

UNIVERSITÀ DI PISA



Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea Magistrale in Fisica
Curriculum: Fisica Medica

TESI DI LAUREA MAGISTRALE

Dinamica molecolare delle rodopsine con modelli multi-scala

Relatore:
Dr. Valentina Tozzini

Laureando:
Francesco Tavanti

Anno Accademico 2012/2013

Sommario

Capitolo 1. Introduzione e motivazione del lavoro di Tesi.....	3
Capitolo 2. Le Rodopsine.....	6
2.1 Le proteine: basi del funzionamento cellulare	6
2.2 Le rodopsine, batterica e dei mammiferi.....	7
2.2.1 <i>Struttura delle rodopsine</i>	9
2.2.2 <i>Zoom sui Retinali e sul sito attivo</i>	13
2.2.3 <i>I Fotocicli delle rodopsine</i>	14
2.3 Scopo del lavoro di tesi e scelte metodologiche	19
Capitolo 3. Modelli e metodi di simulazione	21
3.1 Dinamica molecolare classica.....	22
3.2 Campi di forze empirici per modelli atomistici	25
3.3 Modelli “a grana grossa” (Coarse Grained).....	27
3.3.1 <i>Modelli a rete elastica e dinamica essenziale</i>	29
3.3.2 <i>Estensioni dei modelli a rete: multistabilità</i>	30
3.3.3 <i>Modelli di Gō e modelli a bias parziale</i>	32
3.4 Modello multi-scala per le rodopsine: descrizione sommaria.....	33
Capitolo 4. Risultati: simulazione delle dinamiche molecolari e del fotociclo delle rodopsine	37
4.1 Ottimizzazione del modello multi-scala sulla rodopsina batterica	37
4.2 Simulazione degli stati del fotociclo.....	46
4.3 Modello bi-stabile	48
4.4 Simulazioni delle transizioni tra gli stati.....	50
4.4.1 <i>Simulazione della transizione BR → K</i>	51
4.4.2 <i>Simulazione della transizione K → L e seguenti</i>	52
4.5 Analisi delle componenti principali	56
4.5.1 <i>Transizione BR-K</i>	56
4.5.2 <i>Transizione K-L</i>	57
4.5.3 <i>Transizione L-M</i>	58
4.5.4 <i>Transizione M-N</i>	59
4.5.5 <i>Transizione N-O</i>	60
4.6 Sommario dei risultati	61

Capitolo 5. Conclusioni e possibili sviluppi	63
Appendici.....	67
Appendice A. Scelta delle strutture di riferimento delle rodopsine	67
Appendice B. Descrizione del software utilizzato, modificato e creato	70
<i>B.1 Codici per la generazione dell'input di DL_POLY.....</i>	<i>70</i>
<i>B.2 Modifiche operate a DL_POLY.....</i>	<i>74</i>
<i>B.3. Analisi dell'output.....</i>	<i>76</i>
<i>B.4. VMD.....</i>	<i>77</i>
<i>B.5. Analisi dei modi/componenti principali.</i>	<i>77</i>
Appendice C. Dettagli sulla descrizione del potenziale modello multistabile.....	79
Bibliografia.....	84

Capitolo 1. Introduzione e motivazione del lavoro di Tesi

In questa tesi è presentato uno studio di simulazione al computer della dinamica molecolare delle rodopsine. A differenza di studi precedenti, il problema è qui affrontato usando approcci simulativi differenti, che rappresentano il sistema a diversi livelli di risoluzione e con diversi modelli di descrizione delle interazioni, combinati in maniera innovativa. Il risultato è che, con risorse computazionali modeste, si riesce a coprire nelle simulazioni l'intero ciclo funzionale delle proteine, che si estende su scale temporali che vanno dall'ultraveloce a quella macroscopica, per più di dieci ordini di grandezza. L'approccio qui presentato non è per altro specifico per le rodopsine, ma applicabile a tutti i casi in cui si abbia a che fare con sistemi la cui dinamica si estende su scale spazio-temporali varie, caso comunissimo nei sistemi bio-molecolari, data la loro naturale organizzazione strutturale gerarchica.

Come sarà spiegato nel Capitolo 2, la funzione delle rodopsine è tradurre stimoli luminosi in segnali chimici per comunicazione trans-cellulare. Specificamente, sono state considerate in questo studio due rodopsine, prototipiche di due sottoclassi funzionali: la rodopsina batterica e la rodopsina dei mammiferi. La prima è una pompa di protoni localizzata nella membrana purpurea del batterio *Halobacterium salinarum*; la seconda si trova nelle cellule fotorecetriche della retina dei mammiferi ed ha lo scopo di generare il segnale che avvia il processo della visione. Il loro funzionamento è molto simile: entrambe contengono al loro interno un retinale, una molecola fotoattiva, che, con l'assorbimento di un fotone, cambia conformazione; questa transizione dà il via ad una serie di cambiamenti della struttura globale delle proteine che portano al rilascio del segnale. Il "fotociclo" si conclude con il ritorno delle proteine nello stato iniziale.

Ogni stato del fotociclo è identificabile attraverso un diverso assorbimento ottico. Le due rodopsine hanno anche una struttura globale simile, ma le strutture dei vari stati della batterica sono note sperimentalmente con maggiore accuratezza, mentre per la rodopsina dei mammiferi i dati strutturali sono più frammentari e meno accurati. Si è dunque scelta la rodopsina batterica come prototipo per la costruzione del modello, che poi è stato anche verificato sulla rodopsina dei mammiferi.

Come accennato, un problema fondamentale nella simulazione al computer e nella modellistica di sistemi biologici in generale, sono le diverse scale di tempi coinvolte nei processi. In particolare, il fotociclo delle rodopsine ha una fase iniziale molto veloce (sulla scala delle centinaia di fs) che coinvolge il cambiamento strutturale del retinale (struttura interna alla proteina responsabile della sua foto-attività) conseguente all'assorbimento del fotone, seguita da diverse fasi caratterizzate da variazioni strutturali più globali e lente, che collegano i vari stati e che si attuano sulle scale dei μ s e ms. Come verrà illustrato nel Capitolo 3, le simulazioni di dinamica molecolare classica "standard" rappresentano un sistema molecolare al livello atomistico: ogni atomo del sistema è rappresentato tramite le sue coordinate cartesiane e le interazioni sono descritte da hamiltoniane

empiriche (o “Campi di Forze”, CF), rappresentate da somme di termini analitici. Poiché i parametri di queste hamiltoniane sono stati ottimizzati nel corso degli ultimi 50 anni per proteine e acidi nucleici, ormai si tratta di una metodologia in grado di descrivere la dinamica del sistema con buona accuratezza. Sistemi hardware di uso routinario (*workstations*) sono in grado di raggiungere nelle simulazioni di singole proteine idratate le scale di (centinaia di) ns, quindi certamente di rappresentare bene le fasi iniziali del fotociclo, mentre le scale del μ s richiedono sistemi pesantemente paralleli, che comunque riescono appena a raggiungere il ms. Quindi rappresentare l'intero fotociclo con simulazioni atomistiche non è attualmente possibile.

D'altra parte, sono stati di recente considerati modelli a bassa risoluzione (Coarse Grained, CG), in cui i gradi di libertà espliciti rappresentano gruppi di atomi anziché atomi singoli, con notevole risparmio di tempo di calcolo. Questo tipo di rappresentazioni, dipendentemente dalla risoluzione, è in grado di raggiungere tempi macroscopici nelle simulazioni e quindi in linea di principio di rappresentare le fasi lente del fotociclo. Naturalmente la semplificazione del modello comporta una diminuzione di accuratezza, che però può venire compensata con una adeguata cura nella parametrizzazione delle interazioni.

Il modello sviluppato in questa tesi, descritto nel Capitolo 4, rappresenta la proteina a due diversi livelli di risoluzione (modello multi-scala): per il retinale e le molecole di acqua interne alla proteina si è usata una rappresentazione quasi atomistica, in cui i gradi di libertà espliciti sono le coordinate degli atomi “pesanti” (C, N e O); in questa rappresentazione, gli atomi di idrogeno ad essi legati sono trattati implicitamente. Questo è il livello minimo di risoluzione necessario per il retinale, perché la sua foto-isomerizzazione è molto spazialmente localizzata e specifica. Per tutto il resto della proteina si è usato un modello “minimalista” in cui un amminoacido è descritto tramite un singolo centro interattivo localizzato sull'atomo più rappresentativo, il carbonio alfa ($C\alpha$). In questo modo la parte peptidica della proteina è rappresentata da una catena lineare di centri interattivi (chiamati *bead*). Questo modello misto con parte peptidica a bassa risoluzione e sito attivo (retinale + acqua) ad alta risoluzione, consente di rappresentare con buona accuratezza le fasi iniziali veloci del fotociclo che coinvolgono la transizione strutturale del retinale, e contemporaneamente mantiene un basso costo computazionale, che permette la rappresentazione dell'intero fotociclo.

La parametrizzazione dell'hamiltoniana di questo modello è il primo risultato di questo lavoro di tesi. Lo scopo perseguito nell'ottimizzazione della parametrizzazione è stato quello di riprodurre accuratamente sia la struttura e la dinamica di equilibrio dei vari stati, sia la cinetica e l'energetica dei passaggi tra di essi. Allo scopo di rappresentare le transizioni tra gli stati, il campo di forze deve includere multistabilità, caratteristica non comune e difficile da implementare soprattutto nei modelli a bassa risoluzione. Dunque, si è proceduto prima ad ottimizzare i parametri del modello per ciascuno degli stati separatamente, in modo da riprodurre le proprietà strutturali e vibrazionali. Successivamente questi sono stati composti a due a due, inserendo informazioni sperimentali sulla stabilità relativa dei vari stati. Si sono poi studiate, tramite simulazione, la dinamica, la cinetica ed i profili energetici delle transizioni,

simulando l'intero fotociclo. Queste simulazioni e la loro analisi costituiscono il secondo risultato di questo lavoro di Tesi.

La rappresentazione multi-scala consente di eseguire queste simulazioni su calcolatori monoprocesso, senza l'utilizzo di risorse computazionali massive. Per le simulazioni si è usato un pacchetto software per dinamica molecolare classica molto diffuso, DL_POLY, scelto per la sua flessibilità. DL_POLY è infatti "multi-purpose", cioè non implementa specificamente un determinato modello ma è utilizzabile per sistemi diversi, in stato solido, chimica inorganica, organica e biochimica. Inoltre è *open source*, cioè consente di modificare i file di codice sorgente, scritti in linguaggio Fortran. Questa è una caratteristica molto importante, che ha consentito alcune necessarie modifiche al codice per implementare le forme funzionali del modello multi-stabile. Queste modifiche sono state eseguite in maniera da poter essere utilizzate anche per sistemi diversi dalle rodopsine, seguendo la filosofia generale della programmazione di DL_POLY. Inoltre, sono stati programmati piccoli codici accessori per creare gli input della simulazione per le rodopsine. Questi software e i moduli modificati di DL_POLY sono un valore aggiunto di questo lavoro di tesi, descritti in Appendice, e possono essere resi disponibili su richiesta.

La struttura della tesi è già stata delineata: il prossimo capitolo è dedicato alla descrizione delle caratteristiche delle rodopsine e del loro fotociclo. Segue un capitolo che introduce alcuni concetti generali sulle simulazione di dinamica molecolare classica e sui modelli atomistici, a bassa risoluzione e multiscala. Il Capitolo 4 contiene tutti i risultati, simulazioni ed analisi, e discussione. Il Capitolo 5 un riassunto, conclusioni e possibili sviluppi. Dettagli sulle rodopsine e descrizione del software, e dettagli aggiuntivi sul modello multi-stabile sono riportati nelle Appendici.

Capitolo 2. Le Rodopsine

2.1 Le proteine: basi del funzionamento cellulare

Il funzionamento delle cellule viventi è deputato ad un gran numero di molecole di diverso tipo. Il contenuto di una cellula è tipicamente costituito da 70% di acqua, circa 8% di carboidrati e lipidi (la maggior parte costituente la membrana) e piccole molecole di vario tipo. Il restante 22% è composto dalle due classi di “molecole informative”, responsabili delle funzioni superiori della cellula: Acidi Nucleici (DNA e RNA, per 1% e 6% rispettivamente), deputati a conservazione, utilizzo e trasmissione dell'informazione genetica (e regolazione di questa funzione), e proteine, il restante 15%, che svolgono tutti gli altri compiti, intervenendo anche in funzioni in parte sovrapposte con quelle degli acidi nucleici. Anche se nella storia dell'evoluzione della vita non è probabilmente sempre stato così (esiste l'ipotesi che vi siano state forme di vita basate solamente su di RNA prima della formazione degli attuali organismi viventi basati soprattutto sul DNA¹), si può dire che, attualmente, le proteine svolgono il grosso del lavoro per il funzionamento dei sistemi viventi.

Dal punto di vista biochimico, le proteine sono etero-polimeri (polipeptidi) i cui monomeri sono gli amminoacidi (AA). La reazione di polimerizzazione avviene in uno speciale macchinario cellulare (il ribosoma) che “monta” il polipeptide pezzo per pezzo seguendo l'istruzione contenuta nel DNA. Il risultato è una catena di amminoacidi, scelti tra 20 tipi diversi (ciascuno caratterizzato da una diversa “catena laterale” o “residuo”), uniti uno all'altro dai legami peptidici in una sequenza precisa, chiamata struttura primaria, che caratterizza ciascuna proteina.

La struttura primaria determina il ripiegamento della proteina in diverse possibili “strutture secondarie” stabilizzate da legami ad idrogeno, che si possono raggruppare in due grandi classi: le strutture elicoidali (di cui la più diffusa è l'elica α) e le strutture a foglietto (β parallelo o antiparallelo). Queste a loro volta si organizzano in strutture super-secondarie, terziarie e quaternarie, stabilizzate da interazioni più deboli e/o meno locali (forze di Van Der Waals, elettrostatica, idrofobicità). È la struttura tridimensionale finale che determina la corretta funzionalità della proteina: proteine mal ripiegate non sono funzionali, e anzi, lo scorretto ripiegamento o in generale la distruzione della struttura terziaria o quaternaria della proteina, dovuta a fattori ambientali oppure a mutazioni della sequenza, determina l'insorgere di malattie generalmente molto gravi come la malattia di Alzheimer² o come l'encefalopatia spongiforme bovina, noto come morbo della mucca pazza³. Nella Fig. 1 è riportato uno schema riassuntivo della biochimica di base e struttura delle proteine.

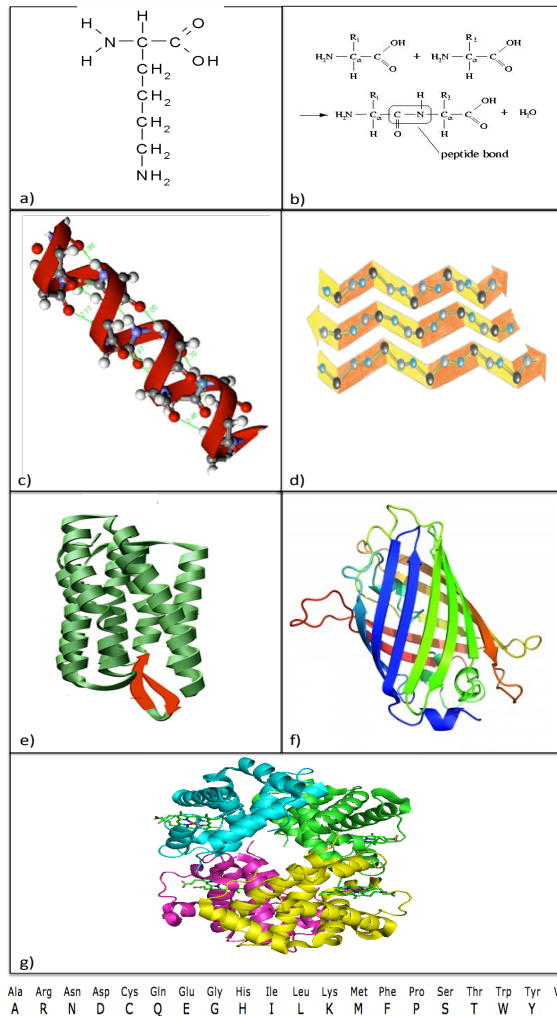


Figura 1: a) la struttura chimica di un amminoacido (Lisina); b) la struttura primaria di un dipeptide; c) la struttura secondaria di un'elica e d) di un foglietto; in e) la struttura terziaria della rodopsina batterica dove le eliche sono in verde ed il foglietto in rosso; in f) la struttura terziaria della Green Fluorescent Protein, detta “barilotto beta”; in g) la struttura quaternaria dell'emoglobina. In calce, la lista dei codici a tre lettere (sopra) e ad una lettera (sotto) degli AA. Le rappresentazioni in (c)-(g) sono in stile “cartoons”, in cui le strutture secondarie vengono schematicamente rappresentate come elementi macroscopici (vere e proprie eliche o nastri che compongono i foglietti).

2.2 Le rodopsine, batterica e dei mammiferi

Le rodopsine sono proteine fotoattive, specializzate nell'utilizzo di fotoni a specifiche lunghezze d'onda nel visibile. In particolare, oggetto di questo studio sono la rodopsina batterica (indicata con BR) dell'*Halobacterium salinarum* (Fig. 2) e la rodopsina dei mammiferi (MR). Nonostante la similitudine strutturale, BR e MR hanno funzioni diverse.

La BR è una pompa protonica azionata dalla luce, che il *salinarum* utilizza, come altri batteri alofili, per regolare e mantenere un gradiente elettrochimico da cui ricava l'energia per le proprie funzioni vitali. Questo efficiente modo di sfruttare l'energia solare in mancanza di altre fonti, consente ad

Halobacterium salinarum di sopravvivere in ambienti a salinità estrema, come saline e laghi salati, non tollerabili per altri organismi.⁴

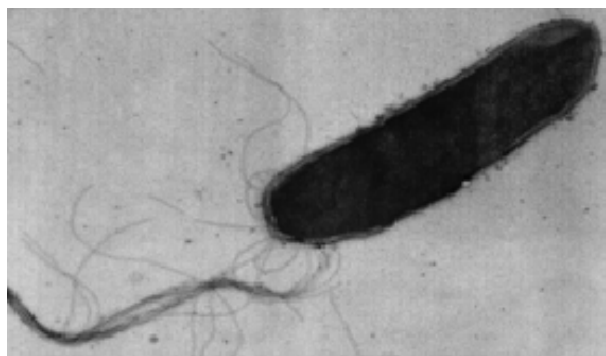


Figura 2: immagine ottenuta tramite microscopia elettronica dell'*Halobacteria salinarum*, che mostra la tipica forma a barretta e i flagelli, con cui il batterio si sposta.

Il *salinarum* esprime in realtà quattro tipi di rodopsine, tutte caratterizzate dal funzionamento basato sul cofattore retinale, un composto derivato dalla vitamina A (Fig. 3) che agisce come cromoforo, ed è responsabile dell'assorbimento del fotone. Di queste, la BR in esame in questa Tesi, viene estratta dalla membrana viola del batterio, che può essere facilmente isolata e ne permette la produzione in grandi quantità⁵. Le altre rodopsine di *H. salinarum* sono l'alorodopsina, una pompa per il cloro, e due rodopsine sensorie (I e II) coinvolte nelle risposte fotofile e fotofobiche alle varie parti dello spettro solare.

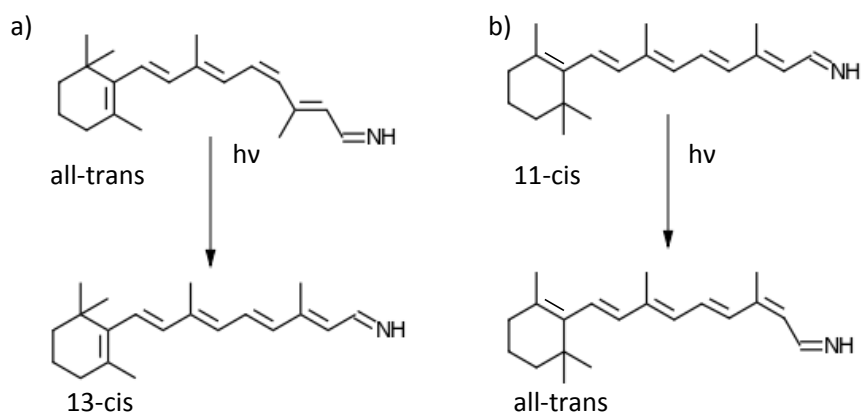


Figura 3: a) isomerizzazione del retinale *cis-trans* della rodopsina umana; b) isomerizzazione del retinale *trans-cis* della rodopsina batterica. Da questa si notano i doppi legami C=C della catena del retinale e la base di Schiff che crea il doppi legame C=N in fondo alla catena del retinale.

La MR, invece, si trova nella membrana delle cellule fotorecetriche della retina dei mammiferi, ed è alla base del meccanismo molecolare della visione. In particolare la MR presa in esame in questo studio è quella presente nei bastoncelli (fotorecettori responsabili per la visione di luce debole) della retina bovina. È considerata il paradigma per le rodopsine dei mammiferi e molto simile a quella umana. Anche la MR è una proteina di membrana e come la BR fornisce una risposta dopo uno stimolo luminoso, anche se, nei due casi, gli stimoli danno vita a differenti cascate di reazioni biochimiche.

Come nella BR la parte fotoattiva è costituita dal retinale. In entrambi i casi in seguito all'assorbimento di un fotone la coda aromatica del retinale, caratterizzata da alternanza di singoli e doppi legami carbonio-carbonio (C-C e C=C, Fig. 3), cambia conformazione. Specificamente, avviene una isomerizzazione cis-trans (o trans-cis) di uno dei doppi legami. Tuttavia mentre nel caso della MR il fotone induce una trasformazione da 11-cis ad all-trans (trasformazione da cis-trans del legame C=C n. 11, Fig. 3(a)), nel caso della BR si passa da all-trans a 13-cis (trans-cis del doppio legame C=C n. 13, Fig. 3(b)). Di queste isomerizzazioni fotoindotte si parlerà più estesamente nelle prossime sezioni.

Nella MR il cambiamento conformazionale del retinale provoca una variazione strutturale nella proteina, che a sua volta attiva una proteina di membrana segnalatrice (la G-protein, che deriva il nome dalla capacità di legare le molecole energetiche guanosin-difosfato (GDP) o trifosfato (GTP)), che amplifica il segnale e alla fine genera l'impulso nervoso che giunge al cervello attraverso il nervo ottico generando un'immagine⁶. Nella BR, invece, il cambiamento strutturale della proteina conseguente a quello del retinale attiva il passaggio protonico verso la superficie extracellulare⁷.

2.2.1 Struttura delle rodopsine

BR e MR fanno parte di una grande super-famiglia strutturale, le opsine caratterizzate da una struttura terziaria con 7 eliche che attraversano la membrana cellulare (7TM) e sono rappresentative delle due principali sottofamiglie, opsine tipoI e opsine tipoII, caratteristiche rispettivamente dei procarioti (come i batteri) e degli eucarioti (come i mammiferi). Il meccanismo di azione della MR la colloca anche nella famiglia dei recettori accoppiati alle proteine G di membrana (GPCR). Questa famiglia riveste un interesse particolare in biologia molecolare, nanomedicina e farmacologia: le GPCR realizzano i collegamenti molecolari tra segnali extracellulari e processi intracellulari, e sono pertanto coinvolte in un grandissimo numero di processi di segnalazione in risposta a stimoli sia esterni, come quelli provocati dalla luce o da sostanze chimiche, che interni, come quelli provocati da ormoni e neurotrasmettitori^{8,9}. Di conseguenza, possono essere utilizzati come potenziale bersaglio per sviluppi di farmaci contro allergie alimentari, disordini della circolazione, disordini del sistema nervoso centrale e anche per l'AIDS. La BR è costituita da sette eliche transmembrana, quindi è un recettore 7TM, ma data la sua diversa funzione, non è un GPCR.

In generale, la struttura delle proteine può essere ottenuta o con diffrazione a raggi X (o più raramente di neutroni), o con spettroscopia NMR o criomicroscopia elettronica. Le tecniche diffrattive richiedono ovviamente che i campioni siano in qualche forma cristallina, e non presentano sostanziali differenze da quelle usate per la determinazione della struttura dei cristalli non biologici, eccetto per il fatto che il contenuto delle celle elementari è enormemente più complesso: si tratta infatti di una intera proteina oppure di un oligomero di proteine organizzate in specifiche simmetrie, generalmente circondate da acqua residua del processo di cristallizzazione, per un totale di

decine o centinaia di migliaia di atomi. Questo comporta una fase di rielaborazione (*post-processing*) dei dati sperimentali piuttosto complessa¹⁰. Il risultato finale è comunque un modello tridimensionale della proteina in cui sono generalmente risolte tutte le posizioni atomiche (se la tecnica è la diffrazione dei neutroni) o almeno di tutti gli atomi “pesanti” (se la tecnica è la diffrazione a raggi X, che non risolve la posizione degli atomi di idrogeno, per via della loro bassa densità elettronica). La risoluzione raggiunta con questa tecnica è dell'ordine di pochi Å. Le tecniche diffrattive, oltre alle posizioni atomiche, forniscono anche le fluttuazioni quadratiche medie per ogni atomo, il cosiddetto fattore B

$$B = \frac{8\pi^2}{3} \langle u^2 \rangle \quad (2.1)$$

B si misura in Å², ed è sostanzialmente il parametro dell'esponenziale nel fattore di Debye-Waller. Ovviamente B dipende dalla temperatura, ed infatti viene chiamato anche fattore di temperatura.

La necessaria cristallizzazione dei campioni per le tecniche diffrattive crea due serie di problemi. In primo luogo, la struttura cristallina per le proteine non è fisiologica (viene infatti ottenuta in condizioni di disidratazione estrema e generalmente a temperature criogeniche). Questo può in certi casi creare conformazioni innaturali dovute all'impacchettamento cristallino. Inoltre alcune proteine sono refrattarie alla cristallizzazione. È questo proprio il caso delle proteine di membrana come BR e MR, che sono instabili fuori dalla membrana cellulare, ma d'altra parte difficilmente impacchettabili in un reticolo se immerse in tratti di membrana. Per questo motivo, la MR è stata cristallizzata per la prima volta nel solo 1998¹¹ mentre la struttura cristallografica della BR risale appena 18 prima (1980). Questo studio valse ad Hartmut Michel il premio Nobel per la chimica nel 1988 per «*la determinazione della struttura tridimensionale di un centro di reazione fotosintetico*»¹².

Le strutture terziarie delle due rodopsine sono riportate in Fig. 4. In entrambe le proteine, le sette eliche attraversano la membrana cellulare, mentre i tratti destrutturati di collegamento (*loop*) si trovano generalmente nelle regioni extra o intra-cellulari. Le eliche sono indicate con le lettere da A a G nella BR e da 1 a 7 nella MR e hanno lunghezze diverse nelle due proteine (Fig. 5). Inoltre in MR è presente una piccola elica posta sulla superficie della membrana cellulare. Il numero totale di amminoacidi di MR è 348, 100 in più della BR (composta da 248 amminoacidi).¹³

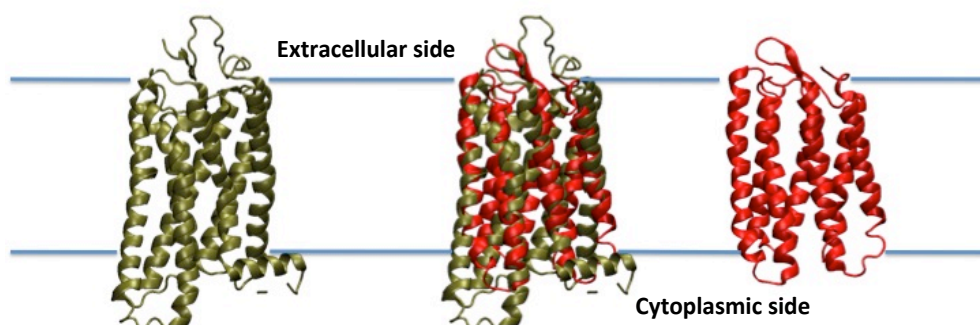


Figura 4: Struttura della MR (in bronzo) e della BR (in rosso) a confronto (al centro la sovrapposizione delle due strutture). Le linee orizzontali rappresentano la membrana cellulare.

Il retinale occupa lo spazio interelocale, trasversalmente rispetto agli assi delle eliche, ed è legato all'elica 7 nella MR e all'elica G nella BR, e in corrispondenza del legame le eliche presentano una distorsione^{14,15}.

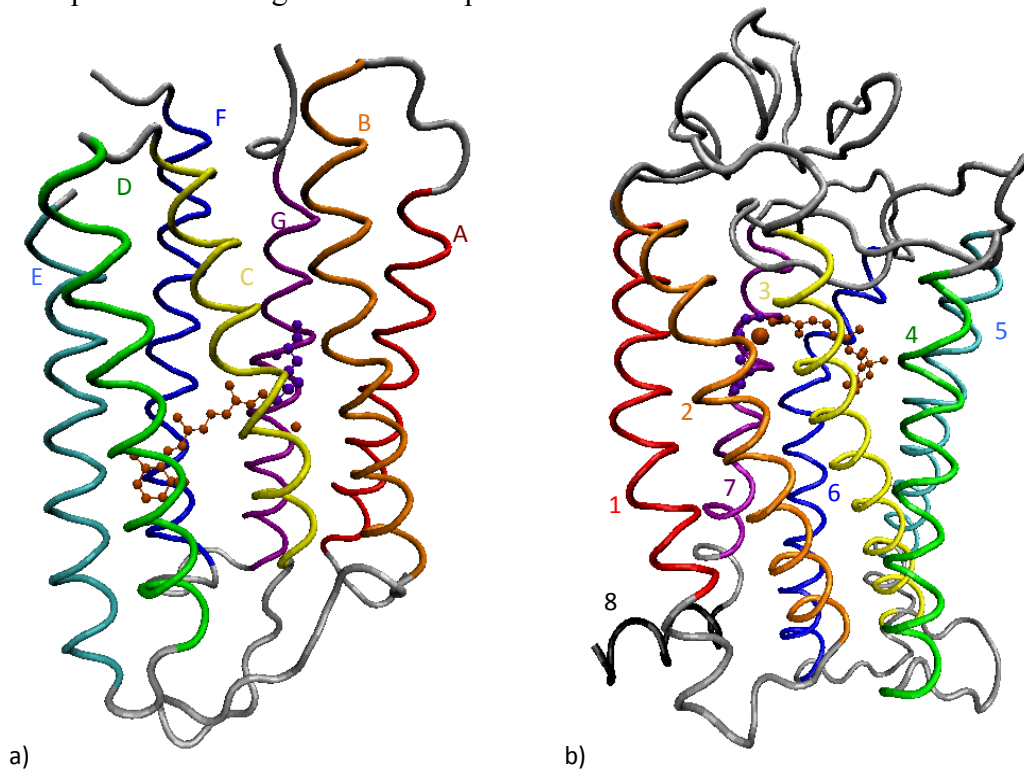


Figura 5: Struttura terziaria della BR (a), e della MR (b). Le eliche hanno accanto il loro nominativo e sono colorate a seconda del numero dell'amminoacido partendo dal rosso fino ad arrivare al violetto per la BR ed al nero per la MR. Il retinale e la molecola di acqua W402 sono rappresentati in bronzo, la catena laterale di Lys216 e di Lys296 in violetto scuro.

Il legame si forma in corrispondenza della catena laterale dell'amminoacido lisina (specificamente Lys216 in BR e Lys296 in MR): una reazione di disidratazione forma una base di Schiff protonata e un legame covalente tra l'ultimo carbonio del retinale e l'azoto della lisina¹⁶ (Fig. 6). Mentre in BR il retinale si lega alla proteina una volta per tutte dopo il ripiegamento¹⁷, nella MR il legame/distacco del retinale ha un ruolo funzionale, e quindi costituisce una delle fasi del foto-ciclo della proteina (si veda più avanti).

