesi di lavoro da ottimizzare riguardanti tutti gli step che costituiscono il meccanismo di reazione che si vuole investigare.

In generale, infatti, intermedi e stati di transizione differiscono solo per il posizionamento di gruppi funzionali, di protoni, di molecole di acqua, per l’ambiente di coordinazione o lo stato di ossidazione di ioni metallici [9].

Tutte le geometrie delle strutture ottimizzate riportate in questa tesi e i relativi calcoli di frequenza sono state ottenuti con il codice di calcolo Gaussian03 [17] utilizzando, nell’ambito della teoria del funzionale della densità, il funzionale B3LYP descritto nel capitolo precedente.

Per i calcoli relativi all’ Insulyn-degrading Enzyme (IDE) e alla Mitochondrial Processing Peptidase (MPP) è stato utilizzato il set di base 6-31G(d,p) di qualità doppia ζ per gli atomi non direttamente coinvolti nella reazione. Invece, gli atomi appartenenti alla prima shell di coordinazione dello ione metallico sono stati descritti dal set di base 6-31+G(d,p).

Per la Cadmium Carbonic Anhydrase (CDCA1) grazie al ridotto numeri di atomi che compongono il modello studiato, è stato usato il set di base 6-311G(d,p) di qualità tripla ζ .

Tutti e tre gli enzimi studiati contengono uno ione metallico. Per il trattamento di questi ioni, in tutti e tre i casi, si è fatto ricorso all’*effective core potential* (ECP) di Stuttgart-Dresden (SDD) [18-19].

Nella costruzione dei modelli di partenza sono solitamente conservate solo le catene laterali dei residui amminoacidici simulate con frammenti di molecole o ioni che conservano le funzionalità in esse presenti. I dettagli di queste simulazioni verranno forniti in seguito caso per caso.

Seguendo una procedura molto comune in tale tipi di studi, gli atomi al livello dei quali avviene il taglio, in origine direttamente legati alla catena polipeptidica, vengono mantenuti fissi nelle loro posizioni cristallografiche durante le ottimizzazioni. Questo accorgimento consente di riprodurre in una certa misura la rigidità posseduta dall’enzima (stato entatico) e

di evitare espansioni irrealistiche del modello. In altre parole, la struttura completa del sito catalitico esercita delle restrizioni steriche sul moto dei residui che la compongono e di cui si tiene in questo modo conto [12].

Tuttavia, se il numero di atomi bloccati è elevato o se essi sono troppo vicini al sito di reazione, il modello può risultare eccessivamente irrigidito, portando a valori errati per le energie calcolate. In questi casi allontanare dal centro reattivo il punto di taglio del residuo dalla catena polipeptidica, ovvero aumentare le dimensioni del modello in modo da mantenere più distante l'atomo bloccato, può alleviare questa fonte di errore [12].

In ogni caso, le conseguenze del mantenere bloccati questi atomi durante le ottimizzazioni hanno un impatto generalmente trascurabile sulla termodinamica della reazione, purché il modello utilizzato non sia eccessivamente piccolo.

I risultati delle ottimizzazioni, comunque, possono essere visualizzati, ad esempio, con GaussView: si possono così misurare con buona precisione i parametri geometrici delle strutture e visualizzare i relativi moti vibrazionali, avendo avuto cura di effettuare il corrispondente calcolo delle frequenze.

In aggiunta, risulta molto utile, specialmente per la preparazione di figure e di presentazioni power point, anche il software XYZ-Viewer²⁷ che, oltre a consentire di salvare le immagini con specificati i parametri geometrici, consente di salvare anche delle animazioni in formato gif dei moti vibrazionali di interesse.

Nella stessa linea si collocano altri software come l'ISIS Draw che consente di creare formule di struttura ex novo o a partire da strutture già presenti nelle librerie del programma ed utilizzato per la creazione delle ipotesi degli schemi di reazione.

Infine, software come Origin [20] o Excel possono essere impiegati per la creazione dei grafici relativi ai pathways energetici dei meccanismi vagliati o alla variazione di grandezze come ad esempio cariche elettriche o lunghezze di legame al variare delle strutture.

3.3 I calcoli in solvente

I risultati ottenuti dalle ottimizzazioni delle strutture basate sul modello dell'enzima costituiscono solo un'indicazione dell'andamento reale del pathway energetico relativo al meccanismo preso in esame poiché l'uso stesso di un modello rappresenta un'approssimazione. Nella realtà l'enzima è presente in toto e il processo catalitico avviene in un ambiente fisiologico diverso da quello gassoso per la presenza di un solvente.

²⁷ XYZ-viewer è un'applicazione basata sulla piattaforma indipendente JAVA e sviluppato da Sven de Marothy per la manipolazione e la visualizzazione di modelli molecolari.

Il solvente può influenzare le strutture degli intermedi di reazione, l'energia di attivazione, l'energia di legame e di conseguenza gli spettri rotazionali, vibrazionali ed elettronici.

Il solvente, in una reazione chimica simulata, può essere preso in considerazione in maniera esplicita o implicita [21-25].

Gli approcci più comuni per la sua trattazione in chimica computazionale sono quelli della supermolecola, i modelli del continuo, la meccanica molecolare (MM) e varie costruzioni ibride (supermolecola + modello del continuo, QM (quantomeccanica) + MM, QM + MM + modello del continuo).

Nell'approccio della supermolecola, attorno al soluto vengono esplicitamente considerate un certo numero di molecole di solvente ed il sistema così ottenuto viene trattato a livello quantomeccanico. I risultati migliorano via via che il numero di molecole di solvente considerato cresce con conseguente aumento del costo computazionale.

Nei metodi basati sulla MM, semplici campi di forza ottenuti classicamente permettono di includere nel calcolo un numero elevato di molecole di solvente. Tuttavia, la semplicità di questo approccio non dà una descrizione adeguata di molti processi come ad esempio la rottura dei legami nelle reazioni chimiche.

I modelli del continuo rappresentano un vero successo. Il campo di applicabilità è però limitato. Comunque, laddove questi metodi possono essere applicati, danno una stima degli effetti del solvente abbastanza buona e con costi computazionali convenienti. Il solvente è descritto come un mezzo uniforme caratterizzato da qualche proprietà macroscopica.

I metodi ibridi QM/MM sono gli approcci più recenti e promettenti. Tuttavia, non è facile applicarli e sono computazionalmente costosi.

Nel presente lavoro di tesi, il ruolo del solvente è stato investigato utilizzando i metodi del continuo noti come *polarizable continuum model* o PCM [26-29], e la sua variante *conductor polarizable continuum model* o CPCM [30-33], come implementati nei codici di calcolo Gaussian03 [17] e Gaussian09 [34].

Secondo tali modelli, il solvente è trattato come un continuo polarizzabile avente una data costante dielettrica [25] ed il soluto è posto dentro una cavità all'interno del solvente.

Molti sono i parametri che influenzano questo modello ma il più importante è senza dubbio è la definizione della superficie della cavità che ospita il soluto [21-22].

Esistono diverse modalità di costruzione di questa superficie: per i modelli presi in esame, cioè il PCM e il CPCM, essa è ottenuta come involucro di sfere centrate su atomi o gruppi di atomi.

Nel Gaussian03 l'impostazione di default è quella che utilizza il modello topologico UAO basato sui raggi ricavati dal campo di forza UFF. Secondo tale modello gli atomi di idrogeno sono inglobati negli atomi pesanti a cui sono legati.

Nel Gaussian09, invece, l'impostazione di default è quella basata sui raggi ricavati dal campo di forza UFF con sfere esplicitate per ogni atomo presente, idrogeni compresi.

Esistono, tuttavia, anche altre opzioni disponibili che consentono di utilizzare valori differenti per i raggi delle sfere²⁸ o di modificarne il posizionamento.

Il valore dei raggi utilizzati per la costruzione delle sfere è in realtà ottenuto moltiplicando il valore dei raggi del modello topologico utilizzato per un fattore di scala α , settato a 1.0 per Gaussian03 e a 1.1 per Gaussian09. Questo offre un'ulteriore possibilità di intervento per la modifica della superficie della cavità.

Esistono anche diverse modalità per la costruzione della cavità: nel Gaussian03 il default la superficie è ottenuta secondo lo schema SES (*solvent excluding surface*) secondo il quale il solvente è impossibilitato a penetrare nella cavità la cui superficie è resa più regolare grazie all'aggiunta di sfere supplementari (Fig. 4); nel Gaussian09, invece, il default è costituito dalla *Van der Waals surface* che impiega i raggi di Van der Waals scalati e senza l'aggiunta delle sfere supplementari.

È possibile, inoltre, costruire superfici anche secondo altri schemi, ad esempio aggiungendo il raggio del solvente²⁹ ai raggi non scalati di atomi e gruppi (SAS: *solvent accesible surface*; Fig. 4) o costruire superfici di isodensità in maniera statica (IPCM) o autoconsistente (SCIPCM).

Ad ogni modo, una volta che è stata definita la cavità, la sua superficie è mappata in maniera più dettagliata attraverso piccole regioni denominate *tessere*, ognuna delle quali è caratterizzata dalla propria area e dall'aver un vettore elettrostatico normale alla superficie passante per il suo centro [32].

Tali tessere sono ottenute proiettando sulla superficie di ogni sfera le facce di un poliedro, il pentakisdodecaedro (avente 60 facce), inscritto in essa. Le tessere derivanti da sfere differenti e che si sovrappongono completamente sono scartate, mentre quelle che si sovrappongono solo in parte sono rimpiazzate da poligoni adeguati.

²⁸ Ad esempio si può utilizzare il modello topologico UAHF che fa uso di raggi ottimizzati per il livello di teoria HF con basi 6-31G(d) o il modello topologico UAKS che fa uso, invece, di raggi ottimizzati il funzionale di scambio e correlazione PBE0 con basi 6-31G(d) o scegliere tra i raggi di Van der Waals o i raggi di Bondi o, ancora, i raggi di Pauling, ecc.

²⁹ Le molecole di solvente sono assimilate a delle sfere aventi come raggio il raggio di Van der Waals: lo scorrimento di una sfera lungo la superficie della cavità definisce una nuova superficie data dalle posizioni assunte dal centro della sfera e oltre la quale esso non può andare.

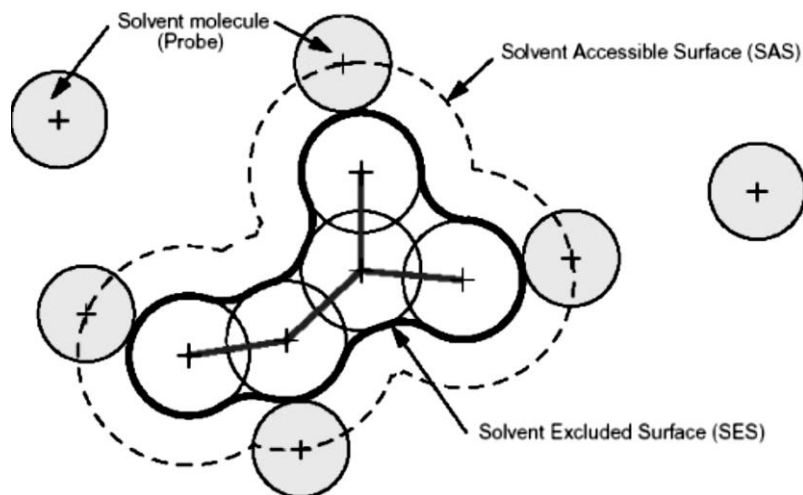


Fig. 4: Rappresentazione della superficie della cavità nel solvente secondo i modelli SES e SAS [25].

Il vettore elettrostatico associato ad ogni tessera, invece, deriva dalla polarizzazione indotta nelle molecole del solvente dai dipoli molecolari del soluto: il potenziale elettrico generato dal continuum in risposta alla polarizzazione (*reaction field*) è uguale ad una distribuzione continua di carica fittizia, distribuita sulla superficie della cavità molecolare [24]. Nella realtà, la distribuzione continua è approssimata da una distribuzione discreta di cariche puntuali poste sulla superficie della cavità: ad ogni tessera è associata una carica e , quindi, un vettore elettrostatico.

Nel modello CPCM, in aggiunta, le cariche sono determinate in modo che il potenziale totale si cancelli sulla superficie.

L'energia di solvatazione può essere approssimata come somma di contributi [22,24]:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{cav} + \Delta G_{dis} + \Delta G_{rep}$$

Il primo termine, ΔG_{el} , è la componente elettrostatica dovuta all'interazione soluto-solvente calcolata usando il metodo *self-consistent reaction field* o SCRF.

Il campo elettrico indotto nel solvente, infatti, interagisce con la densità elettronica del solvente che così cambia. Ciò equivale ad introdurre un termine perturbativo che tenga conto delle interazioni elettrostatiche con il solvente nell'Hamiltoniano che descrive lo stato elettronico del soluto

Poiché esiste una dipendenza reciproca tra densità elettronica del soluto e campo elettrico indotto del solvente, il calcolo di ΔG_{el} può avvenire solo attraverso una procedura iterativa.

In breve, le cariche indotte dalla densità elettronica del soluto sono calcolate in maniera autoconsistente ed il potenziale ad esse associato utilizzato per la risoluzione dell'Hamiltoniano perturbato: si ottiene così un nuovo valore per la densità elettronica da utilizzare per ricalcolare le cariche indotte ed il potenziale ad esse associato e così via fino all'ottenimento della piena convergenza [24,25].

Gli altri termini dell'equazione relativa a ΔG_{solv} rappresentano i contributi non-elettrostatici. In particolare: ΔG_{cav} è il termine di cavitazione, ovvero il lavoro che bisogna compiere per creare la cavità nel solvente; ΔG_{dis} è il termine di dispersione dovuto alle forze di dispersione di London tra soluto e sovente; ΔG_{rep} è il termine di repulsione quantomeccanica tra soluto e solvente [24].

Se i calcoli in solvente forniscono energie piuttosto differenti rispetto a quelle ottenute da calcoli in fase gassosa oppure se tali energie dipendessero significativamente dal valore di ϵ scelto, allora i calcoli effettuati in fase gassosa risultano poco attendibili [11,12].

In questo caso bisognerebbe testare un modello più grande: l'inaccuratezza insita nel modello PCM diminuisce all'aumentare delle dimensioni del modello [12].

Inoltre, se i siti di taglio del modello sono lontani dal centro reattivo, l'effetto della polarizzazione su di esso è in gran parte già considerato nella trattazione quantomeccanica [12].

La bontà dei calcoli teorici effettuati è fornita dalla loro capacità di riprodurre i dati sperimentali, in particolare della costante cinetica k_{cat} per quel che riguarda le reazioni enzimatiche.

Le barriere energetiche ottenute da calcoli in solvente possono essere messe in relazione con le costanti cinetiche della reazione attraverso la relazione

$$k = \frac{K_B T}{h} e^{-\frac{\Delta G^\ddagger}{RT}}$$

dove: k è la costante cinetica

K_B è la costante di Boltzmann

h è la costante di Planck

T è la temperatura assoluta

R è la costante dei gas

ΔG^\ddagger è la differenza di energia libera di Gibbs tra lo stato di transizione ed i reagenti

Data la dipendenza esponenziale tra la barriera energetica calcolata e il valore della costante cinetica ad essa associato, piccoli errori nel calcolo delle prime si traducono in grandi errori nei valori delle seconde (per un errore di 1.4 kcal/mol sulla barriera energetica si compie un errore di un ordine di grandezza nella k).

In pratica, in genere le indicazioni che si possono trarre dai calcoli hanno una valenza solo qualitativa.

Bibliografia

- [1] D.J. Tantillo, J. Chen, K.N. Houk – *Curr. Opinion Chem. Biol.* – 1998, 2, 743
- [2] R.A. Friesner, M.D. Beachy – *Curr. Opinion. Struct. Biol.* – 1998, 8, 257
- [3] T.C. Bruice, K. Kahn - *Curr. Opinion Chem. Biol.* – 2000, 4, 540
- [4] V. Gogonea, D. Suarez, A. van der Vaart, K.M. Merz jr. – *Curr. Opinion. Struct. Biol.* – 2001, 11, 217
- [5] M.J. Field – *J. Comput. Chem.* – 2002, 23, 48
- [6] A. Warshel – *Annu. Rev. Biophys. Biomol. Struct.* – 2003, 32, 425
- [7] T. Lovell, F. Himo, W. Han, L. Noodleman – *Coord. Chem. Rev.* – 2003, 238-239, 211
- [8] A.J. Mulholland – *Drug Discoveri Today: Biosilico* – 2005, 10, 1393
- [9] M. Sommerhalter, R.L. Lieberman, A.C. Rosenzweig – *Inorg. Chem.* – 2005, 44, 770
- [10] P.E.M. Siegbahn – *J. Biol. Inorg. Chem.* – 2006, 11, 695
- [11] M.J. Ramos, P.A. Fernandes – *Acc. Chem. Res.* – 2008, 41, 689
- [12] K.H. Hopmann, F. Himo – *Quantum Chemical Modeling of Enzymatic Reactions – Applications to Epoxide-Transforming Enzymes in Comprehensive Natural Products II. Chemistry and Biology* – 2010, L. Mander & H.-W. Liu Eds, Elsevier, Oxford
- [13] K.R. Brown, K. Djinovic-Carugo, T. Haltia, I. Cabrito, M. Saraste, J.J.G. Moura, I. Moura, M. Tegoni, C. Cambillau – *J. Biol. Chem.* – 2000, 275, 41133
- [14] B.D. Dunietz, M.D. Beachy, Y. Cao, D.A. Whittington, S.J. Lippard, R.A. Friesner – *J. Am. Chem. Soc.* – 2000, 122, 2828
- [15] W. Humphrey, A. Dalke, K. Schulten – VMD – Visual Molecular Dynamics – *J. Molec. Graphics* – 1996, 14, 33
- [16] GaussView 3.07, Gaussian, Inc., Carnegie Office Park – Building 6, Pittsburg PA 15106 USA
- [17] Gaussian 03, Revision C.02, M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, 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, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas,

- D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, and J.A. Pople, Gaussian, Inc., Wallingford CT, 2004
- [18] M. Dolg, U. Wedig, H. Stoll, H. Preuss – *J. Chem. Phys.* – 1987, 86, 866
- [19] D. Andrae, U. Haeussermann, M. Dolg, H. Stoll, H. Preuss – *Theor. Chem. Acc.* – 1990, 77, 123
- [20] Origin: OriginLab – Northampton (MA), USA
- [21] W.-H. Mu, G.A. Chasse, D.C. Fang – *Int. J. Quantum Chem.* – 2008, 108, 1422
- [22] Y. Takano, K.N. Houk – *J. Chem. Theory Comput.* – 2005, 1, 70
- [23] M. Wang, C.F. Wong, J. Liu, P. Zhang – *Chem. Phys. Lett.* – 2007, 442, 464
- [24] I.N. Levine – *Quantum Chemistry* – 5th Ed. – 2000, Prentice Hall, Upple Side River, NJ, USA
- [25] J. Tomasi, B. Mennucci, R. Cammi – *Chem Rev.* – 2005, 105, 2999
- [26] S. Miertuš, E. Scrocco, J. Tomasi – *J. Chem. Phys.* – 1981, 55, 117
- [27] R. Cammi, J. Tomasi – *J. Chem. Phys.* – 1994, 100, 7495
- [28] M. Cossi, G. Scalmani, N. Rega, V. Barone – *J. Chem. Phys.* – 2002, 117, 43
- [29] V. Barone, R. Improta, N. Rega – *Theor. Chem. Acc.* – 2004, 111, 237
- [30] A. Klamt, G. Schüürmann – *J. Chem. Soc. Perkin Trans. 2* – 1993, 799
- [31] J. Andzelm, C. Kölmel, A. Klamt – *J. Chem. Phys.* – 1995, 103, 9312
- [32] V. Barone, M. Cossi – *J. Chem. Phys. A* – 1998, 102, 1995
- [33] M. Cossi, N. Rega, G. Scalmani, V. Barone – *J. Comput. Chem.* – 2003, 24, 669
- [34] Gaussian 09, Revision A.2, 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.

CAPITOLO IV: GLI STUDI EFFETTUATI

4.1 Le Inverzincine

Molti e importanti processi biologici, fra tutti il metabolismo, che hanno luogo all'interno degli organismi viventi coinvolgono l'idrolisi di legami peptidici [1].

Rivestono, pertanto, grande importanza le *peptidasi* ovvero gli enzimi che catalizzano tale reazione. In quanto promotori di una reazione di idrolisi, tali enzimi costituiscono una sottoclasse delle *idrolasi*. Inoltre, a seconda se idrolizzano legami peptidici terminali o interni sono denominati, rispettivamente, esopeptidasi o endopeptidasi.

La reazione di idrolisi di un legame peptidico presenta degli ostacoli, dovuti alla stabilità del legame ammidico e alla scarsa nucleofilicità dell'acqua: entrambi questi ostacoli possono essere superati dagli enzimi grazie ad una catalisi elettrofila, ricorrendo, cioè, all'ausilio di uno ione metallico.

Lo zinco, in particolare, dopo il ferro risulta essere il metallo di transizione più abbondante negli organismi viventi [2]. Una spiegazione di ciò può risiedere nella grande flessibilità che i suoi ioni mostrano nell'adottare alcune geometrie di coordinazione e nella stabilità chimica rispetto alle reazioni di ossidazione.

Le metallo peptidasi, come vengono comunemente indicati tali enzimi, possono essere classificati sulla base di similarità strutturali, come ad esempio il sito di coordinazione del metallo.

In particolare, un ampio numero di enzimi contenenti ioni Zn^{2+} sono accomunati dal condividere per il sito di coordinazione dello ione metallico la sequenza HExxH, dove H è l'amminoacido istidina, E è l'amminoacido acido glutammico e x indica un amminoacido generico. Questi enzimi vengono indicati con il nome di *zincine*.

Circa 30 anni fa, tuttavia, fu isolata e descritta [3-5] una metalloendopeptidasi di *E. coli* contenente ioni zinco, la pitrilisina, che presentava, per il sito di coordinazione del metallo, un motivo perfettamente speculare rispetto quello già noto delle zincine: HxxEH anziché HExxH.

Successivamente molte altre metalloendopeptidasi che presentavano questa caratteristica furono isolate in piante, animali, lieviti e batteri e caratterizzate. A tali enzimi fu attribuito il nome di *inverzincine* [6,7].

Sia nelle zincine che nelle inverzincine il sito di coordinazione del metallo è localizzato in una α -elica destrorsa e di lunghezza simile, ma nelle inverzincine tale struttura ha un'orientazione opposta rispetto alla struttura delle zincine.

A dispetto della differenza di struttura primaria le zincine e le inverzincine mostrano una disposizione spaziale simile attorno al sito catalitico. Le zincine, tuttavia, mostrano una certa

variabilità per quanto concerne il terzo ligando dello ione Zn^{2+} che può essere un residuo di acido glutammico, di acido aspartico, di istidina o di cisteina, mentre per le inverzincine questo è costituito esclusivamente da un residuo di acido glutammico [8].

Inoltre, zincine e inverzincine mostrano specificità e risposte agli inibitori differenti; in particolare le zincine mostrano la caratteristica unica di essere inibiti dagli inibitori delle cisteinoproteasi [9].

Le inverzincine costituiscono la superfamiglia M16 che conta tre distinte sottofamiglie: la M16A, la M16B e la M16C. L'insulin-degrading Enzyme (IDE) e la Mitochondrial Processing Peptidase (MPP), oggetto di studio della presente tesi, sono rappresentanti, rispettivamente, delle famiglie M16A e M16B.

4.2 L'Insulin-degrading enzyme

L'insulin-degrading enzyme o IDE (EC 3.4.24.56) è un omodimero di circa 220 kDa di massa [7,10,11] espresso principalmente nel citosol cellulare, anche se è stato ritrovato in molti organuli e strutture subcellulari, nei compartimenti extracellulari e anche in cellule che risultano non sensibili all'insulina³⁰, uno dei suoi substrati naturali [12-16]. Più specificatamente l'IDE è stato trovato sulla superficie delle cellule endoteliali vascolari del cervello umano [17]. I suoi substrati mostrano poca similarità tra loro in quanto a struttura primaria. L'IDE, tuttavia, non risulta essere una peptidasi generale poiché processa numerosi substrati, ma discrimina tra substrati simili: ad esempio processa l'insulina [18-20] ma non il suo precursore pro-insulina [7,8,21-23].

L'idrolisi dei substrati operata dall'IDE avviene in maniera casuale, anche se con una preferenza per siti idrofobici ingombri. Sembra plausibile, quindi, l'ipotesi che l'IDE riconosca elementi delle strutture secondaria e terziaria dei substrati [16].

Dettagliate analisi della struttura dell'IDE, culminate con la determinazione cristallografica della struttura dell'enzima umano libero [11] e di suoi complessi con alcuni substrati [24], hanno evidenziato la presenza di due distinti domini, uno C-terminale ed uno N-terminale, legati tra loro da un loop di 28 a.a. e da interazioni come legami ad idrogeno [10,11,22,24,25].

Entrambi i domini sono necessari per la catalisi. In particolare, il dominio C-terminale è responsabile della oligomerizzazione dell'IDE, della sua regolazione allosterica [10,26-30] nonché di importanti interazioni con il substrato [23,24]. Il dominio N-terminale, invece,

³⁰ Questo ha fatto ipotizzare che IDE sia coinvolto anche in processi differenti dall'idrolisi peptidica, come ad esempio la differenziazione cellulare.

ospita il sito di coordinazione dello ione Zn^{2+} , costituito dalle catene laterali dei residui di His-108, His-112 e Glu-189 [8,21,31], oltre al residuo di Glu-111, fondamentale per la catalisi [22]. I due domini dell'IDE formano una cavità piuttosto ampia ($\sim 16.000 \text{ \AA}^3$) capace di ospitare substrati lunghi fino a circa 70 amminoacidi [23,24].

La parete formata dal dominio C-terminale risulta essere largamente positiva mentre quella del dominio N-terminale per lo più neutra o di carica negativa. Le dimensioni e la distribuzione di carica, dunque, possono costituire due discriminanti per la specificità dei substrati mostrata dall'IDE.

Il dominio N-terminale contiene, sempre all'interno della camera catalitica, un exosito per l'ancoraggio del substrato: questo non richiede sequenze specifiche nel substrato poiché avviene tramite interazioni tra il foglietto β dell'estremità N-terminale del substrato ed un foglietto β presente all'interno della camera catalitica.

Le strutture cristallografiche dei complessi enzima-substrato hanno evidenziato come il substrato legato all'enzima effettui una transizione da una struttura ad α -elica o a random coil ad una struttura a foglietto β . La flessibilità dei substrati sembra, quindi, costituire un'altra discriminante per la specificità. In effetti, è stato proposto che l'IDE riconosca ed idrolizzi i suoi substrati in base alla loro capacità di formare depositi amiloidi³¹ [16].