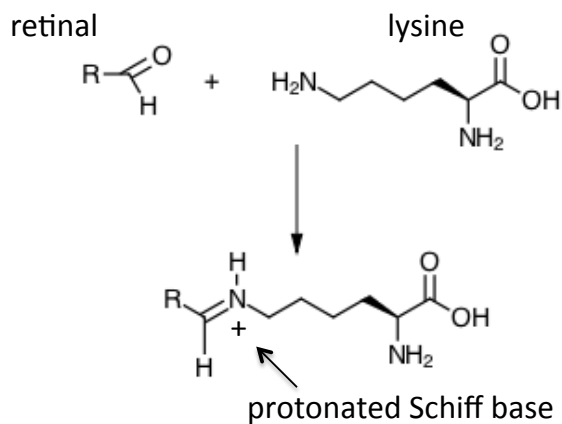


Figura 6: Schematizzazione della formazione della legame tra lisina e retinale e formazione della base di Schiff. Questa avviene in condizioni di pH fisiologico spontaneamente e rapidamente¹⁸. Dopo la disidratazione la base di Schiff risulta protonata, cioè l'atomo di azoto ha una carica positiva.

La struttura quaternaria della BR è un trimero, cioè la proteina si organizza spazialmente in tre monomeri disposti ai vertici di un triangolo. La MR, invece, ha una struttura quaternaria non ancora chiarita; infatti ci sono studi che sostengono che la proteina sia un monomero,^{19,20} mentre altri studi indicano come struttura quaternaria un oligomero a simmetria composizione e simmetria non ben chiarita.²¹

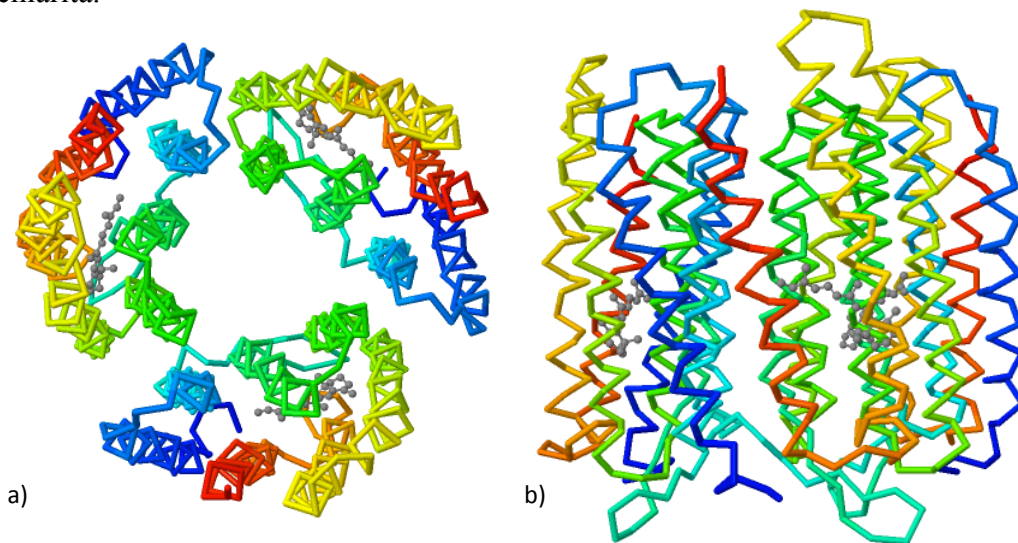


Figura 7: struttura quaternaria della rodopsina batterica organizzata in un trimero. In a) la vista del trimero dall'alto dove è visibile al centro il retinale in grigio, in b) la vista del trimero di lato.

2.2.2 Zoom sui Retinali e sul sito attivo

La formula chimica del retinale è $C_{20}H_{28}O$, quella del suo precursore, retinolo (vitamina A) è $C_{20}H_{30}O$, e i due interconvertono tramite una reazione di idrogenazione/deidrogenazione del carbonio terminale della coda polienica. Nel caso della MR questa interconversione avviene come parte delle reazioni coinvolte nel processo della visione. Tuttavia questo accade nelle fasi del fotociclo in cui il retinale è staccato dalla proteina. Invece nella BR il retinale rimane sempre all'interno della proteina e nella sua forma deidrogenata.

Come già detto, all'interno della proteina, una reazione di disidratazione unisce l'estremità carbossilica della catena ad una lisina (della catena G in BR o 7 in MR), formando un legame a base di Schiff (Fig. 6). L'altra estremità del retinale è formata da un anello ionone. La parte caratterizzante la funzione del retinale è la catena polienica, formata da atomi di carbonio legati da legami singoli e doppi alternati. La lunghezza e la conformazione di questa catena, nei retinali e in altri composti simili con proprietà fotoattive o cromatiche (come ad esempio i caroteni) determinano le proprietà ottiche.

In particolare, i retinali di BR e MR presentano alcune differenze a questo riguardo. Come anticipato, la fotoreazione di base è in entrambi i casi una trans-isomerizzazione, ma coinvolge due legami $C=C$ diversi (il n. 11 nella MR e il 13 nella BR) e in senso opposto (cis-trans in MR e trans-cis in BR). Se in prima approssimazione la lunghezza di assorbimento del retinale dipende dal numero di atomi di carbonio nella catena polienica, la delocalizzazione elettronica degli orbitali π è anche dipendente dalla conformazione cis o trans dei legami, e dalla distorsione indotta dalla proteina sul retinale stesso. Per questo motivo, a diverse conformazioni del retinale e/o del suo intorno proteico corrispondono diversi assorbimenti ottici. Specificamente, il retinale 11-cis della MR assorbe a 400nm nel suo stato fondamentale, e subisce uno shift di 20nm dopo l'isomerizzazione, assorbendo a 380nm nello stato all-trans. Viceversa, il retinale all-trans dello stato fondamentale della BR stabilisce il suo assorbimento a 380nm, che diventa 360nm dopo l'isomerizzazione allo stato 13-cis.

Oltre al retinale, anche alcune molecole di acqua interne partecipano al funzionamento delle rodopsine e che possono quindi essere considerate, insieme al retinale e alle lisine ponte, facenti parte del sito attivo della proteina. Queste molecole di acqua creano una rete di legami a idrogeno con le catene laterali di vari residui e con l'atomo di azoto della lisina. Il sito attivo della BR e MR e la rete di legami a idrogeno sono mostrati in Fig. 8. L'isomerizzazione del retinale porta alla distruzione o riposizionamento di alcuni legami a idrogeno, anche questi eventi rilevanti alla funzione delle due rodopsine.

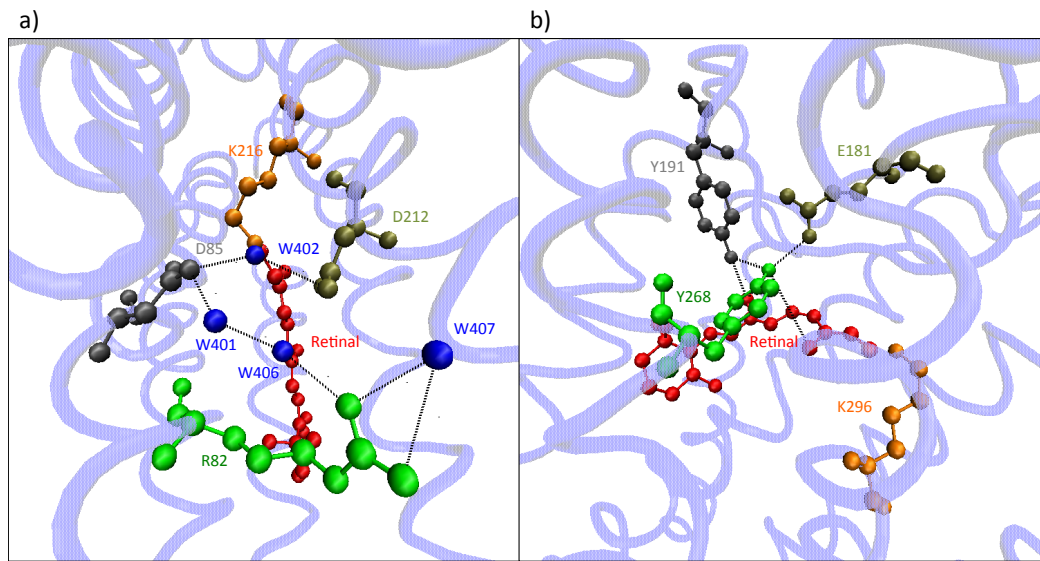


Figura 8: rete di legami a idrogeno di (a) rodopsina batterica dove le eliche sono in blu, il retinale in rosso unito alla lysina 216 in arancione, le molecole di acqua in blu ed i legami a idrogeno in nero; (b) rodopsina umana dove in rosso il retinale è legato alla lysina296 in arancione e le linee tratteggiate rappresentano i legami a idrogeno.

2.2.3 I Fotocicli delle rodopsine

L'assorbimento del fotone dà l'avvio ad una serie di variazioni strutturali delle proteine, che avvengono su diverse scale temporali. A partire dal suo stato fondamentale, la proteina entra in un ciclo all'interno del quale svolge la sua funzione per poi tornare al suo stato di partenza. Dal momento che il preciso stato strutturale della proteina influenza l'energia di eccitazione, i vari stadi si possono identificare con la loro lunghezza d'onda di assorbimento. Essi sono riportati nella Tabella 1^{22,23,24,25} e anche schematicamente nella Fig. 9, insieme ai loro tempi di vita (tempo medio di transizione allo stato successivo).

In riferimento alla Tabella 1 e alle Fig. 9-10, i passaggi fondamentali del fotociclo della BR sono:

1. assorbimento del fotone con conseguente isomerizzazione del retinale e deprotonazione della SB e protonazione del residuo vicino Asp85, e successivo movimento verso il basso del residuo Arg82
2. espulsione del protone dalla superficie extracellulare da parte del gruppo di rilascio del protone vicino a Glu194/Glu204
3. deprotonazione di Asp96 e riprotonazione della SB arrivando allo stato N
4. riprotonazione di Asp96 dalla superficie citoplasmatica, reisomerizzazione termica del retinale in configurazione all-trans verso lo stato O
5. deprotonazione di Asp85 e riprotonazione del gruppo di rilascio del protone rigenerando lo stato fondamentale

BR						
Id	λ_{ass} [nm]	Tempo di transizion e	E kcal/M (pH=5)	E kcal/M (pH=7)	E kcal/M (pH=9)	E kcal/M teorico
BR	570		0	0	0	0
BR*			50.5	50.5	50.5	50.5
BR*→K		5ps	0	0	0	0
K	590		11.7	11.7	11.7	11.7
K→L		2 μ s	20.7	20.6	20.6	
L	550		-0.3	-2	-0.9	5.5
L→M		100 μ s	17.4	10.6	16.7	
M	412		1.7	3.2	6.5	1.25
M→N		10ms	18.1	9.3	26.2	
N	560		34.4	-6.7	-4	0.08
N→O		5ms	2.5	22.8	14	
O	640		16.3	16	13.3	2.75
O→BR		5ms				

MR		
Sigla identificativa	λ_{ass} [nm]	Tempo di transizione
RHO/RHO*	498	
		3ps
BATHO	543	
		1ns
LUMI	497	
		1 μ s
META I	485	
		1ms
META II	380	
		1min
OPSINA		

Tabella 1: nella tabella sono riportati i nomi degli stati del fotociclo per ciascuna rodopsina, lo stato del fotociclo, la lunghezza d'onda di assorbimento di quello stato ed il tempo di transizione tra due stadi successivi.^{26 27} Nella prima colonna, il valore della lunghezza d'onda di assorbimento si riferisce al BR e RHO nello stato fondamentale, mentre il valore del tempo di vita è riferito allo stato elettronico eccitato dal quale avviene la transizione. Per la BR nelle quattro colonne successive sono riportate le energie libere relative allo stato fondamentale dei vari stati e le barriere di potenziale che vengono superate durante le transizioni ottenuti per tre diversi valori del pH e da uno studio teorico. I valori sono espressi in Kcal/mol.^{24,28,29,30,31}

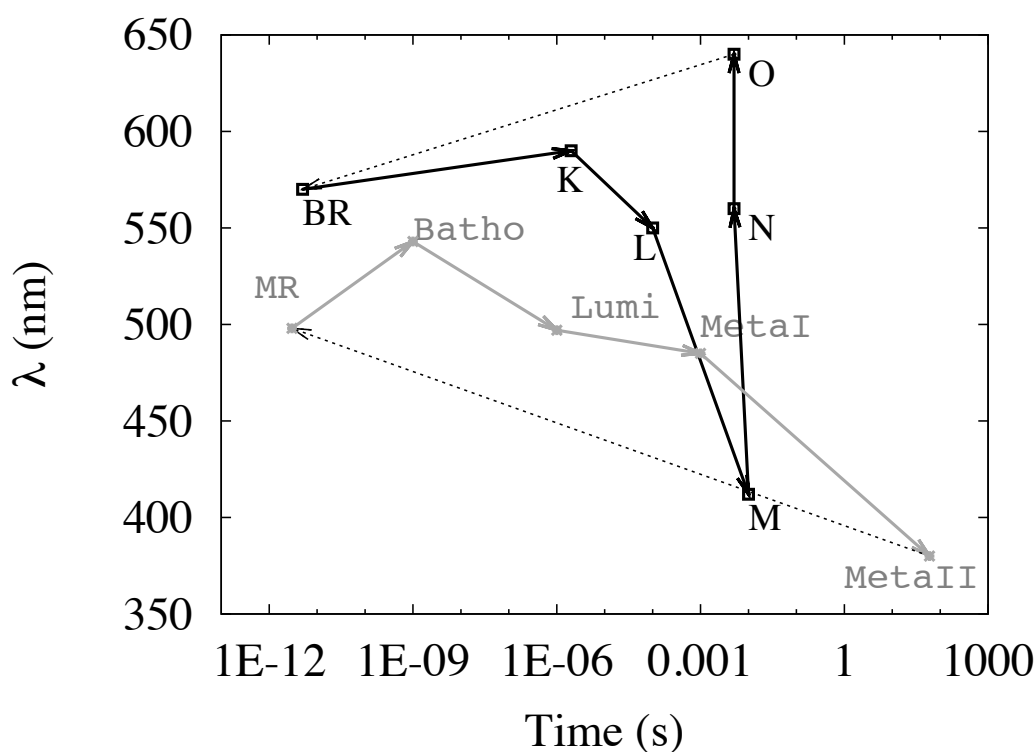


Figura 9: Lunghezza d'onda di assorbimento in funzione dei tempi di vita media di ciascun stato della BR, in nero, e della MR, in grigio. Le linee tratteggiate corrispondono alla transizione che riporta le proteine nello stato fondamentale.

La MR ha un diverso fotociclo, sempre però composto da 6 stati³², (Tab 1, Fig. 9-10). I passaggi fondamentali sono:

1. assorbimento del fotone e isomerizzazione del retinale
2. rilassamento termico lento (circa 1ms) attraverso una serie di stati fotointermedi: Batho, Lumi e MetaI dove avvengono degli spostamenti dell'anello ionone del retinale
3. nello stato Meta II si ha un movimento rigido dell'elica F e la base di Schiff è deprotonata
4. formazione di due aperture tra le eliche 1 e 7 e l'altra tra le eliche 5 e 6.
5. espulsione del retinale dall'apertura tra le eliche 1 e 7. La proteina opsina non è più legata al retinale.
6. la proteina opsina aspetta di legarsi nuovamente con un retinale in configurazione cis per tornare allo stato fondamentale e dare il via ad un nuovo fotociclo. Il nuovo retinale entra nella proteina attraverso la fessura creata tra le eliche 5 e 7.

In questo caso, come precedentemente osservato, il rilascio del retinale da parte di MR e conseguente trasformazione in opsina fa parte del fotociclo della MR.

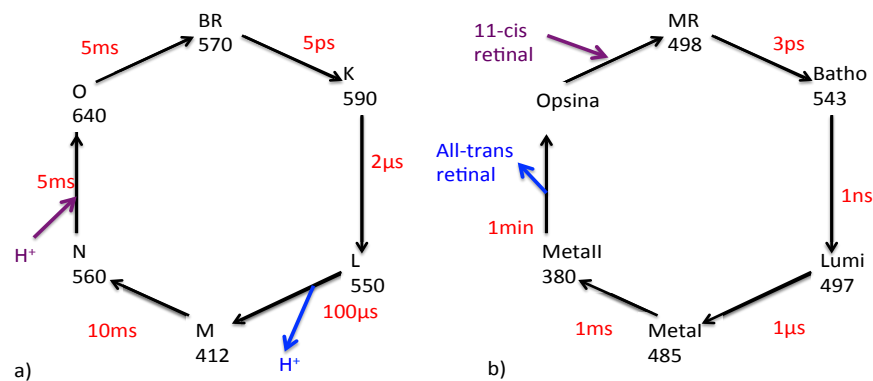


Figura 10: a) schematica rappresentazione del fotociclo della rodopsina batterica. In blu il protone rilasciato verso la superficie extracellulare durante lo stato M intermedio, in viola il protone acquistato dalla proteina durante la transizione N - O dalla superficie citoplasmatica da Asp96. b) rappresentazione del fotociclo della rodopsina umana dove si nota dopo lo stato Meta II il rilascio del retinale *all-trans* ottenendo la proteina opsina senza il ligando. A questo punto la MR è in grado di legarsi con un nuovo retinale in configurazione *cis*. In rosso i tempi di transizione tra gli stati ed in nero sotto ai nomi degli stati i massimi di assorbimento espressi in nm.

Si può osservare (Fig. 10) che per entrambe le rodopsine i tempi di transizione da uno stato all'altro aumentano dalla prima transizione all'ultima passando dalla transizione iniziale ultraveloce ad una finale sulla scala del ms (BR) fino al minuto (MR).

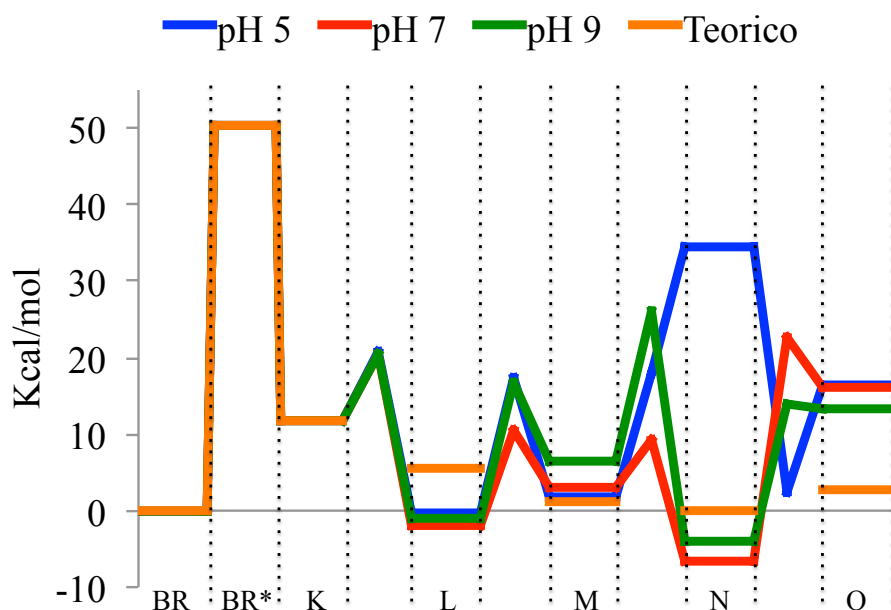


Figura 11: livelli energetici della BR. Le righe blu, rossa e verde sono ottenute da dati sperimentali al variare del pH, mentre le righe arancio sono ottenute dai dati teorici. Lo stato BR è stato convenzionalmente scelto come riferimento delle energie

I dati dei livelli energetici e barriere per la BR della Tabella 1 sono anche riportati graficamente in Fig. 11. Essi sono stati estratti da vari lavori di letteratura, sia sperimentali che teorici^{24,28,30,31,32}. Partendo dallo stato fondamentale convenzionalmente ad energia 0, l'assorbimento del fotone porta la

proteina in uno stato eccitato, BR*, con una transizione verticale di circa 50Kcal. Da questo stato il retinale isomerizza e la proteina va nello stato K spontaneamente, usando quindi l'energia precedentemente fornita dal fotone. Gli altri passaggi avvengono tutti nello stato fondamentale. Come si può osservare dal grafico in Fig. 12 il tempo di transizione tra uno stato e l'altro è correlato all'altezza della barriera tra due stati successivi ed in particolare all'aumentare del tempo di transizione aumenta anche l'altezza della barriera di potenziale.

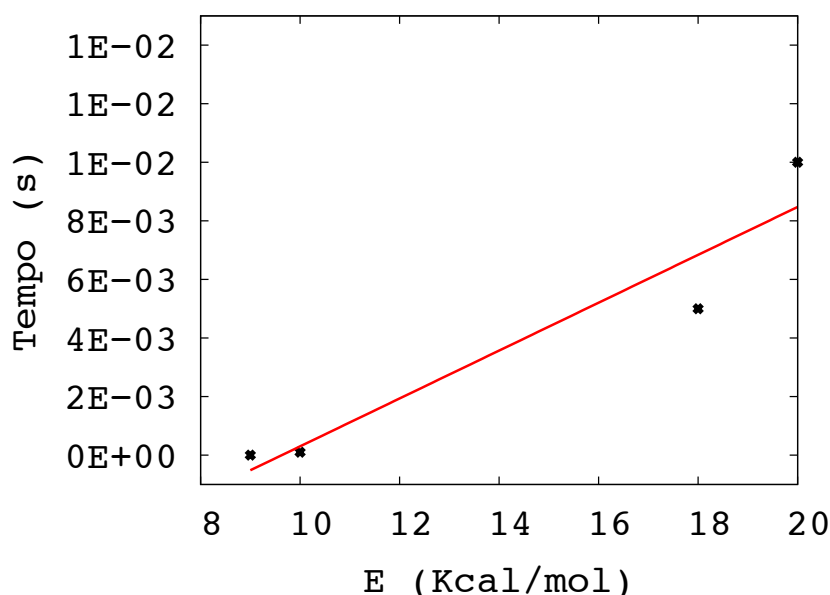


Figura 12: grafico dell'energia della barriera di potenziale tra due stati successivi rispetto al tempo di transizione. I dati delle energie sono relativi allo studio sperimentale a pH=9. La linea rossa è la retta di regressione lineare.

Dalla Fig. 11 si può notare come sia i dati sperimentali a pH 7 e 9 e i dati teorici siano in accordo tra di loro, mentre i dati a pH 5 si discostano soprattutto nello stato N dagli altri valori. Questa differenza deriva dal fatto che non sono riusciti a determinare bene il rate a $\text{pH} = 5^{30}$. I valori delle barriere sono disponibili soltanto per i dati sperimentali e sono diversi a seconda del valore del pH. Si nota come lo stato N sia quello ad energia più bassa, mentre lo stato O sia lo stato intermedio ad energia più alta. L'energia dello stato O, inoltre, non è attendibile come quella degli altri stati soprattutto per lo studio teorico in quanto la struttura utilizzata non era completa. Lo studio è stato effettuato tramite una simulazione di dinamica molecolare utilizzando il FF AMBER su tutti gli stati del fotociclo. Per lo stato O è stata utilizzata la struttura 1JV7 che non è completa e per questo è stato utilizzato un ibrido tra questa struttura ed i gruppi sul lato extracellulare dello stato M^{32} , che non rappresenta esattamente lo stato O.

Lo stesso tipo di analisi non è stato possibile per la MR poiché non vi sono dati sulle energie dei vari stati del fotociclo.

2.3 Scopo del lavoro di tesi e scelte metodologiche

Da quanto detto fino a qui, risultano evidenti due caratteristiche dei processi di cui le rodopsine sono protagoniste: 1. La presenza di tempi caratteristici che spaziano su diversi ordini di grandezza e 2. La presenza di molti stati stabili o metastabili. In realtà la multi-scalarità (generalmente estesa non solo al dominio temporale, ma anche a quello spaziale) e la multi-stabilità sono caratteristiche comuni a praticamente tutti i processi biologici, perché la materia vivente è per sua natura organizzata gerarchicamente³³. La situazione in cui una reazione chimica relativamente veloce e localizzata nel sito attivo di una proteina dà vita ad una cascata di eventi lenti su scala cellulare o addirittura tissutale o estesa all'intero organismo, non solo è frequente, ma è una caratteristica intrinseca del funzionamento dei sistemi viventi.

Chiaramente queste due caratteristiche comportano enormi difficoltà negli approcci teorici, e specificamente in quelli simulativi. Infatti, per tenerne conto in maniera accurata, è necessario l'utilizzo congiunto di diverse tecniche e approcci modellistici, ciascuno specializzato per determinate scale spazio-temporali. Inoltre, anche fissata una determinata scala e quindi un determinato approccio, la multi-stabilità e il superamento di barriere che separano i diversi stati è un problema che affligge i simulatori, perché coinvolge la rappresentazione di un cosiddetto "evento raro".

In questa tesi si affrontano entrambi i problemi in un approccio modellistico e simulativo unitario. Il problema delle scale temporali multiple è affrontato utilizzando una rappresentazione multi-scala di una singola proteina trattata a due risoluzioni spaziali diverse (alta risoluzione per il sito attivo e bassa risoluzione per il resto della proteina). Un terzo livello di risoluzione, il "continuo" è sottinteso nel trattamento implicito di solvente e membrana, trattati come mezzi estesi che producono determinati effetti sulle interazioni. La separazione delle scale spaziali consente di avere un guadagno notevole nelle scale temporali, e quindi di simulare le fasi dell'intero fotociclo. Il problema della multi-stabilità viene affrontato con la combinazione di modelli per ciascuno degli stati del fotociclo in un unico modello multi-stabile, che consente quindi di rappresentare il passaggio da uno all'altro.

Il problema qui affrontato è generale, ma come esempio applicativo sono state scelte le rodopsine per vari motivi. Anzitutto, la multi-stabilità e multi-scalarità temporale del loro fotociclo è paradigmatica, e rappresenta quindi un caso ideale per affrontare il problema. In secondo luogo, entrambe le proteine sono molto studiate, sia dal punto di vista sperimentale che teorico, e quindi ci sono dati con cui confrontare i risultati ottenuti in questa tesi.

Esiste anche un terzo motivo, di natura più prettamente tecnica. Come sarà più chiaro nel prossimo capitolo, l'approccio qui usato necessita dati strutturali da usare come riferimento. L'accuratezza e la quantità di questi incidono sull'accuratezza e sulla predittività del modello. Le rodopsine, nonostante le difficoltà nella loro cristallizzazione comuni a tutte le proteine di membrana, sono studiate da tempo, per cui esistono molte determinazioni sperimentali della loro

struttura, con accuratezza ormai sufficientemente elevata. Questo è soprattutto vero per la BR, per la quale sono disponibili sul database PDB strutture sperimentali per tutti gli stati del fotociclo. Per questo motivo la BR è stata utilizzata come caso test dell'approccio metodologico. Il modello è stato poi trasferito alla MR, che però rappresenta un banco di prova più stringente perché per questa proteina esistono un numero inferiore di determinazioni sperimentali, che sono mancanti per alcuni degli stati. Tuttavia la MR appartiene ad una famiglia, quella dei GPCR, che sono interessanti bersagli farmacologici, per cui ci si aspetta che lo studio su MR possa avere maggiore interesse applicativo di quello sulla BR.

Una parte preliminare di questo lavoro di tesi è quindi consistita nell'analisi e scelta, tra tutte le strutture disponibili sul PDB per le due rodopsine, quelle più adatte ad essere usate come base per i modelli. I criteri su cui la scelta si è basata sono la risoluzione (più alta possibile) e la completezza della struttura. Nel caso in cui la statistica delle strutture disponibili era abbondante, la scelta è ricaduta sulle più recenti, che presentano solitamente un'accuratezza maggiore. Per alcuni degli stati non esisteva nessuna singola struttura che soddisfacesse il criterio della completezza. In quel caso si è costruita una struttura completa utilizzando frammenti di diverse strutture opportunamente sovrapposti e allineati. Il risultato finale è una serie di strutture rappresentative degli stadi del fotociclo, che saranno la base del modello multistabile, costruito come spiegato nel prossimo capitolo. I dettagli della procedura di selezione delle strutture sono riportati in Appendice A.

Capitolo 3. Modelli e metodi di simulazione

In questo capitolo sono descritti i metodi di simulazione ed i modelli utilizzati in questa tesi. I metodi di simulazione si classificano generalmente in base al modo in cui il sistema viene descritto, che si esplica nella scelta dei gradi di libertà e delle interazioni tra di essi, e al modo in cui il sistema viene evoluto nel tempo, ovvero come vengono integrate le equazioni del moto. Riguardo al secondo punto, si distingue generalmente una dinamica completamente deterministica, e in tal caso si parla di dinamica molecolare (MD) “standard”, o stocastica (ad esempio, nei metodi Monte Carlo³⁴), oppure con equazioni parzialmente stocastiche (dinamica di Langevin o Browniana³⁵). In questa tesi si parte da un approccio base puramente MD, quello più robusto e utile per analizzare le proprietà energetiche e strutturali del sistema, e considerato come punto di partenza per successive analisi con dinamiche più sofisticate.

Riguardo al tipo di descrizione del sistema e alle interazioni, queste sono generalmente commensurate al tipo di processo che si vuole simulare, all'accuratezza richiesta, e soprattutto alle scale temporali e spaziali coinvolte. In particolare, quando è necessario simulare fenomeni molto veloci e localizzati spazialmente, è opportuno adottare una rappresentazione atomistica del sistema in modo da avere alta risoluzione spaziale e temporale. Come già commentato in precedenza, le hamiltoniane per la descrizione atomistica dei sistemi biologici sono state parametrizzate accuratamente negli ultimi decenni e quindi danno una descrizione sufficientemente accurata del sistema, almeno per quanto riguarda strutture e dinamiche localizzate nello spazio e nel tempo. Viceversa, le descrizioni atomistiche non sono appropriate per simulare processi che si svolgono su tempi macroscopici, a causa dei limiti imposti dalla velocità dei calcolatori. Per una proteina delle dimensioni delle rodopsine rappresentata al livello atomistico, si può raggiungere la scala del microsecondo solo usando la parallelizzazione massiccia. Inoltre su queste scale emergono anche inconsistenze ed errori sistematici dei campi di forze empirici per dinamica classica comunemente usati³⁶ che spesso falliscono nella valutazione delle energie relative delle diverse strutture secondarie.

In questa tesi si è adottata una rappresentazione multi-scala del sistema in esame, che risolve entrambi i problemi relativi alle rappresentazioni esclusivamente atomistiche. Poiché il passo ultraveloce del fotociclo (isomerizzazione) è localizzato nel retinale, mentre il resto della proteina partecipa a transizioni più globali e lente, si è adottata una rappresentazione (quasi)atomistica per il retinale, mentre il resto della proteina è rappresentato ad un livello più grossolano (Coarse Grained, CG), detto “minimalista”, cioè con un singolo centro interattivo per amminoacido^{37,38}. Il primo effetto di questa rappresentazione è che con un costo computazionale molto contenuto (simulazioni su singoli processori) è possibile rappresentare l'intero fotociclo.

Questo guadagno è ottenuto alle spese di uno sforzo preliminare di parametrizzazione della parte CG del modello, poiché non esistono, a differenza

del caso atomistico, hamiltoniane standard per i modelli CG. Tuttavia questo fatto è stato mutato in un vantaggio: sfruttando il relativamente basso numero di parametri del modello CG è stato possibile ottimizzarli in modo da riprodurre i valori sperimentali delle energie relative dei vari stati, ottenendo in ultima analisi una maggiore accuratezza rispetto ad una analoga eventuale rappresentazione atomistica.

In questo capitolo vengono riportati gli elementi base degli approcci metodologici adottati. Le modifiche specifiche a tali approcci operate come lavoro di tesi sono riportate invece nel capitolo successivo, che contiene i risultati di questo lavoro di Tesi.

3.1 Dinamica molecolare classica

Simulare la dinamica molecolare significa risolvere numericamente le equazioni dinamiche del moto per ottenere l'evoluzione temporale delle variabili interne $x_i(t)$ (normalmente le coordinate cartesiane dei centri interattivi). Il caso più semplice è la dinamica Newtoniana

$$F_i = m_i \ddot{x}_i \quad (3.1)$$

dove m_i sono le masse associate alle variabili ed F_i le forze totali agenti su di esse. F_i è il negativo del gradiente dell'energia potenziale totale $U(\{x_i\})$

$$F_i = - \frac{\partial U(\{x_i\})}{\partial x_i} \quad (3.2)$$

Come già detto, nel caso della dinamica molecolare classica, U viene rappresentata come somma di termini analitici opportunamente parametrizzati (si veda sezione successiva). Una volta fissate U e le condizioni iniziali (posizioni e velocità) la traiettoria si può ricavare con una procedura numerica ricorsiva: note le posizioni e le velocità ad un tempo t , quelle al tempo $t+\Delta t$ si ottengono con uno sviluppo in serie di Taylor

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2} \frac{F(t)}{m} \Delta t^2 + O(\Delta t^3) \quad (3.3)$$

In linea di principio con questa procedura si può ottenere una traiettoria lunga a piacere discretizzata nel tempo con Δt (passo di simulazione, o *timestep*). Nella pratica ci sono vari fattori che limitano la lunghezza massima della traiettoria, tra cui il fatto che ad ogni passo viene sommato un errore numerico, che va tenuto sotto controllo. A tale scopo, ad esempio, è spesso usato un algoritmo di integrazione alternativo, quello di Verlet

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{F(t)}{m} \Delta t^2 + O(\Delta t^4) \quad (3.4)$$

per il quale l'errore numerico è al quarto ordine in Δt anziché al terzo, e quindi più piccolo.

La limitazione dell'errore numerico spinge ovviamente nella direzione di mantenere basso il Δt . Tuttavia, il costo computazionale di una simulazione è direttamente proporzionale al numero di passi necessari per realizzarla, e quindi, fissata la lunghezza della traiettoria in termini di tempo, il costo (e di conseguenza il tempo "reale" di calcolo) è inversamente proporzionale al valore di Δt . Pertanto, è necessario scegliere il Δt maggiore possibile compatibilmente con errori numerici accettabili. A tale scopo si sceglie solitamente Δt uno o due ordini di grandezza minore del periodo della più alta frequenza del sistema

$$\Delta t \ll \frac{2\pi}{\omega_{\max}} \quad (3.5)$$

Nel caso di sistemi atomistici con idrogeno trattato esplicitamente queste sono le frequenze di vibrazione dei legami C-H, N-H o O-H che si aggirano intorno ai 3000-3500 cm^{-1} , e necessitano tempi di integrazione inferiori al fs. Se i legami con l'idrogeno sono bloccati o per modelli con idrogeno implicito, va considerato il gruppo di frequenze immediatamente inferiore, che riguarda la vibrazione dei legami doppi C=C, C=N o C=O, tra i 1700 e i 2000, per i quali un Δt di 1-2 fs è sufficiente. Per i modelli CG il passo di integrazione può essere molto maggiore, dipendentemente dal livello risoluzione. Per i modelli minimalisti qui adottati si aggira intorno alle 20-50 fs. Per modelli multi-scala in cui il sistema è rappresentato a diversi livelli di risoluzione, il Δt è generalmente lo stesso per tutto il sistema, e quindi è la parte a risoluzione più alta che lo limita.

Le equazioni 3.3 e 3.4 descrivono una dinamica del sistema ad energia totale fissata, quindi campionano, per traiettorie sufficientemente lunghe, l'insieme statistico microcanonico, in cui sono costanti il numero totale di particelle, il volume del sistema e l'energia totale (in gergo, simulazioni NVE). Questa è generalmente una situazione poco realistica, in quanto normalmente le coordinate termodinamiche costanti sono temperatura (simulazione NVT, insieme canonico), solitamente anche pressione (simulazione NPT), e a volte anche potenziale chimico (insieme gran-canonico). Le simulazioni NVT si realizzano accoppiando il sistema ad un termostato. L'algoritmo più semplice consiste nel riscaldare le velocità ad ogni passo secondo la relazione $v' = \lambda v$ con

$$\lambda^2 = \frac{T_0}{T} \quad (3.6)$$

in modo che la temperatura venga corretta verso il valore di riferimento T_0 ⁴⁰. Questo metodo può a volte generare forti oscillazioni della temperatura intorno a T_0 , non facilmente controllabili e innaturali. Un metodo più flessibile è il termostato di Berendsen³⁹ che consiste nell'accoppiare il sistema ad un bagno termico attraverso un parametro che corrisponde ad un tempo di rilassamento τ della temperatura. La temperatura diventa una variabile indipendente

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad (3.7)$$

Si può dimostrare che questo metodo è numericamente equivalente al precedente se si pone

$$\lambda^2 = 1 + \frac{\Delta T}{\tau} \left\{ \frac{T_0}{T(t - \frac{\Delta t}{2})} - 1 \right\} \quad (3.8)$$

Per $\tau = \Delta t$ si ottiene il riscaldamento semplice, mentre il valore di τ regola il rilassamento del sistema verso T_0 ⁴⁰. Infine il termostato di Nose-Hoover, specificamente pensato per il corretto campionamento dell'insieme canonico, consiste in un accoppiamento lagrangiano del sistema ad un bagno termico secondo la relazione

$$L = \frac{1}{2} m_i s^2 \dot{v}_i^2 - U(r_i) + \frac{1}{2} Q \frac{\dot{s}^2}{s^2} - gKT_0 \ln(s) \quad (3.9)$$

dove s il fattore di riscaldamento, assume il ruolo di variabile lagrangiana addizionale, Q è la sua inerzia, e g è la costante di accoppiamento⁴¹. Questo tipo di approccio è generalizzabile anche all'introduzione di un barostato per simulazioni a pressione costante⁴².

I termostati (e barostati) fin qui descritti, sebbene di natura completamente deterministica, introducono elementi esterni al sistema (il bagno termico) che permettono all'energia del sistema di fluttuare a causa di un "ambiente", esemplificato da un grado di libertà addizionale che in realtà rappresenta un intero insieme statistico. In generale l'effetto di gradi di libertà eliminati può essere tenuto in conto anche attraverso equazioni esplicitamente non deterministiche. È questo il caso della dinamica di Langevin, descritta dall'equazione

$$m_i a_i = F_i - m_i \gamma_i \dot{x}_i + F_i^S \quad (3.10)$$

in cui F_i^S rappresenta una forza stocastica e $m_i \gamma_i \dot{x}_i$ rappresenta la forza di attrito. Il coefficiente di attrito è correlato alle fluttuazioni della forza stocastica dalla relazione

$$\langle F_i^S(0) F_i^S(t) \rangle = 2m_i K_B T \gamma_i \delta_{ij}(t) \quad (3.11)$$

valida per forze stocastiche scorrelate temporalmente (processi "senza memoria"; nel caso più generale al secondo membro si può avere una funzione di memoria al posto della delta di Dirac). Poiché la forza stocastica compie lavoro sul sistema, mentre il termine dissipativo sottrae energia, la dinamica di Langevin agisce come un termostato, infatti ci si riferisce ad essa anche come "termostato di Langevin"⁴³. Un caso tipico in cui si usa questa equazione è la dinamica in solvente implicito: invece di immergere la molecola di soluto in una scatola di molecole di acqua, che generalmente consumano dal 70% al 90% del tempo di calcolo, la si immerge in un continuo con proprietà dielettriche che tengono conto degli effetti di polarizzazione, mentre gli effetti dinamici degli urti delle molecole di acqua sul soluto sono tenuti in conto dall'eq 3.11. La dinamica Browniana viene spesso considerata come il caso sovrasmorzato della dinamica di Langevin, in cui si può trascurare il termine inerziale (quello proporzionale a m_i) e ottenendo

un'equazione differenziale del primo ordine, la cui integrazione numerica è tra l'altro molto più veloce⁴⁴.

Da quanto fin qui detto, risulta chiaro che le dinamiche stocastiche (Langevin o Browniana) sono le più adatte per simulare in maniera cinematicamente realistica i sistemi rappresentati con modelli a bassa risoluzione. Infatti i gradi di libertà interni eliminati nel processo di semplificazione del modello si comportano esattamente come quelli del solvente, con impulsi stocastici ed effetti di smorzamento. In pratica, usando la dinamica newtoniana pura su un sistema descritto con un modello a bassa risoluzione, si ottiene un effetto noto come “accelerazione fittizia”, dovuto al fatto che l'hamiltoniana semplificata ha molti meno minimi da esplorare e quindi il sistema esplora più velocemente i pochi minimi rimasti. La dinamica stocastica corregge questo problema, “rallentando” il sistema e riportandolo ad una cinetica realistica⁴⁵. Tuttavia in molti casi, quando lo scopo della simulazione è l'esplorazione dello spazio delle configurazioni del sistema anziché la riproduzione della corretta cinetica, l'accelerazione fittizia è un vantaggio, e quindi in quel caso è appropriato usare la dinamica newtoniana.

3.2 Campi di forze empirici per modelli atomistici

Nell'ambito delle tecniche convenzionali di dinamica molecolare classica, l'energia potenziale U del sistema viene definita tramite la somma di termini analitici, anche chiamato campo di forze (*Force Field*, FF). Parlando ad esempio del caso dei modelli atomistici, nei FF comunemente usati per proteine e acidi nucleici si distinguono due classi di termini

$$U_{\text{tot}} = U_{\text{bonded}} + U_{\text{non-bonded}} \quad (3.12)$$

dove U_{bond} contiene la parte di termini che descrivono i legami covalenti, mentre $U_{\text{non-bonded}}$ contiene la parte delle interazioni cosiddette “non legate”. Precisamente

$$U_{\text{bonded}} = U_{\text{bonds}}(R_i) + U_{\text{angles}}(\theta_i) + U_{\text{dihedrals}}(\phi_i) \quad (3.13)$$

dove R_i rappresentano tutti i possibili legami chimici del sistema, θ_i tutti i possibili “angoli di legame” formati da triplette di atomi legati a due a due, e ϕ_i tutti i possibili angoli diedrali di formati da quattro atomi legati. Le forme funzionali scelte sono le più semplici possibile: potenziali armonici per U_{bonds} e U_{angles} e coseni per $U_{\text{dihedrals}}$ (per tenere conto della periodicità del diedro). Chiaramente la costruzione di questo termine presuppone la conoscenza della topologia di legame del sistema in esame, che fa parte del modello stesso.

Il termine di interazioni non legate viene calcolato tra tutte le coppie non coinvolte in una qualche interazione non legata, e si divide in una parte elettrostatica e una di Van Der Waals

$$U_{\text{non-bonded}} = U_{\text{electrostatic}} + U_{\text{VanDerWaals}} \quad (3.14)$$

La prima è un'interazione coulombiana pura oppure schermata tramite dielettrico in caso si consideri un solvente implicito. Per valutare questo termine è necessario attribuire agli atomi delle cariche efficaci che, a parità di elemento chimico, tengono conto dell'intorno chimico e del trasferimento di carica ai primi vicini. Quindi questo termine dipende anche dalla topologia di legame del sistema. Similmente accade, anche se in misura minore, per i parametri del termine $U_{\text{VanDerWaals}}$, che descrivono gli effetti di volume escluso e quelli di dispersione. In conclusione i valori dei parametri del modello vengono assegnati con la seguente procedura: a partire dalla topologia della connettività chimica del modello si calcola il numero dei termini "bonded"; i valori dei parametri vengono poi assegnati a ciascun termine sulla base del tipo di atomi coinvolti nel legame. Il "tipo" dipende sia dall'elemento, che dal suo contorno chimico. Ad esempio esistono molti diversi tipi di atomi di carbonio, a seconda della loro ibridizzazione e dei tipi di ligandi che hanno. I parametri "non bonded" vengono assegnati solo sulla base del tipo di atomo. Durante l'assegnazione non si deve prendere nessuna decisione arbitraria: la connettività è determinata dalla struttura iniziale, i valori numerici dei parametri (in numero di diverse migliaia) sono fissati e caratterizzano il campo di forze (CF). Risulta quindi che l'accuratezza di un modello atomistico, e quindi dell'intera simulazione, è in ultima analisi dovuto all'accuratezza del CF.

Esistono diversi campi di forze ottimizzati per bio-molecole. I più usati sono Amber⁴⁵, CHARMM⁴⁶, GROMOS⁴⁷, ciascuno con diverse versioni, che vengono continuamente aggiornate e migliorate, in un processo che dura da diversi decenni. I metodi usati per la parametrizzazione coinvolgono il fit dei vari termini del CF su dati strutturali, vibrazionali ed elettrostatici ottenuti da calcoli quantistici su molecole piccole, dal confronto con dati strutturali sperimentali ad alta risoluzione, e con dati termodinamici di vario genere. Le prestazioni relative dei vari CF variano a seconda del tipo di simulazione e di osservabili calcolate. Per questo studio ho utilizzato il software GROMACS⁴⁸ per creare delle simulazioni all-atom delle due rodopsine. In tali simulazioni ho utilizzato il campo di forze OPLS (Optimized Potentials for Liquid Simulations)⁴⁹ dove i termini U_{bonds} e U_{angles} sono dati da potenziali armonici, il termine U_{dihedral} da una somma di coseni ed il termine $U_{\text{non-bonded}}$ è descritto nell'Eq. 3.15.

$$U_{\text{non-bonded}} = \sum \left(\frac{A}{r^{12}} - \frac{B}{r^6} + \frac{q_i q_j e^2}{4\pi\epsilon_0 r} \right) \quad (3.15)$$

dove i termini con r^{12} e con r^6 rappresentano le interazioni VdW e l'ultimo termine le interazioni elettrostatiche.

Di particolare interesse per questa Tesi è il cosiddetto approccio ad "atomi uniti" (united atoms, UA)⁵⁰, che consiste in una rappresentazione quasi-atomistica, in cui gli atomi di idrogeno non sono rappresentati esplicitamente, ma implicitamente nell'atomo "pesante" a cui sono legati (Fig. 13). Poiché gli atomi di idrogeno sono numericamente prevalenti nelle bio-molecole, questo approccio consente di risparmiare una grande quantità di tempo di calcolo, ed era molto popolare negli anni 70-80, quando ancora la potenza di calcolo non consentiva rappresentazioni interamente atomistiche delle proteine.

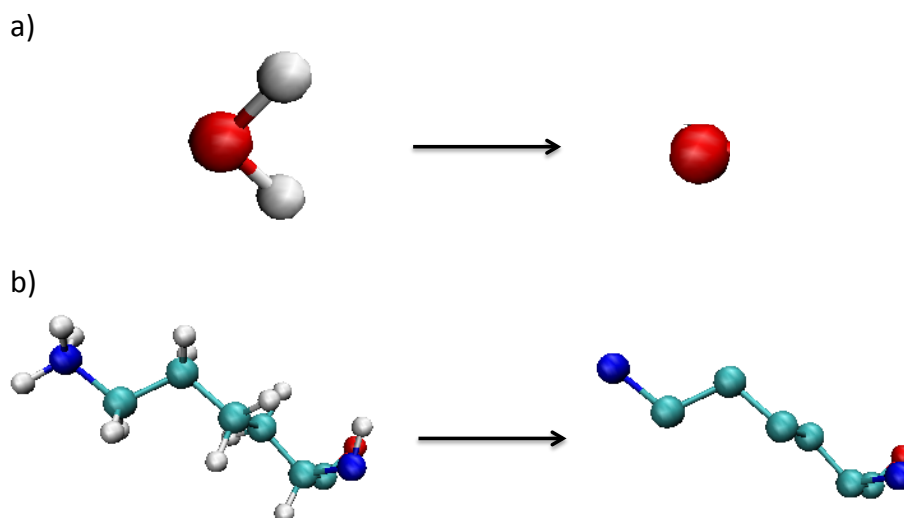


Figura 13: Passaggio dal modello all-atom (sinistra) al modello United Atom (destra) per una molecola di acqua in a) e per la lisina in b).

L'interesse di tale approccio per questo lavoro è dovuto al fatto che si vuole fornire la rappresentazione “minimalista” che consenta di descrivere un certo fenomeno. Da questo punto di vista, un approccio interamente atomistico del retinale è superfluo. Inoltre la rappresentazione UA ha il vantaggio di avere meno parametri, che torna utile perché alcuni di essi devono essere ri-ottimizzati per integrare tale rappresentazione insieme a quella CG nell'approccio multi-scala. Tutto considerato, una rappresentazione UA del sito attivo appare più appropriata di una completamente atomistica, pur rimanendo sufficiente a rappresentare la sua transizione strutturale del retinale.

3.3 Modelli “a grana grossa” (Coarse Grained)

Il “*coarse graining*”, ovvero la riduzione della risoluzione nella rappresentazione di un sistema, consiste nell'inglobare un certo numero di gradi di libertà in uno singolo, e ridurre così il numero totale di gradi di libertà del sistema, in linea di principio senza limiti sul livello di riduzione operata. Infatti in letteratura si trovano esempi di *coarse graining* operati a moltissimi diversi livelli di risoluzione, da quelli “quasi atomistici” (uno ad esempio è il già citato modello UA) a quelli in cui un singolo centro interattivo rappresenta un'intera proteina o un suo dominio, e quindi (decine di) migliaia di atomi. Anche restringendosi a risoluzioni intermedie, per le proteine sono stati adottati nel corso degli anni approcci a diversa risoluzione, generalmente da 1 a 6 centri interattivi (*bead*) per ammino acido (Fig. 14)³⁸. I modelli con rappresentazione esplicita dei gruppi C=O e NH consentono una più semplice descrizione dei legami a idrogeno che stabilizzano eliche e foglietti, mentre la rappresentazione esplicita della catena laterale dell'ammino-acido R consente una semplificazione della parte *non bonded* del potenziale. Tuttavia, il modello a singola *bead*, oltre ad essere tra

questi quello a guadagno massimo in termini di costo computazionale, ha il vantaggio di ridurre il sistema ad una catena lineare, quindi rappresenta quello a topologia connettiva minimale, pur mantenendo la possibilità di descrivere le diverse strutture secondarie.

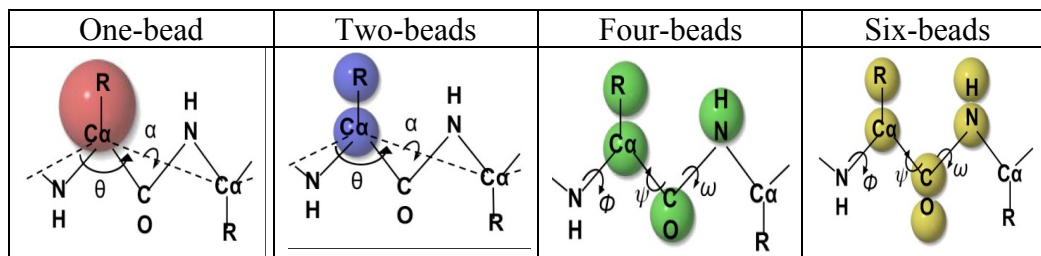


Figura 14: Descrizione sommaria dei più diffusi modelli CG a media risoluzione. Sono indicate le coordinate interne che descrivono la conformazione dello scheletro della proteina e quindi le strutture secondarie.

Un'altra scelta che si deve fare una volta stabilito il livello di risoluzione, è dove localizzare il centro interattivo. Le scelte possibili sono state nel corso degli anni, il centro di massa o il centro geometrico degli atomi inclusi nella *bead*, oppure specifici atomi. Tra questi i più logici sono il $C\beta$, il primo atomo di carbonio della catena laterale, localizzato in un punto solitamente abbastanza centrale rispetto a tutto l'amminoacido, o il $C\alpha$ ³⁸. La scelta fatta in questa tesi è precisamente quest'ultima: la proteina viene rappresentata tramite le coordinate dei suoi $C\alpha$, ciascuno rappresentativo di un amminoacido. Apparentemente questa è la scelta meno razionale, in quanto il $C\alpha$ è decentrato rispetto all'amminoacido, considerando che la catena laterale, specialmente per gli amminoacidi più grandi, può estendersi molto. E in effetti questo comporta qualche difficoltà nella rappresentazione delle interazioni di volume escluso e in generale delle non bonded a medio raggio. Tuttavia questa difficoltà viene ripagata in una semplificazione dei termini del potenziale della parte bonded. Infatti, grazie alla geometria rigida del legame peptidico che vincola il gruppo $C\alpha$ -CO-NH- $C\alpha$ su un piano, la distanza $C\alpha$ - $C\alpha$ risulta fissata e indipendente dalla struttura secondaria e terziaria; le altre due variabili residue che determinano la struttura dello scheletro della proteina sono l'angolo θ e il diedro ϕ che risultano dipendenti solo dalla struttura secondaria e in corrispondenza biunivoca con essa³⁹. In altre parole, nel modello *one-bead $C\alpha$ based*, anche detto minimalista, le variabili interne conformazionali costituiscono un insieme sufficiente e completo per descrivere correttamente le strutture secondarie.

Come già detto, la semplificazione nella parte *bonded* del potenziale si paga con una complicazione della parte non-bonded: i potenziali che descrivono il volume escluso e l'idrofobicità sono necessariamente anisotropi; in aggiunta, in assenza di rappresentazione della catena laterale, le sue variazioni conformazionali ed eventuali interazioni chimiche ed elettrostatiche vanno rappresentate implicitamente in un potenziale unico, che diventa molto complesso. Le soluzioni che sono state trovate in letteratura per affrontare questi problemi sono passate in rassegna qui di seguito.

3.3.1 Modelli a rete elastica e dinamica essenziale

Una soluzione molto elegante e semplice sono i modelli a rete (NM = Network Model), di cui le reti elastiche costituiscono quello base e il più facilmente implementabile⁵¹. Ciascun centro interattivo (Ca in questo caso⁵²) viene considerato il nodo di una rete, legato a tutti gli altri a distanza minore di un determinato raggio di cutoff (r_{cut} , che è un parametro del modello) da un'interazione armonica

$$U = \frac{1}{2}k(R_{ij} - R_{ij}^0)^2 \quad (3.16)$$

dove la costante elastica k è un altro parametro del modello e le distanze di equilibrio di ciascuna coppia di atomi R_{ij}^0 sono estratte da una struttura di riferimento, solitamente sperimentale, che è un input del modello (Fig. 15). Per costruzione, questo modello è il più semplice possibile che ha come minimo una data struttura di riferimento. I parametri k e r_{cut} vengono aggiustati in modo da far coincidere le fluttuazioni termiche del modello con quelle sperimentali ricavate dal fattore B.

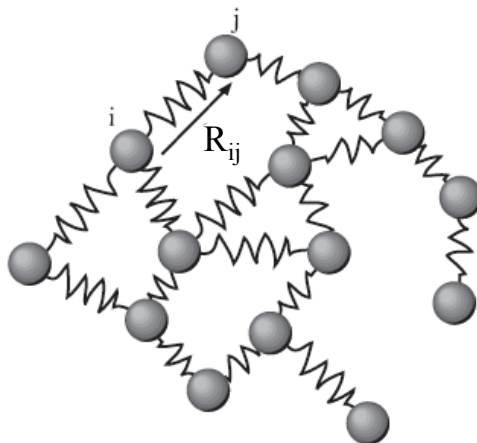


Figura 15: Schematizzazione dell'ENM. Le palline grigie sono i centri interattivi e le molle rappresentano le interazioni. R_{ij} è la distanza tra l'atomo i-esimo e j-esimo.

L'estrema semplicità di questi modelli consente l'analisi dei modi lenti di sistemi anche molto grandi, come interi capsidi di virus. Infatti il calcolo dei modi a bassa frequenza con metodi e modelli convenzionali costituisce un grosso problema. L'inversione diretta della matrice dinamica è computazionalmente costosa per un modello atomistico, e se viene fatta in maniera numericamente approssimata introduce errori specialmente nei modi lenti, che sono anche generalmente i più anarmonici. Un'altra possibilità è l'analisi dello spettro vibrazionale ottenuto da una simulazione, che però deve essere sufficientemente lunga da campionare i moti lenti, il che ripropone il problema del costo di calcolo per i modelli atomistici. I modelli CG a rete risolvono il problema, permettendo sia l'inversione diretta della matrice dinamica a costo computazionale molto basso, sia la simulazione diretta di traiettorie molto lunghe. Sotto l'ipotesi di fluttuazioni gaussiane isotrope ("Gaussian network model" GNM) è possibile calcolare la matrice di covarianza delle fluttuazioni

$$\langle \Delta R_i \Delta R_j \rangle = \frac{k_B T}{\kappa} [\Gamma^{-1}]_{ij} \quad (3.17)$$

la cui traccia restituisce le fluttuazioni quadratiche medie. Γ è la matrice dei contatti (o matrice di Kirchhoff), che per $i=j$ restituisce il numero totale dei legami (contatti) dell'atomo i , vale -1 se la coppia $i \neq j$ è legata, e zero altrimenti, k_B è la costante di Boltzmann, T è la temperatura e κ è il parametro del potenziale di Hooke.⁵³ Il GNM è stato anche esteso per permettere fluttuazioni anisotrope nelle tre direzioni spaziali⁵⁴. La diagonalizzazione della matrice di covarianza fornisce i modi principali del sistema, che in approssimazione armonica coincidono con i modi normali⁵⁵. Gli autovalori sono le ampiezze quadratiche di vibrazione del modo, quindi ordinando i modi secondo l'autovalore, i primi risultano essere i più ampi e quindi quelli più rilevanti alla dinamica del sistema. Si dice che i primi ~10 modi principali costituiscono la “dinamica essenziale” del sistema. Per ottenere i modi principali del sistema in esame ho utilizzato il software ProDy descritto in Appendice B5.

All'equilibrio termico, il teorema di equipartizione dell'energia implica che, per ogni modo, si ha $kT \propto \langle \Delta u^2 \rangle \propto \omega^2$ quindi i modi più ampi sono i più lenti. Questo fatto è solitamente messo in relazione con il fatto che i modi lenti delle proteine sono quelli generalmente legati alla loro specifica funzionalità. Il successo dei modelli EN, nonostante la loro estrema semplicità è in ultima analisi legato alla loro capacità di riprodurre con accuratezza sufficiente i modi funzionalmente importanti⁵⁶. A sua volta, questa capacità è in parte dovuta al fatto che il modello è basato su una singola struttura nota a priori, che viene usata come input, quindi si introduce un alto livello di conoscenza a priori sul sistema.

3.3.2 Estensioni dei modelli a rete: multistabilità

Il difetto principale dei modelli a rete elastica è l'impossibilità di descrivere strutture e dinamiche lontane dalla struttura di riferimento, cioè l'estremo “bias strutturale”, dato sia dall'uso di un singolo dato strutturale nella parametrizzazione, sia dall'estrema semplicità del tipo di potenziale usato. Nel corso degli anni sono state proposte varie estensioni del modello, che agiscono su uno o entrambi questi aspetti.

Una serie di modelli ha proposto l'uso di potenziali a diversi gradi di anarmonicità, sostituendo il potenziale quadratico con forme funzionali tipo il potenziale di Morse, che consente la rottura del legame (reti anarmoniche⁵⁷) e/o l'uso di costanti di forza variabili, dipendenti anch'esse dalla distanza di equilibrio (reti a connettività fisica⁵⁸) o dalla conoscenza a priori di specifici legami (ad idrogeno, interazioni di stacking o ponti disolfuro o salini, reti a connettività “chimica”⁵⁹).

Particolarmente interessanti per questo lavoro di tesi sono i modelli che estendono le reti nel senso della multi-stabilità, includendo due o più stati strutturali di riferimento. Uno di questi è la rete elastica mista (Mixed ENM, MENM⁶⁰) che consiste nel creare un modello a due minimi, corrispondenti a due configurazioni stabili note di una proteina. Indicando con $\{x\}$ l'insieme delle coordinate del sistema, con $\{x_1\}$ e $\{x_2\}$ le coordinate delle due configurazioni, e

con $U_1(x_1)$ e $U_2(x_2)$ le energie potenziali in approssimazione ENM di ciascuno dei due stati, l'energia del sistema viene descritta da

$$U(x) = -\frac{1}{kT_m} \ln \left[e^{-\frac{1}{kT_m}(U_1(x-x_1)+\varepsilon_1)} + e^{-\frac{1}{kT_m}(U_2(x-x_2)+\varepsilon_2)} \right] \quad (3.18)$$

Questa combinazione restituisce il potenziale U_1 (U_2) quando $U_1 \ll U_2$ ($U_2 \ll U_1$), e fornisce un'interpolazione tra i due nella zona in cui sono confrontabili, regolata dal parametro $\beta=1/kT_m$ (T_m viene infatti detta temperatura di mixing: più è alta, minore è la barriera tra gli stati). ε_1 e ε_2 regolano le energie relative dei due stati, che sono un input del modello.

Un metodo di interpolazione alternativo che fornisce un risultato analogo è il cosiddetto il modello a rete a doppia buca (Double Well NM DWNM⁶¹, nel quale le energie dei due stati vengono combinate secondo la seguente regola

$$U = \frac{(U_1 + (U_2 - \delta))}{2} - \frac{1}{2} \sqrt{(U_1 + (U_2 - \delta))^2 + \varepsilon^2} \quad (3.19)$$

La differenza tra le 3.18 e 3.19 è solo formale: in questo caso δ è il parametro che regola l'energia relativa degli stati, e ε quello che regola l'altezza della barriera. Entrambi i modelli creano una connessione tra le due strutture, con barriere e transizioni non banali da uno all'altro. La rete plastica (Plastic NM, PNM.⁶²) dà una prescrizione del tutto analoga per la combinazione dei due stati, introducendo in più la possibilità iterativa di aggiungere altri stati oltre al secondo.

I modelli a rete interpolata (interpolated ENM, iENM) seguono una filosofia diversa da quelli precedentemente descritti⁶³. Date le due strutture di riferimento, si cerca un percorso fisicamente plausibile che interpoli tra le due, oltre che un profilo di energia che descriva la transizione. Come nei casi precedenti, le hamiltoniane corrispondenti agli stati di riferimento sono semplici reti elastiche. Si costruisce un'hamiltoniana unica che interpola tra le due dipendente da un parametro λ che va da 0 a 1. Un termine addizionale di energia U_{coll} (ininfluente negli stati iniziale e finale) che penalizza le interazioni steriche tra residui più vicini di un cutoff (fissato a 4 Å) si rende necessario perché nell'interpolazione tra due strutture molto diverse è possibile che si creino stati stericamente non ammissibili. L'energia potenziale del sistema è descritta da

$$U(\lambda) = U_1 + (1 - \lambda)U_2 + U_{coll} \quad (3.20)$$

che restituisce i modelli a rete degli stati 1 e 2 (e le due strutture di riferimento) per valori estremali di λ , mentre la ricerca del minimo di U per valori intermedi fornisce strutture che, in sequenza, danno un cammino privo di inconsistenze steriche e poco dipendente dai dettagli dei potenziali U_1 ed U_2 . $U(\lambda)$ costituisce un profilo di energia non dissimile da quelli forniti dagli approcci precedenti. A differenza dei modelli precedenti, questo non ha parametri aggiustabili per aggiustare l'altezza della barriera, che è determinata dal percorso stesso.

3.3.3 Modelli di $G\bar{o}$ e modelli a bias parziale

I modelli di $G\bar{o}$ ⁶⁴, specificamente intesi per lo studio del ripiegamento delle proteine hanno in comune con i modelli a rete il bias verso una struttura di riferimento (in questo caso la struttura correttamente ripiegata, o “nativa”). Tuttavia in questo caso il campo di forze contiene, in linea con quelli di tipo atomistico, termini “conformazionali” dipendenti dalle variabili interne R , θ e ϕ , come in Eq. 3.13, con

$$U_R = \sum_{i=1}^n \frac{1}{2} k_R (R - R_0)^2 \quad (3.21)$$

$$U_\theta = \sum_{i=1}^n \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (3.22)$$

$$U_\phi = \sum k_\phi (1 + \cos n(\phi - \phi_0)) \quad (3.23)$$

dove i parametri di equilibrio R_0 , θ_0 e ϕ_0 sono estratti dalla configurazione nativa. Il termine non bonded è separato in una parte “locale” e una “non locale”, tramite un raggio di cutoff analogo a quello dei modelli a rete. La parte locale è descritta tramite un potenziale attrattivo di tipo Lennard Jones o Morse

$$U_1 = \sum \varepsilon \left[\left(e^{-k(r_{ij} - r_{ij}(0))} - 1 \right)^2 - 1 \right] \quad (3.24)$$

in cui i parametro $r_{ij}(0)$ sono estratti dalla struttura nativa. Il termine non locale è puramente repulsivo. Questo modello ha come minimo la struttura nativa, e per strutture diverse descrive una transizione senza barriera a tale struttura, pertanto viene detto modello per il ripiegamento a frustrazione minima. Viene utilizzato per analizzare l’ordine di formazione dei contatti nella struttura nativa e come prima approssimazione per la cinetica del ripiegamento.

Il difetto principale del modello di $G\bar{o}$ è il bias totale verso la struttura nativa, che implica frustrazione minima e assenza della rappresentabilità di stati intermedi diversi da essa. Per questo motivo miglioramenti del modello tendono ad introdurre frustrazione tramite termini del potenziale indipendenti dalla struttura di riferimento.