Substrati per l'IDE sono, oltre all'insulina, anche l'amilina [32], il peptide natriuretico atriale [33], il glucagone [34], i peptidi ABri e ADan [35], i peptidi β -amiloidi 1-40 e 1-42 [16,23,36-45]: tutti composti in grado di formare depositi amiloidi³².

4.2.1 Il meccanismo catalitico

Le strutture cristallografiche dell'enzima ottenute in assenza [11] ed in presenza [24] di substrato hanno rivelato da un lato come l'IDE prediliga assumere una conformazione chiusa (IDE^c) e dall'altro l'assenza di una qualsiasi apertura che garantisca al substrato l'accesso nella cavità catalitica.

Nella conformazione chiusa l'IDE è in grado di condurre la sua azione catalitica ma non di rilasciare i prodotti della reazione né di garantire l'accesso del substrato al sito di reazione. Di conseguenza, perché possa svolgere il suo compito, l'IDE deve subire un cambiamento

³¹ I peptidi amiloidi si caratterizzano per la lenta formazione di depositi insolubili formati per aggregazione degli stessi peptidi in seguito ad una transizione della struttura a foglietto β [16].

³² Depositi amiloidi sono stati riscontrati in pazienti diabetici nei punti dove si iniettavano l'insulina; l'amilina è responsabile della formazione delle isole amiloidi pancreatiche in soggetti malati di diabete mellito di tipo II o insulino-resistente; ABri e ADan sono responsabili di alcune forme di demenza ereditaria; i peptidi β -amiloidi 1-40 e 1-42 sono invece associati con il Morbo di Alzheimer. Esistono, inoltre, numerose evidenze, anche di tipo genetico, per una correlazione tra diabete mellito di tipo II, Morbo di Alzheimer e IDE.

conformazionale in modo da trovarsi nello stato aperto (IDE^o) che garantisce l'accesso alla cavità interna [11] (Fig. 5).

Il meccanismo con cui avviene questo cambiamento conformazionale e cosa lo inneschi sono a tutt'oggi sconosciuti. Questo comportamento, però, ha suggerito per l'IDE il soprannome di enzima pac-man.

In ultima analisi, il ciclo può così riassumersi: inizialmente l'IDE è in una conformazione chiusa; in seguito a perturbazioni termodinamiche o all'interazione con qualche effettore l'IDE assume la conformazione aperta grazie alla quale può inglobare il substrato, dopodiché riassume la conformazione chiusa, nella quale il substrato subisce il cambio conformazionale a foglietto β e, successivamente, l'idrolisi; una seconda transizione dell'IDE verso la conformazione aperta permette il rilascio dei prodotti: a questo punto l'enzima è pronto a ricominciare un nuovo ciclo oppure a transire verso lo stato iniziale inattivo.

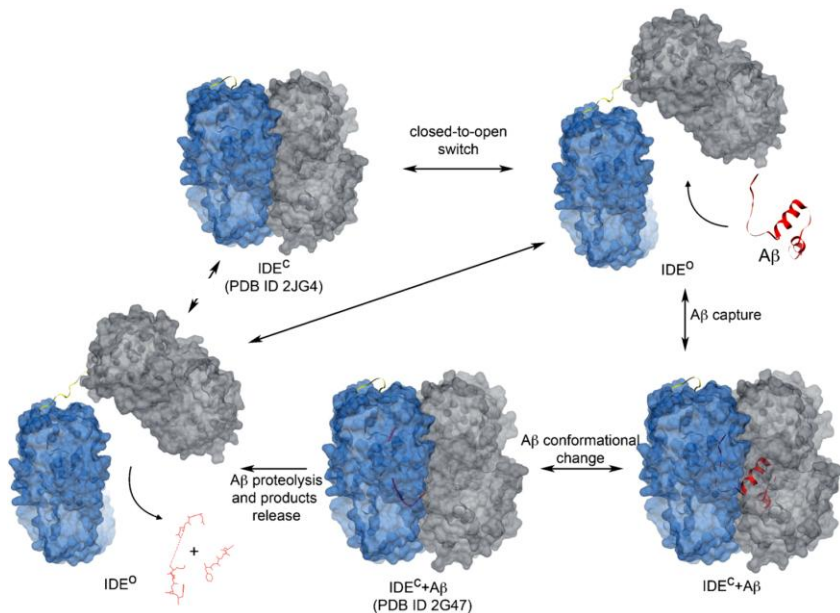


Fig. 5: Equilibrio conformazionale dell'IDE tra la forma chiusa IDE^c e la forma aperta IDE^o [22]. Per i dettagli si veda il testo

Il meccanismo esatto con cui avviene l'idrolisi ad opera dell'IDE non è noto, tuttavia sulla base delle similarità con altre endopeptidasi contenenti ioni Zn^{2+} ne è stato proposto uno (Schema 1) secondo il quale il residuo di Glu-111 agisce da base generale deprotonando una molecola di acqua coordinata allo ione Zn^{2+} : lo ione idrossido così generato conduce un attacco nucleofilo sul carbonile producendo un intermedio tetraedrico.

In un secondo tempo il Glu-111, attraverso un meccanismo concertato, trasferisce l'idrogeno acquisito dall'acqua al gruppo uscente, estraendo nel contempo l'altro atomo di idrogeno dallo ione idrossido, provocando la rottura del legame peptidico.

4.2.2 I modelli studiati

Il meccanismo di idrolisi ipotizzato per l'IDE è stato oggetto di uno studio computazionale (vedi Articolo I).

Sulla base della struttura cristallografica dell'enzima con la catena β dell'insulina (PDB code: 2G54) e delle indicazioni sperimentali sono stati costruiti 3 diversi modelli del sito catalitico di dimensioni crescenti, denominati rispettivamente modello A, B e C.

Il modello A (Fig. 6), il più piccolo dei tre con i suoi 111 atomi, è composto dai residui His-108, His-112, Glu-189 ed una molecola di acqua come ligandi dello ione Zn^{2+} , dai residui Arg-824 e Tyr-831, formanti legami ad idrogeno con il substrato costituito dai residui 15-17, rispettivamente Leu, Tyr e Leu, della catena β dell'insulina.

Uno dei siti di idrolisi del substrato è localizzato tra il residuo di Tyr-16 e il residuo di Leu-17.

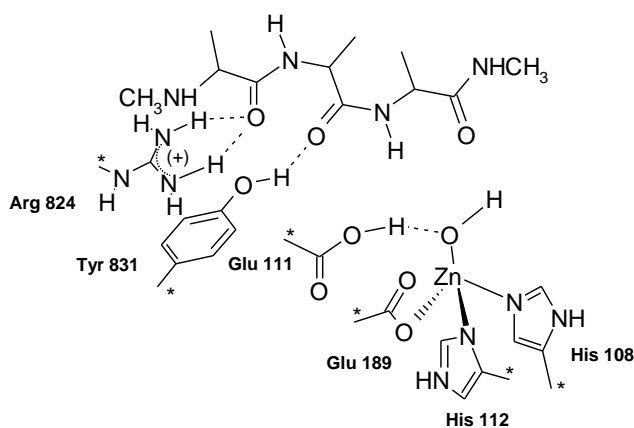
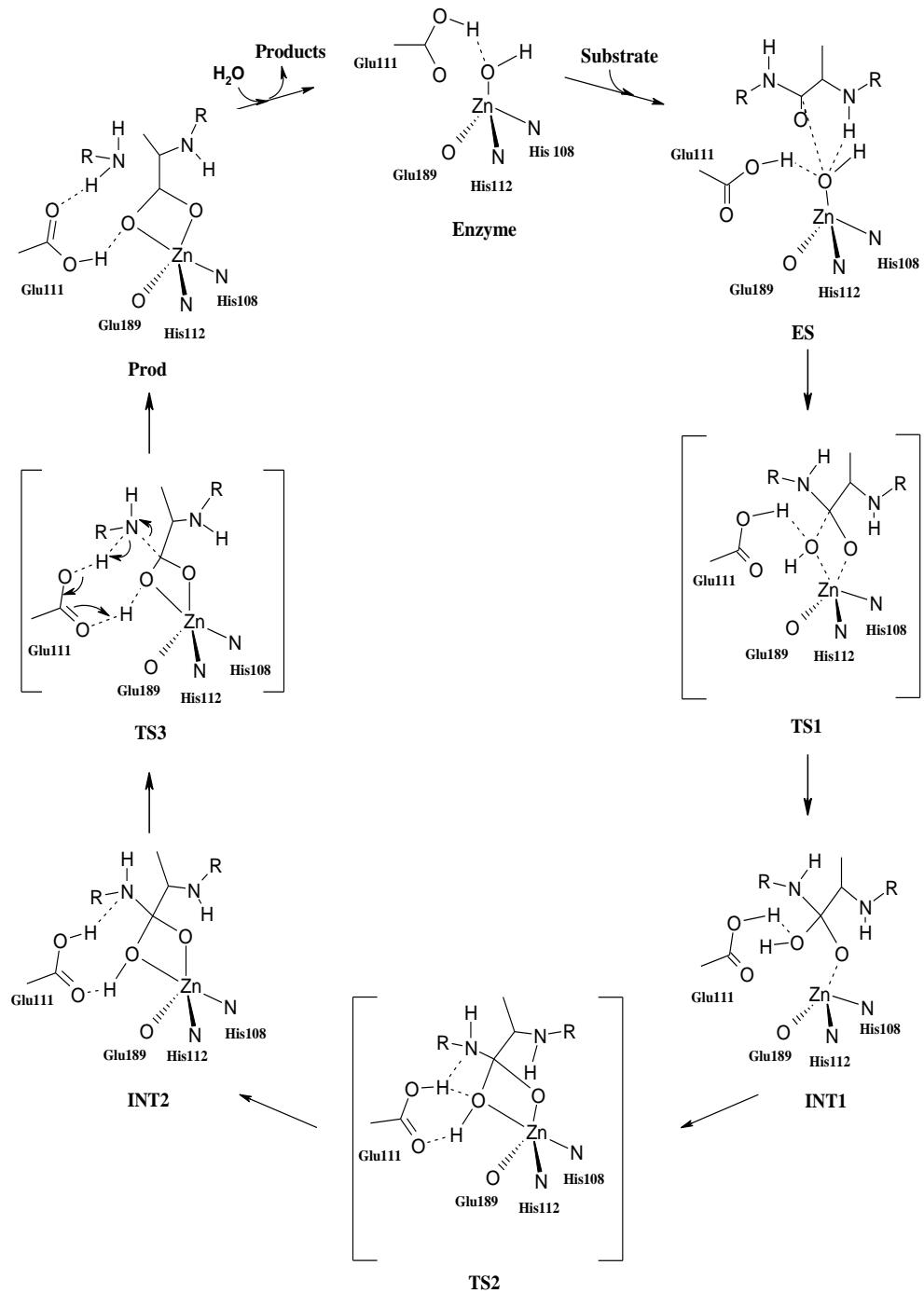


Fig. 6: Rappresentazione del modello A utilizzato nello studio del meccanismo catalitico dell'IDE. Gli atomi contrassegnati con l'asterisco sono quelli mantenuti bloccati durante le ottimizzazioni

Nel modello A, che ha carica totale +1 e molteplicità 1, i residui amminoacidici di His sono stati simulati con metil-imidazoli, i Glu con ioni acetato, l'Arg con lo ione N-metilguanidinio, la Tyr con p-metilfenolo. Tutti gli atomi di carbonio dei gruppi metile di tali residui sono stati tenuti fissi nelle loro posizioni cristallografiche durante le ottimizzazioni. Il

substrato è stato approssimato con il peptide trialanina la cui estremità C-terminale è sotto forma di N-metilammide, mentre quella N-terminale è N-metil sostituta.



Schema 1: Ciclo catalitico proposto per la Insulin-degrading Enzyme

Il modello B (Fig. 7), invece, consta di 130 atomi, ha carica nulla e molteplicità 1. Rispetto al modello A sono inclusi il Glu-182 e il legame peptidico compreso tra i residui di Thr-220 e di Gly-221, appartenenti alla seconda sfera di coordinazione dello ione Zn^{2+} . In particolare, da dati di letteratura, il Glu-182 risulta essere un residuo conservato e appare avere un ruolo di regolatore dell'attività enzimatica; inoltre esso va ad interagire con legame idrogeno con il residuo di Arg-824. L'interazione con la catena principale compresa tra Thr-220 e Gly-221, invece, pare non essere mai stata evidenziata in precedenza. Anche in questo caso i residui sono stati simulati da uno ione acetato e da N-metilacetammide, rispettivamente. In aggiunta, oltre agli atomi già bloccati nel modello A, anche gli atomi di carbonio dei gruppi metili dei residui aggiunti sono stati mantenuti fermi.

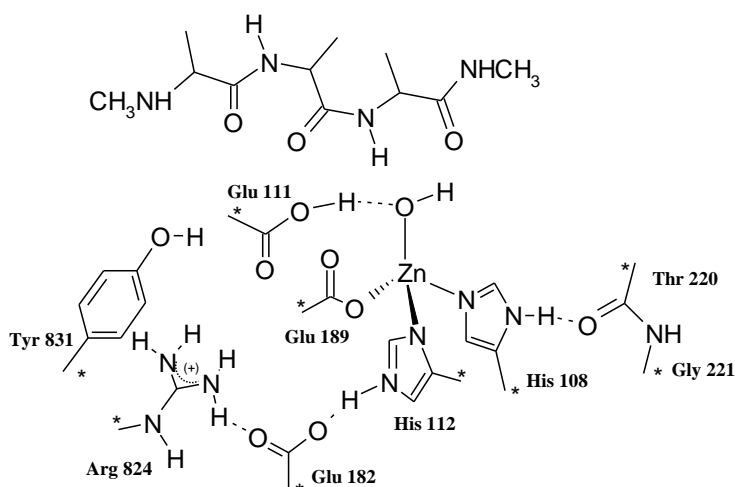


Fig.7: Rappresentazione del modello B utilizzato nello studio del meccanismo catalitico dell'IDE. Gli atomi contrassegnati dall'asterisco sono quelli mantenuti bloccati durante le ottimizzazioni

Il modello C (Fig. 8), infine, conta 159 atomi e, come il modello B, ha carica nulla e molteplicità 1. Tale modello si differenzia dal precedente in quanto sono state reintrodotte le catene laterali del substrato reale.

4.2.3 I risultati ottenuti per l'IDE

I calcoli effettuati hanno avuto come punto di partenza una struttura in cui la molecola di acqua è già deprotonata dallo ione acetato che simula il residuo Glu-111.

Lo studio condotto sul modello A è stato rapidamente abbandonato. La sola quantificazione della barriera energetica³³ per l'attacco nucleofilo ad opera dello ione OH legato allo Zn²⁺ sul carbonile (oltre 30 kcal/mol) ha infatti dimostrato l'inadeguatezza del modello.

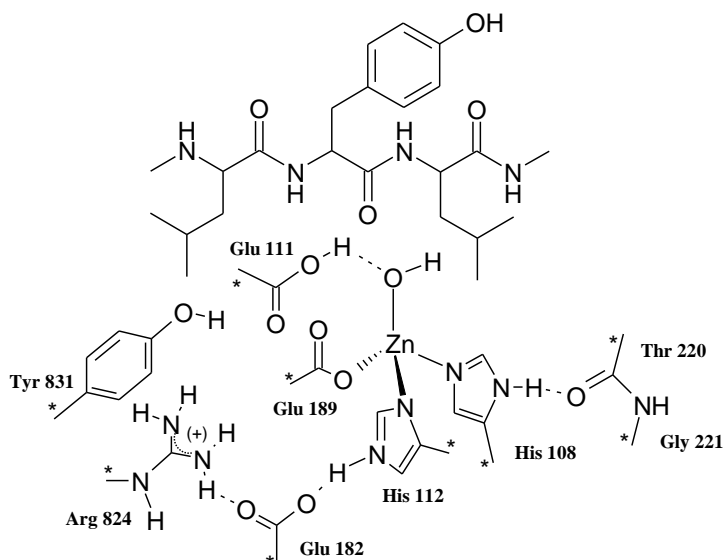


Fig. 8: Rappresentazione del modello B utilizzato nello studio del meccanismo catalitico dell'IDE. Gli atomi contrassegnati dall'asterisco sono quelli mantenuti bloccati durante le ottimizzazioni

L'inclusione nel modello della seconda sfera di coordinazione dello ione metallico (modello B) ha condotto a risultati soddisfacenti: il rate limiting step presenta una barriera di poco meno di 16 kcal/mol (Fig. 9).

In breve, lo ione OH legato allo Zn²⁺ conduce un attacco nucleofilo sul carbonile generando un primo stato di transizione (TS1) che evolve in un primo intermedio (INT1) in cui il residuo Glu-111 interagisce tramite un legame a ponte idrogeno con il gruppo OH. A questo intermedio segue un secondo stato di transizione (TS2) in cui il suddetto Glu-111 compie un movimento di pseudo rotazione per andare a formare un nuovo legame idrogeno con il gruppo uscente e ristabilire quello con il gruppo OH (INT2). Questo moto richiede circa 3.5 kcal/mol. All'INT2 fa seguito un terzo stato di transizione (TS3) nel quale si osserva l'avvenuta cessione del protone al gruppo uscente e uno shift del protone legato al gruppo OH del substrato al Glu-111 con contemporaneo stretching del legame C-N. Da un punto di vista energetico, il TS3 si posiziona la di sotto del precedente intermedio di poco più

³³ Tutti i valori energetici riportati si riferiscono a calcoli single-point effettuati in solvente utilizzando il modello CPCM con $\epsilon=4$ su strutture ottimizzate in fase gassosa

di 3 kcal/mol indicando come questa seconda parte del cammino energetico risulti piuttosto piatta se non proprio discendente.

L'ultimo intermedio (PROD) presenta l'estremità C-terminale del substrato, ormai lontana da quella N-terminale, ancora legata allo ione metallico. Il residuo Glu-111 rimane in uno stato protonato. Il ripristino dell'enzima attivo richiede che qualche altro residuo presente acquisti il protone da Glu-111. Questo aspetto, tuttavia, non è stato investigato poiché lo scopo dello studio è stato quello di elucidare il meccanismo dell'idrolisi.

In tutti i casi, la reazione risulta esotermica per circa 13 kcal/mol.

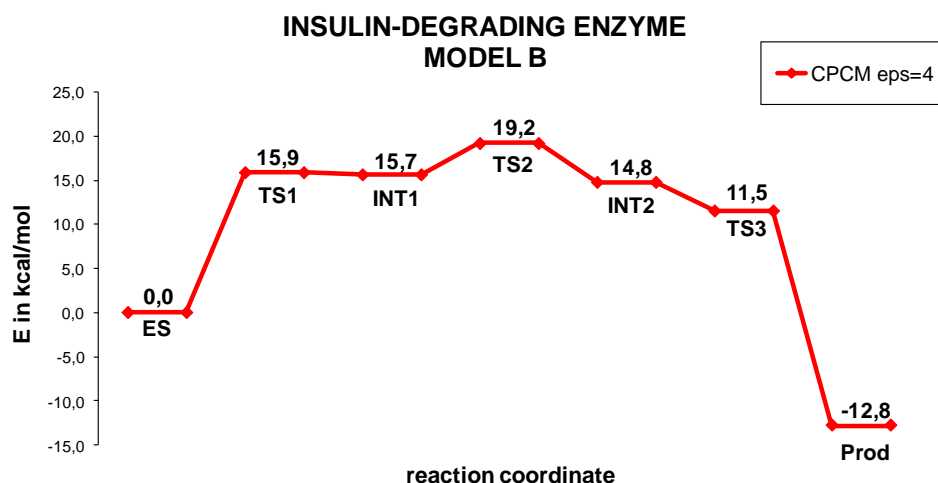


Fig. 9: Profilo energetico per la reazione di idrolisi effettuata utilizzando il modello B

La descrizione dettagliata del meccanismo catalitico e le figure relative alle geometrie di equilibrio possono essere trovate nell'articolo I.

Il meccanismo ottenuto con il modello C è risultato molto simile a quello ottenuto con il modello B sia in termini di geometrie di equilibrio che di profilo energetico (Fig. 10).

In definitiva, si possono trarre le seguenti conclusioni: 1) in tutti i modelli studiati il residuo di Arg-824 serve all'ancoraggio del substrato; 2) il residuo di Tyr-831 non solo ancora il substrato ma concorre a stabilizzare la carica negativa che si genera in seguito all'attacco nucleofilo; 3) il residuo di Glu-111 agisce non solo come base generale ma anche come shuttle protonico risultando indispensabile per la catalisi; 4) l'effetto della seconda sfera di coordinazione dello ione Zn^{2+} si esplica non solo in termini elettronici ma anche in termini di carica del sistema che passa da +1 a 0; 5) l'approssimazione fatta per il substrato (modello B) risulta valida ai fini dell'energetica di reazione e della determinazione strutturale dei punti

stazionari: ciò consente di operare un risparmio sui tempi di calcolo; 6) l'enzima lavora altrettanto bene con due tripeptidi differenti, pur simili per natura idrofobica, riproducendo il dato sperimentale che vede l'IDE non richiedere specifiche sequenze di riconoscimento del substrato; 7) la reazione è, in ogni caso esotermica e lo stadio che determina la cinetica di reazione è l'attacco nucleofilo.

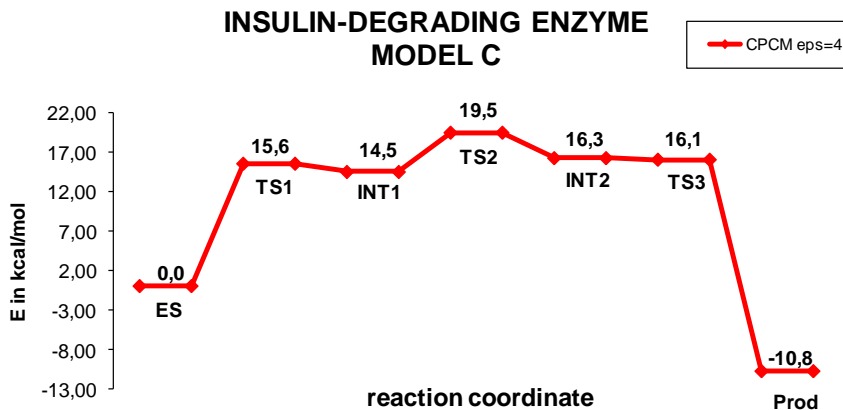


Fig. 10: Profilo energetico per la reazione di idrolisi effettuata utilizzando il modello C

4.3 La Mitochondrial Processing Peptidase: il suo ruolo nei mitocondri

I mitocondri [46,47] possono essere considerati, a tutti gli effetti, le centrali energetiche delle cellule. Essi, infatti, sono l'unico organulo cellulare che consuma ossigeno (nella cosiddetta catena respiratoria) per formare molecole di ATP ad alta energia.

Il mitocondrio, inoltre, è la sede in cui avvengono altri importanti processi biochimici quali il ciclo di Krebs, il ciclo dell'urea e il metabolismo degli acidi grassi, solo per citarne alcuni.

La loro diffusione è praticamente ubiquitaria: si ritrovano, infatti, anche se in misura differente, in tutte le cellule, sia che appartengano al regno animale, sia che appartengano al regno vegetale o a quello dei funghi o dei batteri.

I mitocondri sono costituiti da una membrana esterna, regolare e abbastanza porosa, e da una membrana interna che presenta delle profonde increspature (le cosiddette creste) e altamente selettiva. Si vengono così a determinare tre differenti ambienti: il citosol esternamente ad essi, uno spazio intermembrana ed una soluzione acquosa interna nota come matrice.

Sebbene dotati di codice genetico e degli apparati necessari per la sintesi proteica, la maggior parte delle proteine necessarie per il funzionamento dei mitocondri sono codificate nel nucleo della cellula e sintetizzate nel citoplasma.

Queste proteine, così come accade per quelle destinate ad altri organelli cellulari, sono indirizzate ai mitocondri grazie ad un segnale peptidico, noto come presequenza, posto all'estremità N-terminale della catena proteica e successivamente importati nella matrice mitocondriale attraverso delle traslocasi di membrana (Fig. 11) [48].

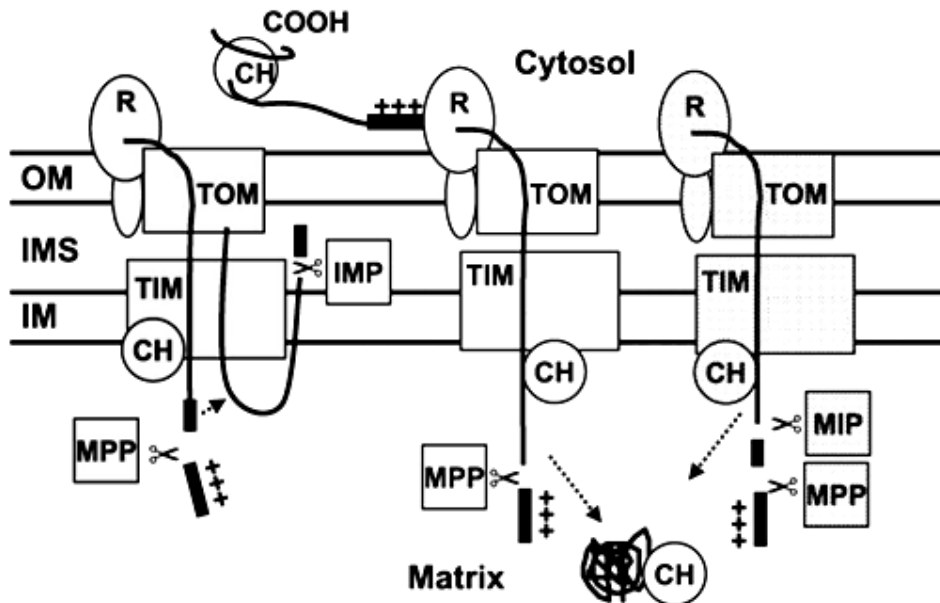


Fig. 11: Importazione e processo di proteine citoplasmatiche nei mitocondri. Le proteine sintetizzate nel citoplasma sono condotte ai mitocondri da alcune chaperonine (CH) e, attraverso i recettori (R) della membrana esterna (OM) e alle traslocasi della membrana esterna (TOM) e di quella interna (TIM), introdotte nella matrice dove la MPP taglia via la presequenza. Le proteine ottenute possono essere indirizzate attraverso la membrana interna (IM) verso lo spazio intermembrana (IMS), dove la peptidasi della membrana interna (MIP) le porta a maturazione, oppure trovarsi nella forma attiva o in una forma intermedia maturata poi dalla peptidasi mitocondriale intermedia (MIP). (adattato da rif. 47)

Giunte nella matrice, mentre iniziano a ripiegarsi [49], le proteine provenienti dal citoplasma sono idrolizzate ad opera della Mitochondrial Processing Peptidase o MPP (EC 3.4.24.64) che taglia via la presequenza [7,48,50].

Molte malattie dell'uomo sono correlate a mancanza di proteine mitocondriali codificate dal nucleo della cellula [48,51]. Un'alterazione della funzionalità della MPP porta ad un accumulo dei peptidi precursori nella matrice che ha come conseguenza un rallentamento della crescita cellulare e, infine, la morte [52-55].

La necessità di eliminare la presequenza dalle catene peptidiche è richiesta per un corretto folding della proteina: in caso contrario gli enzimi risultano meno attivi e meno stabili termodinamicamente [48,52].

La MPP, dunque, svolge un ruolo fondamentale per la vita, idrolizzando praticamente tutte le proteine mitocondriali provenienti dal citoplasma³⁴.

4.3.1 Le caratteristiche strutturali della MPP

La MPP è stata isolata e caratterizzata in numerosi organismi [53-65]. Essa risulta composta da due differenti unità, indicate come α e β , la cui localizzazione differisce a secondo dell'organismo considerato: nelle piante fanno parte del complesso bc1 della catena respiratoria; nel lievito una subunità è localizzata nella matrice mentre l'altra continua a far parte del suddetto complesso; negli animali e nei funghi entrambe le subunità sono presenti nella matrice, mentre nel complesso bc1 sono presenti delle subunità omologhe che hanno perso la capacità di idrolisi³⁵.

Ogni subunità è sintetizzata nel citoplasma come precursore più grande ed importata nei mitocondri dove in maniera autocatalitica o, più probabilmente, grazie alla presenza di MPP preesistente viene attivata [54-56,58,60].

Le subunità, in aggiunta, risultano essere altamente omologhe tra di loro [56,58,60] e con i domini N-terminale e C-terminale dell'IDE [24].

Ognuna di essa, presa singolarmente, non possiede attività di peptidasi: perché l'enzima sia attivo sono, infatti, richieste entrambe le subunità [55,58,65-67].

La MPP, quindi, è un eterodimero avente una massa molare di circa 100-110 kDa.

La subunità α , in particolare, ha una massa molare di circa 52-57 kDa. Il suo ruolo sembra essere quello di riconoscere le presequenze, specie quelle più lunghe, attraverso molteplici siti di legame carichi negativamente presenti nel suo dominio C-terminale [56,68-72] e tramite una regione altamente conservata ricca di residui di glicina e di istidina [7,48,50,71].

La subunità β , invece, ha una massa molare di circa 50-55 kDa e contiene la caratteristica sequenza HXXEH tipica delle inverzincine [71].

Per svolgere la sua attività catalitica la MPP richiede, infatti, uno ione Zn^{2+} in mancanza del quale l'enzima risulta inattivato. Tuttavia, altri metalli come Mn^{2+} , Co^{2+} , Mg^{2+} , Fe^{2+} , Ni^{2+}

³⁴ In realtà, molte delle proteine destinate alla membrana esterna e le proteine carriers della membrana interna non presentano presequenze da eliminare [48].

³⁵ In un caso, tuttavia, ciò è causato dalla presenza di una presequenza che funge inibitore [65].

[56-58,60,65,71] sembrano svolgere un ruolo catalitico anche se per questo non c'è evidenza diretta.

La subunità β , inoltre, contiene numerosi residui amminoacidici acidi conservati che risultano importanti per il riconoscimento del substrato [72-75].

Nel complesso, quindi, le due subunità formano una cavità catalitica di carica negativa [7,68,71]. In aggiunta, le pareti della cavità contengono delle strutture β -sheet che costituiscono un'impalcatura per il legame del substrato tramite legami idrogeno [71,75].

Questo ben si accorda con il fatto che le presequenze presentano numerosi residui sia basici (carichi positivamente) che idrossilati, ma molto raramente acidi (carichi negativamente).

Le sole interazioni elettrostatiche, tuttavia, non spiegano la specificità dell'azione catalitica della MPP: si stima che nei mammiferi la MPP processi ad un singolo e ben definito sito oltre 1000 presequenze che hanno poco in comune in termini di lunghezza (variano da 8 a circa 69 amminoacidi) e di struttura primaria.

La MPP, pertanto, deve riconoscere elementi strutturali [56,57]. In particolare, risultano importanti la presenza di residui basici distali e di uno prossimale al sito di taglio, collegati da una regione di lunghezza variabile ma piuttosto flessibile [56,61,74-78].

Molte pre-proteine mostrano, per il sito di taglio, un motivo costituito dalla sequenza $Rx\downarrow\Phi\Psi\Psi$, dove \downarrow indica il legame idrolizzato, R un residuo di arginina, Φ un residuo idrofobico ingombrato e Ψ un residuo idrofilico [61,62,76,77]; in alcuni casi il residuo di arginina è localizzato tre amminoacidi prima del sito di taglio, mentre in altri casi è posizionato ben dieci residui distante da questo³⁶; in circa un terzo dei casi, infine, esso manca del tutto, indicando come la presenza di tale amminoacido possa essere importante ma non sufficiente a guidare l'azione della MPP [48,80].

Le estese interazioni tra substrato ed enzima, infatti, consentono di processare con alta specificità anche quelle pre-proteine che non soddisfano appieno tutti i requisiti richiesti per la catalisi [75].

4.3.2 Il meccanismo ipotizzato

La specificità ed il meccanismo di azione della MPP rimangono enigmatici, soprattutto alla luce della variabilità delle caratteristiche strutturali e di sequenza mostrate dai suoi substrati.

³⁶ In questi casi, in realtà, la MPP idrolizza la pre-proteina ad una forma intermedia che è successivamente trasformata nella proteina matura dalla peptidasi mitocondriale intermedia (MIP) che taglia via un octapeptide dall'estremità N-terminale [79].

La determinazione delle strutture della MPP libera e della MPP legata ad un substrato [71] ha permesso di effettuare un paragone con le strutture di altri metallo-enzimi afferenti sia alla famiglia delle inverzincine che ad altre famiglie e di confrontarne le ipotesi di meccanismo. In particolare, le strutture cristallografiche mostrano la presenza di una molecola di acqua come quarto ligando dello ione Zn^{2+} .

Tale molecola, inoltre, si colloca alla giusta distanza per la formazione di un legame idrogeno con il residuo di Glu-73 che così può agire come base generale.

È così plausibile ipotizzare che il sito attivo contenga in realtà uno ione idrossido coordinato al sito metallico (Schema 2) o, quantomeno, una molecola di acqua fortemente attivata.

Il meccanismo ipotizzato, pertanto, richiama da vicino quello proposto in precedenza per l'IDE: lo ione idrossido conduce un attacco nucleofilo sul carbonile del substrato formando un intermedio tetraedrico che evolve poi in secondo intermedio in cui il residuo di Glu-73 ha trasferito un protone al gruppo uscente; la deprotonazione del gruppo OH legato al carbonile sempre ad opera del Glu-73 porta infine alla rottura del legame C-N ed alla formazione dei prodotti.

4.3.3 Il modello utilizzato

L'ipotesi di meccanismo catalitico formulata per la Mitochondrial Processing Peptidase è vagliata attraverso uno studio computazionale.

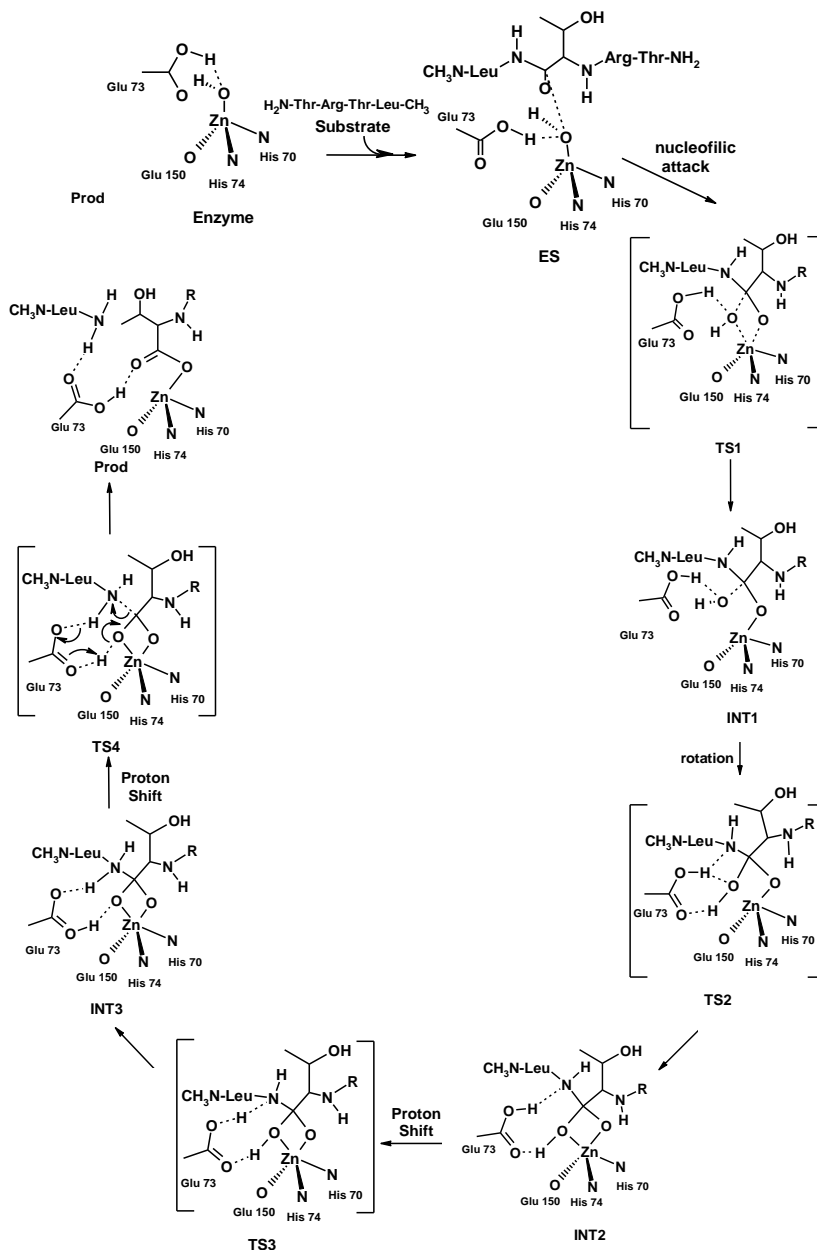
Nell'effettuare lo studio si è fatto ricorso ad un modello di sito catalitico ricavato dalla struttura cristallografica del complesso enzima-substrato dove, più specificatamente, l'enzima è quello del lievito e il substrato è costituito dal peptide 2-25 della subunità IV della citocromo ossidasi (COX IV; PDB code: 1HR8).

Nella costruzione del modello si è tenuto conto delle indicazioni sperimentali ricavate dalla determinazione della suddetta struttura e dai numerosi studi anch'essi sperimentali che hanno interessato la MPP.

Il modello in questione (Fig. 12) è stato ottenuto prendendo in considerazione la sola subunità β in quanto è questa che contiene il sito di coordinazione dello ione Zn^{2+} , mentre la subunità α non interagisce con la presequenza della COX IV.

Nello specifico, nel modello sono stati inclusi: due residui di istidina (His-70 e His-74), uno di glutammato (Glu-150) ed una molecola di acqua in quanto ligandi dello ione Zn^{2+} , un residuo di glutammato (Glu-73) che agisce come base generale, un altro (Glu-160) che forma

un ponte salino con il residuo di arginina del substrato ed un terzo glutammato³⁷ (Glu-143) interagente con His-74. Completano il modello la porzione di un foglietto β posizionato immediatamente sopra in sito catalitico ed interagente con il substrato nonché il substrato stesso.



Schema 2: Ciclo catalitico proposto per la Mitochondrial Processing Peptidase

³⁷ Dati di letteratura indicano come il Glu-143 eserciti un'influenza sulla capacità di His-74 di ritenere lo ione metallico e di riflesso sulla catalisi [72].

In particolare, i residui del foglietto β considerati sono quelli di una treonina (Thr-103), di una tirosina (Tyr-102) e parte di una alanina (Ala-101). L'estremità C-terminale di Thr-103 è in forma ammidica, così da simulare il legame peptidico, mentre la Tyr-102 è modellata con un residuo di alanina e l'Ala-101 è troncata al carbonio α . Durante le ottimizzazioni l'atomo di azoto legato all'estremità C-terminale della Thr-103 e il carbonio metilico rappresentante il carbonio α di Ala-101 sono stati mantenuti fissi nelle loro posizioni di origine.

Il substrato, infine, è composto da un tetrapeptide formato dai residui di Thr-15, Arg-16, Thr-17 e Leu-18, essendo il legame interessato dall'idrolisi quello tra la Thr-17 e la Leu-18. L'estremità N-terminale del substrato è in forma amminica libera mentre quella C-terminale è in forma metil-ammidica. Non sono state introdotte approssimazioni per quanto riguarda le catene laterali del tetrapeptide.

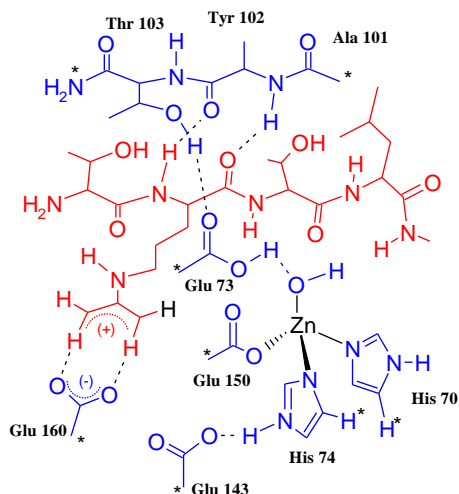


Fig. 12: Modello del sito catalitico della MPP. In rosso il substrato, in blu l'enzima. Gli atomi contrassegnati dall'asterisco sono quelli mantenuti bloccati durante le ottimizzazioni.

Tutti i residui di glutammato considerati nel modello, e cioè Glu-73, Glu-143, Glu-150, Glu-160, sono stati simulati con ioni acetato, mentre la His-70 e la His-74 con molecole di imidazolo. Il carbonio metilico degli ioni acetato e gli atomi di idrogeno degli imidazoli in posizione 4 dell'anello sono stati mantenuti fissi nelle loro posizioni cristallografiche.

In definitiva, il modello consta di 161 atomi per una carica totale pari a -1 e una molteplicità pari ad 1.

Un attento confronto tra il modello utilizzato per lo studio della MPP ed i modelli impiegati nello studio dell'IDE consente di porre in rilievo alcune differenze relativamente all'organizzazione strutturale del sito catalitico.

Nonostante la disposizione dei residui attorno allo ione metallico sia simile nei due enzimi, nel caso della MPP si nota l'assenza del residuo di tirosina o di qualche altro amminoacido che ne svolga le funzioni: stabilizzare la carica negativa che si genera al carbonile in seguito all'attacco nucleofilo operato dallo ione idrossido legato allo Zn^{2+} .

4.3.4 I risultati dello studio sulla MPP

L'ottimizzazione del complesso enzima-substrato della MPP ha permesso di metterne in rilievo alcune peculiarità strutturali quali la breve distanza, circa 1.37 Å, intercorrente tra lo ione idrossido ed il protone estratto ad opera del glutammato, distanza che fa pensare più ad una molecola di acqua attivata che ad un vero e proprio ione idrossido, e la presenza di un legame idrogeno lungo, poco più di 2,10 Å, tra l'atomo di ossigeno del carbonile e l'atomo di idrogeno del legame ammidico formato dai residui di Leu-18 e Cys-19 del substrato.

La descrizione dettagliata del meccanismo catalitico, le figure relative alle strutture del complesso enzima-substrato e dei successivi punti stazionari nonché i parametri geometrici più significativi relativi ad esse saranno presentati in un articolo che attualmente è in preparazione.

Successivamente al complesso enzima-substrato si colloca il primo stato di transizione (TS1) in cui lo ione idrossido conduce un attacco nucleofilo sul carbonile della Thr-17 del substrato che risulta essere ora coordinato allo ione metallico. Lo ione idrossido si allontana sensibilmente dallo ione Zn^{2+} (circa 0,9 Å) e dal Glu-73 (di circa 0,2 Å). La distanza tra l'ossigeno del carbonile e l'idrogeno ammidico si accorcia fino a 1.93 Å, suggerendo un possibile effetto stabilizzante dell'incipiente carica negativa che si forma e svolgendo un ruolo simile a quello della Tyr-831 nell'IDE.

Da un punto di vista energetico³⁸ il TS1 si pone 20,37 kcal/mol al di sopra del complesso enzima-substrato (Fig. 13).

Il TS1 evolve, quindi, nel primo intermedio (INT1) strutturalmente simile ad esso ma più stabile per circa 4 kcal/mol.

All'INT1 fa seguito il secondo stato di transizione (TS2) in cui il legame a idrogeno tra il Glu-73 ed il gruppo OH legato al carbonile si allenta mentre se ne cominciano a formare altri due nuovi, uno tra il gruppo OH di Glu-73 e l'atomo di azoto del gruppo uscente ed uno tra il

³⁸ Tutti i valori energetici riportati si riferiscono a calcoli single-point effettuati in solvente utilizzando il modello CPCM con $\epsilon=4$ su strutture ottimizzate in fase gassosa

gruppo OH legato al carbonile e l'atomo di ossigeno dello stesso Glu-73. Questa struttura risulta essere circa 7.7 kcal/mol al di sopra del precedente INT1.

Nell'INT2 successivo al TS2 la rete di legami a idrogeno tra il Glu-73 ed il substrato viene consolidata. L'ossigeno del carbonile, tuttavia, non interagisce più con l'idrogeno ammidico del legame tra Leu-18 e Cys-19 ma forma un nuovo legame a idrogeno con la catena laterale di Thr-17. Energeticamente questa struttura risulta più stabile del precedente TS2 di circa 13,9 kcal/mol.

Il successivo step del meccanismo prevede uno shift protonico dal Glu-73 all'atomo di azoto del gruppo uscente per formare il corrispondente intermedio (INT3), passando attraverso un terzo stato di transizione (TS3).

Ogni tentativo di determinare il TS3, tuttavia, non ha portato alla sua individuazione.

In particolare, operando una scansione sulla coordinata di reazione, ovvero lungo la distanza di legame tra l'atomo di idrogeno e l'atomo di ossigeno di Glu-73, si è trovata una struttura con un massimo di energia di appena 0,3 kcal/mol rispetto il precedente INT2.

In aggiunta, il tentativo fatto utilizzando un livello di teoria più basso (HF invece che DFT), in modo da ricavare le costanti di forza del moto molecolare da utilizzare nella successiva ottimizzazione DFT, ha consentito l'individuazione del TS3 a livello HF ma non a livello DFT.

È noto che il funzionale B3LYP, utilizzato nelle ottimizzazioni DFT, tende a sottostimare le barriere energetiche relative a trasferimenti protonici: questa osservazione, unita alla piccola barriera energetica ricavata dalla scansione, può spiegare il perché della mancata localizzazione del TS3.

In ogni caso, nel pattern di reazione, in luogo della struttura del TS3 pienamente ottimizzata, è stata inserita la struttura corrispondente al massimo energetico della scansione.

Riprendendo la descrizione del meccanismo catalitico, il TS3 evolve nell'INT3 più stabile di questo di oltre 6.3 kcal/mol.

Dall'INT3 si arriva, infine ai prodotti, dove il legame peptidico è ormai idrolizzato, passando attraverso il TS4 in cui il legame C-N è quasi completamente rotto (2,08 Å), mentre il protone è ceduto dal gruppo OH del carbonile al Glu-73. In aggiunta, nel TS4, il legame a idrogeno tra la catena laterale della Thr-17 e l'ossigeno del carbonile si allenta, per rompersi definitivamente nei prodotti.

Dal punto di vista energetico il TS4 risulta essere circa 3.3 kcal/mol al di sopra dell'INT3, mentre i prodotti sono più stabili del TS4 di circa 16.8 kcal/mol.

In base ai risultati ottenuti si possono trarre le seguenti conclusioni: 1) la reazione risulta essere esotermica per poco meno di 10 kcal/mol; 2) lo stadio lento è rappresentato

dall'attacco nucleofilo, anche se il valore energetico della barriera è al limite dell'evento catalitico; 3) il Glu-73, così come accade nel caso dell'IDE, agisce sia da base generale promuovendo la formazione di uno ione idrossido o quantomeno attivando la molecola di H₂O, sia da shuttle protonico favorendo l'allontanamento del gruppo uscente; 4) il ruolo svolto nell'IDE dalla Tyr-831 può essere svolto nella MPP dalla catena principale del substrato o dalla catena laterale della Thr-17: si potrebbe speculare che la richiesta per un amminoacido idrofilico in posizione P₂' o P₃' del substrato, cioè distante due o tre amminoacidi in direzione C-terminale rispetto al sito di taglio, sia dettata dalla necessità di stabilizzare la carica negativa al carbonile.

Nel caso della COX IV tali amminoacidi sono, rispettivamente, cisteina e serina: nelle coordinate cristallografiche del complesso enzima-substrato, tuttavia, la Cys-19 risulta scarsamente visibile, mentre i residui 20-25 non sono determinati [71].

Lo sviluppo di nuovi funzionali esplicitamente progettati per la determinazione delle barriere energetiche uniti ai limiti riconosciuti al funzionale B3LYP per questo e per altri aspetti, come già esposto nel capitolo 2, e all'ottenimento di una barriera energetica per lo stadio lento del meccanismo della MPP al limite dell'evento catalitico, hanno suggerito nuovi calcoli con il funzionale MPWB1K.

Il pattern energetico così ottenuto è riportato in grafico (Fig. 13): tutti i punti stazionari ad eccezione del TS4 e dei prodotti risultano sistematicamente stabilizzati di circa 8-9 kcal/mol; solo il TS4 è stabilizzato di circa 4 kcal/mol mentre l'esotermicità della reazione risulta ridotta di meno di 1 kcal/mol.

Ad un'attenta analisi si può osservare una piccola variazione rispetto al meccanismo proposto per l'IDE.

Successivamente alla formazione dell'intermedio tetraedrico (INT1) e alla pseudo-rotazione del Glu-73 per formare il secondo intermedio (INT2) in cui si ha la formazione di un legame a idrogeno tra l'atomo di azoto del gruppo uscente ed il gruppo OH di Glu-73 e di un secondo legame idrogeno tra il gruppo OH legato al carbonile e l'atomo di O dello stesso Glu-73, si situa un terzo intermedio (INT3) in cui il Glu-73 ha trasferito l'atomo di idrogeno: si ha quindi l'aggiunta di un ulteriore intermedio e del relativo stato di transizione che conduce ad esso (TS3).

Nel meccanismo ipotizzato per l'IDE, successivamente all'INT2 si collocava uno stato di transizione che conduceva ai prodotti ed in cui la cessione dell'idrogeno al gruppo uscente era contemporanea alla rottura del legame ammidico e alla deprotonazione del gruppo OH legato al carbonile.

Eccettuato questo passaggio supplementare, il resto del meccanismo ipotizzato è simile a quello proposto per l'IDE.

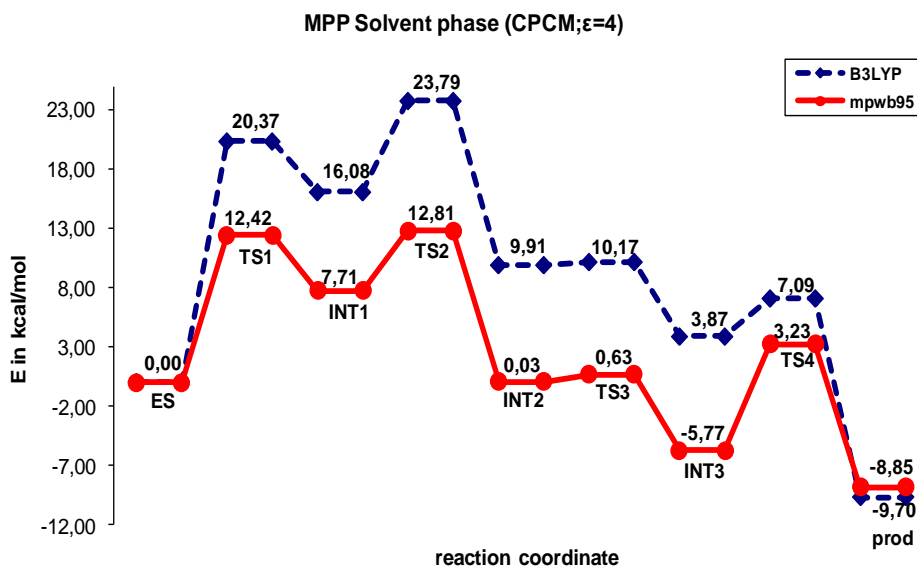


Fig. 13: Confronto tra i pattern energetici relativi al meccanismo catalitico della MPP ottenuti con i funzionali B3LYP e MPWB1K in solvente utilizzando il modello CPCM ed un valore di $\epsilon=4$

4.4 Le Anidrasi Carboniche

L'anidride carbonica (CO_2) è una piccola molecola apolare presente come gas nell'atmosfera in piccole quantità, appena lo 0.04 % in volume. Eppure, questa piccola molecola riveste una grande importanza e non solo per il sempre attuale problema del surriscaldamento globale.

La CO_2 , infatti, è una specie comune ai processi di fotosintesi e di respirazione, essendo un reagente per il primo ed un prodotto per il secondo.

La sua disponibilità ed il suo trasporto, pertanto, risultano essere importanti per tutti gli organismi viventi.

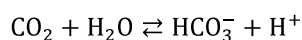
Come gas, la CO_2 diffonde piuttosto lentamente, mentre in soluzione tende a sommare una molecola di acqua per produrre acido carbonico che, a seconda del pH, può cedere uno o due protoni per formare, rispettivamente, ioni bicarbonato e carbonato oppure restare in forma neutra [81]. L'idratazione è una reazione di equilibrio lenta a differenza delle reazioni di deprotonazioni che sono considerate istantanee [82].

Le condizioni di pH dei mari fanno sì che quasi tutta la CO₂ disciolta in essi sia presente come ione bicarbonato e come tale non può attraversare la membrana cellulare a meno che non siano presenti dei sistemi di scambio.

In aggiunta, anche nei cloroplasti le condizioni ambientali sono tali da favorire la formazione dello ione bicarbonato, mentre l'enzima fotosintetico Rubisco ha come substrato la CO₂: è richiesta, pertanto, una "riconversione" dello ione bicarbonato in CO₂ [83].

Non sorprende, quindi, che praticamente tutti gli organismi viventi, sin dai primordi [84], per poter effettuare il processo di fotosintesi [82] o il processo della respirazione [85] abbiano dovuto sviluppare un sistema per la conversione, il trasporto e l'accumulo della CO₂.

L'anidrasi carbonica (CA) (EC 4.2.1.1) sembra svolgere egregiamente questo compito. Esso è un metallo-enzima che catalizza l'idratazione reversibile della CO₂ in soluzione.