Questa filosofia è anche quella che caratterizza una classe di modelli minimalisti cosiddetti a bias parziale specificamente sviluppati nel corso degli ultimi anni per riprodurre la dinamica su larga scala spazio-temporale e con forti fluttuazioni fuori dall’equilibrio⁶⁵. Il concetto base è di mantenere il bias verso una struttura di riferimento in alcuni dei termini del campo di forze, solitamente quelli più locali, e rilasciarlo negli altri, che vengono invece parametrizzati fittando su un set di dati strutturali generali. I termini locali mantengono l’accuratezza strutturale, riproducendo correttamente le interazioni più complesse, come legami ad idrogeno ed effetti sterici; viceversa, la generalità e accuratezza dei termini a medio-lungo range conferisce al sistema la possibilità di esplorare configurazioni alternative e diverse da quella di riferimento. Il livello di bias è

regolabile e se mantenuto sufficientemente basso può conferire al sistema sufficiente potere predittivo: il modello applicato alla proteasi di HIV-1⁶⁴ è stato in grado di esplorare la conformazione aperta della proteina, completamente diversa da quella chiusa usata come riferimento. Il livello di bias riguarda principalmente il termine $U_{\text{non bonded}}$, che in questi modelli viene separato in due parti

$$U_{\text{non bonded}} = U_{\text{local}} + U_{\text{non local}} \quad (3.25)$$

di cui U_{local} mantiene il bias. La definizione delle interazioni contenute in U_{local} può essere di tipo geometrico, basata su un cutoff sulle distanze tra atomi, oppure includere informazioni sulla struttura secondaria (indicazione di quali coppie formano un legame ad idrogeno) o su altri legami chimici. In questo lavoro di tesi ci si riferisce ad uno di questi modelli adottato per l'integrasi di HIV-1⁶⁶, in cui la parametrizzazione è semplificata al massimo. In questo modello le forme funzionali per i termini legati sono le più semplici possibili (armoniche)

$$U_b = \frac{1}{2} k(r - r_0)^2 \quad (3.26)$$

$$U_\theta = \frac{1}{2} k(\theta - \theta_0)^2 \quad (3.27)$$

$$U_\phi = \frac{1}{2} k(\phi - \phi_0)^2 \quad (3.28)$$

mentre per le parti non bonded, sia quella locale che quella non locale si sono usati dei potenziali di morse

$$U_{nl} = \varepsilon \left[(1 - e^{-k(r-r_0)})^2 - 1 \right] \quad (3.29)$$

in cui il criterio per definire le interazioni locali è il più semplice possibile (basato su un cutoff sulle distanze), e per entrambi i termini, locale e non locale le costanti di forza sono dipendenti dalla distanza (più forti per i legami più corti). Per il termine non locale r_0 è indipendente dalla struttura, mentre per il termine locale $r_0 = r_{0\ ij}$, presi da una struttura di riferimento, come per i modelli a rete elastica. Ma a differenza che in questi, i legami locali in questo modello possono rompersi e formarsene diversi, sotto la spinta del termine non locale e degli altri termini privi di bias.

3.4 Modello multi-scala per le rodopsine: descrizione sommaria

Il modello presentato in questa tesi combina caratteristiche dei vari modelli fin qui descritti in uno solo:

1. La multi-stabilità viene inclusa in maniera simile al modello a rete plastica, combinando modelli relativi ai vari stati della rodopsina a due a due in maniera sequenziale. Tuttavia, la relazione 3.19 viene usata per

combinare tra loro singoli termini del potenziale, anziché l'energia totale, scelti di volta in volta tra quelli che più significativamente contribuiscono alla transizione. In questo modo si introduce il livello di arbitrarietà e bias minimo indispensabile a riprodurre il passaggio tra gli stati. In questa fase si possono anche introdurre dati sperimentali sulle energie relative dei vari stati, in modo da rendere il modello più energeticamente accurato possibile. Un approccio tipo rete interpolata è anche usato per determinare i percorsi di reazione, verificati poi anche con simulazioni dirette.

2. I modelli per ciascuno stato sono basati su strutture di riferimento sperimentali, e si ispirano al modello a bias parziale per l'integrasi di HIV-1. Rispetto a questo sono stati operati alcuni miglioramenti che riguardano le forme funzionali dei vari termini e l'ottimizzazione della parametrizzazione. Il risultato finale è una maggiore accuratezza nella riproduzione della struttura e delle fluttuazioni termiche.
3. Infine, la multi-scalarità, come già accennato, è stata introdotta al fine di poter simulare accuratamente eventi su scale temporali molto diverse. A tale scopo il sito attivo (retinale, lisina ponte e molecole di acqua interne) è stato rappresentato al livello "atomi uniti", riadattato alla filosofia della parametrizzazione con bias locale. Il resto della proteina, quella strettamente amminoacidica è rappresentata come già detto al livello di modello minimalista, cioè con un singolo sito interattivo per ogni amminoacido, posizionato sul C α .

La combinazione di tutte queste caratteristiche in un modello singolo è il compromesso ottimale tra semplicità e flessibilità. Il limitato numero di gradi di libertà consente di simulare l'intero fotociclo su scale temporali macroscopiche, includendo però anche l'evento ultraveloce dell'isomerizzazione, grazie alla rappresentazione del retinale ad alta risoluzione. Il bias locale è stato ridotto al minimo in modo da conservare la possibilità di uscire dai minimi di riferimento in maniera realistica, preservando però l'accuratezza strutturale e quindi la possibilità di descrivere correttamente il cammino di reazione. Infine, la combinazione multi-stabile è stata realizzata in modo da introdurre anche dati macroscopici e quindi riprodurre accuratamente anche l'energetica della reazione.

Lo sviluppo del modello e i risultati delle simulazioni sono descritti in dettaglio nel prossimo capitolo. È qui di seguito riportato invece un accenno all'implementazione del modello descritta poi dettagliatamente in appendice B.

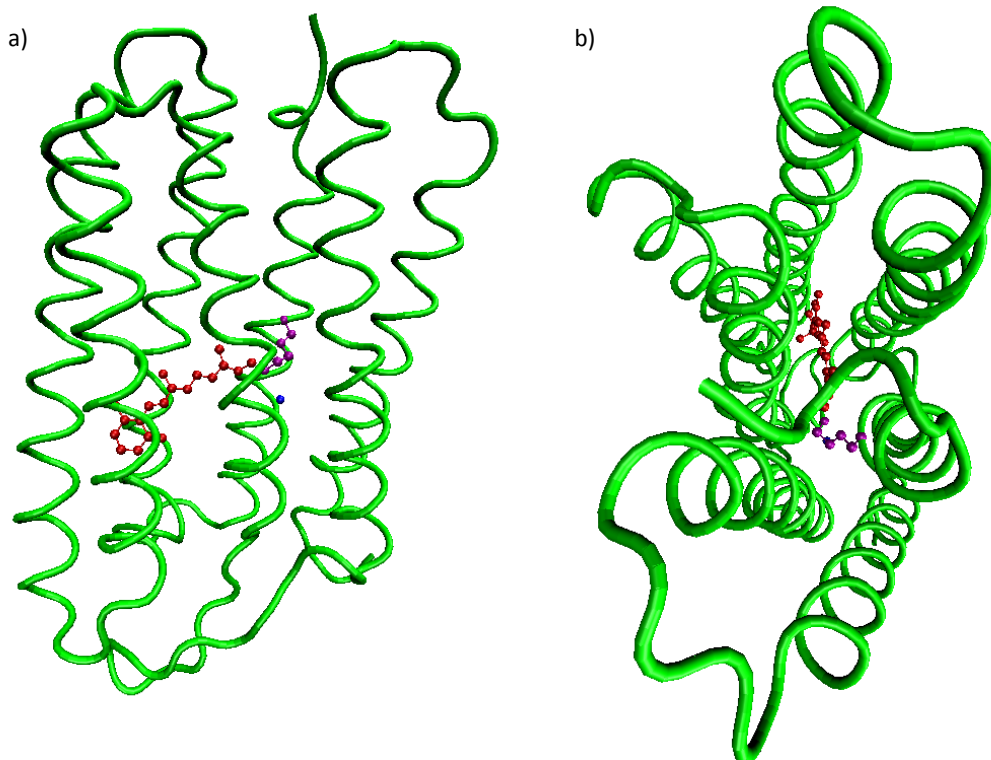


Figura 16: rappresentazione multi-scala della BR vista di lato in a) e vista dall'alto in b); in verde la struttura elicale e dei loop, in rosso il retinale, in viola la catena laterale del residuo Lys216 ed in blu la molecola di acqua W402.

Le simulazioni sono state realizzate con il programma DL_POLY⁶⁷, un pacchetto di software per dinamica molecolare classica generico, scelto per la sua flessibilità. Esso infatti non ha modelli pre-implementati ed è quindi adattabile a qualsiasi modello, ha un grande numero di forme funzionali per i termini del campo di forze, e consente di introdurne aggiuntive. L'implementazione dello specifico modello e campo di forze però ha richiesto un lavoro preliminare che si può schematizzare nei seguenti passi:

1. creazione dei file di input che descrivono il modello in questione. Essi sono principalmente due, il CONFIG, in cui sono definiti i centri interattivi contenuti nel modello e le loro posizioni iniziali, e il FIELD, che contiene la topologia del modello le forme funzionali e i parametri del campo di forze. Questi due file definiscono interamente il modello in analisi. Per la loro costruzione corretta è stato necessario programmare un codice accessorio, che di fatto implementa il modello in DL_POLY. Un terzo file che contiene i parametri di controllo della simulazione (timestep, temperatura ecc) viene preparato a mano.
2. Alcune delle forme funzionali necessarie per il modello, specificamente quelle che descrivono i potenziali a doppia buca per il passaggio tra gli stati, non erano presenti nel codice DL_POLY, per cui è stato necessario modificarne alcuni moduli per aggiungerle. La versione modificata di DL_POLY conserva tutte le funzionalità precedenti ed è maggiormente flessibile.

Le modifiche al codice DL_POLY, rese disponibili su richiesta, costituiscono un valore aggiunto a questo lavoro di tesi, in quanto non solo rendono il modello

fruibile da altri, ma sono anche abbastanza generali da potersi utilizzare per modelli completamente diversi (anche nelle simulazioni di sistemi non biologici), in tutti i casi in cui si devono descrivere transizioni tra due stati, di ciascuno dei quali è noto il campo di forze.

Capitolo 4. Risultati: simulazione delle dinamiche molecolari e del fotociclo delle rodopsine

Questo capitolo riporta tutti i risultati originali di questo lavoro di Tesi, che si articolano in diverse fasi. La prima ha riguardato l'ottimizzazione del modello multi-scala della rodopsina batterica. I parametri del modello sono stati ottimizzati in modo da riprodurre le fluttuazioni termiche dei singoli stati del fotociclo. Il campo di forze è unico per tutti gli stati, anche se come detto in precedenza la connettività e la parte locale del campo dipendono dalla struttura di riferimento, che introduce le differenze tra stato e stato. I parametri di forza e tutti i parametri dei termini non locali sono generali. Quindi il modello è utilizzabile per qualsiasi proteina, data una struttura di riferimento. È qui riportato anche un test che ne mostra le prestazioni sulla rodopsina bovina.

La seconda fase è consistita nella costruzione del modello multistabile, necessario nella simulazione del passaggio tra i vari stati. Esso è stato costruito combinando il campo di forze ottenuto per i vari stati dalla fase precedente, e inserendo informazioni sperimentali sull'energia relativa dei vari stati. Il risultato di questa fase è stato un campo di forze in grado di riprodurre il profilo di energia lungo la coordinata di reazione dell'intero ciclo.

Infine si è proceduto alla vera e propria simulazione della dinamica del passaggio tra gli stati. Data l'altezza delle barriere, si sono utilizzati metodi di dinamica non convenzionale in grado di superarle. Questa fase ha permesso non solo di ottenere una "traiettoria" che descrive l'intero fotociclo, ma anche, dall'analisi dei singoli passaggi, di rilevare movimenti funzionali specifici.

4.1 Ottimizzazione del modello multi-scala sulla rodopsina batterica

La multi-scalarità del modello implica che sono presenti tre classi di parametri: 1. i parametri del campo di forze dell'opsina, puramente "coarse grained", 2. i parametri puramente "quasi-atomistici" del retinale e acqua, 3. i parametri di interfaccia tra le due parti, che contengono sia termini di tipo "non bonded" (parametri Van der Waals ed elettrostatici) che parametri di tipo "bonded", che descrivono il legame della base di Schiff. A partire da un set di parametri preesistente, disponibile in letteratura per le varie classi, si è proceduto ad un aggiustamento dei parametri in varie fasi, prima quelli dell'opsina, poi quelli del retinale e dell'interfaccia, utilizzando come osservabile di riferimento l'andamento delle fluttuazioni quadratiche medie (RSMF), ottenute dai dati sperimentali del fattore di temperatura (B a 100K) e da simulazioni atomistiche specificamente realizzate allo scopo, a 300K, con il campo OPLS (si veda cap 3) e idratando completamente la proteina con molecole di acqua esplicite (in una cella

di simulazione cubica di lato 78Å). In pratica questi due dati per l'RSMF sono utilizzati come osservabile da riprodurre nelle simulazioni CG, cambiando i parametri con una procedura adattativa razionale, riportata qui di seguito:

1. Come parametri di partenza si sono adottati, per la parte CG quelli utilizzati per un modello CG minimalista della proteina integrasi del virus HIV-1⁶⁴, per il retinale e per l'acqua quelli del modello Atomi Uniti⁴⁸ (UA). Si noti che nel modello UA i parametri sono indipendenti dalla struttura di riferimento e adottano forme funzionali differenti da quelle del modello minimalista. Ad esempio, sono utilizzate forme tipo Lennard Jones per le interazioni Van der Waals, anziché potenziali di Morse (Eq. 3.28 cap 3). Per consistenza con il resto del campo, tutte le interazioni sono state uniformate usando le forme funzionali dell'opsina e usando parametri coerenti.
2. Nel modello dell'integrasi, le costanti di forza per i termini angolari U_{angles} (vedere Eq. 3.26 cap. 3) sono dipendenti dalla struttura secondaria. In questo lavoro la dipendenza è stata accentuata e resa più dettagliata, e introdotta anche nel termine diedrale $U_{\text{dihedrals}}$ (Eq. 3.27 cap. 3) per riprodurre più accuratamente le fluttuazioni soprattutto nei loop interelicali. Il termine U_{bond} qui è trattato come un semplice vincolo sulla distanza tra $C\alpha$ consecutivi. Le differenze tra i parametri dell'integrasi e quelli ottimizzati per la rodopsina sono riportate in Tabella 2.
3. I termini U_{local} (Eq. 3.24 cap. 3) per la parte dell'opsina sono trattati con un potenziale di Morse con parametri energia di legame ϵ e larghezza di buca α dipendenti dalla distanza di equilibrio r_0 secondo una legge esponenziale shiftata, come nel caso dell'integrasi. Anche in questo caso si è proceduto ad una ottimizzazione, che non ha prodotto modifiche sostanziali (si veda Tabella 2).
4. Il termine U_{local} del retinale è stato costruito in conformità con la filosofia del modello a bias parziale: la forma funzionale usata è il Morse, con distanze di equilibrio estratte dalle strutture di riferimento. I parametri iniziali α , ϵ sono stati presi dal modello UA.
5. I termini d'interfaccia sito attivo/opsina (termini misti UA/CG) necessitano una parametrizzazione *de novo*. Queste interazioni sono termini U_{local} tra retinale/acqua e opsina, termini U_{θ} e U_{ϕ} in prossimità del legame della base di Schiff, e termini $U_{\text{non-local}}$.
 - a. I termini di legame chimico del retinale-base di Schiff e legame con la lisina sono stati parametrizzati combinando la filosofia del modello a bias parziale con quella del modello "atomi uniti". In particolare, come distanze/angoli/diedri di equilibrio si sono usati quelli della struttura di riferimento, mentre le costanti elastiche sono quelle del campo UA⁴⁸.
 - b. I termini U_{local} sono trattati in conformità con il resto del campo di forze con potenziali Morse. Per determinare i parametri, si sono calcolati i profili di energia basati su rappresentazioni atomistiche (CF OPLS) in funzione della variabile indipendente del potenziale (ad esempio, distanza O- $C\alpha$ per l'interazione tra acqua e opsina). I parametri sono stati ottenuti fittando il Morse su questi dati. L'ottimizzazione tramite confronto con il dato sperimentale a 100K ha mostrato che gli andamenti dei parametri ϵ e α in funzione della distanza di equilibrio r_0 seguono abbastanza bene l'andamento generale dei parametri per la parte di opsina, se le leggi

esponenziali shiftate vengono moltiplicate per un fattore di scala (pari a 005, vedere tabella). Questo fattore si giustifica posteriori considerando il passaggio di risoluzione tra CG e quasi atomistico: nella parte CG un singolo legame in realtà descrive implicitamente diverse interazioni che invece sono rappresentate esplicitamente nella parte UA, quindi le interazioni della parte UA.

- c. Dai numerosi studi sulla BR sappiamo che la molecola di acqua W402 forma tre legami a idrogeno rispettivamente con la base di Schiff, ed i residui Asp85 e Asp212, che sono stati inseriti tra i legami U_{local} , con diversi tipi di potenziale a seconda della funzione (donore/accettore) della molecola: in particolare, il legame risulta più forte se l'acqua funge da donore di protoni (si veda tabella per i valori numerici). I valori numerici sono stati ottenuti con la procedura descritta in Figura 17. Si è successivamente osservato nelle simulazioni (vedere sotto) che l'ottimizzazione di questi parametri è importante per la corretta descrizione del comportamento dell'acqua nel sito attivo.
6. Il termine $U_{non-bonded}^{non\ local}$, per la parte di opsina, in conformità con i lavori precedenti, utilizza un potenziale medio, indipendente dall'amminoacido⁶⁵. Per quanto riguarda i parametri della parte opsina-acqua e opsina-retinale, essi sono stati parametrizzati in conformità con quelli U_{local} , calcolando i parametri con la procedura di fit descritta sopra.
7. I valori dei cutoff tra parte non-bonded locale e non-bonded non locale sono 8.5Å per la parte di opsina, valore che ha fornito buoni risultati per l'integrasi⁶⁴ e per la proteasi di HIV⁶³. Lo stesso valore è stato usato per il retinale/acqua e per l'interfaccia, per conformità. Il controllo a posteriori non ha necessitato ulteriori aggiustamenti.

Termine del potenziale	FF iniziale		FF ottimizzato per rodopsine	
	Forma analitica	parametri	Forma analitica	parametri
CG				
U_{local}	armonico	$k = 500$	vincolo olonomo	
U_{θ}	coseno armonico	$k = \begin{cases} 10 & \text{foglietti} \\ 30 & \text{eliche} \end{cases}$	coseno armonico	$k = \begin{cases} 110 & \theta < 97 \\ 20 & 97 < \theta < 115 \\ 70 & 115 < \theta < 130 \\ 90 & 130 < \theta < 150 \\ 20 & \theta > 150 \end{cases}$
U_{ϕ}	coseno	$A=30$	coseno	$A = \begin{cases} 25 & \phi < 80 \\ 5 & \phi > 80 \end{cases}$
$U_{non-bonded}^{local}$	Morse	$\varepsilon = 8e^{-\left(\frac{x}{6.3}\right)^4} + 0.05$ $\alpha = 0.4e^{-\left(\frac{x}{6.3}\right)^4} + 0.707$ $r_{cut} = 8.5$	Morse	$\varepsilon = 3.9e^{-\left(\frac{x}{6.15}\right)^6} + 0.05$ $\alpha = 2e^{-\left(\frac{x}{6.2}\right)^8} + 0.7$ $r_{cut} = 8.5$
$U_{non-bonded}^{non-local}$	Morse	$\varepsilon = 0.01$ $\alpha = 0.707$ $r_0 = 9.55$	Morse	$\varepsilon = 0.01$ $\alpha = 0.7$ $r_0 = 9.5$
$U_{tethering}$	no		armonico	$k=2$
retinale – acqua (sito attivo)				
U_{local}			vincolo	
U_{θ}			armonico	$k = 63$
U_{ϕ}			coseno	$A = 2$
$U_{non-bonded}^{local}$			morse	$\varepsilon = 0.12$ $\alpha = 2e^{-\left(\frac{x}{6.2}\right)^8} + 0.7$
$U_{non-bonded}^{non-local}$			12-6	$A=538 \ B=11.9$ $A=1285145 \ B=673.2$ $A=372.2 \ B=8.6$ $A=142.5 \ B=9.2$ $A=712422 \ B=494$ $A=205 \ B=6.4$
C(ret) – C(ret) C(ret) – H ₂ O C(ret) – lisina H ₂ O – H ₂ O H ₂ O – lisina lisina – lisina				
interfaccia opsina/sito attivo.				
U_{local}			vincolo	
$U_{non-bonded}^{local}$			Morse	$\varepsilon = \left(8e^{-\left(\frac{x}{6.3}\right)^4} + 0.05\right) * 0.05$ $\alpha = 0.4e^{-\left(\frac{x}{6.3}\right)^4} + 0.7071$
$U_{non-bonded}^{non-local}$			Morse 12-6 12-6	$\varepsilon = 0.1 \ \alpha = 0.72217 \ r_0 = 6.5$ $A = 695928 \ B = 560$ $A = 815008 \ B = 461$
BR – C(ret) BR – H ₂ O BR – lisina			Morse 12-6 12-6	$\varepsilon = 1.9 \ \alpha = 2 \ \text{donore}$ $\varepsilon = 5 \ \alpha = 1.85 \ \text{accettore}$
U_{h-bond}			Morse	

Tabella 2: valori dei parametri dei FF del modello CG, UA e dell'interfaccia. I valori di k sono espressi in Kcal · mol/Å². $x=r_0$, la distanza di equilibrio del potenziale di morse. Dove non indicato, è quella nella struttura di riferimento. I valori di ε per i Morse ed i valori A e B per il potenziale 12-6 sono espressi in Kcal/mol. La sigla C(ret) indica un atomo di carbonio del retinale. Il coseno armonico è descritto dall'equazione $h \cos(x) = \frac{k}{2} (\cos(\theta) - \cos(\theta_0))^2$, la funzione coseno da $c(x) = A[1 + \cos(m\phi - \delta)]$ e la funzione di Morse da $m(x) = \varepsilon \left[\left(1 - e^{-\alpha(r-r_0)}\right)^2 - 1 \right]$.

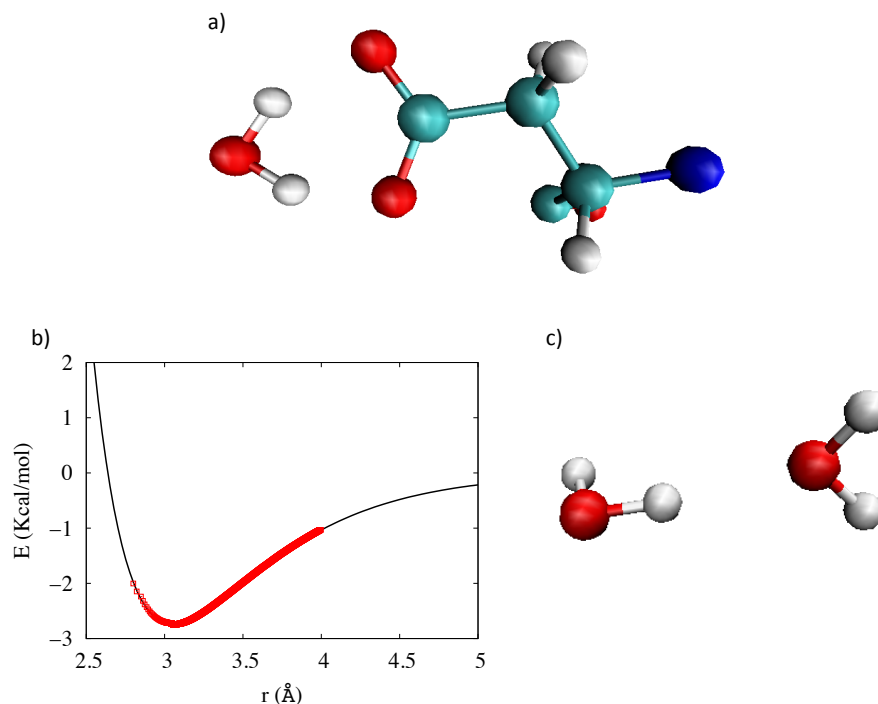


Figura 17: Illustrazione della procedura per parametrizzare i termini di potenziale U_{local} e $U_{\text{non-local}}$. (a) Viene creata una rappresentazione atomistica completa (inclusi atomi di idrogeno) semplificata del sistema. Un caso esemplare è il sistema Asp-acqua, rappresentato in (a). Esso è anche particolarmente rilevante per il sistema in quanto l'acqua W402 presente all'interno del sito attivo interagisce precisamente con Asp85. Vengono create molte copie del sistema in cui la distanza che costituisce la variabile del potenziale (in questo caso la distanza tra O dell'acqua e Ca di Asp) viene tenuta fissa a valori diversi in un intervallo di interesse (in questo caso tra ~ 2.5 e ~ 4 Å). Il sistema viene ottimizzato in modo da consentire agli altri atomi di rilassare nella configurazione di equilibrio e l'energia misurata per ciascuno valore della distanza (punti rossi in b). Dopodiché l'energia viene fittata con la funzione di interesse, in questo caso il Morse (linea nera in b), ottenendo i valori di tutti i parametri. In questo caso in realtà la procedura è stata usata per fittare principalmente il parametro α (larghezza di buca, che risulta essere $\alpha=1.4$), in quanto la distanza di equilibrio e la profondità di buca sono ottenuti in conformità con il resto del campo di forze (si veda testo). Essi risultano tuttavia compatibili con quelli ottenuti dal fit. Oltre al valore della larghezza di buca per i parametri misti retina-opsina e acqua-opsina, si sono ottenuti nello stesso modo anche i parametri Morse per l'interazione atomi uniti acqua-acqua (c). Le simulazioni atomistiche sono state effettuate con DL_POLY, utilizzando per la creazione dell'input il programma DL_FIELD⁶⁸.

Il campo di forze di Tabella 2 dà risultati soddisfacenti considerando il confronto tra la simulazione e il dato sperimentale cristallografico a 100K (Fig. 18, a, confronto tra linea nera continua e rossa tratteggiata). La simulazione non tiene conto, se non in maniera implicita, dell'effetto della membrana. D'altra parte il confronto con il dato cristallografico non è particolarmente dirimente a questo riguardo, in quanto nella proteina cristallizzata, la membrana è presente solo in maniera frammentaria. Si è comunque successivamente incluso l'effetto della membrana nella simulazione vincolando le posizioni dei Ca in contatto con essa (vedere Fig. 19) alla loro posizione tramite un potenziale armonico

$$U_{\text{teth}} = \frac{1}{2} kx_0^2 \quad (4.1)$$

Il valore del parametro k è stato determinato confrontando il valore del B-factor della simulazione a 300K con i dati di uno studio di dinamica molecolare atomistica dove si simula l'effetto della membrana costruendo due piani costituiti da sfere di Van Der Waals ottenendo così un doppio strato lipidico⁶⁹.

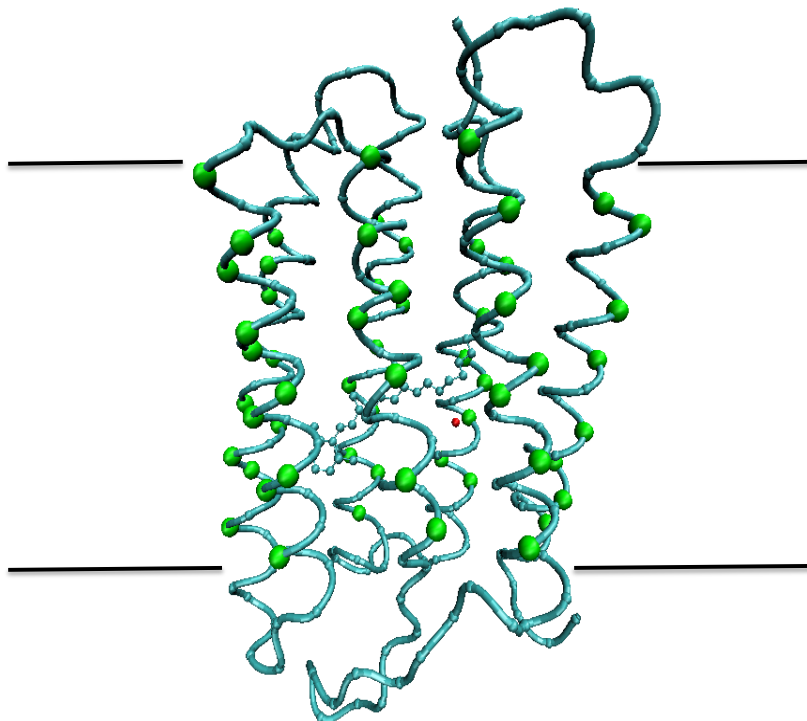


Figura 18: rappresentazione della BR in azzurro nella membrana schematizzata da linee nere. I $C\alpha$ in contatto con la membrana sono dati dalle sfere in verde e le loro posizioni sono state ottenute tramite un plugin di VMD, `vmd_use_pdb_ss.tcl`, che legge le informazioni sulla struttura secondaria dal file `.pdb` e indica quali atomi sono appartenenti alle eliche o ai loop.

Per ottimizzare k si è confrontato specificamente l'RMSF di eliche e loop della simulazione di questo lavoro con quelli della simulazione di letteratura², ottenendo $k=2$. Questo valore, se pur relativamente basso, ha permesso di abbassare l'RMSF delle eliche trans-membrana fino ad un valore realistico, contemporaneamente fornendo una misura del potere contenitivo della membrana. L'accordo tra simulazioni multi-scala e dati sperimentali o dalle simulazioni atomistiche risulta qualitativamente comparabile con quello di precedenti modelli simili. Particolarmente accurato risulta il comportamento della parte di retinale/sito attivo, derivante dalla specifica cura dedicata all'ottimizzazione dei parametri del sito e dell'interfaccia con l'opsina. Importanti sono in particolare i parametri delle interazioni con le molecole di acqua, che se non opportunamente regolati, risultano nell'espulsione delle molecole di acqua interne dalla struttura, durante la simulazione a 300K.

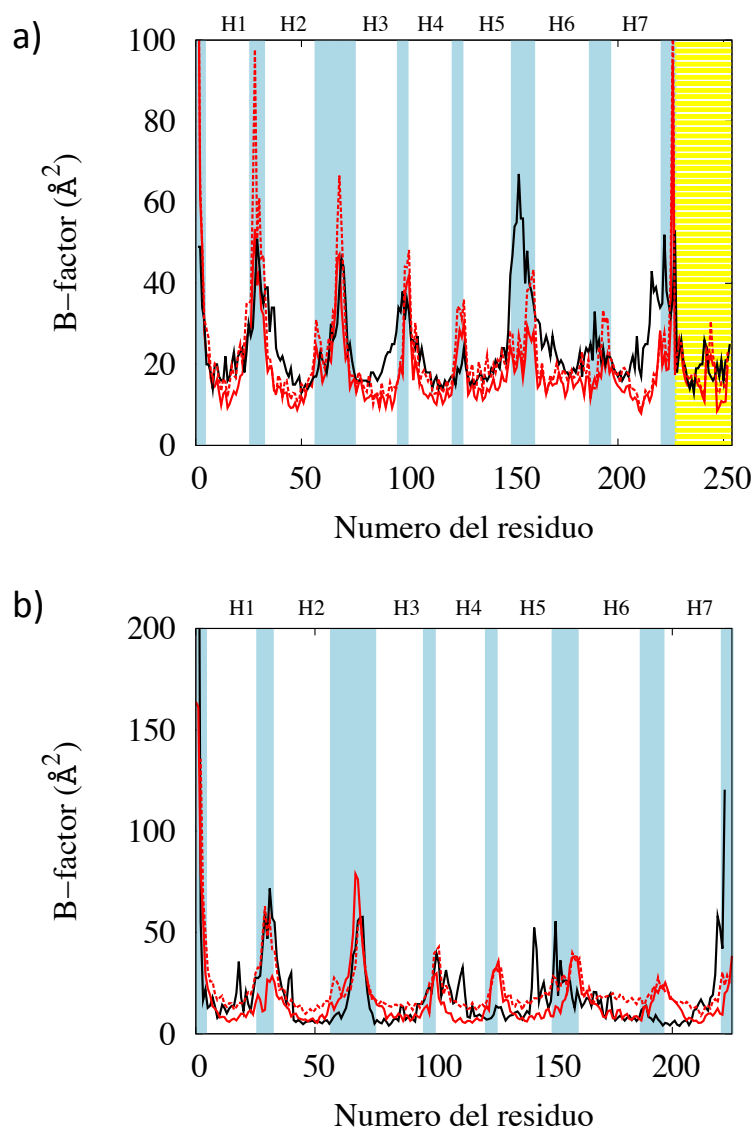


Figura 19: a) grafico dei B-factor alla temperatura di 100K della BR: la linea nera rappresenta i dati sperimentali ottenuti dalla struttura cristallografica, la linea rossa tratteggiata il B-factor della simulazione senza i termini per la simulazione della membrana e la linea rossa continua il B-factor della simulazione con i termini della membrana. In b) grafico del B-factor alla temperatura di 300K: la linea nera rappresenta il B-factor della simulazione atomistica, la linea rossa tratteggiata il B-factor della simulazione CG senza i termini che rappresentano la membrana e la linea rossa continua il B-factor della simulazione con i termini per la rappresentazione della membrana. In a) il numero di residui considerati è 253, cioè tutti gli atomi del modello multi-scala, mentre in b) il numero di residui è dato soltanto dagli atomi dell'opsina, in quanto il retinale non era presente nella simulazione atomistica.