Il ruolo dell'anidrasi carbonica negli organismi viventi non è però limitato al semplice trasporto della CO₂, ma coinvolge, tra le altre cose, anche funzioni di regolazione del pH, di trasporto di ioni e di omeostasi della CO₂. Nei mammiferi, ad esempio, la CA è coinvolta in numerosi altri processi quali la gluconeogenesi, l'ureagenesi, la lipogenesi, la calcificazione ed il riassorbimento osseo, la riproduzione, la visione, la digestione e molti altri ancora [83,85-87]. L'interesse per la CA, comunque, non è dettato solo dal suo coinvolgimento in così tanti processi biologici ma anche perché implicata in molte malattie umane, incluse alcune forme tumorali. Inoltre, la sua presenza in organismi quali funghi e batteri la rendono un possibile target nello sviluppo di farmaci contro le infezioni [84,86].

La diffusione ubiquitaria della CA tra gli organismi viventi e il suo diffuso coinvolgimento nei loro cicli biologici hanno fatto sì che questo enzima sia presente sotto molte isoforme e con differenti strutture.

Dipendentemente dalla sua struttura e dalla diversità dei geni che codificano per essa, la CA può essere ripartita in almeno quattro differenti classi, denominate, rispettivamente, α , β , γ e δ [82-88].

Gli enzimi della classe α sono diffusi in quasi tutti gli organismi viventi, ma costituiscono le uniche CA finora caratterizzate nei mammiferi e come tali sono anche le più studiate: fino ad ora sono state identificate sedici isoforme differenti, di cui tre prive di attività catalitica, ripartite tra forme citoplasmatiche, forme associate a membrane e forme secrete. Ogni isoforma presenta differenti cinetiche di reazione e risposte agli inibitori [89-92].

Gli enzimi appartenenti a questa classe sono presenti per lo più come monomeri di circa 30 kDa e presentano tutti un sito attivo contenente tre residui di istidina quali ligandi dello ione metallico.

Gli enzimi della classe β sono anche essi ampiamente espressi nei vari organismi viventi, anche se meno ampiamente degli enzimi di classe α dai quali differiscono sia per il sito attivo, il quale presenta due residui di cisteina ed uno di istidina come ligandi dello ione metallico, sia per la tendenza a formare oligomeri a base di dimeri. Gli enzimi appartenenti a questa classe sono anche gli unici, fra tutte la CA, a mostrare una regolazione allosterica [93,94].

La classe γ conta, in realtà, un solo enzima, denominato Cam e presente come omotrimerico nell'archeobatterio *Methanosarcina thermofila*. Nonostante le differenze strutturali (contiene ad esempio eliche β sinistrorse), il sito catalitico richiama quello degli enzimi di classe α : delle tre istidine, però, due provengono da un monomero mentre la terza è fornita da un monomero adiacente. In aggiunta, lo ione metallico coordina due molecole di acqua e non solo una come avviene invece per gli enzimi delle altre classi.

Anche la classe δ si compone di un solo enzima, la TWCA1, isolato dalla diatomea *Thalassiosira weissflogii*. Questo enzima presenta delle caratteristiche strutturali molto simili a quelle degli enzimi di classe α , con un sito catalitico contenente i tre residui istidinici [95], ma con una sequenza genica differente da tutte quelle riferibili agli enzimi della altre classi.

Tutte le CA, a prescindere dallo loro inclusione in una classe piuttosto che un'altra, richiedono come cofattore lo ione Zn^{2+} . L'enzima Cam, tuttavia, presenta delle velocità di idratazione più elevate nel caso in cui lo ione Zn^{2+} è sostituito dallo ione Fe^{2+} o dallo ione Co^{2+} [88], mentre l'attività degli enzimi della classe α [89] e dell'enzima TWCA1 [96] contenenti ioni Co^{2+} risulta paragonabile a quella degli enzimi nativi.

L'effetto degli ioni dei metalli di transizione diversi dallo Zn^{2+} sull'attività delle CA varia a seconda del metallo considerato e del suo stato di ossidazione: in generale gli ioni bivalenti di tali metalli tendono ad inibire fortemente l'azione catalitica [89,97]. Tale inibizione può derivare dalla competizione per il sito di coordinazione dello ione metallico o per l'interazione con altri siti di coordinazione presenti nell'enzima.

In alcuni casi, inoltre, la presenza di un particolare ione metallico può anche promuovere l'espressione genica di enzimi altrimenti rimasti inespressi: è il caso, ad esempio, della *Thalassiosira weissflogii* che per esposizione a ioni Cd^{2+} promuove l'espressione di una nuova forma di CA, la CDCA1 [98].

4.4.1 La CDCA1

Il Cadmio è un metallo pesante noto per la sua tossicità [99-101].

La sua concentrazione nei mari, tuttavia, varia come quella di altri nutrienti algali: bassa concentrazione sulla superficie marina e rigenerazione sui fondali, comportamento questo che può essere, in parte, spiegato con l'assunzione da parte di microorganismi.

In particolare, la *Thalassiosira weissflogii* in condizioni di bassa concentrazione di CO₂ e di carenza di ioni Zn²⁺, può sostenere la sua crescita esprimendo un enzima dotato di attività di anidraasi carbonica contenente ioni Cd²⁺ anziché ioni Zn²⁺, la CDCA1, riflettendo la capacità unica degli organismi marini all'adattamento di condizioni estreme [98].

Ciò sorprende ancora di più se si pensa come le CA, in esperimenti in vitro, abbiano mostrato attività significativamente ridotte in presenza di ioni Cd²⁺ [102].

Un'analisi genetica più attenta, tuttavia, ha rivelato come l'adattamento mostrato dalla *T. weissflogii* non costituisca un fenomeno isolato, ma bensì così ampiamente diffuso tra le diatomee da indurre ad inserire questi enzimi in una nuova classe di CA: la ζ [103].

L'unico enzima della classe ζ fino ad oggi caratterizzato rimane, tuttavia, la CDCA1 [104].

La CDCA1 ha una massa molare di poco meno di 68 kDa e si compone di tre unità, aventi sequenze primarie identiche all'85 %. Ogni subunità, inoltre, è dotata di attività catalitica e presenta un sito catalitico contenente uno ione Cd²⁺ coordinato a due residui di cisteina e ad uno di istidina. La coordinazione dello ione metallico è completata da due molecole di acqua, delle quali una è presumibilmente uno ione idrossido. Il sito catalitico è posizionato in fondo ad una fenditura a forma di imbuto che attraversa l'enzima.

La fenditura divide l'enzima in due lobi, uno contenente il sito di coordinazione dello ione metallico e l'altro recante il presunto sito di legame del substrato. Tali lobi, in aggiunta, sono strutturalmente sovrapponibili alla β-CA del *Pisum sativum* che è, però, un omodimero e contiene ioni Zn²⁺. In più, confrontando le strutture primarie degli enzimi delle classi ζ e β sono state evidenziate ulteriori similitudini [103].

Le somiglianze strutturali con le β-CA sono ancora più rafforzate dal fatto che la CDCA1 è in grado di funzionare altrettanto bene utilizzando ioni Zn²⁺ anziché Cd²⁺: esso è, dunque, un enzima "scambiatore" [104].

Questo comportamento è reso possibile dalla flessibilità mostrata dall'octapeptide compreso tra il residuo di istidina ed uno dei residui di cisteina coordinanti lo ione metallico. In tal modo, infatti, la conformazione dell'apo-enzima, nella quale i ligandi dello ione risultano dislocati dalla posizione normalmente occupata nell'enzima legato allo ione metallico, è

stabilizzata: una qualsiasi mutazione che comporti un'alterazione della flessibilità di questo peptide ha come conseguenza l'abolizione di questa insolita caratteristica.

4.4.2 Il meccanismo catalitico delle CA

Il confronto tra le strutture dei siti attivi delle CA appartenenti alle varie classi fino ad oggi riconosciute ha messo in luce come in tutti i casi esaminati la forma attiva dell'enzima³⁹ presenti uno ione metallico coordinato ad uno ione idrossido, portando ad ipotizzare un meccanismo d'azione simile per tutti gli enzimi [82,85,86,88,89,94,104,105].

Questa ipotesi è supportata dai dati ricavati dalla misura delle cinetiche di reazione, le quali hanno evidenziato per tutti gli enzimi caratterizzati un meccanismo del tipo a "ping-pong", così detto in quanto la catalisi avviene in due stadi.

Nel primo stadio lo ione idrossido⁴⁰ legato al centro metallico reagisce con la CO₂ formando uno ione bicarbonato, successivamente spostato da una molecola di acqua che si lega allo ione metallico generando una specie enzimatica inattiva.

Nel secondo stadio, invece, quello cineticamente lento, la molecola di acqua legato allo ione metallico cede uno ione H⁺ in modo da riformare la specie enzimatica attiva.

In questo caso, il valore di k_{cat} riflette maggiormente la velocità di trasferimento di H⁺, mentre il valore di k_{cat}/K_M è un indice dell'efficienza catalitica connessa all'interconversione tra CO₂ e ione bicarbonato.

Molta attenzione è stata spesa nella determinazione del meccanismo e dei residui coinvolti nel secondo stadio del ciclo catalitico [89,93,94,107,108] senza, peraltro, giungere a delle conclusioni definitive.

Anche per quanto riguarda l'interconversione tra CO₂ e ione bicarbonato sono stati effettuati numerosi studi che hanno portato alla formulazione di due ipotesi (Schema 3).

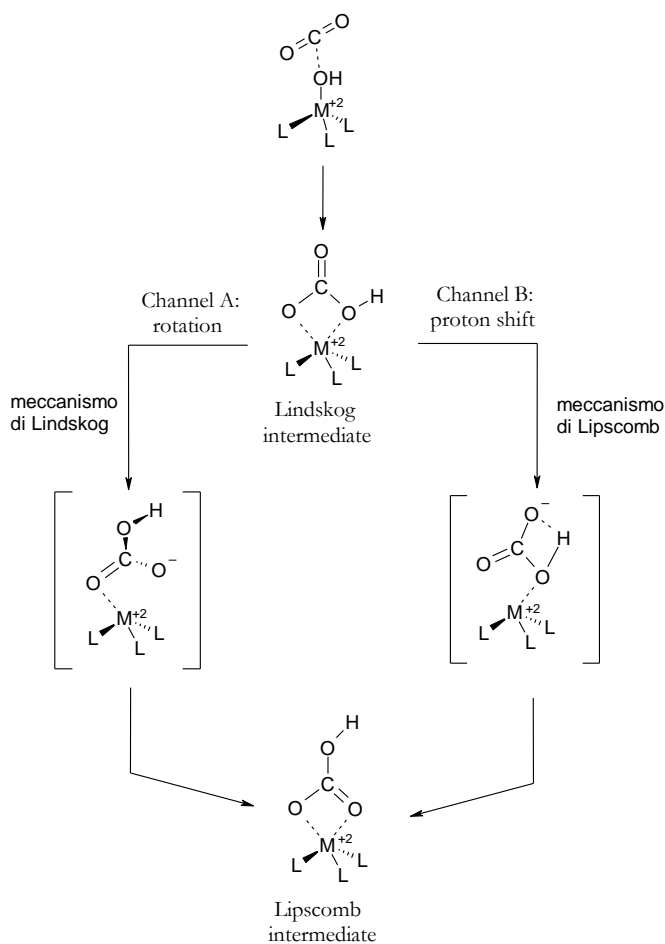
Entrambe prevedono un iniziale attacco nucleofilo sulla CO₂ da parte dello ione idrossido legato allo ione metallico con formazione di un primo intermedio, denominato intermedio di Lindskog [109], che evolve poi in un intermedio più stabile, l'intermedio di Lipscomb [110], o attraverso una rotazione interna (meccanismo di Lindskog; Channel A, Schema 3) o

³⁹ Alcuni enzimi della classe β , quelli che mostrano regolazione allosterica, possono assumere anche una conformazione inattiva in cui il quarto ligando dello ione metallico è costituito dalla catena laterale di un residuo di acido aspartico [93,94].

⁴⁰ Lo ione metallico agisce come acido di Lewis abbassando il pK_a dell'acqua da circa 14 a circa 7 nell'enzima nel caso di Zn²⁺. Ciò è in parte dovuto anche alle interazioni tra la molecola di acqua ed i residui presenti nel sito catalitico [2,89].

attraverso uno shift protonico dal gruppo OH dello ione bicarbonato a un altro atomo di ossigeno del bicarbonato stesso (meccanismo di Lipscomb; Channel B, Schema 3) [111].

I tentativi fatti per determinare quale delle due ipotesi di meccanismo sia la meno dispendiosa dal punto di vista energetico sono stati, finora, infruttuosi.



Schema 3: Il meccanismo di interconversione della CO_2 in ione bicarbonato ipotizzato per le CA

4.4.3 Il modello utilizzato per la CDCA1

La recente caratterizzazione della CDCA1 ha evidenziato la capacità di questo enzima di utilizzare per la catalisi sia ioni Zn^{2+} che ioni Cd^{2+} , a differenza di altre forme di CA che sono fortemente inibite da quest'ultimo [112].

Questa scoperta ha costituito la spinta per intraprendere uno studio computazionale (vedi articolo II) con l'intento di contribuire a chiarire le basi di questo comportamento. Inoltre, sono state vagliate entrambe le ipotesi del meccanismo di interconversione tra CO₂ e ione bicarbonato.

A tal fine, utilizzando come punto di partenza le strutture cristallografiche della CDCA1 recentemente ottenuta (PDB code: 3BOB, 3BOC), è stato costruito un modello del sito catalitico dell'enzima (Fig. 14).

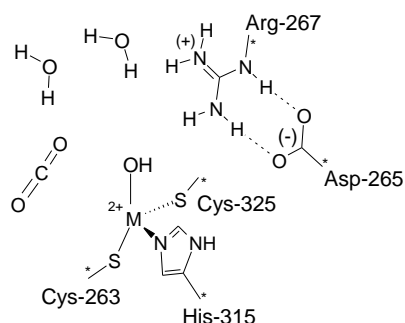


Fig. 14: Modello del sito catalitico della CDCA1. Il simbolo M²⁺ è utilizzato per indicare genericamente gli ioni Cd²⁺ e Zn²⁺. Gli atomi contrassegnati dall'asterisco sono quelli mantenuti bloccati durante le ottimizzazioni.

Il modello, costruito sulle basi delle indicazioni sperimentali disponibili [104], si compone di 54 atomi, ha una carica elettrica totale pari a -1 ed una molteplicità di 1.

In particolare, oltre lo ione metallico, fanno parte del modello i ligandi proteici dello ione stesso, costituiti da una molecola di 4-metilimidazolo e da due di metantiossido e rappresentanti, rispettivamente, le catene laterali dei residui di His-315 e delle Cys-263 e Cys-325; uno ione acetato ed uno ione N-metilguanidinio, rappresentanti la diade Asp-265 e Arg-267; lo ione idrossido legato al centro metallico; il substrato CO₂; due molecole di acqua presenti nelle immediate vicinanze del sito catalitico.

Inoltre, gli atomi di C dei gruppi metilici delle molecole e degli ioni considerati nel modello e contrassegnati con l'asterisco nella Fig. 14, sono stati mantenuti fissi nelle loro posizioni cristallografiche

4.4.4 I risultati ottenuti per la CDCA1

I risultati ottenuti nel modello contenente lo ione Cd²⁺ mostrano lo stesso andamento di quelli ottenuti nel modello con lo ione Zn²⁺ sia dal punto di vista strutturale che da quello energetico.

La descrizione dettagliata e le strutture contenenti i principali parametri geometrici dei punti stazionari determinati sono presentati nell'articolo II.

In entrambi i casi, la prima cosa che si evidenzia è l'assenza di un complesso di Van der Waals: la formazione dello ione bicarbonato è praticamente un processo che non richiede alcuna attivazione. Dal complesso enzima-substrato (ES), in cui lo ione bicarbonato risulta coordinato allo ione metallico attraverso il gruppo OH, attraverso una rotazione (TS1) si arriva all'intermedio di Lindskog (INT1), in cui ad essere coordinato al centro metallico è uno degli atomi di ossigeno della CO₂. Nel caso dell'enzima contenente lo ione Cd²⁺ la struttura del TS1 presenta una pentacoordinazione per lo ione metallico con lo ione bicarbonato che funge da ligando bidentato. Al contrario nella forma enzimatica contenente lo ione Zn²⁺ è presente una tetraordinazione per tutti gli intermedi del pathway di reazione. Questa diversità di comportamento può essere ascritta alle differenti dimensioni dei due ioni metallici. Dal punto di vista energetico, tuttavia, il passaggio all'INT1 risulta leggermente favorito nel caso dello ione Zn²⁺ (Fig. 15b) rispetto quello del Cd²⁺ (fig. 15a) essendo le barriere energetiche in fase proteica 2.6 kcal/mol e 3.4 kcal/mol, rispettivamente.

Anche l'INT1 risulta essere leggermente più stabile nel caso di Zn²⁺ che nel caso di Cd²⁺ (-5.1 kcal/mol vs -4.1 kcal/mol).

Dall'INT1 è possibile arrivare ai Prod (intermedio di Lipscomb nello Schema 3) attraverso uno stato di transizione (TS4) che consiste in una semplice rotazione interna (meccanismo di Lindskog).

Nel caso del Cd²⁺ il TS3 si colloca 1.8 kcal/mol al di sopra dell'ES, mentre nel caso dello Zn²⁺ è 2.0 kcal/mol al di sotto dell'ES corrispondente. Dal punto di vista strutturale le geometrie sono simili in entrambe le forme enzimatiche considerate e richiamano la geometria del precedente INT1.

Nel caso dello ione Cd²⁺ la reazione risulta leggermente esotermica, essendo il Prod più stabile di 2.8 kcal/mol rispetto il corrispondente ES, mentre nel caso dello Zn²⁺ la reazione risulta più esotermica essendo questa volta il Prod stabilizzato di 6.6 kcal/mol rispetto al relativo ES. Ancora una volta, dal punto di vista strutturale i due enzimi risultano simili, con lo ione bicarbonato coordinato in maniera monodentata e leggermente più distante dal centro metallico rispetto all'INT1 (0.06 Å per Cd²⁺ vs 0.08 Å per Zn²⁺), cosa questa che dovrebbe favorire la sua sostituzione con una molecola di acqua.

Nel meccanismo di Lindskog lo shift protonico appare mediato dalla diade Asp-Arg e da una delle molecole di acqua presenti. In particolare, passando per uno stato di transizione (TS2), il residuo di Asp-265 accetta uno ione H⁺ dallo ione bicarbonato, mentre, nello stesso tempo, il residuo di Arg-267 cede uno ione H⁺ alla molecola di acqua con cui interagisce. Questa, a

sua volta, cede uno dei suoi H⁺ all'atomo di ossigeno dello ione bicarbonato situato in posizione distale rispetto al centro metallico. Si ottiene così un intermedio (INT2) in cui l'Asp-265 è protonato e l'Arg-267 deprotonato: la cessione dell'atomo di idrogeno da Asp a Arg (TS3) conduce al Prod, ripristinando il sito catalitico.

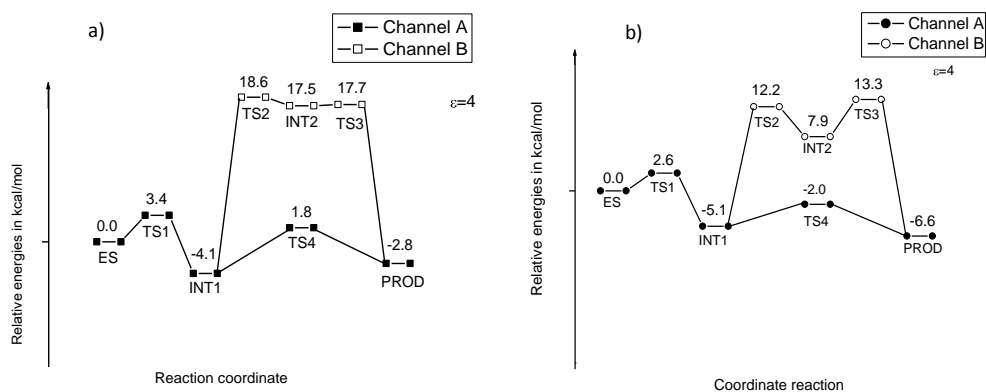


Fig. 15: Pathways energetici in solvente (CPCM, $\epsilon=4$) della CDCA1 relativi al meccanismo di Lindskog (Channel A) e al meccanismo di Lipscomb (Channel B): a) il modello contenente Cd²⁺ e b) per il modello contenente Zn²⁺.

Anche per quanto concerne questa seconda ipotesi, le strutture ottimizzate dei punti stazionari presentano geometrie simili per i due ioni metallici considerati, diversamente da quanto accade invece dal punto di vista energetico. Nel caso del modello contenente lo ione Cd²⁺, infatti, il TS2 si colloca oltre le 18 kcal/mol rispetto all'ES, con una barriera energetica rispetto al precedente INT1 alta circa 23 kcal/mol; INT2 e TS3, invece, sono quasi isoenergetici con il TS2.

Nel caso della Zn-CDCA1, il TS2 si colloca a poco più di 12 kcal/mol rispetto all'ES, con una barriera energetica riferita al precedente INT1 di poco più di 17 kcal/mol; INT2, invece, è più stabile del precedente TS2 di poco più di 4 kcal/mol mentre TS3 è 5.5 kcal/mol al di sopra dell'INT2.

In conclusione, per quanto riguarda sia la Cd-CDCA1 che la Zn-CDCA1 il meccanismo di Lipscomb (shift protonico) risulta sfavorito rispetto quello di Lindskog (rotazione interna).

Nel caso della Cd-CDCA1 lo stadio lento si verifica in corrispondenza del TS4 con una barriera di circa 6 kcal/mol, mentre nel caso della Zn-CDCA1 lo stadio cineticamente lento è rappresentato dal TS1 che presenta una barriera di appena 2.6 kcal/mol. In valore assoluto, per la Zn-CDCA1 la barriera energetica più grande sarebbe quella che conduce al TS4, ma poiché in questo caso tutti i punti stazionari a partire dall'INT1 giacciono sotto l'asintoto dei reagenti, il TS4 costituisce in realtà un ostacolo fittizio.

Dal confronto dei pathways energetici è possibile notare come l'enzima sia abbastanza attivo con entrambi gli ioni metallici, anche se risulta essere leggermente più performante nel caso in cui contenga lo ione Zn^{2+} , riproducendo così anche questo dato sperimentale [104].

Nel tentativo di comprendere se l'aumentata attività della Cd-CDCA1 rispetto altre forme di CA contenenti Cd^{2+} fosse dovuta ad una aumentata nucleofilicità dello ione idrossido dovuta ai ligandi cisteinici dello ione metallico, è stato valutato l'effetto, in termini di carica atomica, della progressiva sostituzione dei ligandi istidinici con ligandi cisteinici su un modello rappresentato dal solo ione metallico e dai suoi ligandi (Fig. 16).

L'effetto della sostituzione risulta essere simile per entrambi i modelli e, pertanto, l'aumentata reattività dello ione Cd^{2+} non può essere ascritta né alla differente polarizzabilità del centro metallico, né ad una maggiore nucleofilicità dello ione idrossido.

Un spiegazione, tuttavia, può essere fornita in termini di teoria Hard-Soft: l'atomo di azoto nelle istidine è classificabile come base intermedia, mentre l'atomo di zolfo nelle cisteine è una base soft, risultando quindi capace di formare legami covalenti più forti con lo ione Cd^{2+} , che è anch'esso una specie soft. In conseguenza di ciò, il legame tra lo ione Cd^{2+} e lo ione idrossido, una base hard, si allunga rendendo quest'ultimo più labile e pronto alla reazione.

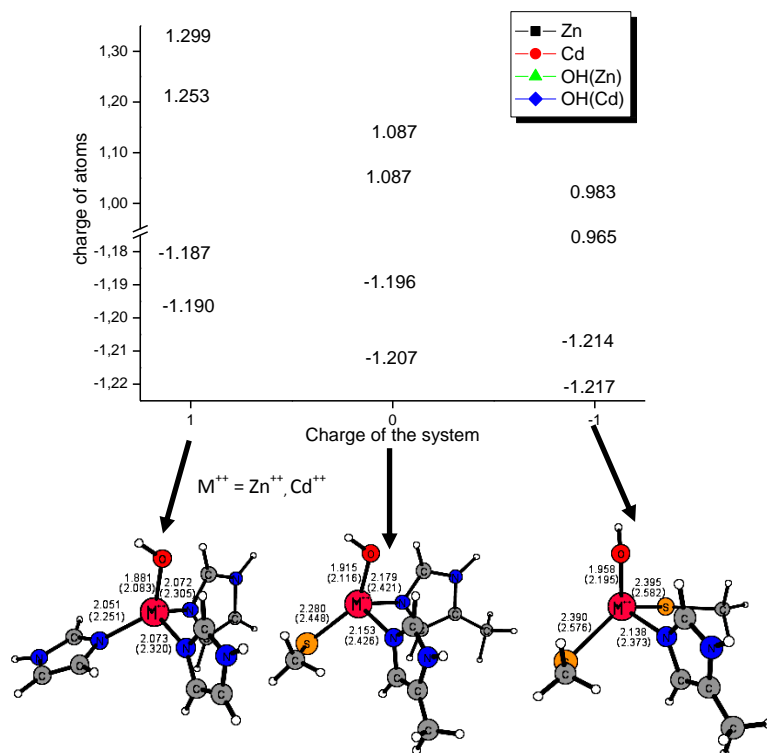


Fig. 16: Effetto della progressiva sostituzione dei ligandi istidinici con ligandi cisteinici sulle cariche dell'atomo di ossigeno del gruppo OH e dello ione metallico. In parentesi le distanze di legame dei ligandi con lo ione Cd^{2+} .

Bibliografia

- [1] M. Comellas-Bigler - *Structural Studies of the Serine-Carboxyl Proteinase Kumamolisin and the Metallopeptidase Peptidyl-Dipeptidase Dcp* – 2004, Max-Planck-Institut für Biochemie, D-82152, Martinsried, Germany
- [2] K.A. McCall, C.-C. Huang, C.A. Fierke – *J. Nutr.* – 2000, 130, 1437S
- [3] Y. S. Cheng, D. Zipser – *J. Biol. Chem.* – 1979, 254, 4698
- [4] K.H. Swamy, A.L. Goldberg – *J. Bacteriol.* – 1982, 149, 1027
- [5] A.B. Becker, R.A. Roth – *Proc. Nat. Acad. Sci.* – 1992, 89, 3835
- [6] N.M. Hooper – *FEBS Lett.* – 1994, 354, 1
- [7] K. Maskos – *Pitriylsins/Inverzincins* in *Handbook of Metalloproteins* – A. Messerschmidt, W. Dode, M. Cygler, eds. – 2004, John Wiley & Sons
- [8] R.K. Perlman, M.R. Rosner – *J. Biol. Chem.* – 1994, 269, 33140
- [9] E. Malito, L.A. Ralat, M. Manalopoulou, J.L. Tsay, N.L. Wadlington, W.-J. Tang – *Biochem.* – 2008, 47, 12822
- [10] E.S. Song, C. Clady, M.G. Fried, L.B. Hersh – *Biochem.* – 2006, 45, 15085
- [11] H. Im, M. Manalopoulou, E. Malito, Y. Shen, J. Zhao, M. Neant-Fery, C.-Y. Sun, S.C. Meredith, S.S. Sisodia, M.A. Leissring, W.-J. Tang – *J. Biol. Chem.* – 2007, 282, 25453
- [12] K.A. Seta, R.A. Roth – *Biochem. Biophys. Res. Commun.* – 1997, 231, 167
- [13] W.C. Duckworth, R.G. Bennett, F.G. Hamel – *Endocr. Rev.* – 1998, 19, 608
- [14] M. Morita, I.V. Kurochkin, K. Motojima, S. Goto, T. Takano, S. Okamura, R. Sato, S. Yokota, T. Imanaka – *Cell. Struct.Funct.* – 2000, 25, 39
- [15] M.A. Leissring, W. Farris, X. Wu, D.C. Christodoulou, M.C. Haigis, L. Guarente, D.J. Selkoe – *Biochem. J.* – 2004, 277, 13389
- [16] I.V. Kurochkin – *Trends Biochem. Sci.* – 2001, 26, 421
- [17] J.A. Lynch, M.A. George, P.B. Eisenhauer, K. Conn, W. Gao, I. Carreras, J.M. Wells, A. McKee, M.D. Ullman, R.E. Fine – *J. Neurosci. Res.* – 2006, 83, 1262
- [18] F.B. Stentz, A.E. Kitabchi, J.W. Schilling, L.R. Schronk, J.M. Seyer – *J. Biol. Chem.* – 1989, 264, 20275
- [19] G. Grasso, E. Rizzarelli, G. Spoto – *J. Mass Spectrom.* – 2007, 42, 1590
- [20] G. Grasso, E. Rizzarelli, G. Spoto – *J. Mass Spectrom.* – 2009, 44, 735
- [21] B.D. Gehm, W. Kuo, R.K. Perlman, M.R. Rosner – *J. Biol. Chem.* – 1993, 268, 7943
- [22] R.K. Perlman, B.D. Gehm, W. Kuo, M.R. Rosner – *J. Biol. Chem.* – 1993, 268, 21538
- [23] E. Malito, R.E. Hulse, W.-J. Tang – *Cell. Mol. Life Sci.* – 2008, 65, 1

- [24] Y. Shen, A. Joachimiak, M.R. Rosner, W.-J. Tang – *Nature* – 2006, 443, 870
- [25] P. Li, W.-L. Kuo, M. Yousef, M.R. Rosner, W.-J. Tang – *Biochem. Biophys. Res. Comm.* – 2006, 343, 1032
- [26] E.S. Song, M.A. Juliano, L. Juliano, L.B. Hersh – *J. Biol. Chem.* – 2003, 278, 49789
- [27] M.C. Camberos, A.A. Pérez, D.P. Udrisar, M.I. Wanderley, J.C. Cresto – *Exp. Biol. Med.* – 2001, 226, 334
- [28] E.S. Song, M.A. Juliano, L. Juliano, M.G. Fried, S.L. Wagner, L.B. Hersh – *J. Biol. Chem.* – 2004, 279, 54216
- [29] T. Saric, D. Mulle, H.-J. Seitz, K. Pavelic – *Mol. Cell. Endocr.* – 2003, 204, 11
- [30] F.G. Hamel, J.L. Upward, R.G. Bennett – *Endocrinology* – 2003, 144, 2404
- [31] E.S. Song, A. Daily, M.G. Fried, M.A. Juliano, L. Juliano, L.B. Hersh – *J. Biol. Chem.* – 2005, 280, 17701
- [32] R.G. Bennett, W.C. Duckworth, F.G. Hamel – *J. Biol. Chem.* – 2000, 275, 36621
- [33] W.C. Duckworth, A.E. Kitabchi – *Diabetes* – 1974, 23, 536
- [34] D. Muller, H. Baumeister, F. Buck, D. Richter – *Eur. J. Biochem.* – 1991, 202, 285
- [35] L. Morelli, R.E. Llovera, L.G. Alonso, B. Frangione, G. de Prat-Gay, J. Ghiso, E.M. Castaño – *Biochem. Biophys. Res. Comm.* – 2005, 332, 808
- [36] V. Chesneau, K. Vekrellis, M.R. Rosner, D.J. Selkoe – *Biochem. J.* – 2000, 351, 509
- [37] D. Edbauer, M. Willem, S. Lammich, H. Steiner, C. Haass – *J. Biol. Chem.* – 2002, 277, 13389
- [38] G. Evin, A. Weidemann – *Peptides* – 2002, 23, 1285
- [39] B.C. Miller, E.A. Eckman, K. Sambamurti, N. Dobbs, K.M. Chow, C.B. Eckman, L.B. Hersh, D.L. Thiele – *PNAS* – 2003, 100, 6221
- [40] L. Zhao, B. Teter, T. Morihara, G.P. Lim, S.S. Ambegaokar, O.J. Ubeda, S.A. Frautschy, G.M. Cole – *J. Neurosci.* – 2004, 24, 11120
- [41] W. Farris, S. Mansourian, M.A. Leissring, E.A. Eckman, L. Bertram, C.B. Eckman, R.E. Tanzi, D.J. Selkoe – *Am. J. Pathol.* – 2004, 164, 1425
- [42] W. Farris, M.A. Leissring, M.L. Hemming, A.Y. Chang, D.J. Selkoe – *Biochemistry* – 2005, 44, 6513
- [43] L.B. Hersh – *Cell. Mol. Life Sci.* – 2006, 63, 2432
- [44] W.Q. Qiu, M.F. Folstein – *Neurobiol. Ag.* – 2006, 27, 190
- [45] B.J. Alper, W.K. Schmidt – *J. Neurosci. Meth.* – 2009, 178, 40
- [46] H. Firket – *La Cellula Vivente* – 1994, Newton Compton ed., Roma
- [47] D.L. Nelson, M.M. Cox – *I principi di biochimica di Lehninger* – 4a Ed. – 2006, Zanichelli, Bologna

- [48] O. Gakh, P. Cavadini, G. Isaya – *Biochim. Biophys. Acta* – 2002, 1592, 63
- [49] A. Mukhopadhyay, P. Hammen, M. Waltner-Law, H. Weiner – *Prot. Sci.* – 2002, 11, 1026
- [50] M. Paetzel, A. Karla, N.C.J. Strynadka, R.E. Dalbey – *Chem. Rev.* – 2002, 102, 4549
- [51] J.A. MacKenzie, R.M. Payne – *Biochim. Biophys. Acta* – 2007, 1772, 509
- [52] A. Mukhopadhyay, C.-S. Yang, B. Wei, H. Weiner – *J. Biol. Chem.* – 2007, 282, 37266
- [53] V. Géli, M. Yang, K. Suda, A. Lustig, G. Schatz – *J. Biol. Chem.* – 1990, 265, 19216
- [54] C. Witte, R.E. Jensen, M.P. Yaffe, G. Schatz – *EMBO J.* – 1988, 7, 1439
- [55] R.E. Jensen, M.P. Yaffe – *EMBO J.* – 1988, 7, 3871
- [56] M. Yang, V. Géli, W. Oppliger, K. Suda, P. James, G. Schatz – *J. Biol. Chem.* – 1991, 266, 6416
- [57] P.C. Böhni, G. Daum, G. Schatz – *J. Biol. Chem.* – 1983, 258, 4937
- [58] J. Kleiber, F. Kalousek, M. Swaroop, L.E. Rosenberg – *PNAS* – 1990, 87, 7978
- [59] V.M. Saavedra-Alanis, P. Rysavy, L.E. Rosenberg, F. Kalousek – *J. Biol. Chem.* – 1994, 269, 9284
- [60] H. Schneider, M. Arretz, E. Wachter, W. Neupert – *J. Biol. Chem.* – 1990, 265, 9881
- [61] M. Arretz, H. Schneider, B. Guiard, M. Brunner, W. Neupert – *J. Biol. Chem.* – 1994, 269, 4959
- [62] H.-P. Braun, U.K. Schmitz – *Int. J. Biochem. Cell. Biol.* – 1997, 29, 1043
- [63] M. Emmermann, U.K. Schmitz – *Plants Physiol.* – 1993, 103, 615
- [64] M. Emmermann, H.-P. Braun, M. Arretz, U.K. Schmitz – *J. Biol. Chem.* – 1993, 268, 18936
- [65] K. Deng, L.Zhang, A.M. Kachurin, L. Yu, D. Xia, H. Kim, J. Deisenhofer, C.-A. Yu – *J. Biol. Chem.* – 1998, 273, 20752
- [66] V. Géli – *PNAS* – 1993, 90, 6247
- [67] M. Yang, R.E. Jensen, M.P. Yaffe, W. Oppliger, G. Schatz – *EMBO J.* – 1988, 7, 3862
- [68] K. Shimokata, S. Kitada, T. Ogishima, A. Ito – *J. Biol. Chem.* – 1998, 273, 25158
- [69] O. Gakh, T. Obsil, J. Adamec, J. Spizek, E. Amler, J. Janata, F. Kalousek – *Arch. Biochem. Biophys.* – 2001, 385, 392
- [70] J. Janata, K. Holá, M. Kubala, O. Gakh, N. Parkhomenko, A. Matušková, E. Kutejová, E. Amler – *Biochem. Biophys. Res. Comm.* – 2004, 316, 211
- [71] A.B. Taylor, B.S. Smith, S. Kitada, K. Kojima, H. Miyaura, Z. Otwinowski, A. Ito, J. Deisenhofer – *Structure* – 2001, 9, 615

- [72] K. Kojima, S. Kitada, K. Shimotaka, T. Ogishima, A. Ito – *J. Biol. Chem.* – 1998, 273, 32542
- [73] S. Kitada, K. Kojima, K. Shimokata, T. Ogishima, A. Ito – *J. Biol. Chem.* – 1998, 273, 32547
- [74] S. Kitada, K. Kojima, A. Ito – *Biochem. Biophys. Res. Comm.* – 2001, 287, 594
- [75] S. Kitada, E. Yamasaki, K. Kojima, A. Ito – *J. Biol. Chem.* – 2003, 278, 1879
- [76] S. Sjöling, A.C. Eriksson, E. Glasier – *J. Biol. Chem.* – 1994, 269, 32059
- [77] C. Rudhe, R. Clifton, O. Chew, K. Zeman, S. Richter, G. Lamppaa, J. Whelam, E. Glaser – *J. Mol. Biol.* – 2004, 343, 639
- [78] K. Kojima, S. Kitada, T. Ogishima, A. Ito – *J. Biol. Chem.* – 2001, 276, 2115
- [79] G. Isaya, F. Kalousek, W.A. Fenton, L.E. Rosenberg – *J. Cell. Biol.* – 1991, 113, 65
- [80] M. Waltner, H. Weiner – *J. Biol. Chem.* – 1995, 270, 26311
- [81] B.J.H. Matthews - *The Rate of Air-Sea CO₂ Exchange: Chemical Enhancement and Catalysis by Marine Microalgae* – 1999
- [82] J.V. Moroney, S.G. Bartlett, G. Samuelsson – *Plant, Cell Environ.* – 2001, 24, 141
- [83] M. Lapointe, T.D.B. MacKenzie, D. Morse – *Plant Physiol.* – 2008, 147, 1427
- [84] K.S. Smith, C. Jakubzick, T.S. Whittam, J.G. Ferry – *PNAS* – 1999, 96, 15184
- [85] A.J. Esbaugh, B.L. Tufts – *Resp. Physiol. Neurobiol.* – 2006, 154, 185
- [86] C.T. Supuran, A. Scozzafava – *Bioorg. Med. Chem.* – 2007, 15, 4336
- [87] P. Hynninen, J.M. Hämäläinen, S. Pastorekova, J. Pastorek, A. Waheed, W.S. Sly, E. Tomas, P. Kirkinen, S. Parkkila – *Reprod. Biol. Endocr.* – 2004, 7, 73
- [88] B.C. Tripp, K. Smith, J.G. Ferry – *J. Biol. Chem.* – 2001, 276, 48615
- [89] S. Lindskog – *Pharmacol. Ther.* – 1997, 74, 1
- [90] T. Kararli, D.N. Silverman – *J. Biol. Chem.* – 1985, 260, 3484
- [91] T.T. Baird Jr., A. Waheed, T. Okuyama, W.S. Sly, C.A. Fierke – *Biochem.* – 1997, 36, 2669
- [92] J.N. Earnhardt, M. Qian, C. Tu, M.M. Lakkis, N.C. Bergenheim, P.J. Laipis, R.E. Tashian, D.N. Silverman – *Biochem.* – 1998, 37, 10837
- [93] A.S. Covarrubias, T. Bergfors, T.A. Jones, M. Högbom – *J. Biol. Chem.* – 2006, 281, 4993
- [94] R.S. Rowlett – *Biochim. Biophys. Acta* – 2010, 1804, 362
- [95] E.H. Cox, G.L. McLendon, F.M.M. Morel, T.W. Lane, R.C. Prince, I.J. Pickering, G.N. George – *Biochem.* – 2000, 39, 12128
- [96] D. Yee, F.M.M. Morel – *Limnol. Oceanogr.* – 1996, 41, 573
- [97] H. Soyut, S. Beydemir, O. Hisar – *Biol. Trace Elem. Res.* – 2008, 123, 179

- [98] T.W. Lane, F.M.M. Morel – *PNAS* – 2000, 97, 4627
- [99] S.A. Hasan, Q. Fariduddin, B. Ali, S. Hayat, A. Ahmad – *J. Environ. Biol.* – 2009, 30, 165
- [100] E.A. Boyle, F. Sclater, J.M. Edmond – *Nature* – 1976, 263, 42
- [101] K.W. Bruland, G.A. Knauer, J.H. Martin – *Limnol. Ocean.* – 1978, 23, 618
- [102] J.E. Coleman – *Curr. Opin. Chem. Biol.* – 1998, 2, 222
- [103] H. Park, B. Song, F.M.M. Morel – *Environ. Microbiol.* – 2007, 9, 403
- [104] Y. Xu, L. Feng, P.D. Jeffrey, Y. Shi, F.M.M. Morel – *Nature* – 2008, 452, 56
- [105] S.A. Zimmerman, J.G. Ferry – *Biochem.* – 2006, 45, 5149
- [106] B.E. Alber, C.M. Colangelo, J. Dong, C.M.V. Stålhandske, T.T. Baird, C. Tu, C.A. Fierke, D.N. Silverman, R.A. Scott, J.G. Ferry – *Biochem.* – 1999, 38, 13119
- [107] H. Shimahara, T. Yoshida, Y. Shibata, M. Shimizu, Y. Kyogoku, F. Sakiyama, T. Nakazawa, S. Tate, S. Ohk, T. Kato, H. Moriyama, K. Kishida, Y. Tano, T. Ohkubo, Y. Kobayashi – *J. Biol. Chem.* – 2007, 282, 9646
- [108] B.C. Tripp, J.G. Ferry – *Biochem.* – 2000, 39, 9232
- [109] S. Lindskog, L. Henderson, K.K. Kannan, A. Liljas, P.O. Nyman, B. Strandberg – *The Enzymes* – vol. 5, 3rd ed., 1971, Academic Press, New York
- [110] J.-Y. Liang, W.N. Lipscomb – *Int. J. Quantum Chem.* – 1989, 36, 299
- [111] W.N. Lipscomb – *Annu. Rev. Biochem.* – 1983, 52, 17
- [112] T. Marino, N. Russo, M. Toscano – *J. Am. Chem. Soc.* – 2005, 127, 4242

APPENDICE

ARTICOLO I

Human insulin-degrading enzyme working mechanism.

Orazio Amata, Tiziana Marino, Nino Russo, Marirosa Toscano

J. Am. Chem. Soc **2009**, *131*, 14804–14811

Human Insulin-Degrading Enzyme Working Mechanism

Orazio Amata, Tiziana Marino, Nino Russo,* and Marirosa Toscano

Dipartimento di Chimica and Centro di Calcolo ad Alte Prestazioni per Elaborazioni Parallele e Distribuite-Centro d' Eccellenza MURST, Università della Calabria, I-87030 Arcavacata di Rende (CS), Italy

Received May 7, 2009; E-mail: nrusso@unical.it

Abstract: The possible mechanism by which the insulin-degrading enzyme (IDE) zinc-binding protease carries out its catalytic function toward two peptides of different length, simulating a portion of B chain of insulin, was investigated on an enzymatic model consisting of 130 /159 atoms, using the density functional theory method and the hybrid exchange–correlation functional B3LYP in gas phase and in the protein environment. Based on the geometry and relative stabilities of minima and transition states on the potential energy profiles, we determined that proteolysis reaction is exothermic and proceeds quickly as the barrier in the rate-limiting step falls widely within the range of values expected for an enzymatic catalysis, both in vacuum and in protein medium.

Introduction

Insulin-degrading enzyme (IDE), also known as insulysin, is a highly conserved M16A Zn²⁺-metalloprotease found in bacteria, fungi, plants, and animals.¹ IDE is unusual in its high affinity for substrates that can be highly diverse in sequence and structure. It also contains an inverted zinc–metalloprotease core motif (HXXE) located within 200 residues of the N-terminal. This is the reason IDE can be numbered among the “inverzincins” family.

IDE was first identified by its ability to degrade the A and B chains of the hormone insulin, mainly at the sites Leu13-Tyr14 and Tyr14-Gln15 (A chain) and Ser9-His10, His10-Leu11, Glu13-Ala14, Tyr15-Leu16, and Phe25-Tyr26 (B chain).² This discovery suggested IDE shares a high degree of homology to pitrilysin, a bacterial protease, and a probable similar proteolytic mechanism.³ In fact, the enzymes of the M16A subfamily are highly related in primary sequence. Human IDE and pitrilysin have, for example, about 30% identity.³ This similarity generally extends throughout the entire length of these proteins but is often higher in the N-terminal half.

Interest in IDE was stimulated by the discovery that it can degrade amyloid β (A β), a peptide involved in the pathogenesis of Alzheimer's disease, which is the leading cause of dementia affecting a large proportion of the elderly population worldwide.^{4–9} In fact, a hypothesized mechanism of this disease assigns a prominent role as causative agent to the hydrophobic peptide A β , which forms quaternary structures that in turn, by an unknown mechanism, cause neuronal death. Numerous in vitro and in vivo studies showed correlations between IDE, A β degradation, and Alzheimer's disease.^{4–9}

Some researchers studied mice that were missing insulysin gene activity.^{10,11} They found that the insulysin enzyme degrades not only insulin but also amyloid β peptides. In addition, even a partial decrease in insulysin activity was found to increase amyloid β peptide levels in the brain.¹⁰

By examining individuals affected by inherited forms of Alzheimer's, a reduction in both IDE expression and catalytic activity was discovered.^{12,13} Today the role of IDE in this disease is quite clear; however, the biological functions of IDE as well as many M16A enzymes remain largely uncharacterized. The fact that IDE is localized in the cytosol, peroxisomes, endosomes, proteasome complexes, and in the surface of cerebrovascular endothelial cells suggests these functions may be varied.^{14,15}

- (1) Mirsky, L. A.; Broth-Kahn, R. H. *Arch. Biochem.* **1949**, *20*, 1.
- (2) Duckworth, W. C. *Endocrinol. Rev.* **1988**, *9*, 319.
- (3) Alper, B. J.; Nienow, T. E.; Schmidt, W. K. *Biochem. J.* **2006**, *398*, 145.
- (4) Morelli, L.; Llovera, R. E.; Alonso, L. G.; Frangione, B.; de Prat-Gay, G.; Ghiso, J.; Castaño, E. M. *Biochem. Biophys. Res. Commun.* **2005**, *332*, 808.

- (5) Farris, W.; Leisring, M. A.; Hemming, M. L.; Chang, A. Y.; Selkoe, D. J. *Biochemistry* **2005**, *44*, 6513.
- (6) Llovera, R. E.; de Tullio, M.; Alonso, L. G.; Leisring, M. A.; Kaufman, S. B.; Roher, A. E.; de Prat-Gay, G.; Morelli, L.; Castaño, E. M. *J. Biol. Chem.* **2008**, *283*, 17039.
- (7) Zhao, L.; Teter, B.; Morihara, T.; Lim, G. P.; Ambegaokar, S. S.; Ubeda, O. J.; Frautschi, S. A.; Cole, G. M. *J. Neurosci.* **2004**, *24*, 11120.
- (8) Qiu, W. Q.; Folstein, M. F. *Neurobiol. Aging* **2007**, *27*, 190.
- (9) Hersh, L. B. *Cell. Mol. Life Sci.* **2006**, *63*, 2432.
- (10) Farris, W.; Mansourian, S.; Chang, Y.; Lindsley, L.; Eckman, E. A.; Froesch, M. P.; Eckman, C. B.; Tanzi, R. E.; Selkoe, D. J.; Gaetette, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4162.
- (11) Miller, B. C.; Eckman, E. A.; Sambamurti, K.; Dobbs, N.; Chow, K. M.; Eckman, C. B.; Hersh, L. B.; Thiele, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6221.
- (12) Cook, D. G.; Leverenz, J. B.; McMillan, P. J.; Kulstad, J. J.; Erickson, S.; Roth, R. A.; Schellenberg, G. D.; Jin, L. W.; Kovacina, K. S.; Craft, S. *Am. J. Pathol.* **2003**, *162*, 313.
- (13) Kim, M.; Hersh, L. B.; Leisring, M. A.; Ingelsson, M.; Matsui, T.; Farris, W.; Lu, A.; Hyman, B. T.; Selkoe, D. J.; Bertram, L.; Tanzi, R. E. *J. Biol. Chem.* **2007**, *282*, 7825.
- (14) Duckworth, W. C.; Bennett, R. G.; Hamel, F. G. *Endocr. Rev.* **1988**, *9*, 608.
- (15) Lynch, J. A.; George, A. M.; Eisenhauer, P. B.; Conn, K.; Gao, W.; Carreras, I.; Wells, J. M.; McKee, A.; Ullman, M. D.; Fine, R. E. *J. Neurosci. Res.* **2006**, *83*, 1262.

From studies addressed to the determination of IDE crystal structure,¹⁶ it was ascertained that the zinc-binding active site stays within a proteolytic cavity formed by defined N- and C-terminal units joined by a loop involving 28 amino acid residues.

The catalytic chamber is large enough to selectively bind and degrade substrates such as insulin and amyloid β ($A\beta$), composed of 51 and 39–43 amino acid residues, respectively. In these cases the enzyme uses an exosite that lies about 30 Å away from the catalytic center to anchor the N-terminus of peptides. This serves to locate suitably the substrate facilitating the cleavage at the catalytic site. However, for substrates less than 12 amino acids long, the anchoring and degrading mechanism is not known since it is impossible for shorter peptides to bind the exosite and the catalytic center at the same time. In any case, recent studies,^{17,18} showed that the rate of catalysis of IDE is higher toward short substrates ($>2000\text{ s}^{-1}$) than toward long substrates such as insulin (about 0.56 min^{-1}).

IDE can exist in two conformations. The first one is an open conformation that allows entry of substrates and exit of products. The other one is a closed conformation, in which the active site, contained within the chamber formed by the two concave domains, is harder to access. The switch from the closed to the open conformation favors the entry of the substrates into the catalytic cavity. In fact, if mutations prevent the closed conformation, a significant increase (about 40-fold) in catalytic efficiency of the enzyme is revealed.¹⁶ This observation suggested a possible therapy for Alzheimer's consisting of changing the conformational preference of IDE to the open state. This should increase $A\beta$ degradation, preventing aggregation and neuronal death.

However, this problem should be addressed carefully since mutation of the active-site residues of IDE can not only reduce catalytic activity but also cause local conformation changes that alter the allosteric properties of the enzyme.¹⁹

Studies present in literature explored many aspects concerning IDE,^{1–24} but none of them deals with the mechanistic details of the proteolysis performed by this enzyme.

In this work, we used density functional theory to explore the catalytic mechanism of IDE in the breaking of two peptides of different lengths, representing a portion of the B chain of insulin hormone. We hoped our work would provide some valuable fundamental insights into the behavior of this enzyme.

Computational Details

Density functional theory (DFT) calculations were carried out with the Gaussian03 program.²⁵ The geometries of reactants,

- (16) Shen, Y.; Joachimiak, A.; Rosner, M. R.; Tang, W. J. *Nature* **2006**, *443*, 870.
- (17) Malito, E.; Ralat, L. A.; Manolopoulou, M.; Tsay, J. L.; Wadlingthorn, N. L.; Tang, W.-J. *Biochemistry* **2008**, *47*, 12822.
- (18) Song, E.-S.; Juliano, A. M.; Juliano, L.; Hersh, L. B. *J. Biol. Chem.* **2003**, *278*, 49789.
- (19) Song, E.-S.; Daily, A.; Fried, M. G.; Juliano, A. M.; Juliano, L.; Hersh, L. B. *J. Biol. Chem.* **2005**, *280*, 17701.
- (20) Caccio, C.; Tundo, G. R.; Grasso, G.; Spoto, G.; Marasco, D.; Ruvo, M.; Gioia, M.; Rizzarelli, E.; Coletta, M. *J. Mol. Biol.* **2009**, *385*, 1556.
- (21) Grasso, G.; Rizzarelli, E.; Spoto, G. *J. Mass Spectrom.* **2009**, *44*, 735–741 (DOI 10.1002/jms.1550).
- (22) Alper, B. J.; Schmidt, W. K. *J. Neurosci. Methods* **2009**, *178*, 40.
- (23) Evin, G.; Weidemann, A. *Peptides* **2002**, *23*, 1285.
- (24) Neant-Fery, M.; Garcia-Ordoñez, R. D.; Logan, T. P.; Selkoe, D. J.; Li, L.; Reinstatler, L.; Leissring, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9582.
- (25) Frisch, M. J., *Gaussian 03*; Gaussian, Inc.: Pittsburgh, PA, 2003.

intermediates, transition states, and products were fully optimized via diverse treatment for atoms belonging to the first and second coordination shells of catalytic zinc ion. The B3LYP/6-31+G(d,p) and B3LYP/6-31G(d,p) levels were used, respectively.^{26–29} The Stuttgart/Dresden (SDD) pseudopotential was applied to describe the metal center.^{30,31} During geometry optimization, some degrees of freedom of the molecular system were constrained to maintain a reasonable approximation to the active-site crystal geometry. This is realized by fixing the atoms to their crystallographic positions where truncation is done. With this procedure the risk exists that the model can become too rigid with significant effects on the energetics. However, it was demonstrated³² that these effects usually do not alter any conclusion concerning the working mechanism of the enzyme.