Per convalidare ulteriormente il modello, esso è stato applicato alla rodopsina bovina (MR, Fig. 20) senza cambiare i parametri. Come nel caso precedente, si sono confrontati i risultati della simulazione multi-scala (senza e con il termine di membrana) con i dati cristallografici a 100K e simulazioni atomistiche (ottenute con la stessa configurazione della BR) a 300K. Il confronto è riportato in Fig. 21.

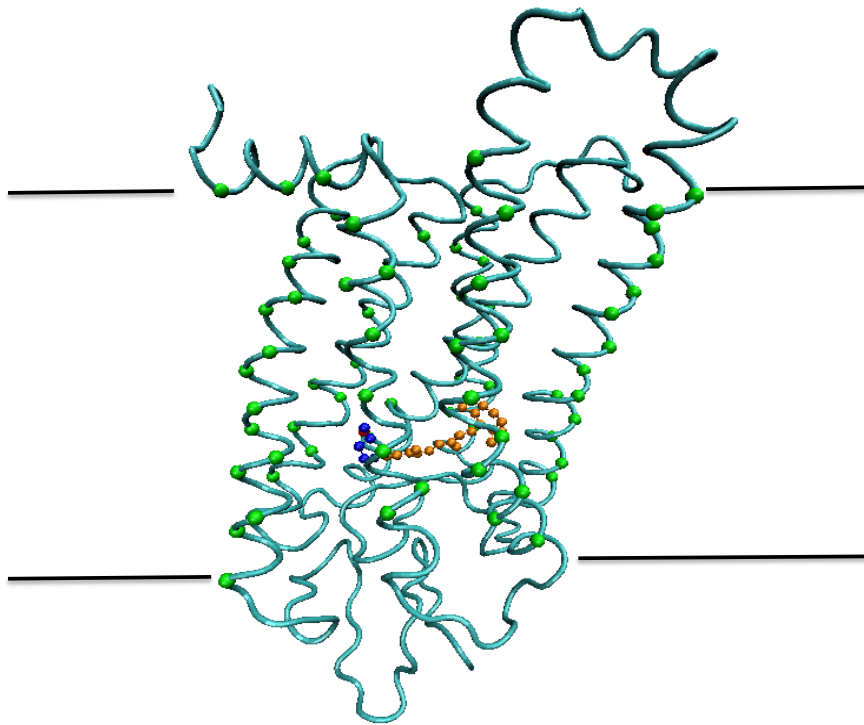


Figura 20: rappresentazione della MR dove si vede in arancione il retinale, in blu il residuo Lys296, in rosso la molecola di acqua W402, in grigio la catena della proteina e in verde i Ca soggetti al potenziale U_{eth} .

Il confronto con il dato cristallografico (Fig. 21(a)) non è particolarmente buono. Tuttavia bisogna notare che il B factor di questa misura risulta insolitamente alto rispetto alla media di altri dati, e potrebbe quindi essere un artefatto sperimentale, legato alla difficoltà nella cristallizzazione della rodopsina dei mammiferi. Per contro, infatti, il confronto con il dato di simulazione a 300K, risulta anche migliore che nel caso della BR, validando il modello in maniera soddisfacente.

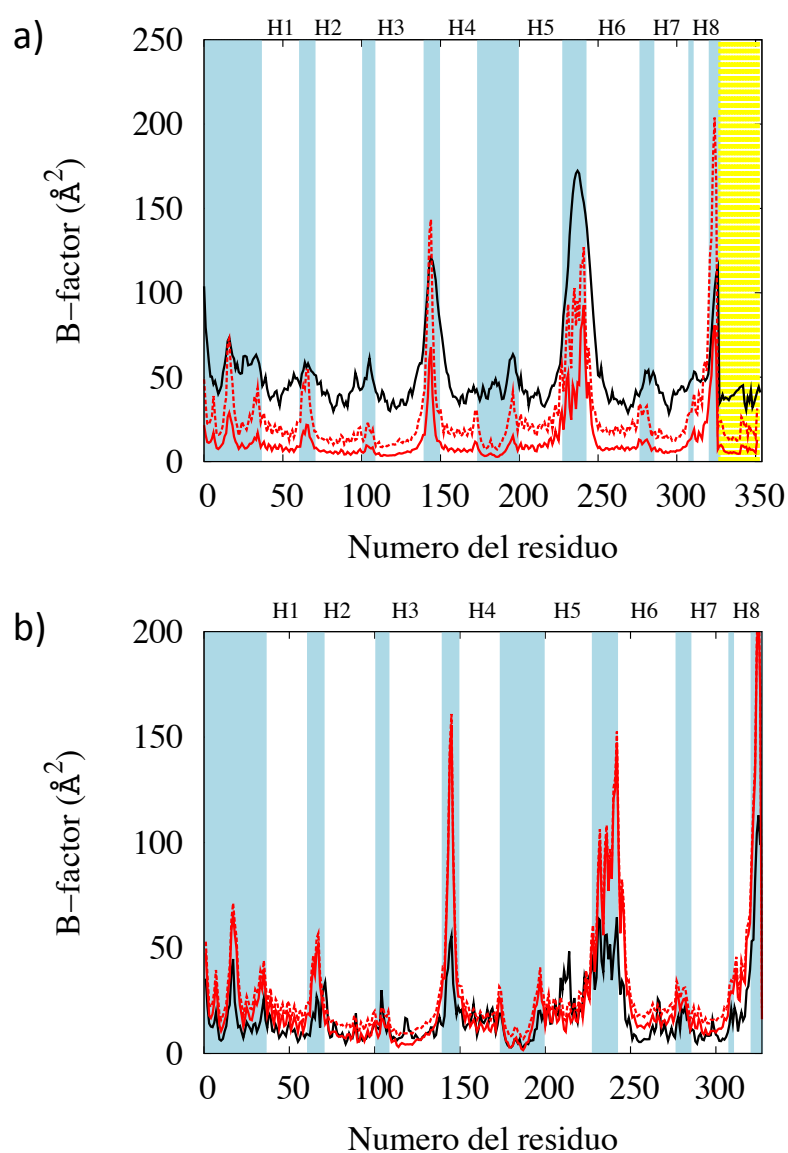


Figura 21: a) Confronto tra i B-factor alla temperatura di 100K della MR: la linea nera rappresenta i dati sperimentali derivanti dalla struttura cristallografica, la linea rossa continua la simulazione dove si tiene conto dell'effetto della membrana e la linea rossa tratteggiata la simulazione senza effetti della membrana. In b) il confronto tra i B-factor a 300K la linea nera rappresenta la simulazione atomistica, la linea rossa continua la simulazione multi-scala con la membrana, la tratteggiata senza la membrana. a) il numero di residui considerati è 352, cioè tutti gli atomi del modello multi-scala, mentre in b) il numero di residui è dato soltanto dagli atomi del modello CG, cioè 325, in quanto la simulazione all-atom è ottenuta soltanto sulla proteina senza considerare il retinale.

4.2 Simulazione degli stati del fotociclo

Il campo di forze ottimizzato nella sezione precedente è stato utilizzato in simulazioni a 300K separate dei singoli stati, per valutarne le proprietà strutturali e dinamiche locali.

Per ciascuno stato il protocollo di simulazione è lo stesso, e descritto dai seguenti passi.

1. A partire dalla struttura di riferimento, ottimizzazione strutturale fino a raggiungere il minimo locale
2. Riscaldamento controllato fino a 300K, con step intermedi a 10K, 20K, 50K, 100K e 200K, ciascuno della durata di 1ns.
3. Simulazione di produzione a 300K della durata di 50ns
4. I parametri di simulazione sono i seguenti:
 - pressione: 1 atm
 - temperatura: 300K
 - ensemble NVT
 - termostato: Berendsen con $\tau=0.001$
 - timestep: 1fs

Il tempo macchina necessario per un run di 50ns (in parallelo su 8 nodi) è di circa 11(x8) ore.

La deviazione standard (Root Mean Squared Deviation, RSMD) da una struttura di riferimento calcolata lungo la simulazione è di solito una misura della stabilità di un modello. La Fig. 22 riporta le RSMD su run a 300K di 50 ns, nel caso dello stato BR, su run di lunghezza macroscopica (μsec). Sono anche riportate le fluttuazioni quadratiche medie dei vari stati lungo la catena.

	RMSD(t)	RMSD medio (Å)	RMSF (Å)
BR		0.942	
K		0.944	
L		0.952	
M		0.942	
N		0.929	
O		0.993	

Figura 22: Deviazioni standard (RSMD) dalla struttura di riferimento degli stati del fotociclo, su run di 50 ns, e nel caso dello stato BR su scala del μ sec; terza colonna: valori medi della RMSD; quarta colonna: RMSF.

Il sistema mostra stabilità anche su scala macroscopica. I valori delle RSMD e le loro fluttuazioni risultano compatibili con quelli usualmente calcolati nelle simulazioni atomistiche, ad ulteriore validazione del modelli. Sono inoltre abbastanza simili tra loro sui vari stati, con l'eccezione dello stato O, il cui RSMD è più alto di circa 8%. Questa differenza può essere attribuita ad una reale maggiore flessibilità dello stato O, testimoniata dalle maggiori fluttuazioni rispetto agli altri stati (si veda Fig. 22) e probabilmente correlata con la maggiore difficoltà di cristallizzazione.

4.3 Modello bi-stabile

I modelli fin qui riportati descrivono correttamente le strutture e le fluttuazioni dei vari stati, ma non sono in grado di descrivere i passaggi di stato, per i quali è necessario costruire modelli multi-stabili. Il modello viene costruito in questo modo. Dati due stati collegati da un passaggio (ad es. K e L) che chiameremo A e B, ad essi sono associati due strutture di riferimento che a loro volta generano due campi di forze FIELD_A e FIELD_B diversi solo per i termini locali (i termini non-locali sono indipendenti dalla struttura). Specificamente, i vari termini conterranno potenziali locali con distanze/angoli/diedri di equilibrio corrispondenti alla struttura di riferimento. Molti di questi termini saranno molto simili in A e B, altri invece varieranno sostanzialmente, e questi sono quelli rilevanti alla descrizione della transizione. In generale comunque è possibile costruire un modello bistabile sostituendo ciascun termine locale con un termine a doppia buca, in cui le due buche corrispondono ai termini di FIELD_A e FIELD_B. In pratica, se $f_A(x)$ e $f_B(x)$ sono le funzioni che descrivono i termini del campo A e B per la variabile interna x ($x=\theta,\phi,r$), il termine corrispondente del campo bistabile è descritto da

$$f(x) = \frac{f_A(x) + (f_B(x) - \Delta)}{2} - \frac{1}{2} \sqrt{(f_A(x) - (f_B(x) - \Delta))^2 + \varepsilon^2} \quad (4.2)$$

dove Δ contribuisce all'energia relativa tra gli stati A e B. Questa formula restituisce per ogni x la minore delle due funzioni $f_A(x)$ e $f_B(x)$ e ε smussa la barriera (che per $\varepsilon=0$ risulta una cuspid) e ne influenza leggermente l'altezza. Le forme funzionali per i vari termini e la loro composizione secondo questa regola sono riportati in Appendice B. Il "merging" dei due FIELD nel FIELD.merged bistabile è operato da un software appositamente creato (merge_field.f, si veda appendice B). L'operazione va ripetuta per ogni coppia di stati consecutivi (a parte eccezioni discusse), generando una serie di campi bistabili in grado di descrivere le transizioni. Essi in linea di principio potrebbero essere composti in serie in modo da generare un unico campo multistabile. L'unica difficoltà in questo senso è la necessità di implementare forme funzionali adeguate nel software DL_POLY. Al momento si sono implementate solo le forme bistabili, anche perché, come si vedrà nel seguito, le simulazioni dei passaggi di stato a due a due appaiono già sufficienti a descrivere adeguatamente il comportamento del sistema.

Il campo bistabile necessita come unici parametri aggiuntivi (per ogni termine) ε , che, come già detto serve principalmente a smussare la cuspid, e a tale scopo si è scelto per esso il valore minimo possibile, generalmente pari a 0.0001. Se necessario, questo parametro però può anche regolare finemente l'altezza della barriera. L'altro parametro è Δ , che regola l'energia relativa degli stati. In pratica, l'energia relativa dello stato A e B (per la quale, ricordiamo, almeno nel caso della BR esistono stime sperimentali) è determinata dalla somma di tutti i Δ .

In linea di principio dunque si avrebbero da fissare N diversi Δ , con N=numero dei termini locali. In realtà si è posto $\Delta=0$ per i termini bistabili che componevano tra loro due potenziali Morse, perché in quel caso la differenza di

energia è già presente e fisicamente sensata, a causa del fatto che Morse locali con distanze di equilibrio diverse hanno profondità diverse. Si noti che è stato necessario inserire dei termini a doppia buca –Morse anche per quei termini per i quali durante la transizione c'era un passaggio del cutoff di 8.5 tra descrizione locale e non locale. Riguardo ai termini bistabili rimanenti (angoli e diedri di legame, perché le distanze C α -C α lungo la catena non variano) si è visto che la maggior parte di essi aveva piccole variazioni, quindi si è attribuito un potenziale bistabile solamente a quelli in cui il valore di equilibrio della variabile cambiava più dell'1% e rappresentando gli altri con termini a singola buca (con valore di equilibrio intermedio tra i due). Il valore di Δ per questi termini è stato scelto uniforme, e regolato in modo che, considerata la variazione prodotta dai termini Morse locali, fosse riprodotto il valore sperimentale della variazione di energia interna. Alla fine di questa operazione si sono ottenuti campi bistabili, uno per ciascuna transizione. Per una lista e una descrizione dettagliata dei termini bistabili nei vari passaggi di stato, si veda l'Appendice C.

Per testare la correttezza dell'implementazione di questo nuovo tipo di potenziale ed il suo corretto funzionamento si è per prima cosa misurato il profilo di energia potenziale lungo la transizione. Per far ciò, per ogni singola transizione si è identificata una singola coordinata di reazione scegliendola tra quelle i cui termini locali del potenziale variavano maggiormente, e si sono create configurazioni intermedie fissando la variabile a valori intermedi tra quelli assunti in A e in B, e ottimizzando la struttura con tutte le altre variabili libere. Questo fornisce l'andamento dell'energia potenziale totale (e termine per termine) lungo una traiettoria che collega A e B. La procedura è mostrata per il passaggio tra i primi due stati BR e K in Fig. 23 e 24.

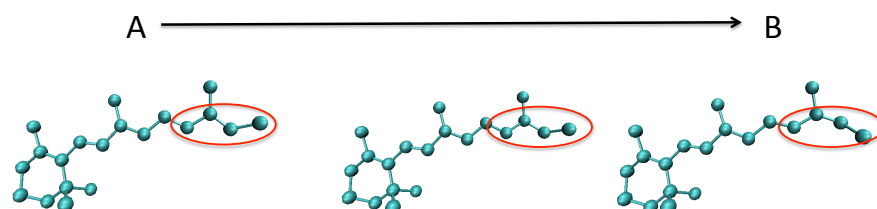


Figura 23: Illustrazione della procedura per identificare il profilo di reazione, sulla prima transizione BR-K. In questo caso la coordinata principale è il diedro corrispondente all'isomerizzazione.

Si osserva che il modello produce sia barriere di energia (principalmente dovute ai termini di angolo e diedro di legame) sia differenze di energie tra gli stati. In questo caso particolare preliminarmente si è posto $\Delta=0$ per evidenziare che il modello, anche senza parametri aggiuntivi, è in grado di riprodurre differenze di energia tra gli stati per effetto delle variazioni fisiche dei termini Morse locali. Δ è tuttavia soggetto ad ottimizzazione, come già spiegato e come si vedrà nel seguito.

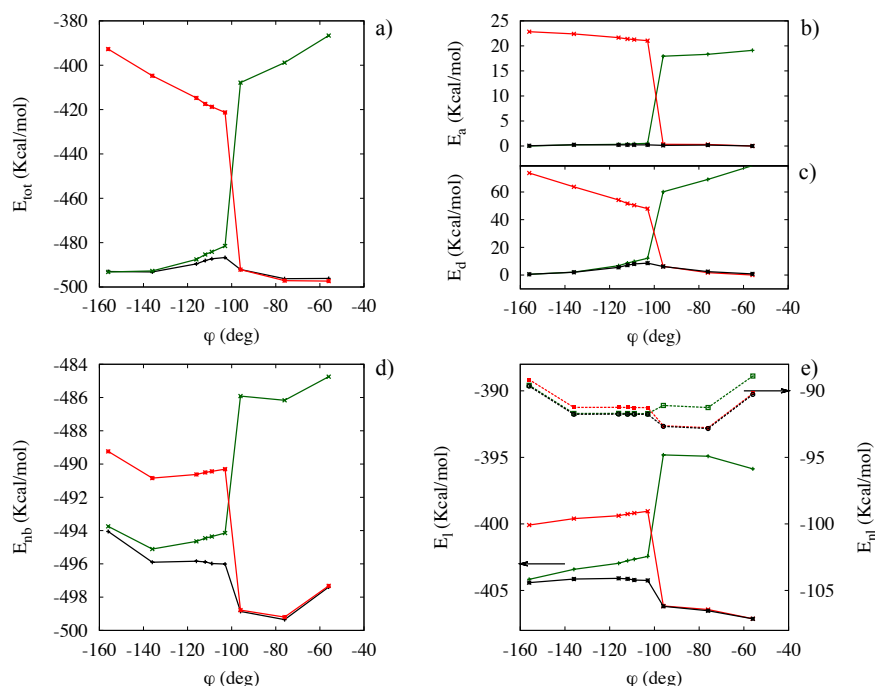


Figura 24: Profili di energia per la transizione BR-K. In rosso, profili misurati sul campo FIELD_A, in verde su FIELD_B e in nero sul FIELD.merged. Per questo calcolo i Δ dei termini bond angle e dihedral (pannelli b e c) sono imposti nulli, mentre per i termini locali e misti locali non locali (pannello e, linee continue scala sulla sinistra e tratteggiate-scala sulla destra rispettivamente), come spiegato nel testo, portano una differenza di energia che dipende da come le distanze di equilibrio variano da stato a stato. Questa, in totale sommata sulla parte non bonded locale e non locale risulta in questo caso di circa 5-6 kcal/mole in favore dello stato K (pannello d).

La procedura è stata ripetuta per tutti gli altri passaggi, che mostrano caratteristiche simili (si veda appendice C per i dettagli).

4.4 Simulazioni delle transizioni tra gli stati

In questa fase sono stati assegnati i valori di delta secondo i criteri spiegati nella sezione precedente, in modo da riprodurre nelle simulazioni il valore sperimentale delle variazioni di energia interna.

Transizione	Δ_{tot}	Δ_{angoli}	Δ_{diedri}	Δ
K-L	12	0	12	12 in C12-C13=C14-C15
L-M	20	0	20	10 in C12-C13=C14-C15 10 in C14-C15=N(lisina)-C(lisina)
M-N	1.2	0	1.2	1.2 in C12-C13=C14-C15
N-O	5.2	0	5.2	2 in C12-C13=C14-C15 2 in C14-C15=N(lisina)-C(lisina) 1.2 in C15=N(lisina)-C(lisina)=C(lisina)

Tabella 3: suddivisione dei valori del parametro Δ per ciascuna transizione. La colonna Δ_{tot} indica il valore del Δ totale per ciascuna transizione che viene suddiviso sui termini diedrali. Le colonne Δ_{angoli} e Δ_{diedri} indicano come viene suddivisa l'energia tra i vari termini del FF. La colonna Δ indica il valore dell'energia per ciascuna transizione ed in quali diedri viene suddivisa. I valori delle energie sono espressi in kcal/mole.

I valori dei parametri sono riportati nella Tabella 3, mentre il profilo dell'energia dei vari stati calcolato da simulazioni a temperatura zero (cioè con ottimizzazioni locali) sugli stati di equilibrio e sugli intermedi è riportato in Fig 25.

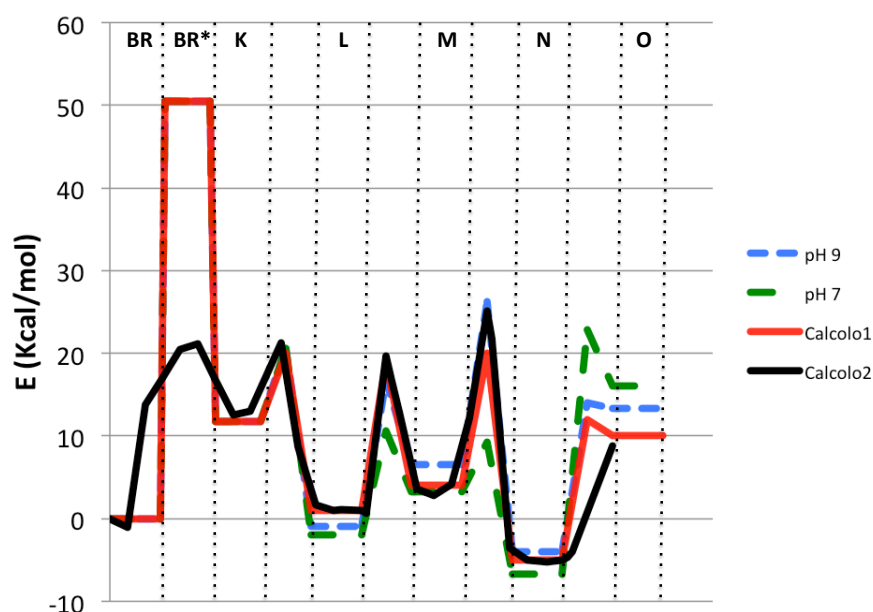


Figura 25: grafico delle energie relative dei vari stati della BR rispetto allo stato fondamentale. In blu i dati sperimentali a pH 9, in verde i dati sperimentali a pH 7, in rosso i dati ottenuti dalle simulazioni (ottimizzazioni locali) utilizzando la funzione doppia buca ed in nero i dati ottenuti dalle simulazioni (ottimizzazioni locali) sia sugli stati fondamentali che su quelli intermedi utilizzando la funzione a doppia buca.

Questi profili sono ottenuti semplicemente con calcoli di ottimizzazione locale a partire da strutture vicine all'equilibrio stabile (o metastabile), e rappresentano l'andamento dell'energia interna. Vere simulazioni di dinamica molecolare di passaggio di stato sono invece necessarie per stimare i modi di passaggio e l'energia libera.

4.4.1 Simulazione della transizione BR→K

In questa sezione si descrivono le simulazioni dei passaggi di stato, operate con i modelli sopra descritti. Il passaggio BR-K avviene in conseguenza ad una eccitazione elettronica dovuta all'assorbimento del fotone. Quindi il processo si può descrivere come una transizione verticale (senza variazione strutturale) che porta BR ad uno stato eccitato BR* ad energia ~50 kcal/mole maggiore. Da BR* il sistema passa a K diseccitandosi e cambiando conformazione del retinale, mentre il resto della proteina subisce solo piccoli riaggiustamenti.

Dato che il passaggio interessante avviene da BR* a K senza barriera in questo caso, l'unico di tutto il ciclo, la descrizione più appropriata non si ottiene con il potenziale bistabile, ma semplicemente usando il campo di forze dello stato K a partire dallo stato BR. I risultati sono riportati in Fig. 26, che mostra il monitoraggio di alcune variabili durante la simulazione. Si osserva che il retinale isomerizza e dopo alcune oscillazioni raggiunge il valore di equilibrio (~-56 gradi)

in circa 3ps. Successivamente il sistema compie oscillazioni intorno al nuovo stato di equilibrio che corrisponde a K.

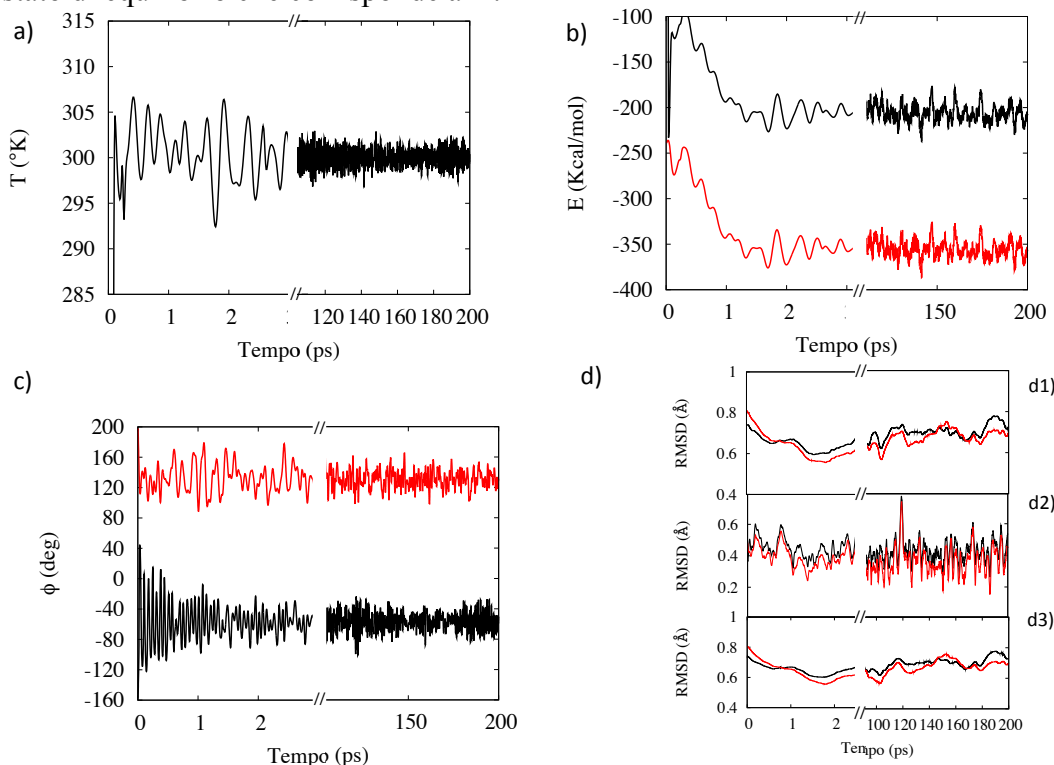


Figura 26: Simulazione della transizione tra BR e K. a) andamento della temperatura in funzione del tempo; b) in rosso l'energia potenziale ed in nero l'energia totale; c) in nero il diedro numero 13 del retinale che si isomerizza dai -156° dello stato BR ai -56° dello stato K ed in rosso il diedro che contiene la base di Schiff che cambia dai -160° dello stato BR fino ai $+130^\circ$ dello stato K; d) RMSD (root mean squared deviation) della simulazione rispetto allo stato BR in nero e rispetto allo stato K in rosso. d1) la RMSD calcolata su tutta la proteina, compreso il retinale, d2) la RMSD calcolata solo sul retinale ed d3) la RMSD calcolata soltanto sull'opsina.

4.4.2 Simulazione della transizione K→L e seguenti

Le transizioni seguenti alla prima avvengono tutte nello stato fondamentale. Gli stati sono separati da barriera ed hanno energie diverse, per cui in questi casi è appropriato descriverle con il potenziale a doppia buca. Queste simulazioni sono state effettuate con i seguenti parametri:

- temperatura: 300K
- ensemble: NVT
- termostato: Berendsen con $\tau = 1\text{fs}$
- timestep: 0.1fs

La transizione da K ad L presuppone il superamento di un barriera di potenziale dell'ordine delle 9 kcal/mole. Alla temperatura di 300K la transizione avviene spontaneamente in circa 10 ps, vedi Fig 27c. Il diedro numero 13 del retinale cambia conformazione in qualche ps, mentre il diedro che comprende la base di Schiff cambia in circa 10ps. L'avvenuta transizione si può osservare dalla

misura della RMSD in Fig. 27d. La RMSD calcolata rispetto allo stato L è minore di quella calcolata rispetto allo stato K, come ci si aspetta, a dimostrazione dell'avvenuta transizione.

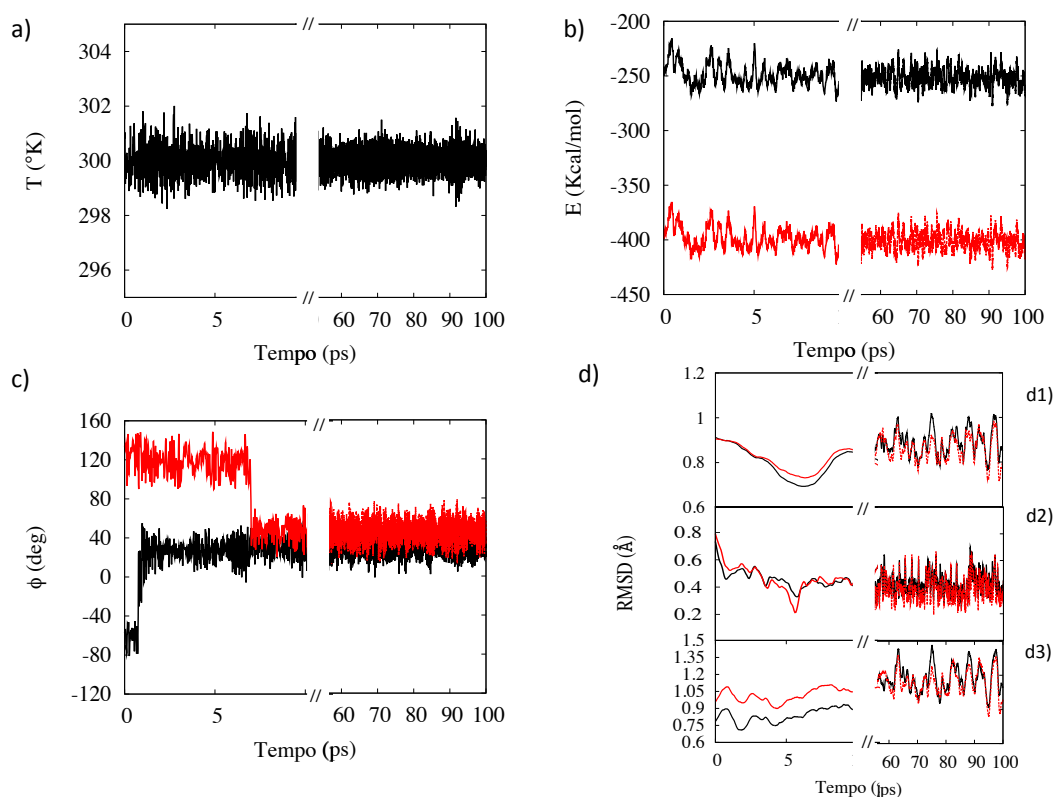


Figura 27: Transizione K-L. a), b) come fig precedente. c) in nero il diedro numero 13 del retinale ed in rosso il diedro che contiene la base di Schiff, d) RMSD rispetto allo stato K in nero e allo stato L in rosso d1) d2) d3) come in fig precedente.