For each species, vibration analyses were performed to obtain the zero-point energies (ZPE) and verify whether each is a minimum or a transition state on the potential energy surfaces (PES). Intrinsic reaction coordinate (IRC) calculations were performed to confirm that a given transition state connects a particular couple of consecutive minima.^{33,34}

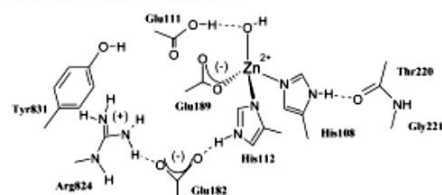
To consider solvent effects, the single-point energies of all structures were calculated with the self-consistent reaction field (SCRF) conductorlike polarizable continuum model (CPCM) method^{35,36} using a dielectric constant value $\epsilon = 4$ that takes into account the coupled effect of the protein itself and the water medium surrounding the protein, according to our previous experience and literature suggestions.^{37–47} United atom topological model (UAO) applied on atomic radii of the UFF force field was used to build the cavity, in the gas-phase equilibrium geometry.⁴⁸

By manual cutting of the crystal structure [PDB code 2G54] of IDE¹⁶ in complex with a portion of insulin B chain, two models for the active site were built up, which consist of 130 and 159 atoms; both models have a total charge equal to zero.

A quite complete identification of zinc ion ligands was done by Perlman and Rosner⁴⁹ in 1994. In this work, on the basis of

- (26) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (27) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (28) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (29) Becke, A. D. *Phys. Rev. B* **1988**, *38*, 3098.
- (30) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866.
- (31) Andrae, D.; Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123.
- (32) Chen, S.-L.; Fang, W.-H.; Himo, F. *Theor. Chem. Acc.* **2008**, *120*, 515.
- (33) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154.
- (34) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523.
- (35) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.
- (36) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.
- (37) Leopoldini, M.; Russo, N.; Toscano, M. *J. Phys. Chem. B* **2006**, *110*, 1063.
- (38) Marino, T.; Russo, N.; Toscano, M. *J. Am. Chem. Soc.* **2005**, *127*, 4242.
- (39) Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* **2000**, *100*, 421.
- (40) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459.
- (41) Leopoldini, M.; Russo, N.; Toscano, M. *J. Am. Chem. Soc.* **2007**, *129*, 7776.
- (42) Leopoldini, M.; Russo, N.; Toscano, M.; Dulak, M.; Wesolowski, A. T. *Chem.—Eur. J.* **2006**, *12*, 2532.
- (43) Leopoldini, M.; Russo, N.; Toscano, M. *Chem.—Eur. J.* **2009**, *15*, 8026–8036.
- (44) Leopoldini, T.; Marino, N.; Toscano, M. *Theor. Chem. Acc.* **2008**, *120*, 459.
- (45) Chen, S.-L.; Marino, T.; Fang, W.-H.; Russo, N.; Himo, F. *J. Phys. Chem. B* **2008**, *112*, 2494.
- (46) Ramos, M. J.; Fernandes, P. A. *Acc. Chem. Res.* **2008**, *41*, 689.
- (47) Russo, N.; Leopoldini, M.; Toscano, M. *Chem.—Eur. J.* **2007**, *13*, 2109.
- (48) Barone, V.; Cossi, M.; Menucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.
- (49) Perlman, R. K.; Rosner, M. R. *J. Biol. Chem.* **1994**, *269*, 33140.

Scheme 1. Active-Site Model of IDE



mutagenesis of some potential zinc-binding residues, it was established that metal center in human IDE is chelated by His108, His112, and Glu189. The more external residues Tyr831, Glu111, Thr220, Gly221, Glu182, and Arg824 were also determined. Among these last residues, mainly Glu182, although not directly involved as metal ligand, plays a fundamental role in influencing metal recognition and binding by zinc proteins.⁴⁹ Glu111 was recognized as the residue that deprotonates the water molecule that completes the coordination sphere of zinc. In fact, in active sites of enzymes, zinc is typically bound to a water molecule involved in nucleophilic attack on the peptide carbonyl. Zinc ion seems to play a double role. It is not only the catalytic center but also contributes to the stabilization of tertiary structure of the proteins by bridging groups into a correct reciprocal orientation. This is confirmed by the fact that its removal, performed by chelation or mutagenesis, entails conformational changes and catalytic activity loss.⁴⁹

In the active-site model we used, illustrated in Scheme 1, imidazole ring, acetate group, guanidinium moiety, and *p*-methylphenol substitute for histidine, glutamate, arginine, and tyrosine amino acid residues, respectively. A CH₃CONHCH₃ peptide bond was used to simulate the linking between threonine and glycine. Substrate CH₃NH-Leu-Tyr-Leu-CONHCH₃ present in the PDB was modeled by a CH₃NH-Ala-Ala-Ala-CONHCH₃ peptide in the first model (130 atoms), while it was considered as such in the second one (159 atoms).

Results and Discussion

As mentioned before, in spite of the enormous attention that has been dedicated to the study of IDE over the past decade, several crucial aspects in the catalytic mechanism of this enzyme remain unexplained. IDE cleaves insulin B chain as well as amyloid β at several sites,¹⁶ however, our study seeks to determine not the particular cleavage site but rather how this cleavage occurs. The steps involved in the whole process and determined through our calculations are collected in Scheme 2. Structural details of each intermediate and transition state along the reaction path in the presence of CH₃NH-Ala-Ala-Ala-CONHCH₃ peptide substrate are given in Figures 1–3. Energetic profiles in gas phase and in protein environment can be found in Figure 4a.

The optimized structure of enzyme–substrate complex (ES) is reported in Figure 1. Comparison of some B3LYP interatomic distances with experimental values (given in parentheses in Figure 1) confirms that the computational approach used is quite reliable. We can note, for example, the similarity between theoretical and experimental values of the coordination distances of His108 (2.060 vs 2.003 Å), His112 (1.974 vs 1.998 Å), Glu189 (2.033 vs 2.047 Å), and deprotonated water (1.963 vs 2.090 Å) to zinc ion.^{16,49} Furthermore, as expected, the role of Glu111 as general base is highlighted by the presence of its hydroxyl hydrogen atom placed almost halfway between this residue and the zinc-bound hydroxide (O₁H–O₂H = 1.483 Å).

Substrate, in ES, is kept in an orientation suitable so that a carbonyl group of the peptide can undergo nucleophilic attack

by the hydroxide linked to the zinc ion. This orientation is allowed by the hydrogen bonds that Arg824, Tyr831, and Glu189 residues establish with two carbonyl oxygens and one amide hydrogen of substrate. This agrees with the role assigned to these amino acids in previous experimental works.^{16,50}

The other two hydrogen bonds, formed by Glu182 with His112 and by the pair Thr220–Gly221 with His118, have a stabilizing effect on the whole structure of enzyme–substrate complex. The distance between the nucleophile and the carbonyl attack site, in ES, is still quite long (C₁–O₁ = 2.894 Å) but in any case of the same order of those computed for other enzymes^{38,41,51} in which the nucleophilic attack is performed by a zinc-bound hydroxide. In fact, zinc(II) is an intermediate acid in Pearson's hard and soft acid and bases (HSAB), which usually imparts to the hydroxyl oxygen a moderate nucleophilicity that prevents the immediate activation of the substrate.

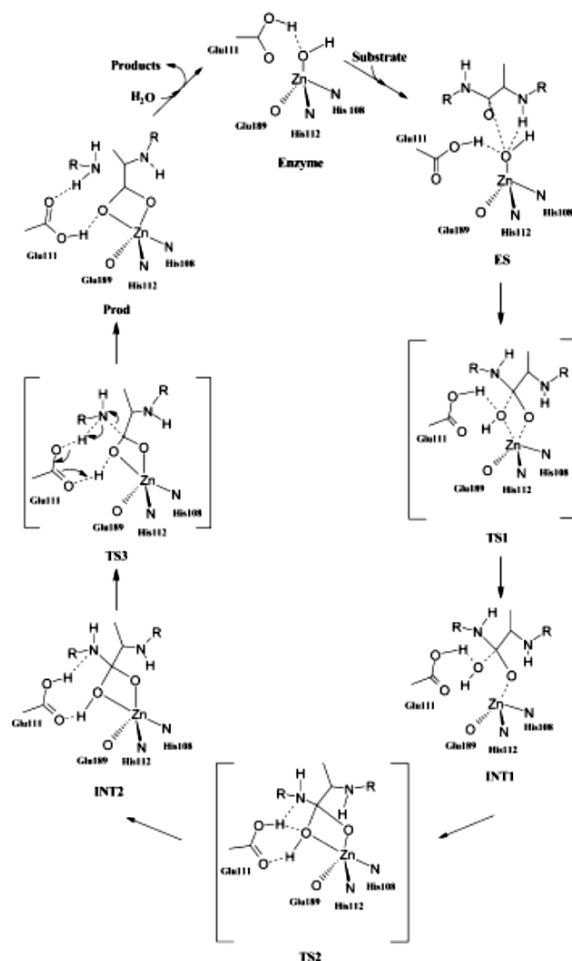
The first step of the mechanism that we propose consists of the mentioned nucleophilic attack that materializes in the intermediate INT1 through the transition state TS1, whose optimized structure is presented in Figure 2. TS1 is characterized by a pentacoordinated zinc ion because of the incipient formation of a bond with the nearest carbonyl oxygen of the substrate (Zn²⁺–O₂ = 2.225 Å). Simultaneously, the pre-existing Zn²⁺–O₁H bond lengthens by 0.237 Å with respect to its value in ES complex. A lengthening by about 0.08 and 0.07 Å can be observed also for C₁–O₂ carbonyl and the C₁–N₁ distance in the substrate, respectively. TS1 transition state was confirmed to be a first-order saddle point with an imaginary frequency of 107i cm⁻¹ that is associated with the stretching vibrational mode of the incoming Zn²⁺–O₂ bond. Hydrogen-bond interactions present in the enzyme–substrate complex are all retained, although with values different from those obtained in the previous step (see Figures 1 and 2). TS1 lies at 17.2 kcal/mol above ES, as can be observed from Figure 4a. The barrier for nucleophilic attack in the gas phase is thus quite in line with those encountered in enzymatic catalysis. Upon addition of protein environment effects, the barrier becomes slightly lower (15.9 kcal/mol), suggesting a slight stabilization due to the low polarity of solvent.

In the next step, the intermediate INT1 is formed (see Figure 2). In this species, we can note that the zinc-bound hydroxide is completely transferred on the carbonyl carbon of substrate (C₁–O₁H = 1.540 Å) as a consequence of the Zn²⁺–O₁H (2.405 Å) bond breaking. From the energetic point of view, INT1 is found at about 0.1 and 0.2 kcal/mol below TS1 in gas phase and in solvent, respectively. This light difference in the energy values can be due to the great resemblance of two stationary points. In fact, the change of geometrical parameters in going from TS1 to INT1 is on the whole very small.

Moving along the gas-phase potential energy profile, a new transition state TS2 is encountered that lies at 4.2 kcal/mol above INT1 (see Figure 4a). In TS2, the imaginary frequency of 181i cm⁻¹ describes the rotation of Glu111 residue that assumes the right disposition to form two hydrogen bonds with the amide nitrogen (HN₁–HO₃ = 2.238 Å) and the –OH group (C₁O₁–HO₄ = 2.120 Å) linked to the carbon atom of substrate, thus behaving as hydrogen donor and acceptor, simultaneously. The formation of the second cited bond favors the re-establish-

(50) Maskos, K. In *Handbook of Metalloproteins*; Messerschmidt, A., Dode, W., Cygler, M., Eds.; John Wiley & Sons: New York, 2004; pp 190–198.

(51) Bottoni, A.; Lanza, C. Z.; Miscione, G. P.; Spinelli, D. *J. Am. Chem. Soc.* **2004**, *126*, 1542.

Scheme 2. Chemical Steps in the Proposed Mechanism of Proteolysis by IDE

ment of the Zn²⁺-O₁H bond (2.285 Å) broken previously at the INT1 level. The peptide N₁-C₁ bond of the substrate continues gradually lengthening in all the stationary points along the reaction path. The hydrogen bonds between the Arg824 residue and the terminal CO group of the substrate assume in TS2 values quite different from those observed in the other structures. The same occurs for the hydrogen bond between the hydrogen of His108 and the carbonyl oxygen of the simple moiety used to simulate Thr220 and Gly221 amino acids (see Figure 2).

Solvation effects do not significantly change the energetics of the step that leads from INT1 to TS2. The barrier of 4.2 kcal/mol calculated in gas phase is only 0.7 kcal/mol lower in a protein environment.

The intermediate INT2 (see Figure 3) that originates from TS2 shows the strengthening of the hydrogen bonds between Glu111 and the substrate. HN₁-HO₃ and C₁O₁-HO₄ distances are now 1.756 and 1.722 Å long, respectively. The peptide N₁-C₁ bond that must be broken in the substrate further lengthens by 0.05 Å. All other hydrogen bonds are retained with values that in some case are slightly different from those present in TS2 because of the reorganization of amino acid residues.

In gas phase, INT2 is found to be more stable than TS2 by about 2.0 kcal/mol. On the contrary, in a protein medium, it becomes 4.4 kcal/mol more stable than TS2. A look at the previous part of the path, in which the relative differences in energy between the various minima and transition states appear to be quite small, indicates now a moderately larger solvent

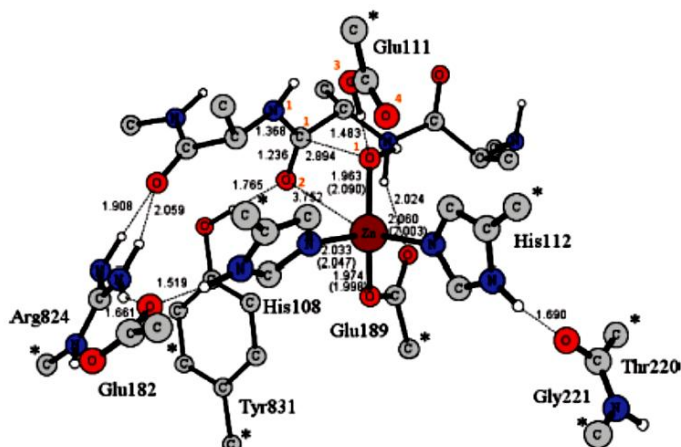


Figure 1. Optimized structure of the enzyme- $\text{CH}_3\text{NH-Ala-Ala-Ala-CONHCH}_3$ substrate complex. Only the most significant hydrogen atoms are reported. Stars indicate the atoms kept fixed to their crystallographic positions during optimizations. Atoms directly involved in the catalytic reaction are numbered in orange (in the other figures this numeration is not reported but it is the same). Distances are in angstroms. Values in parentheses come from refs 16 and 48.

effect. Large solvent effects usually occur when charges are created or quenched in the proximity of the edge of the quantum active-site model. So, in the present case, we could attribute them to the major negative charge distributed on the Glu111 residue oxygens in INT2 rather than in INT1 and due to the incipient hydroxyl proton transfer on N_1 atom of substrate.

The following point on the reaction path is the transition state TS3 (see Figure 3). The most significant peculiarities in this species are the definitive shift of the proton from Glu111 to the amide nitrogen of the substrate and the consistent lengthening (about 0.2 Å) of peptide $\text{N}_1\text{-C}_1$ bond. Besides, both these processes are accompanied by the breaking of the $\text{Zn}^{2+}\text{-O}_2\text{C}_1$ bond that restores the tetrahedral environment around the catalytic ion.

Imaginary frequency visualization program shows that the three chemical events mentioned above occur simultaneously at $158i\text{ cm}^{-1}$. All other stabilizing interactions are always present and the values of distances can be found in Figure 3.

As in other catalytic cycles^{43,45} described by flat energetic profiles involving species of comparable energy, we have found that TS3 lies 1.4 kcal/mol below INT2, in gas phase. When the solvent effect is added, the difference in energy becomes 2.2 kcal/mol, with TS3 again more stable than INT2. Usually, on the basis of similar results, conclusions were drawn that hypothesize that nucleophilic addition and the protonation of peptide amide nitrogen may occur as a single step in the catalytic process. In other words, we could say that the protonation of amide nitrogen is a very fast process.

The final species on the path is the product PROD (see Figure 3). As a consequence of transfer of the proton of Glu111 onto the amide nitrogen of substrate that occurred in TS3, the peptide $\text{N}_1\text{-C}_1$ bond is broken, as is evident from the distance value of 4.290 Å.

One of the peptide fragments (C-terminus) remains neutral and anchored to the Arg824 residue, while the other one (N-terminus) is still linked to the zinc ion. In PROD, Glu111 is in the protonated state. To restore the active form of the enzyme, some residue

present in the protein environment should snatch the proton of Glu111 before the entry of a new water molecule in order to allow this amino acid to carry out its base function. However, this aspect is not thus important in our investigation, since our aim was to describe in detail the peptide hydrolysis phase.

PROD lies 7.0 kcal/mol below the enzyme-substrate complex so that the reaction is exothermic. The exothermicity increases in solvent by 5.8 kcal/mol (see Figure 4a).

A look at the whole reaction path indicates that the rate-determining step in this process is the nucleophilic addition and that after this point the catalytic event should proceed without particular obstacles. Considering the particular location of TS3, that is, the moment in which the protonation of amide nitrogen of peptide substrate occurs, and the small energetic expense to cross TS2, we could hypothesize after TS1 a descendent pathway.

Results just described concern our first model in which the $\text{CH}_3\text{NH-Leu-Tyr-Leu-CONHCH}_3$ substrate present in the PDB was simulated by a $\text{CH}_3\text{NH-Ala-Ala-Ala-CONHCH}_3$ peptide, having replaced all side chains by methyl groups. In order to verify that such a simplification does not affect the specificity of the substrate recognized by the enzyme, all calculations were redone removing this simplification and reintroducing all the side chains of substrate amino acids residues.

Results indicate that reaction occurs following the same steps and the same mechanism as in the previous case. All stationary points of the new energetic path (reported in Figure 4b) show geometrical parameters that, although similar to those of the corresponding structures obtained in the case of the smaller peptide substrate, present some slight differences due essentially to the higher steric hindrance. The optimized structure of the enzyme- $\text{CH}_3\text{NH-Leu-Tyr-Leu-CONHCH}_3$ substrate complex and geometrical features of its stationary points can be found in Figures S1-S3 in Supporting Information.

As can be noted, the energetic paths of Figure 4a,b for the two examined peptides substrates are evidently comparable. TS1, which also in this second case represents the rate-determining step of the reaction, is lower by only 1.5 kcal/mol (in gas phase)

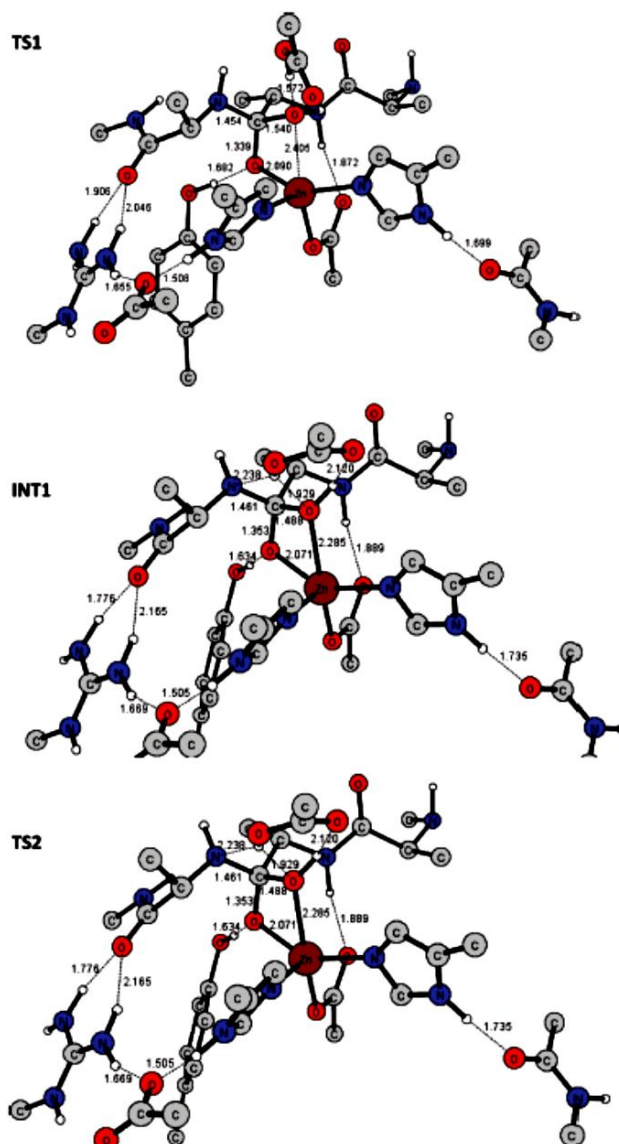


Figure 2 Stationary points (TS1, INT1, and TS2) on the IDE energetic profile for $\text{CH}_3\text{NH-Ala-Ala-Ala-CONHCH}_3$ substrate.

and 0.3 kcal/mol (in proteinlike medium) with respect to the corresponding TS1 obtained with the substrate model. The vibrational frequency of $120i\text{ cm}^{-1}$ appears to be associated to the stretching mode of the $\text{Zn}^{2+}-\text{O}_2$ bond as before.

The similarity between the relative energy values of TS1 and the following INT1 found in the path of Figure 4a is proposed

again in the path of Figure 4b and can be explained still by their structural resemblance.

The rotation of Glu111 residue occurs at a frequency value of $153i\text{ cm}^{-1}$ and requires an expense of 6.3 and 5 kcal/mol in gas and in protein phase, respectively. These values are slightly

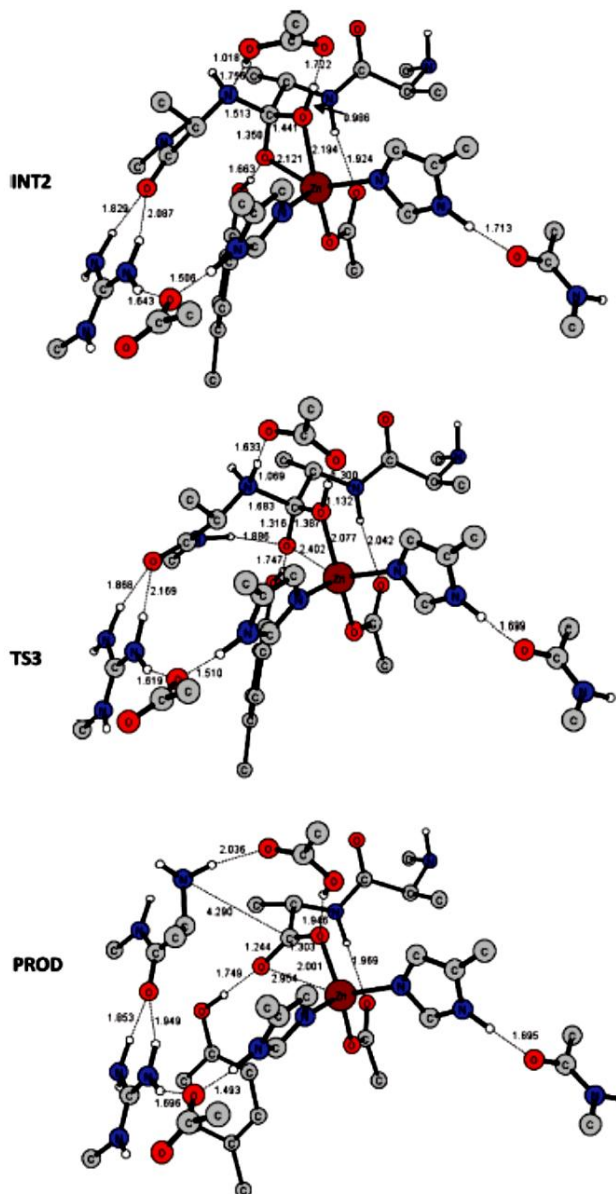


Figure 3. Stationary points (INT2, TS3, and PROD) on the IDE energetic profile for $\text{CH}_3\text{NH-Ala-Ala-Ala-CONHCH}_3$ substrate.

higher than those obtained with the smaller substrate but it seems quite reasonable that the side chains present in the larger substrate can to some extent block Glu111 to realize the proper orientation to form the two hydrogen bonds with the amide

nitrogen and the $-\text{OH}$ group linked to the carbon atom of substrate. This in turn, as mentioned before, allows re-formation of the $\text{Zn}^{2+}-\text{O}_\text{H}$ bond and the lengthening of the peptide N_1-C_1 bond of the substrate.

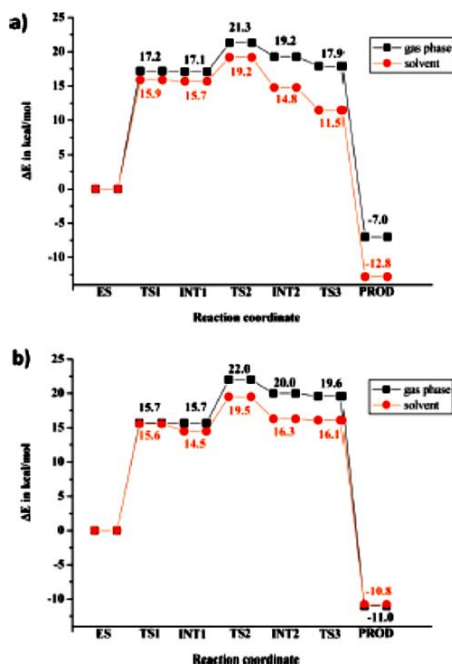


Figure 4. Energetic profile of proteolysis mechanism of IDE toward (a) CH₃NH-Ala-Ala-Leu-CONHCH₃ and (b) CH₃NH-Leu-Tyr-Leu-CONHCH₃ substrates.

The energy difference between TS2 and the next INT2 in gas phase is practically the same as found in the previous case (2.0 vs 2.1 kcal/mol), but now it is more restrained in solvent (3.2 vs 4.4 kcal/mol).

TS3 lies below INT2 as before, but the presence of the larger substrate reduces the energy difference between them from 1.1 to 0.4 kcal/mol in gas phase and from 3.3 to 0.2 kcal/mol in protein environment. The calculated frequency of 88i cm⁻¹ describes the proton shift from Glu111 residue to the amide nitrogen atom of the substrate and the simultaneous lengthening of peptide N₁-C₁ bond.

Finally, the product is obtained exothermically with an energy gain of 11.0 and 10.8 kcal/mol in the two phases, respectively. These two values are much more similar than those obtained in the path of Figure 4a, probably because in the presence of the bulky aliphatic side chains of substrate the active site is less accessible to the solvent.

The concordance of the results in the two examined cases, especially as far as the evaluation of the energetic barrier in the rate-limiting step of the reaction is concerned, suggests that reliable information can be obtained also by using a model for the substrate. In fact, at least in the present case, the removal of the side chains from substrate present in the PDB does not entail significant variations in the catalytic behavior of the enzyme. However, we think that this is no accident but depends on the manner in which the model of substrate is realized. If

this is made in the right way, apart from the reliability of the results that is the most important goal, the additional advantage of a computational cost reduction can be reached.

Conclusions

In this work, we have reported a B3LYP investigation of the possible reaction mechanism of hydrolysis of two short peptides by insulin-degrading zinc enzyme, whose active-site model was constructed on the basis of crystal structure. The model, which consists of 130 or 159 atoms depending on the peptide length and has a total charge equal to zero, was built up large enough to allow reliable results. All stationary points along the reaction path were localized and characterized in gas phase and in protein environment. Results common to both examined cases can be summarized as follows:

The enzyme carries out its catalytic function through three fundamental steps that involve first a nucleophilic attack by a zinc-bound hydroxide to a substrate carbon. After a structural rearrangement of the key Glu111 residue that works as a shuttle, a proton is transferred onto an amide nitrogen of substrate and the peptide C-N is broken.

The role of amino acid residues suggested by previous experimental studies was fully confirmed by our results, from which can be deduced their active participation in the catalytic process or their presence as anchorage points for substrate and/or stabilization factors.

The rate-determining step of the reaction is the nucleophilic addition that implies an expense of 17.2 (15.7) kcal/mol, which is reduced to 15.9 (15.6) kcal/mol in protein medium when the shorter (longer) peptide substrates are considered. Calculations indicate that the next chemical events along the path are fast and lead to a product more stable than the starting enzyme-substrate complex by about 7.0 (11.0) kcal/mol in gas phase and 12.8 (10.8) in solvent for the first and second substrate, respectively.

The reliability of the model used to simulate the substrate was ascertained by calculations taking into account the real substrate as in the PDB structure. The similarity of the obtained results indicates that approximation of side chains with methyl groups is valid and allows a significant reduction of computational time.

The agreement with the unfortunately scarce experimental information concerning the working mechanism of this enzyme can be considered a sufficient validation for the theoretical approach used here, together with the fact that it was widely used previously for studying enzymatic reactions.

To our knowledge, this is the first investigation providing details on the mechanistic working of these types of enzymes. Thus, we hope that it will be used to stimulate future theoretical and experimental improvements.

Acknowledgment. We gratefully acknowledge the Dipartimento di Chimica, Università della Calabria, for financial aid.

Supporting Information Available: Three figures showing optimized structure of the enzyme-CH₃NH-Leu-Tyr-Leu-CONHCH₃ substrate complex and its stationary points (TS1, INT1, TS2, INT2, TS3, and PROD) on the IDE energetic profile. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA9037142

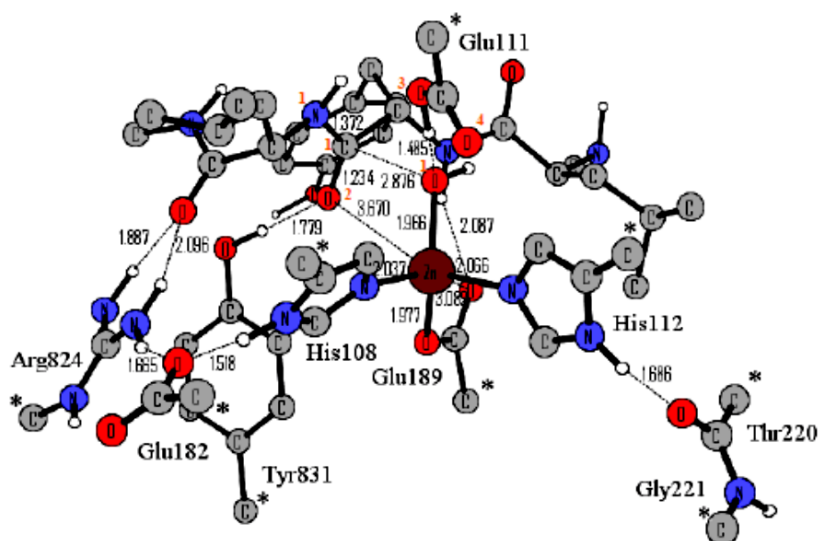


Figure S1: Optimized structure of the enzyme-CH₃NH-Leu-Tyr-Leu-CONHCH₃ substrate complex. Only the most significant hydrogen atoms are reported. Stars indicate the atoms kept fixed to their crystallographic positions during optimizations. Atoms directly involved in the catalytic reaction are numbered in orange (in the other figures this numeration is not reported but it the same). Distances are in Å.

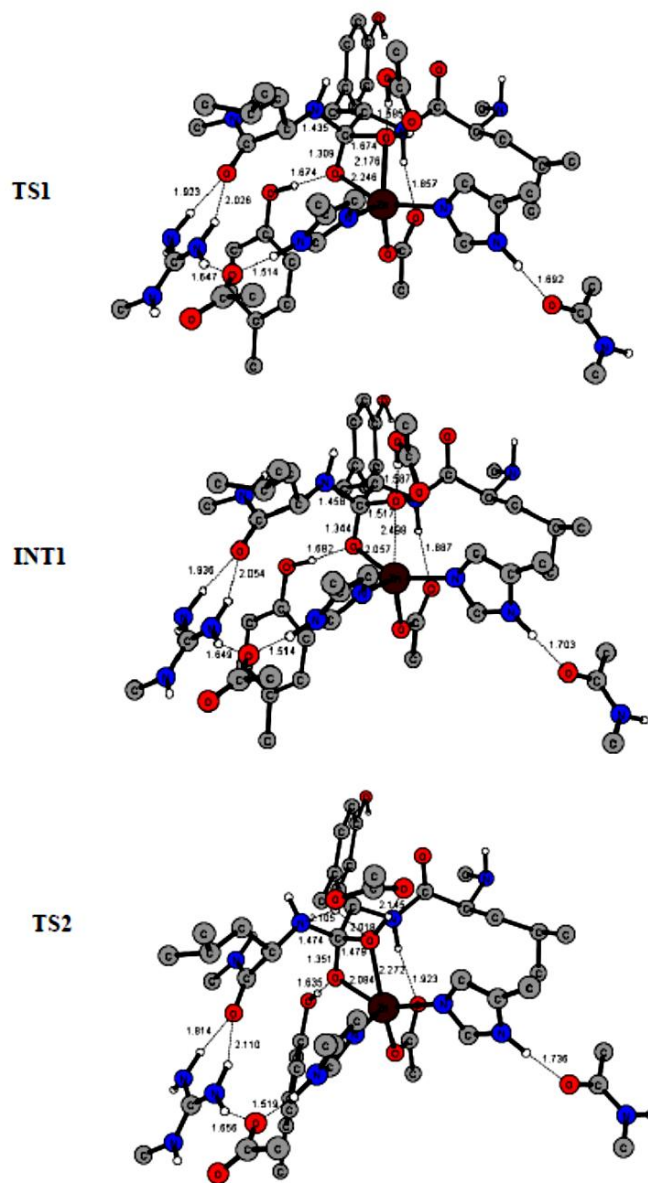


Figure S2: Stationary points (TS1, INT1 and TS2) on the IDE energetic profile.

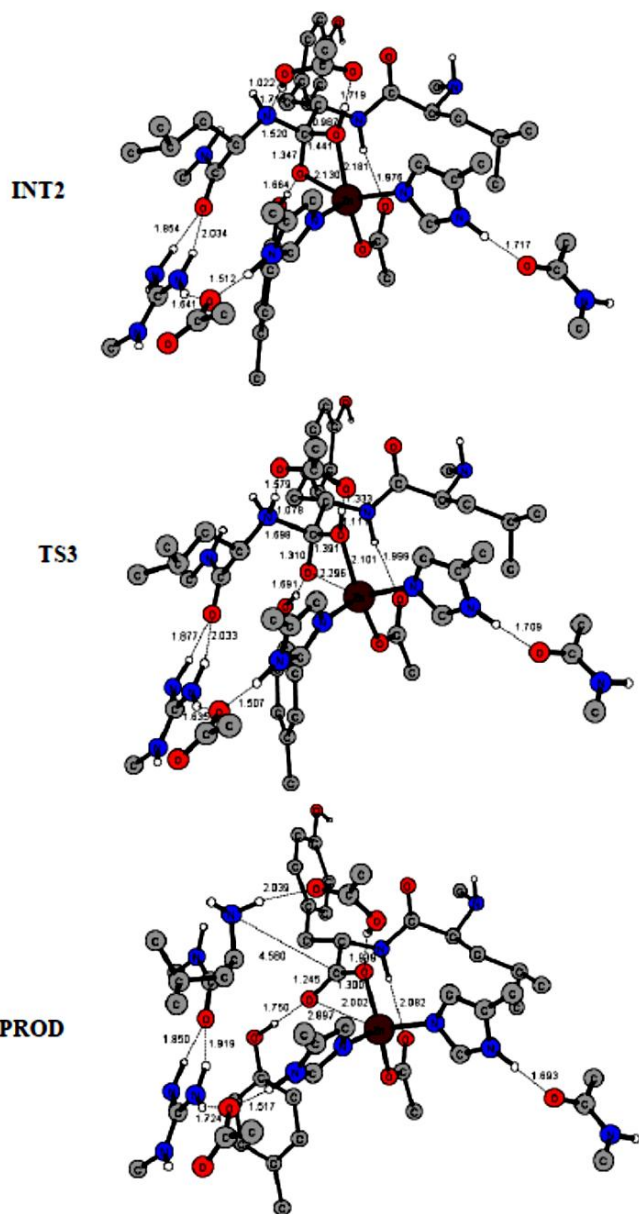


Figure S3: Stationary points (INT2, TS3 and PROD) on the IDE energetic profile

ARTICOLO II

*Catalytic activity of a ζ -class zinc and cadmium containing
carbonic anhydrase. Compared work mechanisms.*

Orazio Amata, Tiziana Marino, Nino Russo, Marirosa Toscano

Phys. Chem. Chem. Phys., DOI:10.1039/C0CP01053G

Catalytic activity of a ζ -class zinc and cadmium containing carbonic anhydrase. Compared work mechanisms

Orazio Amata^a, Tiziana Marino^a, Nino Russo^a and Marirosa Toscano^a

Received (in XXX, XXX) Xth XXXXXXXXX 200X, Accepted Xth XXXXXXXXX 200X

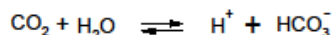
First published on the web Xth XXXXXXXXX 200X

DOI: 10.1039/b000000x

The carbonic anhydrase is the enzyme that catalyzes the reversible hydration of carbon dioxide and represents one of the most ancient proteins to which a plethora of works was devoted. The three main classes rely on zinc ion for activity. Most recently a new class of CA was discovered in marine diatoms to use naturally cadmium ion as catalytic metal. In the present investigation we focused our attention on a carbonic anhydrase *cambialistic* enzyme (CDCA1) belonging to this new class. The study was inspired by the discovery that the replacement of zinc ion with cadmium does not entail significant differences in the catalytic performance of enzyme. Our aim was to give further insights on enzymatic work mechanism. Different possible reaction paths were considered for both metallic forms of enzyme and comparison with previous studies concerning other carbonic anhydrases was made. The effects of solvent on the energetic of catalytic process, was also taken into account by means of polarizable continuum model. Results obtained from density functional calculations, using a well consolidated mixing of exchange-correlation potential and basis set, and performed with a model of the active site designed on the basis of the X-ray crystal structure, proposed for both metal ions similar reaction pathways consisting in the nucleophilic attack by the metal bound hydroxide to the carbon dioxide with bicarbonate formation, in a next internal rotation of this last fragment, and then in the formation of a species ready for the product removal. Similar activation barriers were found in the rate determining steps that confirm the experimental indication concerning the comparable efficiency of enzyme in the presence of zinc or cadmium metal ion.

1. Introduction

Carbonic anhydrase (CA) is a metalloenzyme firstly identified, in 1933, in red blood cells of cows. Since then, it has been found to be abundant in all mammalian tissues, plants, algae and bacteria.¹ CA assists rapid inter-conversion of carbon dioxide and water into carbonic acid, protons and bicarbonate ions:



The characterization of CAs from different beings showed structural variances between these enzymes so they can be allocated in three classes: α -, β - and γ -carbonic anhydrases.

The α -class is common mainly between mammalian where it is present in 11 isoforms, some of that implicated in several diseases.² Up to now, these enzymes, which are monomeric, are the more extensively studied among the three classes as far as the catalytic mechanism is concerned. They present an active site involving three histidine ligands.

The β -class, prevalent in plants, algae, bacteria and archaea domains, differs from the previous one in the active site, made up of two cysteine and one histidine ligands, and in the fact that these enzymes are polymeric³, i.e. they exist like dimer, tetramer, hexamer and so on.

The γ -class is composed, actually, of only one omotrimeric enzyme found in the archaeon *Methanosarcina Thermophila*. The active site of this enzyme is similar to the active site of α -class enzymes, being formed by three histidine ligands, but coming from different monomer.

Despite the structural differences, all the CAs require a Zn^{2+} ion for the catalytic activity: the replacement of this metal ion with other divalent metal ions usually entails a lowering of the activity. Also, the α -class enzymes present an affinity towards Zn^{2+} ion at subpicomolar level.

Moreover, the hydration mechanism follows a three steps single general scheme.

In the first step a water molecule bound to the metal center is deprotonated to form a hydroxide ion that in the second step yields a nucleophilic attack on the CO_2 with formation of the HCO_3^- ion. In the last step a new water molecule approaches the

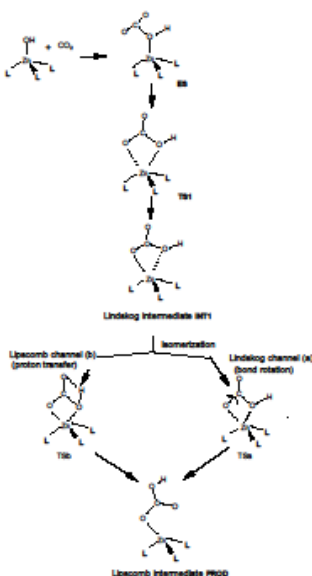
metal ion and displaces the product. According to this scheme the metal ion acts as a Lewis acid lowering the pK_a of water from about 15.7 to about 7.

The catalytic mechanism of CA was the subject of many experimental¹⁸⁻¹⁹ and theoretical²⁰⁻⁴¹ investigations, pointing out that the rate limiting step is the deprotonation of metal bound water molecule, at least for those enzymes having values of $k_{cat} > 10^6$. The catalytic efficiency, instead, reflects mainly the hydration step and is independent on the proton transfer. Many attempts were made even to elucidate the details of the reaction of formation of the HCO_3^- ion.

Presently, there are two hypotheses illustrated in Scheme 1. Both indicate cycle to begin with a nucleophilic attack by the metal bound OH^- ion on CO_2 with the formation of a bicarbonate intermediate (Lindskog intermediate)⁴². This intermediate can evolve toward another more stable intermediate (Lipscomb intermediate)²³ by an internal rotation (Lindskog mechanism) or by a proton transfer from the OH^- to another oxygen atom of the HCO_3^- ion (Lipscomb mechanism)⁴³.

The attempts done to determine the lowest path were unsuccessful: both these mechanisms have comparable energetics so there is no consensus about that preferred by the enzyme.

To give new insights in the CA, other enzymatic forms were characterized from several organisms. However, some of these new forms found had no reference to any of the classes named.



Scheme 1.

In 1997 Morel et al. related the purification of a new form of CA from the diatom *Thalassiosira weissflogii* (TWCA1) with an

active site like that of α -class enzymes, but with a genes sequence different from all those known for this enzyme. They proposed the introduction of a fourth class: the δ -class⁴⁴⁻⁴⁶.

This diatom, further, in event of low concentration of Zn^{2+} ions, like in the sea water, and in presence of Cd^{2+} or low CO_2 pressure, produces another form of CA (CDCA1)^{44,47-50}.

This new CA is composed by three repeats sharing the 85 % of sequence similarity, each of them containing a Cd^{2+} ion and able to catalytic activity. In addition, the active site of CDCA1 resembles that of the β -class enzymes, but there are no structural and genes sequence similarities pointing this enzyme as representative of a new enzymatic class⁵⁰, widespread in diatoms and common in the ambient⁵¹. This new class is named ζ -class.

The discovery of this last enzyme explains the behaviour like an algal nutrient in the sea water of cadmium despite its toxicity for all the living organisms⁵²⁻⁵³.

The crystallographic structure of CDCA1 was recently determined⁵⁴ showing its unique ability to exchange Cd^{2+} with Zn^{2+} keeping with both these metal ions a high catalytic efficiency, while the other known enzymes are severely inhibited by the Cd^{2+} ions, perhaps due to lower acidity of this last metal species and to the different coordination chemistry with respect to Zn^{2+} .

A recent theoretical investigation showed that, in the case of α -class enzymes containing Cd^{2+} , the reaction path was characterized by deep intermediates and very high energy barriers⁵¹.

In the CDCA1 case, instead, the structural and ligands differences should justify the different behaviour suggesting the possibility that another catalytic mechanism can operate.

With the aim to confirm this experimental finding⁵⁴, we have thus performed a density functional study of the mechanistic aspects of the enzyme catalytic pathway in the presence of the zinc and cadmium metal ions.

1. Computational Details

Geometry optimizations were performed using Density Functional Theory (DFT) with the B3LYP functional which includes the Becke's hybrid⁵⁵ exchange and correlation functional of Lee, Yang and Parr^{56,57} as implemented in Gaussian 03 package⁵⁸. Equilibrium structures were obtained utilizing the 6-311+G(d,p) basis set for heavy elements and hydrogen's, while transition metals (Zn and Cd) were described by Stuttgart-Dresden ECP pseudopotentials^{59,60}. Minima and transition states on potential energy surfaces (PES) were confirmed by vibrational analysis. Relative energies were all corrected by zero point vibrational contributions (ZPE).

Analysis of Natural Bonding Orbitals (NBO)⁶¹ was performed to estimate net charges within all stationary points on the PES.

Solvent effects were computed as single points determination on previously optimized geometries applying the Self Consistent Reaction Field Polarizable Continuum Model (SCRF/PCM)^{62,63} using different dielectric constant values (ϵ : =4 recommended by many works⁶⁴⁻⁶⁹ to simulate the bulk of protein environment and ϵ =80 since the active site is solvent exposed). Furthermore, limited to the first two stationary points of the energetic profiles (ES and TS1) of both enzymes, other two sets of calculations using the dielectric constants of 2 and 6 very close to ϵ =4 were performed in order to evaluate the effects deriving from small changes in the dielectric constant values.

Results have demonstrated that variation of relative energy values is less than 1 kcal/mol in going from $\epsilon=2$ to 4 to 6. For this reason, we have chosen to discuss only the data obtained from calculations with $\epsilon=4$ and $\epsilon=80$.

For both considered metal forms of enzyme, active sites were simulated on the basis of the crystallographic structures determined by Morel et al. at 1.45 Å resolution.³⁴ In particular, the Zn-CDCAl [PDB code: 3BOC] shows a metal center coordinated to the oxygen of the water molecule 1002 (2.27 Å) to the sulphur atoms of Cys263 (2.30 Å) and of Cys325 (2.34 Å) and finally to the nitrogen (2.15 Å) of His315. Other two water molecules lying at 2.47 and 2.97 Å from water bound to the zinc ion were retained in the model.³⁴ In the case of Cd-CDCAl [PDB code: 3BOB], a different geometrical rearrangement is observed because of the presence of a further water molecule (1003) coordinated to the cadmium ion at 2.65 Å.³⁴ In both cases the two second shell conserved residues, Asp265 and Arg267, present in the active sites were added to complete the models (see Scheme 2). These two amino acids are considered essential for the catalytic activity, since their mutation entails the inactivation of CDCAl.³⁴

Following a common practice⁶⁴⁻⁶⁹, the ligands were truncated such that histidines were represented by methylimidazoles, aspartates by acetates, cysteines by CH₃S- and arginine by a guanidine derivative. In order to preserve the spatial arrangement of amino acid residues and avoid their unrealistic translations, the atoms where the truncation was done were kept fixed during the geometry optimizations.⁶⁴⁻⁶⁹ The risk that this procedure can produce effects on the energetics of the examined processes is real but it was demonstrated that the eventual differences usually do not alter any conclusions about the working mechanism of the enzyme.^{65,70} As pointed out in recent works,⁷¹⁻⁷⁴ the presence of atoms kept frozen to their X-ray positions during optimization, gives rise to small imaginary frequencies, all lying below 50i cm⁻¹. These do not contribute significantly to the ZPE and thus they usually are ignored. However, it is worth to note that they generate inaccuracy in the calculations of harmonic entropy effects. For this reason, these were not included in the present investigation.

Active site model including the substrate consists of 54 atoms and the total charge is -1.

1. Results and Discussion

In the absence of any proposal of catalytic mechanism for Cd-CDCAl, we have initially thought to explore the possibility that the present enzyme as well as its analogous with zinc (Zn-CDCAl) could follow one of the two cycles proposed by Lipscomb and Lindsag^{62,63} mentioned before. For the nucleophilic attack on CO₂, we have chosen the metal-bound hydroxide rather than the metal-bound water form of the enzyme for both ions. This choice is supported by a series of indications that underline metal-bound hydroxide as the active form of enzyme.^{23,34,40,41,70}

To obtain the initial structure of the enzyme-substrate complex, carbon dioxide was placed following the indications reported by Morel et al.³⁴ about the substrate binding mode. In fact, the authors propose that the acetate, used as substrate analogue in the Cd-CDCAl, replaces one of the two water molecules that were hydrogen bonded to the metal ion in the acetate-free structure.³⁴ However, besides that indicated in the experimental work,³⁴ other orientations of CO₂ with respect to the hydroxyl group bound to the metal ion, were taken into account. After optimization, all

these attempts collapsed in a single enzyme-substrate complex (ES later discussed). Because of the absence of any initial van der Waals complex, the CO₂ activation results to be a barrierless process for both studied metal ions. This is practically in agreement with the findings of a previous theoretical work that, for this step, estimated a barrier of about 0.6 kcal/mol.⁴⁰

In agreement with what experimentally supposed³⁴, we verified that the metal ion exchange does not produce significant differences in the working of the enzyme.

In Figure 1, the mechanisms shared by both cadmium and zinc enzymes which we obtained and we propose to explain how they carry out catalysis, are presented.

Cd-CDCAl mechanism

The energetic values in gas phase and in both condensed media of the stationary points belonging to the reaction path of cadmium enzyme are reported in Figure 2a, while their geometrical main features and the atom labels can be found in Figure 3.

Reaction begins by the nucleophilic attack of the cadmium bound hydroxyl oxygen on the carbon atom of the CO₂ substrate. In the tetrahedral enzyme-substrate complex (ES), that is obtained from separated reactants without energetic barrier, the distance between these two atoms is 1.421 Å.

In confirmation of that, NBO analysis suggests oxygen and carbon to participate to this covalent bond overlapping their *s*(29%) *p*(71%) and *s*(26%) *p*(74%) orbitals, respectively.

ES is characterized by the presence of a hydrogen bonds network that involves water molecules, Arg267 and Asp265 residues, and the substrate. This last one, on the nucleophilic attack, presents one of the two C-O bonds slightly longer than the other (C-Oa = 1.236 Å and C-Ob = 1.253 Å, see Figure 3) and a Oa-C-Ob angle of 130.4°.

The activation phase of CO₂ is followed by a transition state (TS1). In this structure the hydrogen carbonate group acts as bidentate ligand to metallic center (Cd²⁺-OH = 2.593 Å and Cd²⁺-Oa = 2.429 Å) causing a new disposition of the two water molecules and the formation of a new hydrogen bond between them. The imaginary frequency of 176i cm⁻¹ found for this transition state, correlates with the stretching of the OH—Cd²⁺ bond concerted with the rotation of the closest water. The Oa atom goes nearer the metal center as indicated by the Cd-Oa bond shortening of about 1.5 Å with respect to the value calculated in the previous ES adduct. The network of hydrogen bonds between Arg267 and Asp265 residues is retained although some bond distances are slightly different as compared to before due to the appearance of a new hydrogen interaction involving Asp265 and cadmium bound hydroxyl.

Reaching TS1 requires an energetic expense of 9.2 kcal/mol that reduces to 3.4 and 6.2 kcal/mol the $\epsilon=4$ and $\epsilon=80$ dielectrics, respectively (see Figure 2a).

The catalytic event goes on with the pentacoordinated intermediate (INT1) where the Cd²⁺-OH bond assumes the value of 2.832 Å. One of the two water molecules establishes a new hydrogen bond with the Ob atom. From the energetic point of view, INT1 is 1.8 kcal/mol less stable than ES, but its stabilization degree with respect to ES improves in protein and in the polar solvent as can be noted by the values of the relative energies on the paths (-4.1 and -8.9 kcal/mol, see Figure 2a).

The evolution of INT1 towards the Lipscomb intermediate⁽²³⁾ called PROD (see scheme 1 and Figure 3) can occur through two distinct reaction channels, named channel A and channel B (see Figure 2a).

The channel A leads INT1 to PROD simply via an internal rotation of the bicarbonate moiety around the C-Oa bond (TS4,

Lindskog mechanism). This rotation implies an energetic expense of 1.6 kcal/mol in gas phase, 5.9 kcal/mol in protein and 5.8 kcal/mol in water.

The nature of TS4 was confirmed by a vibrational analysis. Computations give an imaginary frequency of $68i \text{ cm}^{-1}$ corresponding to the previously indicated rotation.

The species PROD originated by INT1 via TS4, is now a four-coordinated complex in which the Cd-Oa bond is longer by 0.06 Å than in INT1 (2.390 Å). The Cd-Ob bond is once and for all broken (3.646 Å). PROD presents the bicarbonate fragment to some extent linked to enzyme but the major length of Cd-Oa bond with respect to that present in INT1 (a pentacoordinated complex), should facilitate its expulsion by an incoming water molecule, in order to restore the catalyst.

Reaction appears to be slightly endothermic in gas phase but, in protein and in water it becomes slightly exothermic (1.7 kcal/mol vs -2.8 and -3.5 kcal/mol).

In gas phase and in water, TS1 represents the rate limiting step for the process requiring only 9.2 and 6.2 kcal/mol, respectively. Instead, this role in protein is up to TS4 (5.9 kcal/mol).

These results are significantly different from those obtained in a previous study concerning the catalytic mechanism of carbonic anhydrase (Cd-CA) from *Thalassiosira weissflogii* diatom⁴¹. In that case, very high activation barriers were found suggesting a process improbable from both kinetic and thermodynamic points of view. However, it is worth to note that the events involved in the rate determining step of these two enzymes are different as well as the aminoacid residues coordinating to the metal ion (two cysteines and one histidine in the present case against three histidines in the previous work⁴¹) and the total charge of the active site.