Lo stato L ha una vita media molto lunga, (circa $100\mu\text{s}$, molto più lunga di quella dello stato K, dell'ordine del μs), correlata con una maggiore barriera di potenziale tra L e M, di circa 17 kcal/mole. Con una simulazione partendo dallo stato L servirebbe una simulazione molto lunga, che sarebbe ancora fattibile con questo modello. Tuttavia per accelerare i tempi e mantenerli dentro il progetto di Tesi, si è deciso di partire da una configurazioni intermedie progressivamente sempre più vicine allo stato M, e far superare artificialmente la barriera, rimandando a sviluppi futuri la simulazione realistica. In questo modo, si sono campionati sufficientemente stati da entrambi i lati della barriera. Nella Fig 28 è riportata la simulazione di transizione, ottenuta partendo dallo stato attivato. Come nel caso precedente, si riportano vari parametri che mostrano, con la loro variazione, l'avvenuta transizione. In questo caso è particolarmente interessante il grafico 28d, dove si osserva come la RMSD rispetto allo stato M sia più bassa di quella rispetto allo stato L soprattutto nella parte dove considero solo il retinale. Questo è dato dal fatto che il retinale nei due stati assume due conformazioni molto diverse (il diedro numero 13 passa dai $+30^\circ$ dello stato L ai -157° dello stato M), mentre il resto della proteina è affetto da grandi fluttuazioni termiche che rendono simili le due configurazioni nei due stati. Questo è confermato anche dalle fluttuazioni della temperatura che si osservano in Fig. 28a.

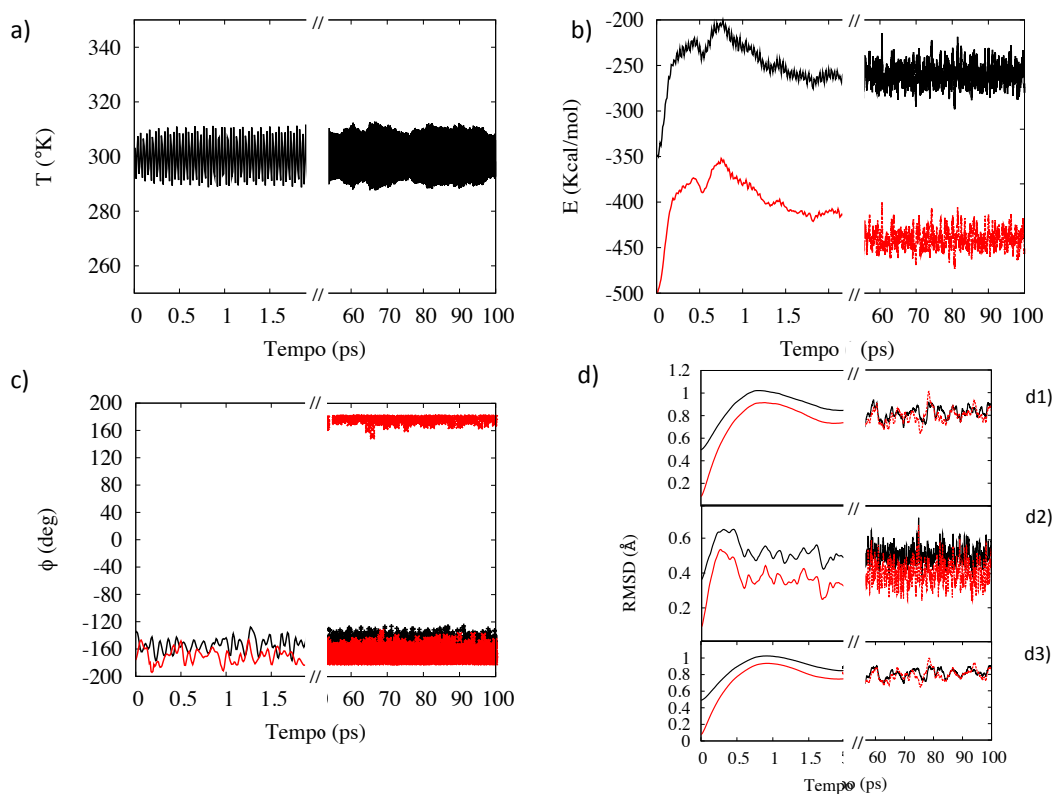


Figura 28: Transizione L-M. a) , b) come fig precedente. c) in nero il diedro numero 13 del retinale ed in rosso il diedro che contiene la base di Schiff. d) RMSD rispetto allo stato L in nero e allo stato M in rosso d1) , d2) , d3) come fig precedente.

Il problema del superamento di barriera si ripropone e maggiore nel caso del passaggio $M \rightarrow N$, che presenta una barriera 20 kcal/mole (che prolunga la vita dello stato M fino a 10 ms). Il problema è risolto nello stesso modo, e la transizione è mostrata in Fig. 25. In questo caso le fluttuazioni termiche sono minori che nel caso precedente ed infatti si distinguono le due curve della RMSD rispetto ai due stati, vedi Fig. 29d. Anche in questo caso la variazione maggiore della RMSD si ha nel retinale il quale ha i cambiamenti maggiori: il diedro numero 13 passa da -157° dello stato M ai $+10^\circ$ dello stato N.

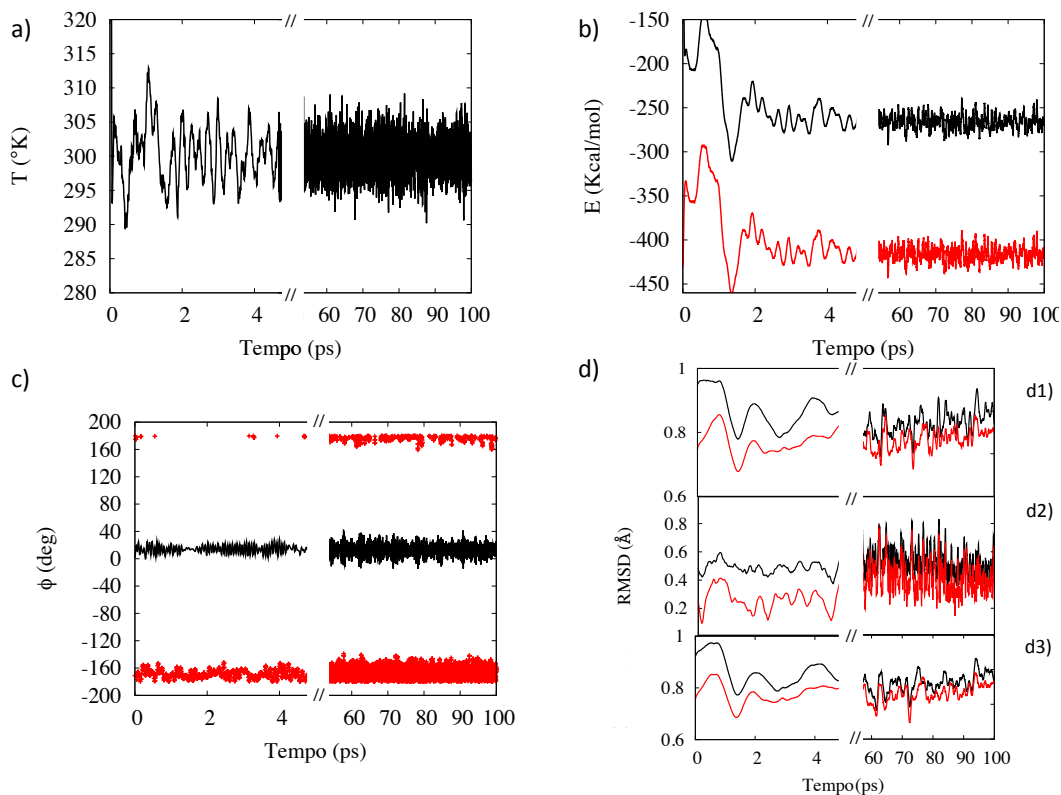


Figura 29: Transizione M-N. a), b) come fig precedente. c) in nero il diedro numero 13 del retinale ed in rosso il diedro che contiene la base di Schiff. d) RMSD rispetto allo stato M in nero e allo stato N in rosso. d1), d2) d3) come fig precedente.

Anche per la transizione dallo stato N allo stato O la barriera di potenziale è dell'ordine delle 20 kcal/mole (vita media di N 5ms). Come nei due casi precedenti per osservare la transizione si è posto il sistema sullo stato attivato. In questo caso risulta più difficile distinguere i due stati strutturali per mezzo delle variabili caratteristiche del cromoforo, perché la sua conformazione non cambia molto nei due stati. Invece la transizione risulta visibile dall'analisi delle variazioni strutturali (RMSD), specificamente quelle dell'opsina, fatto che conferma la presenza di una la transizione strutturale fuori dal retinale.

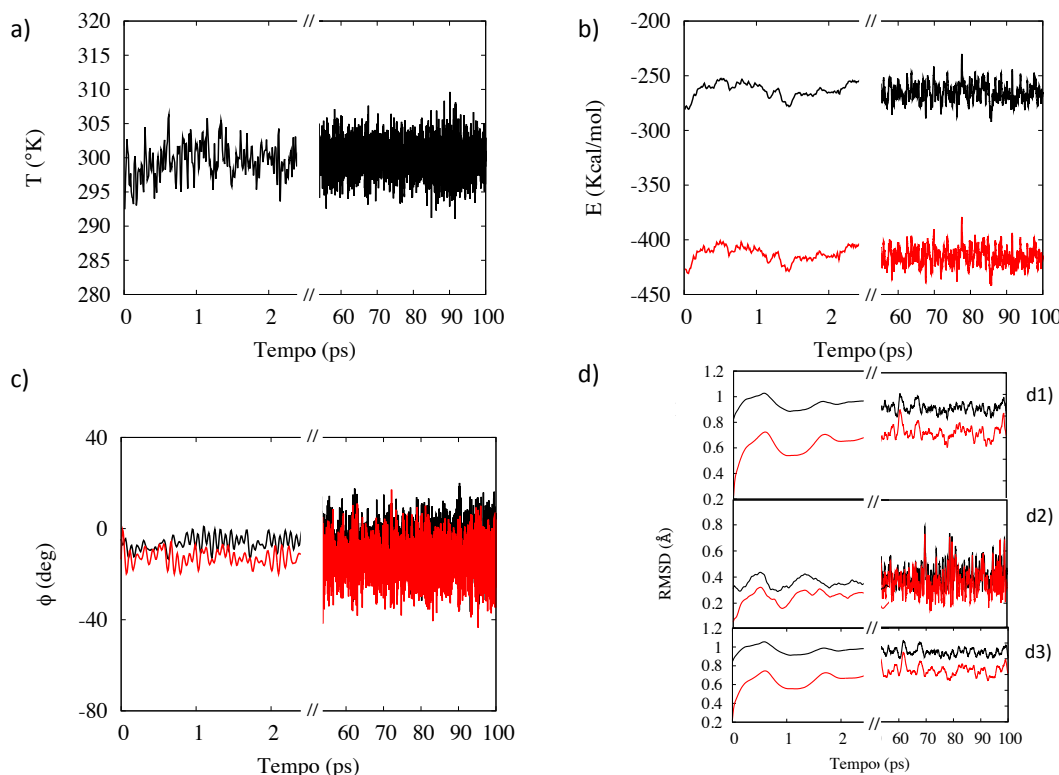


Figura 30: Transizione N-O. a) b) come fig precedente. c) in nero il diedro numero 13 del retinale ed in rosso il diedro che contiene la base di Schiff, d) RMSD rispetto allo stato N in nero e allo stato O in rosso. d1), d2) d3) come fig precedente.

4.5 Analisi delle componenti principali

L'analisi delle componenti principali su di una simulazione di dinamica molecolare restituisce, come spiegato nel capitolo 3 e in Appendice B, i moti fisiologicamente rilevanti del sistema. Se applicata alle transizioni strutturali, in particolare, dà informazioni utili sulla localizzazione e sulla natura delle modifiche strutturali che avvengono durante la transizione. Si è dunque effettuata la PCA (prime 10 componenti) per ciascuna transizione. Dall'analisi è anche possibile calcolare la variazione di energia libera associata alle componenti principali.

4.5.1 Transizione BR-K

In questa transizione si è visto in precedenza che l'evento rilevante è l'isomerizzazione del retinale. La prima componente principale descrive infatti il movimento di isomerizzazione del retinale, come visibile in Fig. 31c e 31d. Il calcolo della RMSF sulla prima componente principale mostra quali sono i residui che hanno le variazioni maggiori. Dalla Fig. 31c si notano due distinte aree di maggiore variazione: il primo si trova nel sesto loop tra le eliche E ed F ed il secondo sul retinale specificamente sulla parte finale della catena del retinale, cioè

al diedro che si isomerizza. Il loop E-F è lo stesso che era mal definito nella struttura cristallografica e che è stato necessario ricostruire. I due fatti sono probabilmente correlati: scarsa definizione nella struttura cristallografica significa maggiore mobilità, come descritto dalla simulazione. Le prime componenti principali descrivono bene quindi il principale evento della transizione tra stato BR e K. Il grafico dell'energia libera, Fig. 31a e 31b, mostra come vi sia un solo minimo che corrisponde allo stato K. Questo accade perché, come detto, in questo caso la transizione è priva di barriera e quindi c'è un solo stato stabile.⁷⁰

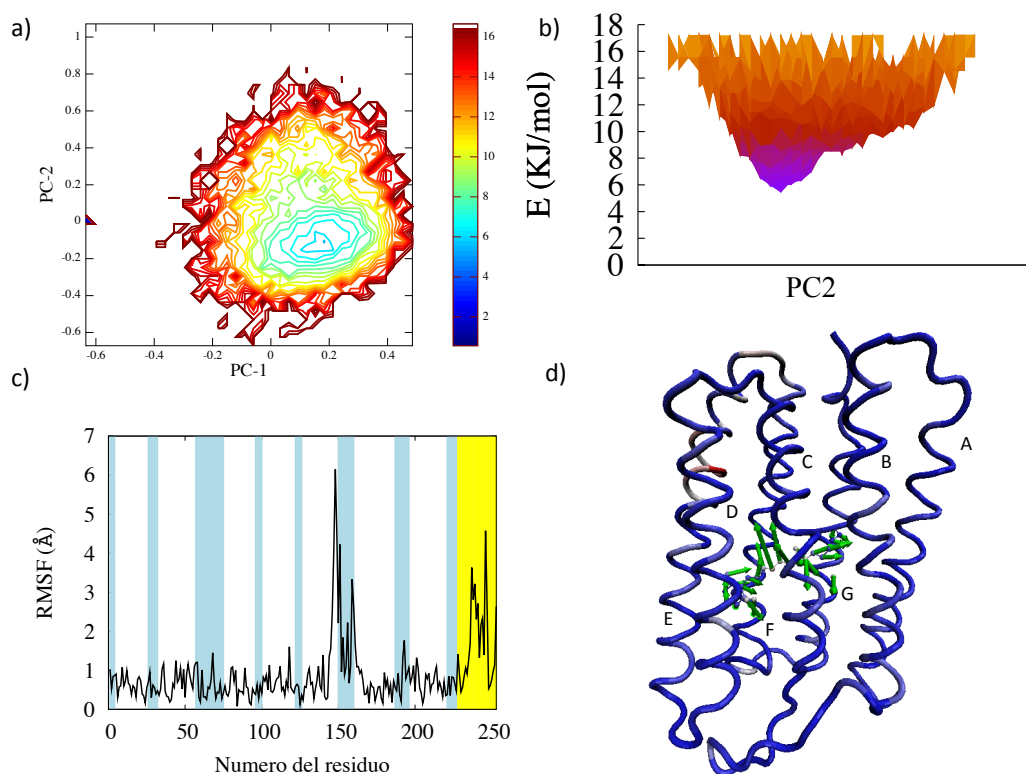


Figura 31: Analisi PCA sulla transizione da stato BR a K. a) il grafico dell'energia libera rispetto alle prime due componenti principali, la scala delle energie riportate a sinistra è espressa in KJ/mol. b) Proiezione dell'energia libera sulla componente PC2; c) la RMSF calcolata sulla prima componente principale. d) Rappresentazione della prima componente principale: le frecce verdi indicano il movimento dei vari residui. In d) la proteina è colorata in modo da far corrispondere alle zone che hanno il maggior spostamento il colore tendente al rosso, mentre le zone che rimangono fisse hanno una colorazione tendente al blu.

4.5.2 Transizione K-L

La seconda transizione del fotociclo, tra gli stati K ed L, interessa di nuovo principalmente il retinale che cambia conformazione. Tuttavia dall'analisi della prima componente principale emerge di nuovo il movimento del loop E-F, mentre il movimento del retinale emerge nella seconda componente, specificamente nella zona della coda verso la base di Schiff, vedi Fig. 32c e 32d. Si osserva che la base di Schiff si avvicina all'elica C, ed in particolar modo al residuo Asp85. Questo movimento è descritto come uno degli eventi del fotociclo che servono al rilascio

del protone sul lato citoplasmatico⁷¹. Un movimento simile si osserva anche nella quarta componente.

La figura riporta il grafico dell'energia sulle prime due componenti. In Fig. 32a e 32b si notano due buche di energia che corrispondono ai due stati, K ed L. In particolare nella buca L si nota una struttura profonda che corrisponde a variabili correlate con i movimenti del retinale.

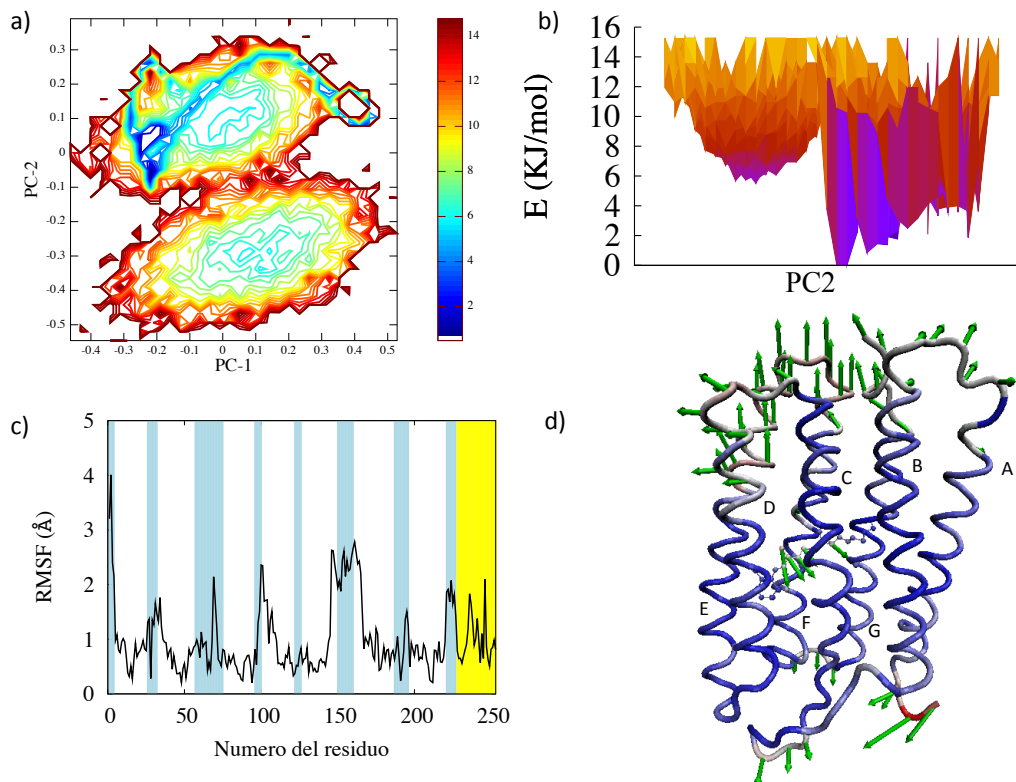


Figura 32: Analisi PCA sulla transizione da stato K a L. a) come fig precedente. b) proiezione dell'energia libera sulla componente PC2; c) la RMSF calcolata sulla PC2 d) come fig precedente, per la PC2.

4.5.3 Transizione L-M

Durante la transizione da L ad M avviene un grande cambiamento del retinale poiché il suo diedro numero 13 cambia dai $+30^\circ$ dello stato L ai -157° dello stato M, quindi una variazione di 180° , cioè una rotazione completa del diedro. Vi sono anche altri piccoli movimenti dati da una serie di cambiamenti dei legami a idrogeno⁷². Dalla Fig. 33c e 33d si vede che la prima componente principale interessa anche alcuni loop ad entrambi i lati della proteina. Anche in questo caso si osservano le due buche corrispondenti ai due stati, nel grafico dell'energia in funzione delle componenti principali. L'altezza della barriera e le differenze di energia non sono riprodotte in maniera molto accurata, ma bisogna ricordare che per una totale riproduzione dell'energia bisogna inserire tutte le componenti principali. Questo aspetto è attualmente in fase di approfondimento.

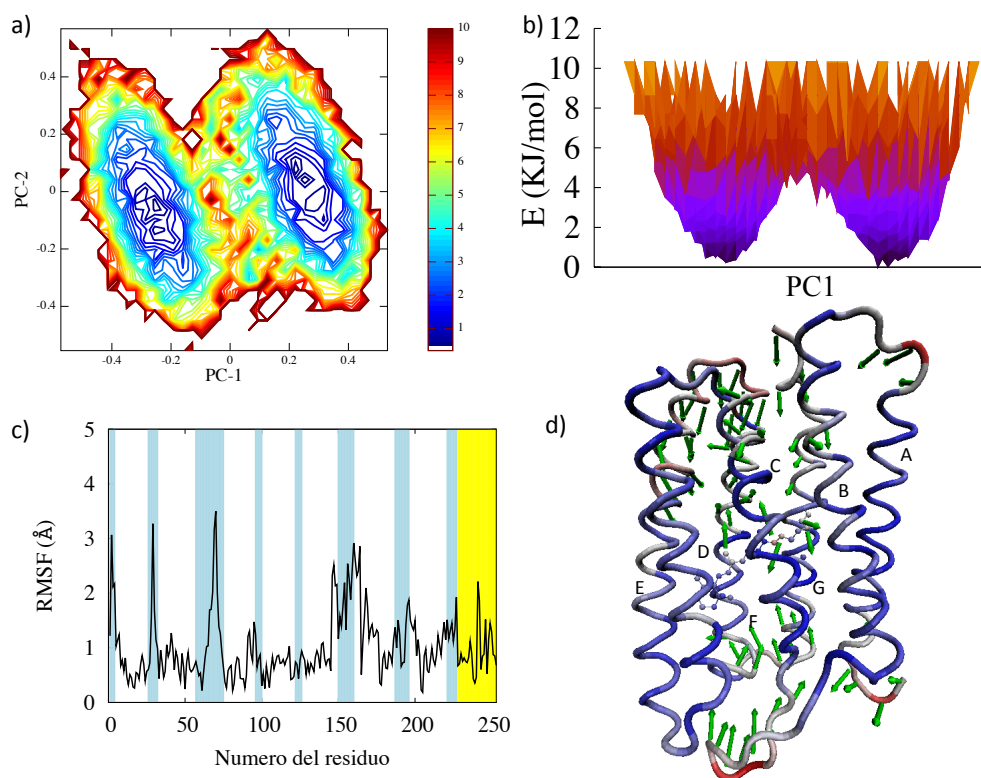


Figura 33: Analisi PCA sulla transizione da stato L ad M. a) come fig precedente. b) proiezione dell'energia libera sulla componente PC1; c) la RMSF calcolata sulla PC1 d) come fig precedente, per PC1.

4.5.4 Transizione M-N

Durante la transizione tra gli stati M ed N avviene l'apertura di un canale verso il lato citoplasmatico creato dal movimento del tratto terminale dell'elica E che si inclina verso l'esterno creando un percorso che porta al centro della proteina⁷³. Questo movimento è descritto dalla seconda componente principale come visibile in Fig. 34c e 34d. Nel grafico della RMSF, infatti, si nota che attorno al residuo 150, cioè il loop E-F, vi è un grande cambiamento. Inoltre dalla rappresentazione in Fig. 34d si nota il movimento di apertura di questo loop sul lato citoplasmatico. In Fig. 34a e 34b si notano tre diversi minimi della superficie di energia che corrispondono a tre diversi stati. Quello più largo corrisponde allo stato M e presenta un'energia maggiore. Gli altri due corrispondono allo stato N ed allo stato attivato, che risulta metastabile. La comparsa dello stato attivato nella superficie di energia è un fatto particolarmente rilevante, in quanto questo stato non è, come gli altri due, inserito per costruzione nel modello, ma emerge spontaneamente.

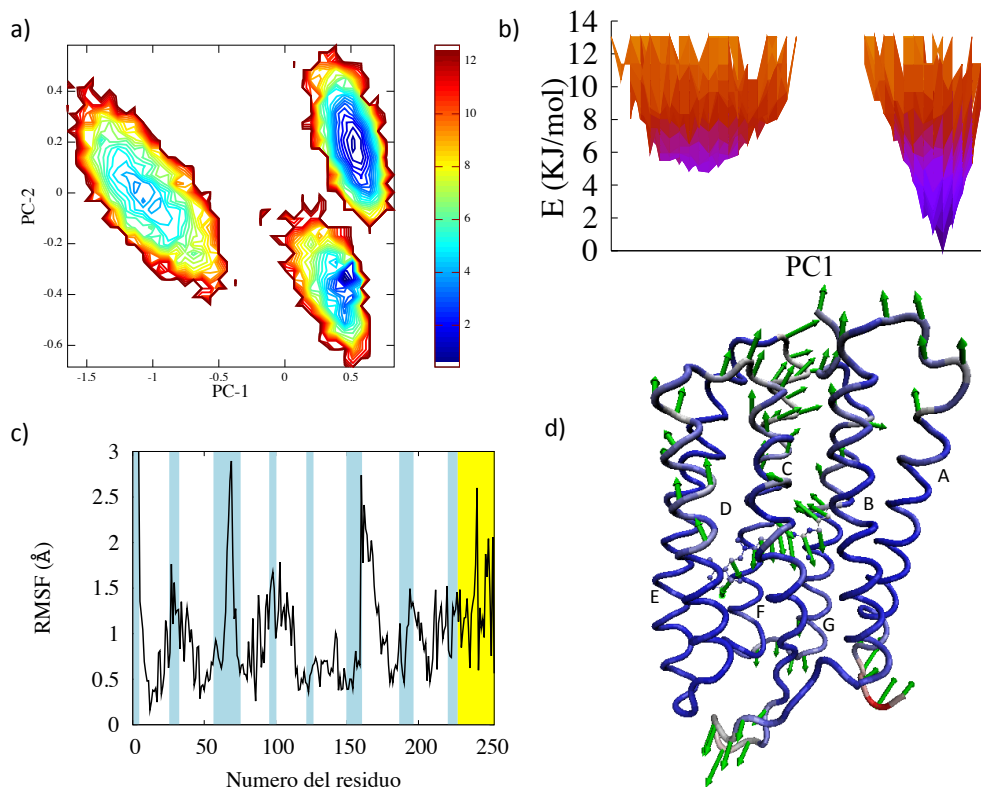


Figura 34: Analisi PCA sulla transizione da stato M ad N. a) come fig precedente b) proiezione dell'energia libera sulla PC1 c) RSMF sulla PC1 d) come fig precedente, sulla PC1.

4.5.5 Transizione N-O

Durante l'ultima transizione una parte dell'elica C sul lato extracellulare si muove verso l'elica G stringendo il canale di rilascio del protone ed espellendo la molecola di acqua. Quest'elica, infatti, funge da valvola grazie alla quale si ottiene l'unidirezionalità del movimento degli ioni⁷⁴. La prima componente tra gli stati N ed O interessa il loop dell'elica C sul lato extracellulare e mostra il movimento sia del loop che del tratto terminale dell'elica C verso l'elica G. Inoltre questo modo indica anche il movimento della molecola di acqua W402 verso il lato citoplasmatico, come se la molecola venisse espulsa, vedi Fig. 35c e 35d. Nella seconda componente principale, invece, si nota molto bene il movimento della molecola di acqua W402 e il cambiamento di conformazione del retinale. Il grafico della superficie di energia libera mostra tre stati. Lo stato N è descritto nella zona centrata in $PC1=0.5$ e $PC2=-0.2$, lo stato O da quella centrata in $PC1=-1$ e $PC2=-0.1$ e lo stato attivato da quella centrata in $PC1=0.2$ e $PC2=0.5$. Anche in questo caso, come nel precedente, lo stato attivato è uno stato emergente del modello. Dalla Fig. 35b si vede che lo stato N ha energia minore rispetto allo stato O in accordo coi dati sperimentali (Tabella 1).

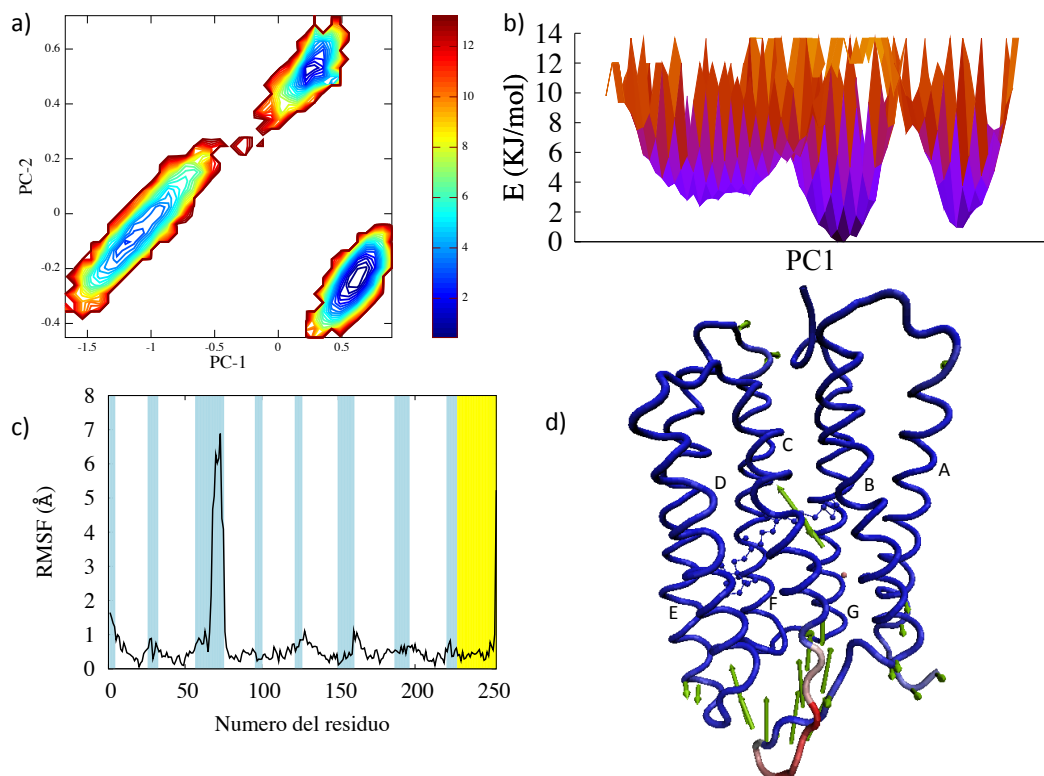


Figura 35: Analisi PCA sulla transizione da stato N ad O. a) Come fig precedente b) Proiezione dell'energia libera sulla PC1 c) RSMF sulla PC1 d) Come fig precedente, sulla PC1.

4.6 Sommario dei risultati

I risultati riportati si possono riassumere nei seguenti punti

1. L'ottimizzazione dei campi di forze per ciascuno stato separato ha necessitato alcune modifiche rispetto ai modelli simili precedenti, riguardanti soprattutto le parti peculiari del modello, ovvero la multi-scalarità, ed in particolare i parametri dell'interfaccia tra parte CG e parte quasi atomistica. Le modifiche operate nel resto del campo di forze sono state minori, essendo già sufficientemente accurati i parametri di letteratura.
2. La costruzione dei campi bistabili ha necessitato l'aggiunta di un singolo parametro aggiustabile, che è stato ottimizzato, caso per caso, sul dato sperimentale di energia relativa.
3. Il modello è in grado di riprodurre:
 - a. Struttura e fluttuazioni termiche degli stati stabili
 - b. Transizioni e profili energetici tra gli stati
 - c. Modi principali delle transizioni con interpretazioni fisiche che riflettono la funzione della proteina
 - d. Stati intermedi metastabili

Tutto ciò considerato, si può attribuire a questo modello l'appellativo di modello minimalista per le transizioni della rodopsina (o di qualsiasi altra proteina i cui stati stabili siano noti, anche a bassa risoluzione).

Non è uno scopo primario di questo lavoro approfondire le caratteristiche molecolari dei singoli passaggi di stato, che, del resto, per il caso particolare della rodopsina batterica sono abbastanza ben note. Qui ci si è concentrati sull'ottimizzazione del modello ed un suo primo test. Dinamiche estensive e mirate saranno oggetto di studi futuri (si veda prossimo capitolo). Tuttavia osserviamo che il modello cattura i due principali movimenti funzionali della proteina, l'isomerizzazione del retinale e l'apertura del canale per il passaggio protonico, e lo fa nelle fasi specifiche del fotociclo e non in altre, in accordo con i dati sperimentali, a validazione dell'accuratezza del modello.

Osserviamo tuttavia anche l'emergenza degli stati intermedi metastabili, che non sono inclusi (a differenza degli stati di partenza ed arrivo) tra le strutture di riferimento del modello. Questo indica le potenzialità di questo approccio, nonostante la sua semplicità e il suo minimo costo computazionale. Questi stati sono particolarmente ben visibili nell'analisi dei modi principali, come minimi aggiuntivi dell'energia libera.

Notiamo che il calcolo dell'energia libera è in genere piuttosto complesso, e coinvolge problemi di campionamento dello spazio delle configurazioni, che dovrebbe assicurare il raggiungimento del limite statistico. Quindi i risultati presentati vanno presi come preliminari. Ciononostante, l'emergenza di stati metastabili intermedi non è sorprendente: il modello ha un bias solo parziale verso i due stati di riferimento, mentre molti dei termini energetici sono privi di bias, e possono tenere conto in maniera predittiva di altri stati.

Capitolo 5. Conclusioni e possibili sviluppi

Il lavoro descritto in questa Tesi riguarda la costruzione di un modello multi-scala e multi-stabile per la descrizione della dinamica molecolare delle rodopsine. I risultati riguardano quindi sia il modello che le dinamiche molecolari proteiche.

La necessità della costruzione di un modello multi-scala nasce da una caratteristica del funzionamento delle rodopsine (che però è in comune con molte altre proteine), ovvero l'esistenza di un passo "trigger" del fotociclo, l'isomerizzazione fotoindotta del retinale, che è ultraveloce e avviene in un'area localizzata della proteina (il sito attivo), seguita da una serie di eventi più estesi e molto più lenti, con il risultato che il fotociclo si estende su molti ordini di grandezza temporali. Conseguentemente, sono necessarie almeno due diverse risoluzioni spaziali: quella quasi-atomistica per descrivere l'isomerizzazione del retinale, e quella Coarse Grained minimalista, per ridurre il costo computazionale del resto della proteina e consentire così di raggiungere la scala temporale macroscopica.

Si è proceduto quindi alla costruzione di un modello multi scala per ciascuno dei singoli stati del ciclo, nella maniera più semplice possibile, ovvero combinando modelli esistenti per le due risoluzioni diverse. Si sono scelti modelli minimalisti, con il numero più piccolo possibile di parametri. Inoltre siccome erano deputati alla descrizione di singoli stati, questi modelli sono caratterizzati da un bias parziale verso strutture di riferimento sperimentali. Il contributo originale a questa parte è una parziale riottimizzazione dei parametri, ma soprattutto la parametrizzazione *de novo* della parte di interfaccia tra le due risoluzioni, operata in maniera più generale possibile, in modo da essere applicabile ad altre proteine. Questa in particolare ha richiesto un approccio creativo, basato sulla riproduzione dei profili di energia appositamente calcolati con modelli ad alta risoluzione. Riassumendo, i risultati di questa parte sono:

1. L'ottimizzazione di un campo di forze per un modello multi-scala per i singoli stati del fotociclo della rodopsina batterica, scelta come caso di studio per la presenza di determinazioni sperimentali accurate dei vari stati stabili.
2. Validazione del campo sulla rodopsina bovina (vicina come classificazione, ma strutturalmente abbastanza diversa), per la quale i parametri non sono stati cambiati.

La seconda serie di risultati, ancora sul lato modellistico, è l'ideazione di un protocollo per generare campi multi-stabili a partire dai campi ottimizzati sui singoli stati. Ancora una volta si è scelto di adottare un approccio minimalista, che introduce un numero minimo di parametri addizionali. Questi non sono in pratica liberi, perché possono venire fissati sulla base del valore dell'energia relativa degli stati (se nota). In questo senso l'approccio è in grado di riprodurre transizioni ed energie libere relative con accuratezza perfino maggiore degli approcci completamente atomistici, che – almeno quelli standard – falliscono esattamente su questo punto: I campi di forze atomistici empirici attuali non hanno, in linea di principio parametri liberi, ma non sono in grado di riprodurre accuratamente le energie relative di stati strutturali molto diversi tra loro.

Anche in questo caso, si tratta di un approccio del tutto generale, estendibile a qualsiasi sistema che presenti multi-stabilità e per il quale esistano modelli per i singoli stati. Inoltre, come accennato nel capitolo precedente, si è implementato qui il caso della bi-stabilità, collegando i vari stati a due a due. Tuttavia il passaggio alla multi-stabilità è formalmente banale, essendo l'unico ostacolo l'implementazione nei codici di simulazioni delle forme funzionali – relativamente più complesse – che descrivono i campi multi-stabili. Questo è un punto che può essere facilmente affrontato nel prossimo futuro.

La terza serie di risultati riguarda le simulazioni del fotociclo della rodopsina batterica. Si sono operate simulazioni di passaggio a due a due tra i vari stati, mostrando che gli elementi fondamentali energetici, strutturali e dinamici sono colti dal modello. Il profilo di energia interna viene correttamente riprodotto. L'analisi dei modi principali, passaggio per passaggio, evidenzia tra i primi di essi i movimenti della proteina correlati con la sua funzione (isomerizzazione del retinale per i primi passaggi, spostamento di specifici loop per gli altri, correlati con l'apertura del canale per il passaggio protonico). In altre parole, il modello coglie la biofisica fondamentale del sistema al livello molecolare. Inoltre è probabile che nessun altro modello più semplice di questo sia in grado di farlo: per rappresentare il cambiamento conformazionale del retinale è necessaria la sua rappresentazione quasi atomistica, mentre rappresentazioni più semplici di quella basata sui C α della proteina non sono in grado di descrivere gli spostamenti delle eliche e i cambiamenti conformazionali dei loop che si osservano in queste simulazioni. Quindi con buone ragioni questo si può definire un modello “minimalista”.

Le simulazioni portano anche alcuni risultati inaspettati, che emergono dal modello: specificamente alcuni degli stati intermedi risultano metastabili, nonostante non siano stati inseriti a priori. Inoltre, risultano particolarmente interessanti le simulazioni sugli stati finali del fotociclo, in particolare lo stato O, che è tra i più elusivi, essendo stato cristallizzato per ultimo e a bassa risoluzione.

Infine, a differenza di tutti gli studi precedenti, in questo lavoro si è analizzato l'intero fotociclo, estendendo la simulazioni su vari ordini di grandezza temporali. In pratica questo studio fornisce la prima rappresentazione continua dell'intero fotociclo della rodopsina batterica, e senza perdere accuratezza sulla parte locale (in senso sia spaziale che temporale).

Ci sono stati poi una serie di risultati collaterali che sono potenzialmente utili alla comunità dei simulatori di biosistemi. Anzitutto, i modelli multi-scala e multi-stabili sono stati implementati in software di supporto al programma principale di simulazione DL_POLY. In pratica, questi software sono in grado di generare il modello per qualsiasi proteina, purché gli input necessari (strutture di riferimento) siano disponibili. Inoltre, il codice DL_POLY stesso è stato manipolato e flessibilizzato, in modo da supportare il modello, includendo le forme funzionali necessarie. Queste tuttavia, sono piuttosto generali, ed adatte anche a rappresentare la multi-stabilità in sistemi di qualsiasi tipo (anche non biologici) e quindi in particolare le transizioni di fase e di struttura.

Tra i possibili sviluppi di questo lavoro, va citato anzitutto uno studio più dettagliato dei passaggi di stato, sia della rodopsina batterica, che della bovina, rappresentante di quella dei mammiferi. Mentre questo lavoro è centrato sull'ottimizzazione del modello, e le simulazioni sulla rodopsina batterica sono

usate per lo più come test, con uno sforzo addizionale non enorme sarà possibile ottenere dinamiche molto accurate e dettagliate dei passaggi di stato. I punti fondamentali sono due: da una parte la descrizione temporalmente realistica della cinetica, che richiede come accennato nel capitolo 3, l'uso della dinamica stocastica (Langevin o Browniana) anziché della usuale dinamica Newtoniana. Con uno sforzo addizionale minimo questo consente di avere tempi di rilassamento e vita media corretti degli stati, aspetto sul quale in questo lavoro non ci si è soffermati. L'altro punto è la descrizione veloce ma al tempo stesso statisticamente accurata del passaggio delle barriere, che richiede o simulazioni più estese di quelle qui operate (per mancanza di tempo) e/o l'uso di specifiche dinamiche accelerate (metadinamica ad esempio) per campionare correttamente i profili di energia libera. Di nuovo, lo sforzo addizionale non è grande: si tratta solo di usare un diverso tipo di dinamica e di campionamento dello spazio delle configurazioni. Il grosso del lavoro (ottimizzazione del modello) è concluso in questa Tesi.

Mentre la rodopsina batterica potrà essere di nuovo usata come caso test per la validazione dei risultati di dinamiche più estese e più statisticamente rilevanti, ci si aspettano risultati biologicamente interessanti specificamente nel caso della rodopsina bovina, che presenta ancora lati oscuri del fotociclo. Ad esempio, un aspetto interessante di questa proteina è che durante la transizione tra lo stato MetaII e lo stato Opsina avviene il rilascio del retinale in configurazione trans, dinamica difficilmente riproducibile in simulazioni atomistiche. Questo modello potrebbe essere adattato all'evento di rilascio del retinale indebolendo le iterazioni tra retinale ed opsina potendo così osservare come avviene il rilascio ed anche un eventuale nuovo accoppiamento tra proteina e retinale in configurazione cis per tornare allo stato fondamentale.

Tra le altre applicazioni, una direttamente collegata alla rodopsina bovina e di particolare interesse sono i recettori accoppiati alle proteine G di membrana (GPCR), una grande classe di proteine di cui la rodopsina dei mammiferi fa parte. Questi vengono utilizzati molto spesso come bersaglio per farmaci proprio perché sono adibiti alla comunicazione cellulare, al bilancio di ioni e ad altre importanti funzioni fisiologiche. Riuscire a predire il loro funzionamento e controllare la loro reazione a specifiche molecole risulta, quindi, molto importante per la progettazione di farmaci. Notiamo che il presente modello, per il suo costo computazionale estremamente basso, è perfettamente adatto a descrivere processi del genere, che coinvolgono non singole proteine, ma complessi macromolecolari proteici molto grandi ed immersi nella membrana.

Quest'ultimo è un altro aspetto interessante che si presta ad una linea di sviluppo separata. Il modello qui descritto rappresenta la membrana implicitamente con dei vincoli armonici, che tengono conto solo dell'effetto contenitivo. Ci sono molti modi diversi con cui questa rappresentazione può essere migliorata. Si possono trattare gli aspetti elettrostatici della membrana trattandola come un continuo con proprietà dielettriche dipendenti dalla posizione; si possono aggiungere proprietà dinamiche finì alla membrana, trattandola come un continuo meccanico; infine, si può descrivere la membrana al livello molecolare, eventualmente anch'essa con modelli coarse grained (ne esistono diversi), che permetterebbe di descrivere accuratamente la diffusione delle

rodopsine al suo interno, e di studiare alcuni aspetti dell'aggregazione in oligomeri.

L'applicabilità ad altre proteine è condizionata solo dalla disponibilità di strutture di riferimento affidabili. Queste strutture possono essere anche a bassa risoluzione o a risoluzione mista, e quindi generate con tecniche sperimentali diverse dalla cristallografia a raggi X, che magari rispettano meglio l'ambiente fisiologico della proteina, e quindi sono più accurate. In linea di principio, non c'è nemmeno un limite stretto sul tipo di biomolecola: con un ulteriore sforzo di parametrizzazione, il modello è facilmente estendibile agli acidi nucleici, per i quali esistono modelli CG e atomistici. Invece per le proteine, non necessita ulteriori modifiche. Considerato che nei sistemi biologici l'estensione dei fenomeni su molti ordini di grandezza spaziali e temporali è la norma, così come sono frequenti variazioni conformazionali, normalmente con importanti significati funzionali e fisiologici, risulta quindi evidente che si apre un di ventaglio di applicazioni, di potenziale interesse biologico, farmacologico e medico.

Per concludere, questo lavoro benché in sé concluso, rappresenta in realtà un punto di partenza sia per studi applicativi che per sviluppi modellistici.

Appendici

Appendice A. Scelta delle strutture di riferimento delle rodopsine

Le strutture sperimentali delle rodopsine sono disponibili sul Protein Data Bank (PDB⁷⁵), la più completa base di dati pubblica per strutture di molecole biologiche. Una ricerca con le parole chiave “bacteriorhodopsin” e “bovine rhodopsin”, ha restituito rispettivamente 102 e 31 strutture, corrispondenti ai vari stati ottenuti con varie determinazioni sperimentali, attraverso diverse tecniche e a diversa risoluzione. La tabella A.1 riporta un sommario delle principali caratteristiche delle strutture della rodopsina batterica.

Per la costruzione del modello serve una singola struttura per ogni stato. Nel caso più di una fosse disponibile, si è proceduto ad una selezione basata sui seguenti criteri: 1. Massima risoluzione possibile (solitamente correlata con l'età della determinazione sperimentale, essendo usualmente le più recenti anche a risoluzione maggiore), 2. Massima completezza (maggior numero di amminoacidi risolti presenti nella struttura).

In alcuni casi le strutture a risoluzione più alta erano incomplete. Si è allora proceduto a completarle ricostruendo gli amminoacidi mancanti sulla base di strutture a risoluzione più bassa appartenenti allo stesso stato. Ad esempio, per lo stato BR si è usata la determinazione 1M0K completata con la struttura 1FBB per gli amminoacidi dal 153 al 166. Il risultato è riportato in figura A.1 (residui provenienti dal 1M0K in blu, gli altri in rosso)

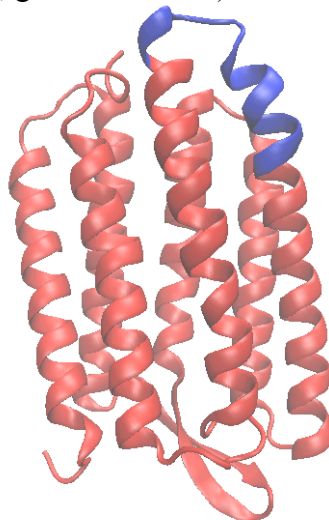


Figura A.1: struttura dello stato fondamentale proveniente dal file 1M0K (rosso) completata con la struttura 1FBB con i residui che vanno dal 153 al 166 (blu).

Il risultato finale di questa operazione sono sei strutture, una per ciascuno stato. Tutte le strutture hanno lo stesso numero di amminoacidi (253), anche se possono avere un numero di atomi variabile, dovuto alla presenza o assenza di molecole di acqua, o protonazioni leggermente diverse. Di queste, per costruire il modello multiscala, vengono selezionati solo gli atomi di Ca e del retinale (idrogeno escluso).

Stato	λ_{ass} nm	Codice PDB	n. res	risol (Å)	Anno pubb.	Stato retinale	Met sp
BR Il retinale è in conformazione all-trans e W402 ha legami a idrogeno con Lys216, Asp85 e Asp212. Asp85 è l'accettore di protoni durante il trasporto. Il legame NH della BS punta nella direzione extracellulare.	570	1C3W	222	1.55	1999	all-trans	rX
		1FBB	248		2000		ME
		1M0K(1)	262	1.43	2002		rX
		1M0M(2)	262	1.43	2002		rX
		1O0A(2)	249	1.62	2003		rX
BR→K: Il retinale isomerizza da all-trans a 13-cis, il legame c15-NZ ruota in senso antiorario. I legami a idrogeno tra Lys216 e Gly220 e tra W402 e Asp85 si rompono.							
K Il retinale è in conformazione 13-cis, la BS, a causa del suo cambiamento di orientazione, disloca W402 e si rompe il legame a idrogeno con Asp85.	590	1IXF	248	2.6	2002	13-cis	rX
		1M0K(2)	262	1.43	2002		rX
		1QKP	248	2.1	1999		rX
		1M0L	262	1.47	2002		rX
K→L: Il retinale va in configurazione 13-cis/15-anti, il legame NH si gira verso la direzione citoplasmatica. W402 si muove verso l'alto e la SB si muove verso Asp85. W402 viene trascinata dal canale di rilascio del protone ad una microcavità creata nel lato citoplasmatico della base di Schiff.							
L Il retinale è in configurazione 13-cis/15-anti e la BS punta verso la direzione citoplasmatica. La BS si muove verso l'accettore di protoni Asp85.	550	1E0P	228	2.1	2000	13cis	rX
		1O0A(1)	249	1.62	2003		15-anti
		1UCQ	249	2.4	2004	rX	
		1VJM	249	2.3	2004	rX	
		2NTU	249	1.53	2007	rX	
L→M: Il legame a idrogeno tra la BS e W402 si rompe. Ciò induce un decremento del pKa della BS e il trasferimento del protone al suo controione Asp85. Le estremità delle eliche F e G sul lato citoplasmatico si spostano. Asp212 rompe il legame con W402, ma ne forma con W401, Trp86 e Tyr185.							
M Aumenta la distanza tra la SB e W402 e si perde il legame a idrogeno tra queste due. Avviene la protonazione di Asp85 ed espulsione del protone sulla superficie extracellulare.	412	1C8R	249	1.8	1999	13-cis	rX
		1F50	227	1.7	2000		15-anti
		1FBK	248		2000	ME	
		1KGB	231	1.65	2001	rX	
		1KG8	231	2.0	2001	rX	
		1M0M(1)	262	1.43	2002	rX	
		1P8H	249	1.52	2003	rX	
		1DZE	248	2.5	2004	rX	
		2ZZL	262	2.03	2009	rX	
M→N: Dalla regione extracellulare avviene la riprotonazione di BS attraverso Asp96 che è posizionato sul lato citoplasmatico della proteina.							
N Le catene laterali amminoacidiche nella regione citoplasmatica tra le eliche F e G sono riorganizzate e l'estremità dell'elica E verso il citoplasma si inclina aprendo un collegamento dall'esterno al centro della proteina	560	1P8U	249	1.62	2003	13-cis 15-anti	rX
N→O: Un protone viene preso dal lato citoplasmatico e avviene la reisomerizzazione del retinale e il trasferimento di un protone da Asp85 al complesso di rilascio del protone. Il retinale torna in configurazione all-trans e la SB si lega a W402, Asp96 è reprotoneato.							
O Deprotonazione di Asp85 e riprotonazione del gruppo di rilascio del protone. Retinale è configurazione all-trans. Una parte dell'elica C sul lato extracellulare si muove verso l'elica G, stringendo il canale di rilascio del protone ed espellendo una molecola di acqua da una microcavità vicina alla BS.	640	1X0I	248	2.3	2005	13-cis	rX
		1JV7	249	2.25	2001		rX
		3VI0	262	2.3	2012		rX
O→BR: A questo punto la proteina torna nella configurazione dello stato iniziale ed è pronta ad una nuova fotoisomerizzazione							

Tabella A.1: Strutture per BR presenti nel database PDB, suddivise per stato del fotociclo, e loro principali caratteristiche: lunghezza d'onda di assorbimento dello stato, codice PDB, n. residui risolti, risoluzione globale, anno di pubblicazione, stato del retinale e tecnica sperimentale di determinazione. Sono riportati anche commenti descrittivi delle varie strutture e del passaggio dall'una all'altra. Abbreviazioni: SB= Base di Schiff; rX= raggi X, ME=microscopia elettronica. In alcuni casi una stessa determinazione sperimentale riporta due diversi modelli corrispondenti a due stati diversi (in quel caso sono contrassegnati con numeri successivi in parentesi vicino al codice pdb)

Nel seguito sono riportati i dettagli delle loro caratteristiche e i criteri di selezione.

Stato BR: struttura 1M0K(1), quella con maggior risoluzione e coerente con le strutture per stati K e M, prodotte dagli stessi autori e con metodologie simili. Completata con i residui dal 153 al 166 della 1FBB ottenendo una struttura di 2109 atomi.

Stato K: struttura 1M0K(2), alta risoluzione e coerente con quella dello stato BR completata con i residui dal 152 al 168 della struttura 1IXF ottenendo una struttura di 2109 atomi.

Stato L: struttura 1O0A(1), miglior compromesso tra completezza e risoluzione. Completata con i residui dal 150 al 168 della struttura 1UCQ ottenendo una struttura di 2108 atomi.

Stato M: struttura 1M0M(1), miglior compromesso tra completezza e risoluzione e più compatibile con quelle degli stati BR e K, perché ottenuta con dagli stessi autori con la stessa tecnica. Completata con la struttura 2ZZL dal residuo 150 al 173 ottenendo una struttura con 2109 atomi.

Stato N: scelta obbligata all'unica struttura esistente per questo stato, la 1P8U, Completata con i residui dal 151 al 163 della struttura 1M0M dello stato precedente, ottenendo 2111 atomi.

Stato O: per questo stato le strutture presenti nel database sono tre. La struttura 1JV7 corrisponde ad una proteina mutante (D85S) stabilizzata nello stato O, ma ci sono delle incongruenze: il lavoro di Rouhani et al.⁷⁶ rileva che le eliche di questa struttura sono di lunghezza diversa rispetto a quelle degli altri stati e ciò potrebbe essere causato dall'impacchettamento della proteina nel cristallo. Questo comporta anche che l'allineamento con le altre strutture e l'eventuale ricostruzione delle parti mancanti sia molto difficoltoso.

La struttura 1X0I, sebbene associata anch'essa allo stato O rappresenta in realtà la sua forma acida (di colore blu) e rappresenta uno stato diverso e più complesso, indotto dalla variazione del pH⁷⁷. Anche in questo caso, infatti, l'allineamento con le altre strutture evidenzia differenze sostanziali. Come nel caso 1X0I ci sono inoltre molti atomi mancanti dovuti anche all'incompletezza di alcuni amminoacidi oltre che alla mancanza di interi amminoacidi. La struttura è il mutante 3VI0 specificamente studiato per allungare la vita media dello stato da 10^{-2} secondi a circa 2.5 secondi in modo da poterne studiare più agevolmente la struttura. La struttura risulta molto simile allo stato M, ma con il retinale in configurazione 13cis/15syn e il residuo Lys216 meno distorto. Questa struttura è quella scelta perché la più completa.

Appendice B. Descrizione del software utilizzato, modificato e creato

Per le simulazioni di dinamica molecolare classica di questo lavoro di tesi si è utilizzato DL_POLY⁶⁵, un codice “*general purpose*”, cioè che in linea di principio permette di implementare qualsiasi modello e/o campo di forze, in contrapposizione ad altri codici, come Gromacs, che implementa solo uno specifico campo di forze per modelli atomistici di proteine. Il codice permette di simulare dinamica molecolare base e con i termostati descritti nel capitolo 3, è molto flessibile riguardo agli algoritmi di integrazione, le condizioni iniziali e al contorno, e ha un’alta scalabilità su sistemi paralleli. Anche se questo aspetto non è stato sfruttato in questo lavoro di tesi, può essere importante per gli sviluppi futuri.

DL_POLY è un codice *open-source* scritto nel linguaggio di programmazione Fortran, a licenza gratuita per uso accademico⁷⁸. Esistono varie versioni, in particolare due classi principali “DL_POLY classic” (ex DL_POLY 2.20) e DL_POLY_4. La seconda è pensata per parallelizzazione massiccia (migliaia di processori) ed è stata pesantemente riscritta rispetto alla versione “classic”, che comunque scala bene fino a centinaia di processori e presenta uno stile di programmazione un pò più semplice e quindi più facilmente modificabile. Per questo motivo (si veda sotto), e poiché nel nostro caso non sono necessarie pesanti risorse parallele, la scelta è ricaduta sul “classic”.

La flessibilità di DL_POLY si paga con il fatto che ciascun modello/campo di forze specifico deve essere implementato dall’utente. Conseguentemente, una parte importante di questo lavoro di tesi è consistita nel programmare codici per generare, a partire dalle strutture di riferimento della rodopsina, i file di input per DL_POLY che implementano il modello CG-multiscala e mutistabile, descritti nella sezione B.1. Il codice DL_POLY stesso ha necessitato alcune modifiche per implementare forme funzionali specifiche per descrivere i passaggi tra gli stati. Le modifiche, che saranno rese disponibili su richiesta, sono descritte in B.2. In B.3 sono descritti alcuni codici di supporto per l’analisi delle simulazioni, esistenti o programmati nell’ambito di questo lavoro di tesi. Infine in B.4 ed in B.5 riporto due software utilizzati per l’analisi delle traiettorie.

B.1 Codici per la generazione dell’input di DL_POLY

I file di input richiesti da DL_POLY per eseguire una simulazione sono tre:

1. CONTROL: descrive i parametri generali della simulazione (temperatura, timestep di integrazione e lunghezza totale della simulazione, raggi di taglio per le interazioni a lungo raggio, condizioni al contorno e parametri di controllo per la generazione dei file di output). È un semplice file di testo che può essere scritto a mano.
2. CONFIG: contiene la lista di centri interattivi del modello (Ca e atomi pesanti del sito attivo, in questo caso), le loro posizioni iniziali (ed eventualmente le velocità).
3. FIELD: contiene la descrizione completa del campo di forze, e quindi

- a. Lista completa dei centri interattivi (coerente con il file CONFIG), loro masse ed eventuali cariche elettriche, più altri parametri che consentono eventualmente di mantenere immobile un sottoinsieme di essi.
- b. Lista completa dei legami contenuti nel termine U_{bond} , vedi Cap. 4.1, forma funzionale e lista dei parametri per ciascuno di essi. Nel presente modello anche tutte le interazioni del termine U_{local} vengono trattate come “legami”, anche se con forma funzionale “separabile” e quindi vanno listati in questa sezione.
- c. Lista completa degli angoli di legame nel termine U_{angle} , vedi Cap. 4.1, forma funzionale e lista dei parametri per ciascuno di essi.
- d. Lista completa dei diedri di legame nel termine U_{dihedral} , vedi Cap. 4.1, forma funzionale e lista dei parametri per ciascuno di essi.
- e. Lista delle interazioni $U_{\text{non_local}}$, vedi Cap. 4.1, forma funzionale e lista dei parametri per ciascuna di esse.

I file CONFIG e FIELD, una volta fissato il campo di forze, sono interamente costruibili a partire da una configurazione iniziale del sistema, che corrisponde ad una delle strutture di riferimento delle rodopsine. Per costruirli, è stato programmato il software “prepare.f” in Fortran che prende in input una struttura in formato .pdb e genera CONFIG e FIELD. Il codice “prepare” è diviso in tre parti

1. Lettura file .pdb e selezione dei siti interattivi del modello (Ca e atomi pesanti del sito attivo)
2. Scrittura del file CONFIG
3. Scrittura del file FIELD, suddivisa in
 - i. Generazione della topologia di legame, per generare la lista dei “bonds”, “bond angles” e dihedrals. Questo è relativamente semplice per la parte proteica che è una catena lineare, più complesso per la parte del sito attivo, per il quale si è usato un criterio che riconosce i primi vicini sulla base della distanza.
 - ii. Assegnazione delle forme funzionali e dei parametri del campo di forze. Questa parte è al momento “hard-coded”, il codice deve essere ricompilato per cambiare la parametrizzazione, ma le modifiche necessarie per leggere i parametri da un file esterno sono molto semplici e verranno operate nelle versioni future.
 - iii. Scrittura del file.

Precisamente PREPARE genera i file per i modelli dei vari stati separati, e quindi dà come output FIELD_X, dove X=BR, K,... cioè un FIELD separato per ciascuno stato. I FIELD sono poi composti a due a due da un altro codice Merge_field.f con la seguente struttura

1. Lettura dei due file FIELD dei due stati di partenza: FIELD A e FIELD B.
2. Scrittura della lista di atomi, che è la stessa per tutti e due i FIELD in quanto non vi sono né aggiunte, né perdite di atomi.
3. Per i constraints viene effettuata la media tra le distanze dei due FIELD
4. Per i bonds ad ogni coppia uguale di atomi vengono riscritti i parametri dei due potenziali ed i parametri ϵ e Δ in modo da creare il giusto input per la forma funzionale della doppia buca. Se una coppia di atomi è in uno solo dei due FIELD significa che nell'altro stato la coppia di atomi è slegata, cioè la distanza è maggiore del raggio di cutoff. In questo caso viene unito il

potenziale della coppia di atomi del FIELD dove sono legati con il rispettivo VdW.

5. Per gli angoli e per i diedri vengono riscritti i parametri dei potenziali dei due FIELD con i parametri ϵ e Δ .
6. I parametri successivi, cioè tethering e VdW, vengono riscritti in quanto sono uguali nei due FIELD.

Il merge_field.f prende in input i due FIELD dei due stati di partenza e genera un unico FIELD AB in cui si hanno tutti i potenziali a doppia buca rispettando i parametri dei due FIELD.

Il risultato finale dopo questi due passaggi sono file FIELD_AB, dove A e B sono gli stati iniziale e finale. In linea di principio è possibile generare FIELD multi-stabili con più di due stati di riferimento, con la stessa procedura applicata iterativamente, a patto che le forme funzionali dei singoli termini del campo di forze siano implementate o implementabili in DL_POLY (si veda prossima sezione).

Lo schema dei codici è riportato nel seguente diagramma di flusso. A seguire il tabulato dei codici.

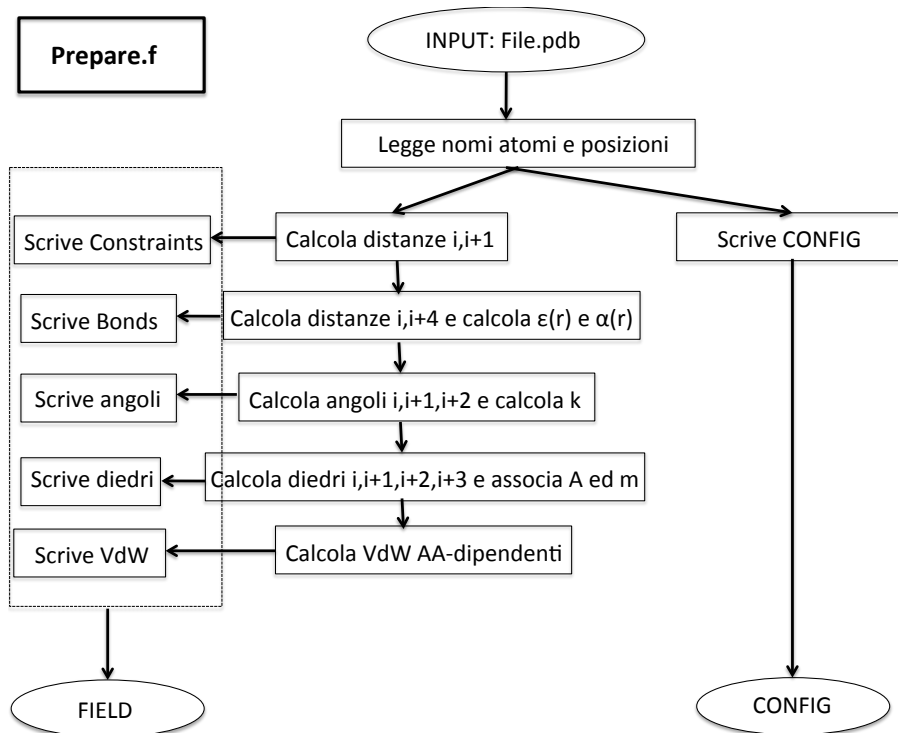


Diagramma 1: diagramma di flusso del file prepare.f. In input prende il file.pdb della struttura desiderata ed in output restituisce il file FIELD ed il file CONFIG.