The alternative reaction path (channel B) mentioned before, proposes after INT1, a double proton transfer (TS2) involving the bicarbonate, the arginine and aspartate residues and a water molecule. TS2 requires the clearing of a barrier of about 13 (in gas phase), 23 kcal/mol (in protein-) and 25.3 kcal/mol (see Figure 2a).

The visualization of the imaginary frequency ($1594i \text{ cm}^{-1}$) in TS2 indicates that three protons are simultaneously migrating from the amino group of Arg267 to the water molecule, from water to the bicarbonate oxygen Ob and from the bicarbonate hydroxyl oxygen to the O2 of Asp265 (see Figure 3 for hydrogen bond lengths). The triple proton transfer gives rise to the intermediate INT2 lying 12.7 kcal/mol above ES, where the aspartate is protonated and the arginine residue is deprotonated (see Figure 3). INT2 is a tetra-coordinated complex with a molecular arrangement of bicarbonate moiety very similar to that present in the final product (PROD) (see Figure 3). A very small barrier of 2.8 kcal/mol (TS3, see Figure 3) must be overcome to transfer back one proton from Asp265 to Arg267, in order to establish again the H-bonds network between the amino acids pair. The imaginary frequency of $279i \text{ cm}^{-1}$ found in this case confirms the proton transfer.

In the protein medium, the TS2→INT2→TS3 transformation determines a region very flat of the path and occurs at energetic values higher than those found in gas phase.

The behaviour in water is quite different since TS2 is strongly destabilized while INT2 and TS3 appear to be much more stable than the corresponding stationary points in both other two media.

According to the path B, TS2 is the rate determining step in all phases.

The channel B (Lipscomb mechanism) results to be characterized by higher barriers than those present in the channel A (Lindskog mechanism) and should not compete with this last one.

Zn-CDCAL mechanism

The zinc containing enzyme follows the same mechanism above discussed in the case of cadmium (Figure 1). The three energetic paths and the structure of stationary points are reported in Figures 2b and 4, respectively.

In principle, neither significant geometrical or energetic differences are encountered in going from cadmium to zinc enzyme except those expected deriving from the different coordination chemistry of the two ions.

Contrarily to that occurred in the human carbonic anhydrase⁴⁰, the adduct (ES) is very similar to that found in the case of cadmium enzyme (see Figure 4). The geometrical parameters calculated for ES suggest that the CO₂ activation occurs without a barrier (Oa-C-Ob= 130.56°, C-OH= 1.426 Å). Instead, Bottoni et al. found a negligible barrier of 0.6 kcal/mol for the nucleophilic attack of the metal bound hydroxide on the CO₂ carbon⁴⁰.

As in the case of cadmium enzyme the molecular orbital responsible for the covalent bond between hydroxyl oxygen and substrate carbon arises from their s and p hybrid orbitals whose composition is the same reported in the previous case.

The rearrangement occurring in the following transition state (TS1) shows a bicarbonate moiety practically monocoordinated to the zinc. The Zn-Oa bond appears significantly shorter than in ES (2.057 Å in TS1 vs 3.740 Å in ES). On the contrary, the Zn-OH lengthens significantly (2.863 Å vs 2.290 Å in ES). The C-OH is now shorter (1.400 Å vs 1.426 Å in ES). The imaginary frequency of $154i \text{ cm}^{-1}$ is mainly associated to the Zn-OH stretching.

The reorganization just described requires an energetic cost of 10.9 kcal/mol in gas phase which undergoes a appreciable reduction in water solvent (6.7 kcal/mol) and one even more drastic in protein medium (2.6 kcal/mol) (see Figure 4).

The rearrangement of bicarbonate moiety begun in TS1, gives rise to the intermediate INT1. In this species, the C-OH bond reaches the length of 1.365 Å. The water molecules separate re-establishing the network of H-bonds present in ES. The -OH orients itself towards one oxygen of Asp265 to form a H-bond 1.785 Å long (see Figure 4).

From the energetic point of view, INT1 lies at 3.5 kcal/mol above ES, in gas phase. The effect of surrounding protein and water results in a considerable stabilization (5.1 and 5.9 kcal/mol, respectively) of INT1 with respect to ES (see Figure 2b).

The reaction path A which leads from INT1 to PROD proceeds through the Lindskog mechanism with the rotation around the C-Oa bond described by TS4 (see Figure 4).

TS4 is geometrically very similar to the corresponding species on the cadmium enzyme energetic path. The beginning of bicarbonate rotation around the C-Oa bond is testified by the large lengthening (about 0.7 Å) of the hydrogen bond between -OH group and Asp265 residue and by the slight increase of the Zn-OH distance (about 0.2 Å). The geometrical changes are described by the calculated imaginary frequency of $66i \text{ cm}^{-1}$. TS4 is found to be almost isoenergetic with INT1 so that, after that TS1 first barrier is cleared, the path can be considered practically descendent.

The final intermediate (PROD) shows the bicarbonate still linked to the enzyme from which it will move away on a water molecule entrance (Oa-Zn 2.180 Å, OH-Zn 3.558 Å).

Contrary to what occurs for cadmium enzyme, the zinc ion prefers to assume a tetra-coordinated geometry in all stationary points of the path.

The reaction appears to be barely less endothermic than that involving cadmium ion. In fact, PROD lies at only 0.7 kcal/mol above ES, in gas phase. As in the cadmium enzyme, protein and water media increase the stability of PROD so that the reaction becomes exothermic by 6.6 and 4.2 kcal/mol, respectively (see Figure 2b).

The TS1 barrier height suffers a significant reduction in going from gas to water to protein phase (10.9 vs 6.7 vs 2.6 kcal/mol). Since the remaining stationary points belonging to the reaction channel A in condensed media lie all below the reactants asymptote, TS1 can be considered the rate limiting step in all the three phases. In fact, the barriers lying below the reactants asymptote represent just fictitious obstacles.

Due to the similarity of the energetic barriers in the rate limiting step, the comparison of the paths A of Zn-CDCA1 and Cd-CDCA1 obtained in gas phase already demonstrates that the two enzymes can effectively compete for the nucleophilic attack on the CO₂, but a more convincing agreement with experimental observation is reached when we consider the protein and aqueous environment. In fact, the energetic values underline a slightly better performance of the zinc containing enzyme.

The difference of exothermicity between the processes catalyzed by the two enzymes in gas and in water is very small while in protein medium, the situation clearly indicates that reaction catalyzed by zinc-enzyme leads to a more stable product. However, this last fact, from the enzymatic catalysis point of view, is a disadvantage since implies a major difficulty in the active form enzyme restoration phase.

The Lipscomb mechanism (channel B) was investigated also in this case. Results indicate that it again should not compete with that of Lindskog (channel A) neither in gas or in other phases. Given to the structural similarity of the stationary points belonging to paths B of two enzymes we have chosen to not discuss further about those of zinc-enzyme, however, their geometrical features can be found in the Figure 4.

A recent work of Dudev and Lim⁷⁵ on metal binding affinity and selectivity in metalloproteins, suggests that into cadmium and zinc enzymes having the same total charge, the substitution of one or more coordinating residues with major or minor affinity towards the central ions, leads to partial charges on metal that are not so different to justify a different catalytic performance.

Using an active site model lacking of the second shell residues, we have ascertained that the progressive replacement of the histidines, normally present in human carbonic anhydrase, by cysteines produces the same effect on the charge of metal centers and enhances the nucleophilicity of the metal bound hydroxide similarly for both ions (see Figure 5).

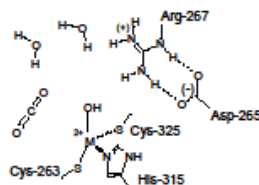
Thus, if unusually the cadmium enzyme now shows an activity comparable to that of zinc enzyme, the reason cannot be searched individually neither into the different polarizability of the metal ions nor into the major nucleophilicity of the hydroxide. Literature works⁷⁵⁻⁷⁷ do not provide conclusions but suggest the probable contribution of several factors among them also the presence of particular second shell residues and the different total charge between present and other carbonic anhydrase enzymes.

However, some indication can derive applying the concepts of HSAB theory. In fact, when the cadmium ion (that is a soft acid) is surrounded by three histidines (classified as intermediate bases), it forms with ligand nitrogen atoms of these residues quite weak bonds. On the contrary, the interaction of this ion with sulphur atoms of cysteines (that are soft bases) is very strong. The covalent nature of the Cd²⁺-S bonds makes the ion less available for the bond with the hydroxide fourth ligand (hard species), thus, as a consequence, we can observe a substantial difference in the Cd²⁺-OH lengths depending on the presence of

histidine or cysteine residues around the metal ion. This is testified by the values of these distances in both substrate free model enzymes of Figure 5 and ES complex of Figure 3 and the M1 complex of Figure 2 in the work of Marino et al.⁴¹

The major weakness of the Cd²⁺-OH bond in the CDCA1 enzyme with respect to that in the conventional carbonic anhydrases can be thus considered as a probable key factor to explain why cadmium can determine a good catalytic efficiency.

Differently from the most part of carbonic anhydrases, the mechanism followed by Cd-CDCA1 and Zn-CDCA1 seems to be well defined because among the two proposed reaction channels, the path B appears quite unfavoured with respect to the A one. Transition states on path B are characterized by hydrogen shifts mediated mainly by Arg267 residue. Just this amino acid represents a particularity which is absent in the most part of carbonic anhydrases. It is characterized by the presence of a positive nitrogen that probably compromises its ability of proton shuttle giving rise to hydrogen bonds weaker and less stabilizing than those obtained in the case of neutral residues.



Scheme 2. Active site model

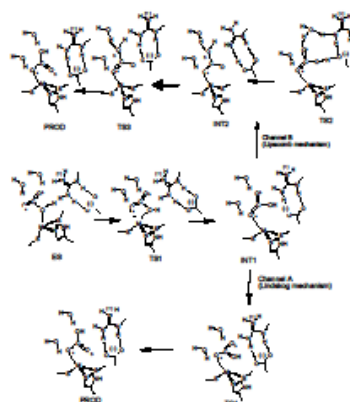


Figure 1. Chemical steps in the proposed mechanism for Cd- and Zn- CDCA1

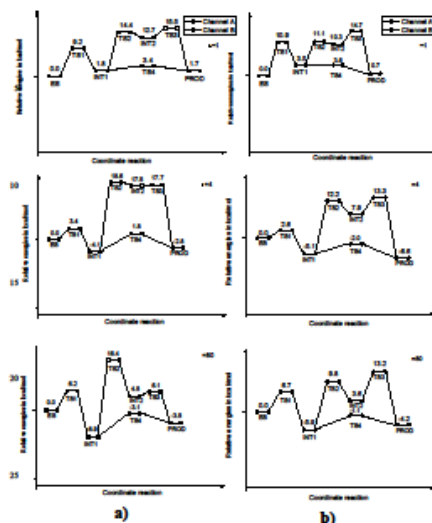


Figure 2. Energetic profiles in gas phase and in condensed media of catalytic mechanism followed by Cd-CDCAI a) and by Zn-CDCAI b).

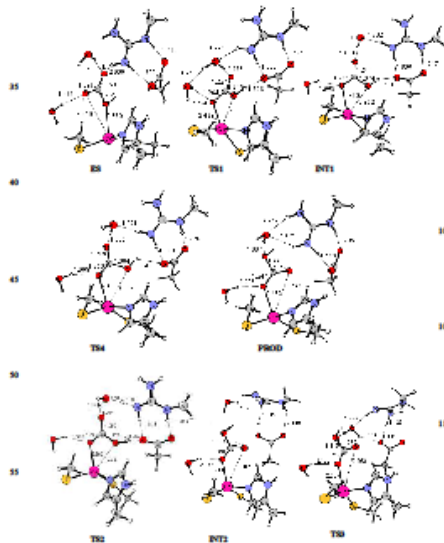


Figure 3. Stationary points on the Cd-CDCAI potential energy profile for both channels A (ES, TS1, INT1, TS4, PROD) and B (ES, TS1, INT1, TS2, INT2, TS3, PROD). Only the most significant hydrogen atoms are reported. Distances are in Å.

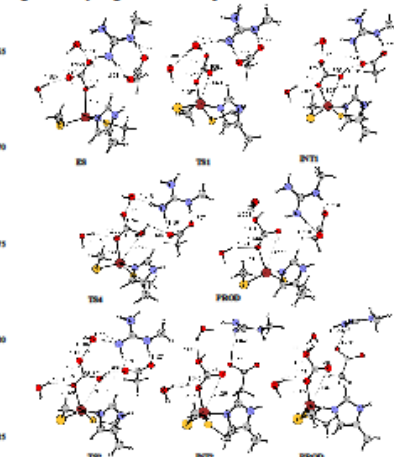


Figure 4. Stationary points on the Zn-CDCAI potential energy profile for both channels A (ES, TS1, INT1, TS4, PROD) and B (ES, TS1, INT1, TS2, INT2, TS3, PROD). Only the most significant hydrogen atoms are reported. Distances are in Å.

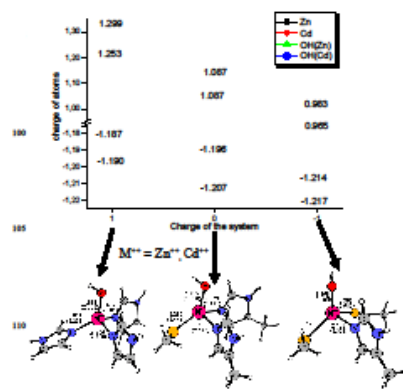


Figure 5. Effect of the gradual substitution of histidines by cysteines on the metal ion and metal bound hydroxide charges.

1. Conclusions

In this paper, we have reported a theoretical examination of the reaction mechanism of a carbonic anhydrase enzyme which can use zinc or cadmium ion to carry out the catalytic conversion of carbon dioxide to bicarbonate. Investigation was performed with an active-site model consisting of 54 atoms and constructed on the basis of X-ray crystal structure. Transition states and intermediates along two different reaction pathways for each ion were located and characterized.

The calculations give confirmation to the experimental information which suggests, contrary to what occurs more frequently, a comparable activity of this enzyme in the presence of cadmium or zinc ion.

The following mechanistic characteristics can be concluded from the calculations:

Both metal forms of enzyme follow preferentially a three steps mechanism which consists in the nucleophilic attack by the electron rich metal bound hydroxide to the carbon dioxide with bicarbonate formation, in a next internal rotation of this last moiety, and then in the formation of a product well arranged to the bicarbonate removal.

A more laborious mechanism involving complicated proton transfers before to reach product was found that requires higher energetic expense for both ions.

The rate determining step in gas phase is represented by the incipient breaking of the bond between metal ion and hydroxide group and the occurrence of a new bond of the central ion with an oxygen atom of the substrate, for both cadmium and zinc. Energy requirement to do this was computed to be 9.2 kcal/mol for cadmium and 10.9 kcal/mol for zinc enzyme. This slight difference in favour of cadmium, can be explained by the major propensity of this metal to adopt five-coordinated geometry or by necessity of zinc to remain tetra-coordinated expending more energy to do it.

In protein and in water environment, both enzymes work better from a kinetic point of view. In particular, cadmium enzyme paths in condensed phases appear to be lowered with respect to that in gas phase and present in the rate limiting step barriers of similar height (5.9 and 6.2 kcal/mol). For zinc enzyme, the energy profiles obtained in proteic ambient and in water indicate that in the rate determining step, barrier of 2.6 and 6.7 kcal/mol must be crossed, respectively. This means that a minor or at most comparable effort must be done to reach the PROD during the catalytic event than that required in the case of Cd-CDCAI enzyme.

Reaction in gas phase appears to be slightly endothermic for both cations but they become exothermic in the protein and in water.

The fact that the catalytic performance of this enzyme is not substantially influenced by the metal exchange can be the result of several factors. We have ascertained that this does not depend on neither polarizability of the metal ions nor on the enhanced nucleophilicity in the presence of electron rich sulphur atoms of cysteines. A probable explanation could derive from HSAB theory concepts. According to this we have found that cysteines coordinating residues rather than histidines makes the Cd^{2+} -OH bond in the CDCAI weak enough to facilitate the catalytic process.

Acknowledgements

We gratefully acknowledge the Dipartimento di Chimica, Università della Calabria for financial aid.

Notes and references

⁴⁵ Dipartimento di Chimica and Centro di Calcolo ad Alte Prestazioni per Elaborazioni Parallele e Distribuite-Centro d' Eccellenza MURST, Università della Calabria al-87030 Arcavacata di Rende (CS), Italy Fax: (+) 39-0984-493390 E-mail: russo@unical.it

2. References

- 1 K. S. Smith, C. Jakubovick, T. S. Whittam and J. G. Farry *PNAS* 1999, 96, 15184-15189.
- 2 K. B. Smith and J. G. Farry *J. Biol. Chem.* 2001, 276, 48615-48618
- 3 M. S. Kimber and E. F. Pai *EMBO J.* 2000, 19, 1407-1418
- 4 B. E. Albar, C. M. Colangelo, J. Dong, C. M. Stalhandske, T. T. Baird, C. Tu, C. A. Fiera, D. N. Silverman, R. A. Scott and J. G. Farry *Biochemistry* 1999, 38, 13119-13128.
- 5 I. Bartini and, C. Luchinat *Acc. Chem. Res.* 1983, 16, 272-279.
- 6 D. N. Silverman and S. Lindskog *Acc. Chem. Res.* 1988, 21, 30-36.
- 7 D. W. Christianson and C. A. Fierke *Acc. Chem. Res.* 1996, 29, 331-336 and references therein
- 8 P. Woolley *Nature* 1975, 258, 677-682.
- 9 A. Liljas, K. K. Kannan, P. C. Burgsten, I. Waara, K. Friberg, B. Strandberg, U. Carlsson, L. Jarup, S. Lovgren and M. Petef *Nat. New Biol.* 1972, 235, 131-137.
- 10 E. A. Eriksson, T. A. Jones, A. Liljas in *Zinc Enzymes* Vol. I (Eds: I. Bartini, C. Luchinat, W. Maret, M. Zappavigna) - Birkhauser - Boston - 1986, p. 317
- 11 K. Hakansson, M. Carlsson, L. A. Svensson and A. Liljas *J. Mol. Biol.* 1992, 227, 1192-1204.
- 12 E. Kimura, *Acc. Chem. Res.* 2001, 34, 171-179 and references therein
- 13 C. Huang, C. A. Leuberg, L. L. Kiefer, C. A., Fierke and D. W. Christianson *Biochemistry* 1996, 35, 3439-3446
- 14 C. A. Leuberg, C. Huang, D. W. Christianson, C. A. Fierke *Biochemistry* 1997, 36, 15780-15791
- 15 X. Zhang, C. D. Hubbard, R. van Eldik. *J. Phys. Chem.* 1996, 100, 9161-9171.
- 16 S. K. Nair, T. L.; Calderone, D. W. Christianson and C. A. Fierke *J. Biol. Chem.* 1991, 266, 17320-17325.
- 17 C. Tu, B. C. Tripp, J. G. Farry, D. N. Silverman *J. Am. Chem. Soc.* 2001, 123, 5861-5866
- 18 S. Thoms *J. Theor. Biol.* 2002, 215, 399-404.
- 19 D. Schroder, H. Schwartz, S. Schenk, E. Anders *Angew. Chem., Int. Ed.* 2003, 42, 5087-5090.
- 20 J.-Y. Liang and W. N. Lipscomb *J. Am. Chem. Soc.* 1986, 108, 5051-5058.
- 21 J.-Y. Liang and W. N. Lipscomb *Biochemistry* 1987, 26, 5293-5301.
- 22 K. M. Marx, R. Hoffman, M. J. S. Dewar *J. Am. Chem. Soc.* 1989, 111, 5636-5649.
- 23 J.-Y. Liang and W. N. Lipscomb. *Int. J. Quantum Chem.* 1989, 36, 299-312.
- 24 O. Jacob, R. Cardenas, O. Tapia *J. Am. Chem. Soc.* 1990, 112, 8692-8705.
- 25 M. Krauss and D. R. Garmar *J. Am. Chem. Soc.* 1991, 113, 6426-6435.
- 26 Y.-J. Zhang and K. M. Marx *J. Am. Chem. Soc.* 1992, 114, 10498-10507.
- 27 M. Sola, A. Llodas, M. Duran and J. Bertran *J. Am. Chem. Soc.* 1992, 114, 869-877.
- 28 M. Sakurai, T. Furuki and Y. Inoue *J. Phys. Chem.* 1995, 99, 17789-17794.
- 29 M. A. Murcko *Theor. Chem. Acc.* 1997, 96, 56-60.
- 30 K. M. Marx and L. Banci *J. Am. Chem. Soc.* 1997, 119, 863-871.

- 31 D. Lu and G. A. Voth *J. Am. Chem. Soc.* 1998, *120*, 4006-4014.
- 32 S. Toba, G. Colombo and K. M. Marz *J. Am. Chem. Soc.* 1999, *121*, 2290-2302.
- 33 V. P. Denisov, B.-H. Jonsson, B. Halle *J. Am. Chem. Soc.* 1999, *121*, 2327-2328.
- 34 M. Mankich, M. Brauner, J. Weston and E. Anders *Chem. Bio. Chem.* (2001), *2*, 190-198.
- 35 M. Brauner, J. L. Porse-Lustras, J. Weston and E. Anders *Inorg. Chem.* 2002, *41*, 1454-1463.
- 36 Z. Szeadarchina, W. Szebrand, A. Fernandez-Ramos and Q. Cui *J. Am. Chem. Soc.* 2003, *125*, 243-251.
- 37 D. R. Gurnar *J. Phys. Chem. B* 1997, *101*, 2945-2953.
- 38 Q. Cui and M. Karplus *J. Phys. Chem. B* 2003, *107*, 1071-1078.
- 39 C. Mignogna *J. Mol. Struct. (THEOCHEM)*, 1999, *461-462*, 439-452.
- 40 A. Bottoni, C. Z. Lanza, G. P. Miscione and D. Spinelli *J. Am. Chem. Soc.* 2004, *126*, 1542-1550.
- 41 T. Marino, N. Russo and M. Toscano *J. Am. Chem. Soc.* 2005, *127*, 4242-4253.
- 42 S. Lindskog, I. Henderson, K. K. Kannan, A. Lijax, P. O. Nyman, B. Strandberg in *The Enzymes*, Vol. 5, 3rd ed (Ed: P. Boyer), Academic Press - New York - 1971, p. 387.
- 43 W. N. Lipscomb *Annu. Rev. Biochem.* 1983, *52*, 17-34.
- 44 B.C. Tripp, K. Smith, J. G. Ferry *J. Biol.* 2001, *276*, 48615-48618.
- 45 S.B. Roberts, T. W. Lase and F. M. M. Meral *J. Physiol.* 1997, *33*, 834-850.
- 46 A. R. Sotgi, et al. *Appl. Environ. Microbiol.* 2006, *72*, 5500-5511.
- 47 N. M. Price and F. M. M. Meral *Nature* 1990, *344*, 658-660.
- 48 F. M. M. Meral, J. R. Reinfelder, S. B. Roberts, C. P. Chamberlain, J. G. Lee and D. Yee *Nature* 1994, *369*, 740-742.
- 49 T. W. Lase and F. M. M. Meral *PNAS* 2000, *97*, 4627-4631.
- 50 T. W. Lase, M. A. Saino, G. N. Garga, I. J. Pickering, R. C. Prince, F. M. M. Meral *Nature* 2005, *435*, 42.
- 51 H. Park, B. Song and F. M. M. Meral *Environ. Biol.* 2007, *9*, 403-413.
- 52 E. A. Boyle, F. Sclater and J. M. Edmond *Nature* 1976, *263*, 42-44.
- 53 K. W. Bruland, G. A. Kanzer and J. H. Martin - *Limnol. Oceanogr.* 1978, *23*, 618-625.
- 54 Y. Xu, L. Feng, P. D. Jeffrey, Y. Shi and F. M. M. Meral *Nature* 2008, *452*, 56-61.
- 55 A. D. Becke *J. Chem. Phys.* 1993, *98*, 5648-5652
- 56 R. G.; Parr and W. Yang *Density-functional theory of atoms and molecules*, Oxford Univ. Press, Oxford, 1989
- 57 C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B* 1988, *37*, 785-789.
- 58 Gensiana 03, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, 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, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Ciolekowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chan, M. W. Wong, C. Gonzalez, and J. A. Pople, Gensiana, Inc., Wallingford CT, 2004.
- 59 M. Dolg, H. Stoll, H.; H. Preuss and R. M. Pitzer, *J. Phys. Chem.* 1993, *97*, 5852-5859.
- 60 M. Dolg, U. Wedig, H. Stoll and H. Preuss *J. Chem. Phys.* 1987, *86*, 866-872.
- 61 E. D. Glendaming, A. E. Reed, J. E. Carpenter and F. Weinhold NBO, version 3.1.
- 62 V. Barone and M. Cossi, *J. Phys. Chem. A* 1998, *102*, 1995-2001
- 63 M. Cossi, N. Rega, G. Scalmani and V. Barone *J. Comput. Chem.* 2003, *24*, 669-681.
- 64 Siqbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* 2000, *100*, 421-437.
- 65 Noodelman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev.* 2004, *104*, 459-508.
- 66 M. J. Ramos and P. A. Fernandes, *Acc. Chem. Res.* 2008, *41*, 689-698.
- 67 M. Leopoldini, T. Marino and M. Toscano, *Theor. Chem. Acc.* 2008, *120*, 459-466.
- 68 M. Leopoldini, N. Russo and M. Toscano, *J. Am. Chem. Soc.* 2007, *129*, 7776-7784.
- 69 S.-L. Chen, Marino, T.; Fang, W.-H.; Russo, N.; Himo, F. *J. Phys. Chem. B* 2008, *112*, 2494-2500
- 70 K. S. Raviat *Biochimica et Biophysica Acta* 2010, *1804*, 362-373.
- 71 Rong-Zhou Liao, Fahmi Himo, Jian-Guo Yu, Rao-Zhang Lin, *J. Inorg. Biochem.* 2010, *104*, 37-46.
- 72 Rong-Zhou Liao, Jian-Guo Yu, Fahmi Himo, *Inorg. Chem.* 2010, *49*, 6883-6888.
- 73 M. E. Alberto, T. Marino, N. Russo and M. J. Ramos *J. Chem. Theory Comput.* 2010, *6*, 2424-2433
- 74 O. Amata, T. Marino, N. Russo and M. Toscano *J. Am. Chem. Soc.* 2009, *131*, 14804-14811
- 75 T. Dudev, C. Lim *Annu. Rev. Biophys.* 2008, *37*, 97-116
- 76 C. Andrusini, I. Bertini, G. Cavallaro, G. L. Holliday, J. M. Thornton *J. Biol. Inorg. Chem.* 2008, *13*, 1205-1218.
- 77 K. J. Waldron, J. C. Rutherford, D. Ford, N. J. Robinson *Nature* 2009, *460*, 823-830.