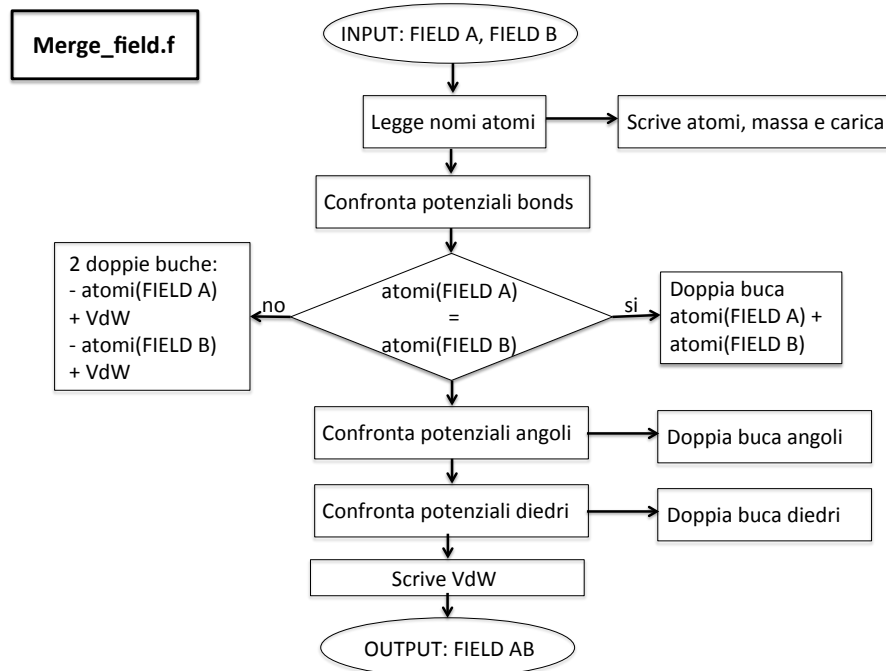


Diagramma 2: diagramma di flusso del Merge_field.f: in input prende due file FIELD e restituisce in output un solo file FIELD AB risultato dalla composizione dei due iniziali.

Di seguito è riportata una parte del codice che utilizzato per creare i file FIELD e CONFIG: nella prima parte c'è la scrittura del file CONFIG secondo la corretta formattazione, nella seconda parte la scelta del parametro k per il potenziale degli angoli al variare del valore dell'angolo.

```

c-----WRITE CONFIG-----

open(10,file='CONFIG',ACCESS='DIRECT',RECL=62,FORM='FORMATTED')
open(12,file='CONFIG.pdb')
write(10,'a58,2a1',REC=1) '1 Bead model GFP - 2 Beads for the chromophore',CHAR(13),CHAR(10)
write(10,'(2I10,a40,2a1)',REC=2) 0,0,stringa_vuota(1:40),CHAR(13),CHAR(10)

j=1
t=0
do i=1,ntot
write(10,110,REC=i+2+t) dlpname(i),j,stringa_vuota(1:42),CHAR(13),CHAR(10)
write(10,120,REC=i+3+t) (pos(k,i),k=1,3),CHAR(13),CHAR(10)
j=j+1

write(12,100) 'ATOM ',atom_count(i),atom_name(i),resname(i),chain(i),nres(i),(pos(k,i),k=1,3)

if(tb(i).eq.1) then

write(10,110,REC=i+4+t) dlpnameb(i),j,stringa_vuota(1:42),CHAR(13),CHAR(10)
write(10,120,REC=i+5+t) (posb(k,i),k=1,3),CHAR(13),CHAR(10)
t=t+2
j=j+1

write(12,100) 'ATOM ',atom_countb(i),atom_nameb(i),resnameb(i),chainb(i),nresb(i),(posb(k,i),k=1,3)

endif

t=t+1
enddo

close(10)

c END CONFIG FILE
c-----

```



```

...
c differenziazione a seconda dell'angolo

```

```

    if(theta(i).le.97)then
      kappatheta=110
    elseif(theta(i).le.115)then
      kappatheta=20
    elseif(theta(i).le.130)then
      kappatheta=70
    elseif(theta(i).le.150)then
      kappatheta=90
    else
      kappatheta=20
    endif

```

Riporto inoltre una parte del file `merge_field.f` dove unisco i potenziali Morse con la doppia buca.

```

c write morse
c-----
write(12,1010),ctemp8,nmor
do i=1,nmor
  if((im11(i).eq.im21(i)).and.(im12(i).eq.im22(i))) then
    if(abs(r01(i)-r02(i)).le.0)then
      write(12,151)morse,im1(i),im2(i),e(i),r0(i),a(i)
    else

      write(12,152)'dblw',im1(i),im2(i),e1(i),r01(i),
$      a1(i),e2(i),r02(i),a2(i),0.1,0.0
      endif

    else
      write(12,151)morse,im1(i),im2(i),e(i),r0(i),a(i)
    endif

  enddo

```

B.2 Modifiche operate a DL_POLY

Il modello bistabile necessita di forme funzionali a doppia buca, per tutti i termini del potenziale (bond, bond angle, dihedral e parte non bonded). Alcune di queste (per bond e bond angle) erano già implementate in DL_POLY, ma in una forma analitica poco flessibile. Pertanto si è preferito implementare per tutti i termini la forma generale

$$f(x) = \frac{1}{2} (f_1(x) + f_2(x) - \Delta) - \frac{1}{2} \sqrt{(f_1(x) - f_2(x) + \Delta)^2 + \varepsilon^2} \quad (\text{C. 1})$$

dove f è un potenziale armonico o Morse se x è la distanza tra due atomi, il coseno armonico se x è l'angolo θ , il coseno se x è il diedro ϕ (vedere fig B.2)

Questo ha necessitato la modifica dei moduli di DL_POLY che calcolano la parte di energia potenziale e le forze dei rispettivi termini del campo, e specificamente

- `Angles_module.f`
- `Bonds_module.f`
- `Dihedrals_module.f`
- `Setup_module.f`

La modifica implica l'aggiunta della nuova forma funzionale che viene poi richiamata nell'input (file FIELD) con una nuova apposita keyword. L'eseguibile

creato con queste modifiche conserva anche tutte le funzionalità precedenti, ed è una sua estensione.

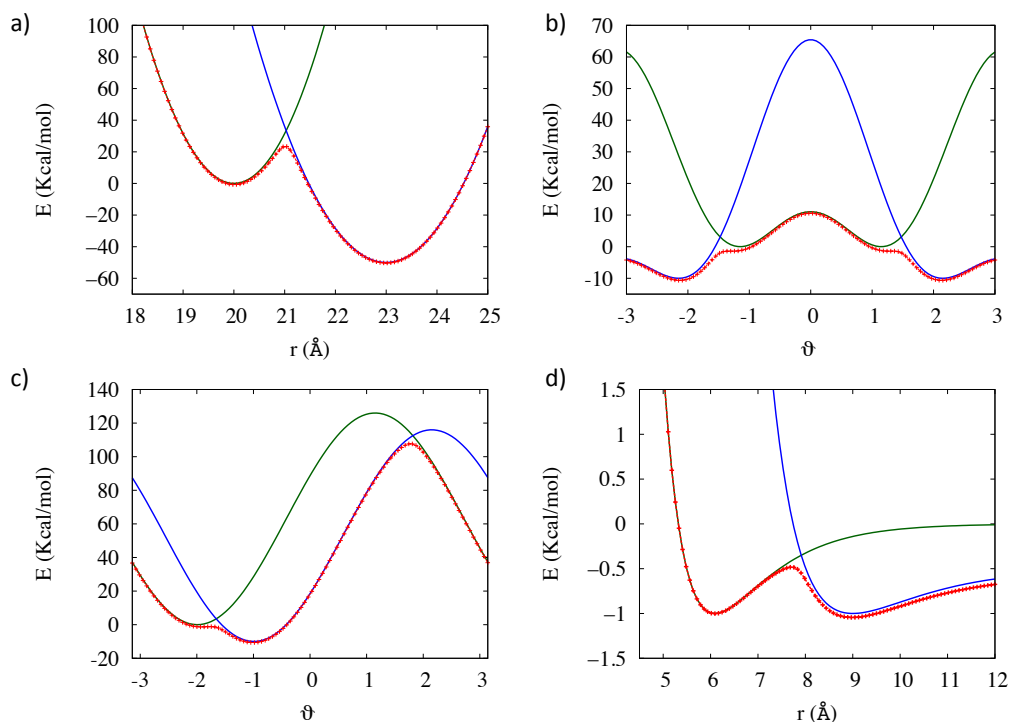


Figura B 2: grafico del potenziale a doppia buca, in rosso tratteggiato, composto da due potenziali di partenza, in verde ed in blu. Nei quattro grafici i potenziali hanno larghezze ed altezze relative diverse. In a) due potenziali armonici, in b) due coseni armonici, in c) due coseni ed in d) due Morse.

I tabulati dei moduli modificati sono riportati qui di seguito **angles_module.f**

```
f1=0.5d0*prmang(kk,1)*(theta-prmang(kk,2))**2
f2=0.5d0*prmang(kk,3)*(theta-prmang(kk,4))**2-prmang(kk,6)
d1=prmang(kk,1)*(theta-prmang(kk,2))/sint
d2=prmang(kk,3)*(theta-prmang(kk,4))/sint

pterm=0.5*((f1+f2)-
x   sqrt((f1-f2)**2+
x   prmang(kk,5)**2))

gamma=0.5*(d1+d2)-
x   0.5*((f1-f2)*
x   (d1-d2)/
x   sqrt((f1-f2)**2+prmang(kk,5)**2))

bonds_module.f

f1=prmbnd(kk,1)*((1-exp(-prmbnd(kk,3)*(rab-prmbnd(kk,2))))**2-1)
f2=prmbnd(kk,4)*((1-exp(-prmbnd(kk,6)*(rab-prmbnd(kk,5))))**2-1)
d1=2.d0*prmbnd(kk,1)*prmbnd(kk,3)*(1.d0-
x   exp(-prmbnd(kk,3)*(rab-prmbnd(kk,2))))*
x   exp(-prmbnd(kk,3)*(rab-prmbnd(kk,2))))*rrab
d2=2.d0*prmbnd(kk,4)*prmbnd(kk,6)*(1.d0-
x   exp(-prmbnd(kk,6)*(rab-prmbnd(kk,5))))*
x   exp(-prmbnd(kk,6)*(rab-prmbnd(kk,5))))*rrab

omega=0.5*((f1+(f2-prmbnd(kk,6)))-
x   sqrt((f1-(f2-prmbnd(kk,6)))**2+
x   prmbnd(kk,5)**2))

gamma=0.5*(d1+d2)-
x   0.5*((f1-f2+prmbnd(kk,6))*
```

```

x      (d1-d2)/
x      sqrt((f1-f2+prmbnd(kk,6))**2+prmbnd(kk,5)**2)

      dihedrals_module.f

      f1=prmdih(kk,1)*(1.d0+cos(prmdih(kk,3)*phi-
x      prmdih(kk,2)))
      f2=prmdih(kk,4)*(1.d0+cos(prmdih(kk,6)*phi-
x      prmdih(kk,5)))-prmdih(kk,8)

      d1=-rpb1*rpc1*rsinp*prmdih(kk,1)*prmdih(kk,3)*
x      sin(prmdih(kk,3)*phi-prmdih(kk,2))
      d2=-rpb1*rpc1*rsinp*prmdih(kk,4)*prmdih(kk,6)*
x      sin(prmdih(kk,6)*phi-prmdih(kk,5))

      dterm=0.5*(f1+f2)-0.5*sqrt((f1-f2)**2+prmdih(kk,7)**2)

      gamma=(0.5*(d1+d2)-0.5*(f1-f2)*(d1-d2)/
x      sqrt((f1-f2)**2+prmdih(kk,7)**2))

      setup_module.f

c-----Modificato Francesco 01/11/2012 per la doppiabuca dei diedri-----
c      ho aumentato il massimo numero di parametri che legge da 5 a 8

      mxpdih=8
c      mxpdih=5

```

B.3. Analisi dell'output

Il software DL_POLY restituisce in output dei file dai quali si può ottenere la traiettoria e l'energetica della simulazione. I file sono tre:

1. OUTPUT: è un file che contiene i dati della simulazione come la temperatura, il numero di passi, il timestep, un campione della configurazione di partenza degli atomi, la replica del file FIELD ed una serie di parametri che riguardano la simulazione in corso. La riscrittura del FIELD è importante in quanto fornisce la prova della corretta lettura di tutti i parametri da parte del software.
2. HISTORY: file che contiene per tutti gli atomi del modello per ciascun passo della simulazione i seguenti parametri: la posizione, la velocità e la forza nelle tre coordinate x, y e z.
3. STATIS: contiene per ciascun passo della simulazione i valori della temperatura e di tutti i termini delle energie dell'intero sistema.

Il file HISTORY può essere letto con il software VMD (Visual Molecular Dynamics⁷⁹), ma la visualizzazione della traiettoria della simulazione non è chiara. Per renderla più comprensibile si deve riformattare il file HISTORY tramite un programma scritto in Fortran: HIS_2traj.f. Questo programma converte il file HISTORY nel file hystory.pdb formattato secondo i criteri della struttura del sito PDB lasciando inalterate le informazioni su posizioni, velocità e forze degli atomi.

Il file history.pdb ottenuto dal file HISTORY consente anche l'analisi della RMSF di ciascun atomo tramite il programma HIS_2rmsf.f. Questo si calcola per ogni atomo la RMSF leggendo le posizioni ad ogni timestep per ciascun atomo e applicando la formula

$$\text{RMSF} = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_i(t_i) - \tilde{x}_i)^2} \quad (\text{C. 2})$$

dove T è il tempo totale della simulazione, $x_i(t_i)$ sono le posizioni di ogni atomo per ciascun timestep ed \tilde{x}_i è la posizione di riferimento dell'atomo i-esimo.

Questo programma restituisce in output il file rmsf.txt dove sono presenti due colonne: la prima indica il numero dell'atomo e la seconda la RMSF di quell'atomo. Da quest'ultimo file ottenuto posso ricavare il valore del B-factor della simulazione applicando la relazione tra RMSF e B-factor, vedi Eq. 2.1.

Per fare ciò si è programmato il codice accessorio B-factor.f che legge il file rmsf.txt e crea il file Bfactor.txt analogo al file rmsf.txt dove al posto del valore della RMSF per ciascun atomo vi è il valore del rispettivo B-factor.

Il file STATIS consente l'analisi delle energie e delle temperature delle simulazioni. Questo file, per essere letto dai più comuni file di analisi dati deve essere convertito in un altro file contenente soltanto le informazioni che vogliamo. Per fare ciò si è utilizzato un programma, anch'esso scritto in Fortran, che legge il file STATIS e riscrive nel file ene.dat i seguenti valori: timestep, energia totale, temperatura, energia potenziale, energia VdW, energia elettrostatica, energia chimica, energia angoli ed energia diedri. Questo nuovo file contiene soltanto le informazioni necessarie all'analisi delle energie della simulazione per questo tipo di studio ed è in un formato più facilmente comprensibile.

B.4. VMD

Per visualizzare le traiettorie delle simulazioni, per poter osservare la struttura delle rodopsine e per fare alcuni calcoli utili a questo studio ho utilizzato un software per la visualizzazione della dinamica molecolare di nome VMD (Visual Molecular Dynamics)⁸⁰. Questo software è disegnato per modellizzare, analizzare e visualizzare sistemi biologici come proteine, acidi nucleici etc. ed è compatibile con i file provenienti dal sito PDB. Tramite VMD ho effettuato i calcoli della RMSF e della RMSD delle simulazioni, le immagini esplicative delle rodopsine e l'analisi delle PCA.

B.5. Analisi dei modi/componenti principali.

L'analisi delle componenti principali (PCA) di una proteina fornisce i modi essenziali del sistema che sono quelli associati alla funzionalità della proteina. Per questo studio ho utilizzato un plug-in di VMD che si chiama ProDy⁸¹ ed è stato creato per l'analisi della dinamica delle proteine. ProDy calcola sulla traiettoria della simulazione la matrice di covarianza data dalla formula

$$\sigma_{ij} = \frac{1}{n} \sum_{h=1}^n (x_{hi} - \mu_j)^2 \quad (C. 3)$$

dove x_{hi} è la posizione dell'atomo i -esimo al tempo h , μ_j è la media delle posizioni e n è il numero di passi della simulazione. Da questa il programma estrae gli autovalori e gli autovettori associati e li scrive in due file separati. Vengono inoltre creati i file contenente le prime 10 componenti principali in formato .nmd ed uno contenente le proiezioni degli autovettori. Da quest'ultimo file si ricava, tramite un software grafico, come Gnuplot⁸² od Octave⁸³, la superficie di energia libera di ciascuna componente principale. Nel software VMD, inoltre, vi è il plug-in NMW (Normal Mode Wizard) che permette di osservare le componenti principali caricando il file .nmd della traiettoria. Tramite l'interfaccia grafica di

NMW si possono selezionare le singole componenti principali visualizzate sia tramite frecce che tramite diverse colorazioni della proteina.

Appendice C. Dettagli sulla descrizione del potenziale modello multistabile

Riporto in Tabella 2 i numeri di doppie buche utilizzate per ciascun termine del potenziale ed i diedri tra i quali ho distribuito l'energia relativa tra i due stati per la creazione dei FIELD intermedi.

Transizione	Bonds ($\Delta r=10\%$)	Angles ($\Delta \theta=0.5\%$)	Dihedrals ($\Delta \phi=0.5\%$)	Totale	Distibuzione energia
K-L	133	212	193	538	• C12-C13=C14-C15
L-M	155	225	191	571	• C12-C13=C14-C15 • C14-C15=N(lisina)-C(lisina)
M-N	280	223	205	708	• C12-C13=C14-C15
N-O	335	244	225	804	• C12-C13=C14-C15 • C14-C15=N(lisina)-C(lisina) • C15=N(lisina)-C(lisina)=C(lisina)

Tabella 4: Numero di potenziali a doppia buca per ciascun termine del potenziale. I valori tra parentesi indicano la variazione percentuale minima oltre la quale utilizzo il potenziale a doppia buca. L'ultima colonna rappresenta i legami (diedri) tra i quali ho distribuito l'energia per ciascuna transizione.

Le energie sono state distribuite soltanto sui diedri interessati dal cambiamento conformazionale tra i due stati e sono diversi per ciascuna transizione.

Nelle figure dalla C1 alla C4 sono riportate le energie delle varie simulazioni rispetto alla coordinata di reazione per le transizioni dallo stato K fino allo stato O. Da queste si può notare come la barriera di potenziale tra due stati aumenti andando avanti nel fotociclo. Questo fatto è dato dall'aumento dei tempi di transizione tra i vari stati, come mostrato in Fig.12.

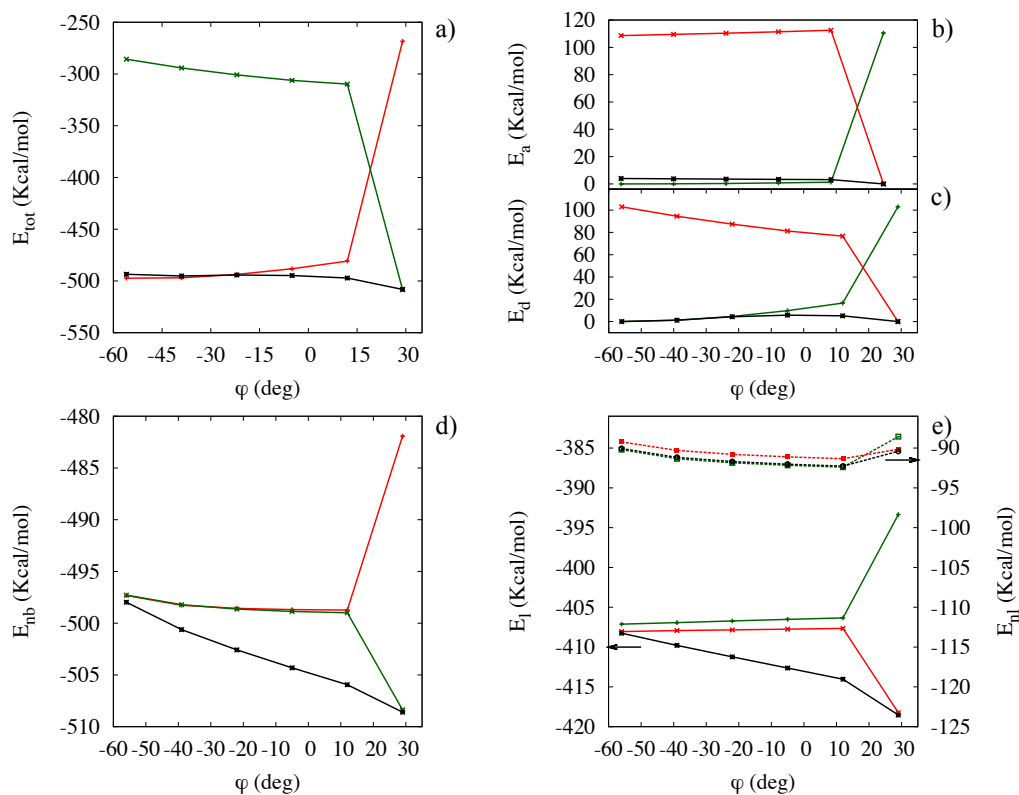


Figura C1: Profili di energia per la transizione K-L. In rosso, profili misurati sul campo FIELD_A, in verde su FIELD_B e in nero sul FIELD.merged. Per questo calcolo i Δ dei termini bond angle e dihedral (pannelli b e c) sono imposti nulli, mentre per i termini locali e misti locali non locali (pannello e, linee continue scala sulla sinistra e tratteggiate-scala sulla destra rispettivamente), come spiegato nel testo, portano una differenza di energia che dipende da come le distanze di equilibrio variano da stato a stato. Questa, in totale sommata sulla parte non bonded locale e non locale risulta in questo caso di circa 10 kcal/mole in favore dello stato L (pannello d).

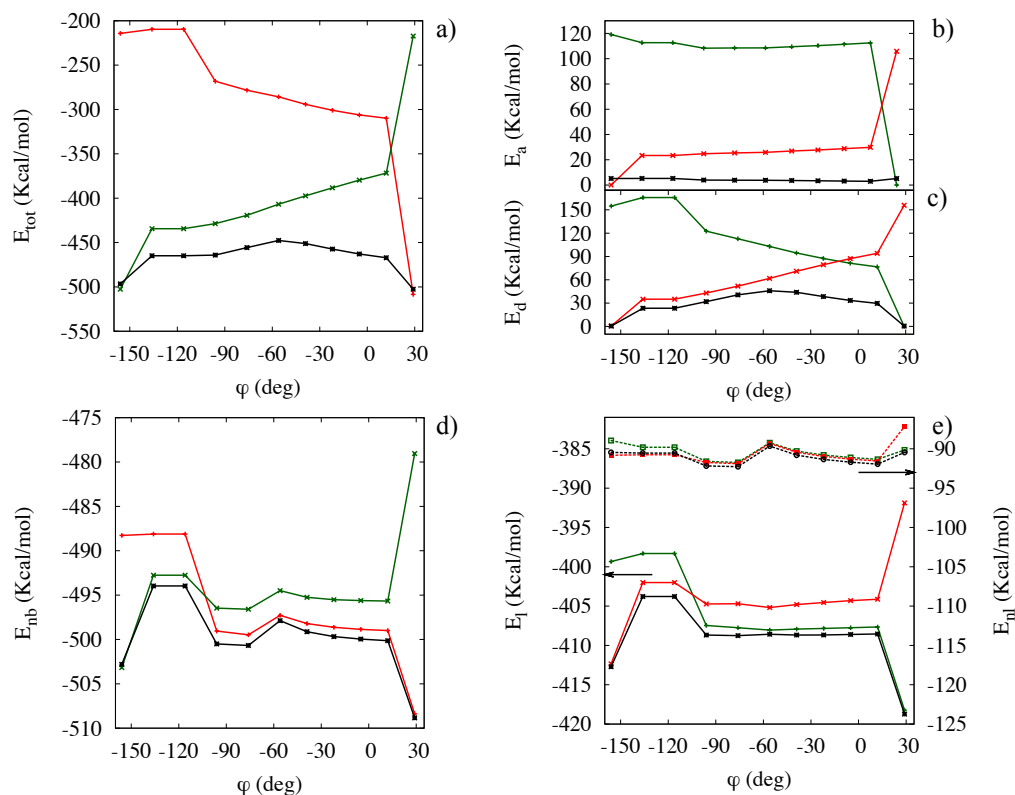


Figura C2: Profili di energia per la transizione L-M. In rosso, profili misurati sul campo FIELD_A, in verde su FIELD_B e in nero sul FIELD.merged. Per questo calcolo i Δ dei termini bond angle e dihedral (pannelli b e c) sono imposti nulli, mentre per i termini locali e misti locali non locali (pannello e, linee continue scala sulla sinistra e tratteggiate-scala sulla destra rispettivamente), come spiegato nel testo, portano una differenza di energia che dipende da come le distanze di equilibrio variano da stato a stato. Questa, in totale sommata sulla parte non bonded locale e non locale risulta in questo caso di circa 5-6 kcal/mole in favore dello stato L (pannello d).

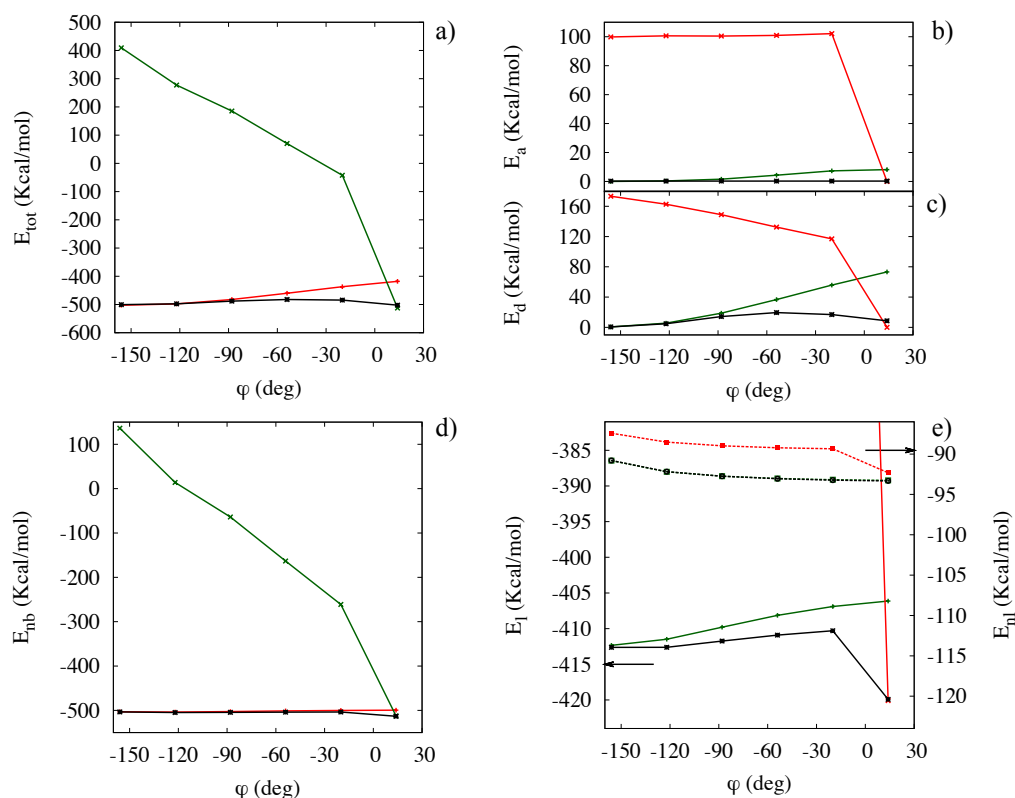


Figura C3: Profili di energia per la transizione M-N. In rosso, profili misurati sul campo FIELD_A, in verde su FIELD_B e in nero sul FIELD.merged. Per questo calcolo i Δ dei termini bond angle e dihedral (pannelli b e c) sono imposti nulli, mentre per i termini locali e misti locali non locali (pannello e, linee continue scala sulla sinistra e tratteggiate-scala sulla destra rispettivamente), come spiegato nel testo, portano una differenza di energia che dipende da come le distanze di equilibrio variano da stato a stato. Questa, in totale sommata sulla parte non bonded locale e non locale risulta in questo caso di circa 7 kcal/mole in favore dello stato N (pannello d).

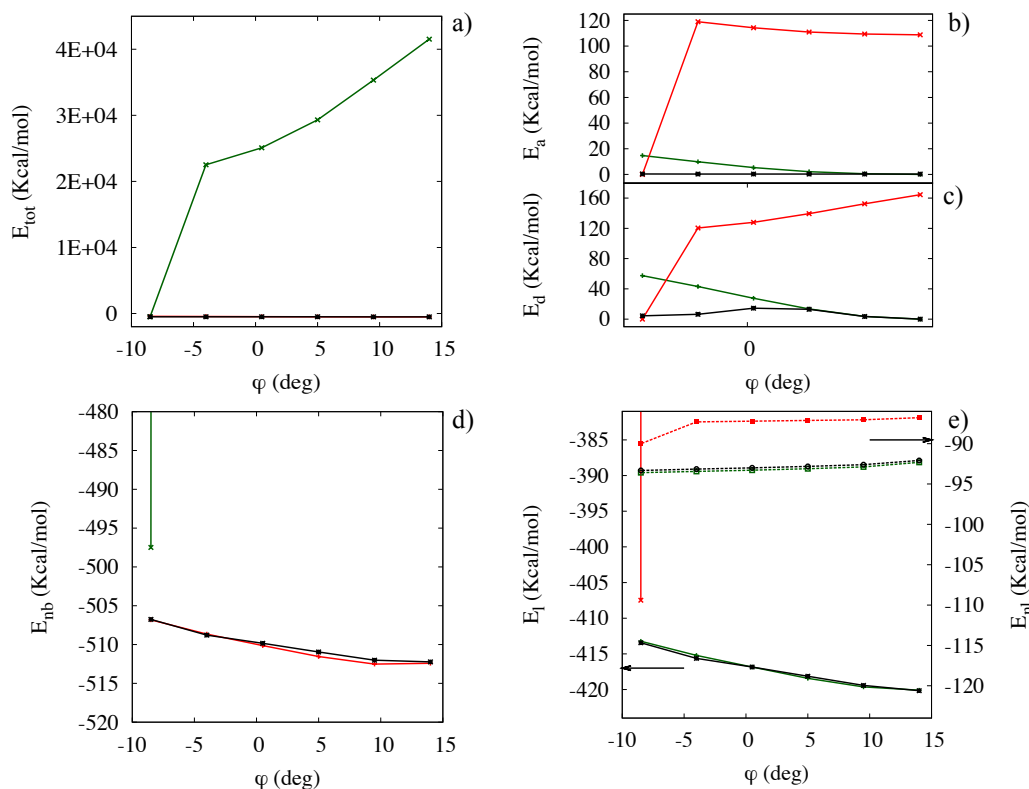


Figura C4: Profili di energia per la transizione N-O. In rosso, profili misurati sul campo FIELD_A, in verde su FIELD_B e in nero sul FIELD.merged. Per questo calcolo i Δ dei termini bond angle e dihedral (pannelli b e c) sono imposti nulli, mentre per i termini locali e misti locali e non locali (pannello e, linee continue scala sulla sinistra e tratteggiate-scala sulla destra rispettivamente), come spiegato nel testo, portano una differenza di energia che dipende da come le distanze di equilibrio variano da stato a stato. Questa, in totale sommata sulla parte non bonded locale e non locale risulta in questo caso di circa 5 kcal/mole in favore dello stato N (pannello d).

Tramite i FIELD utilizzati per il precedente studio ho settato per ciascuno stato le energie relative e le altezze delle barriere modificando i parametri ϵ e Δ . Unendo i dati ottenuti per ciascuna transizione sono riuscito ad ottenere il profilo di energia per tutti gli stati del fotociclo e per le barriere di potenziale. Confrontando questi dati con quelli ottenuti dai dati sperimentali in Tabella 1 osservo che l'accordo è molto buono e che riesco a riprodurre bene tutte le energie in gioco.

Bibliografia

-
- ¹ W. Gilbert (1986). The RNA World. *Nature* 319, 618.
- ² N.C. Berchtold, C.W. Cotman (1997). Evolution in the Conceptualization of Dementia and Alzheimer's Disease: Greco-Roman Period to the 1960s. *Neurobiology of Aging*, 19, 173-189.
- ³ JW Wilesmith, GA Wells, MP Cranwell and JB Ryan (1987). Bovine spongiform encephalopathy: epidemiological studies, *Veterinary Record*, 123, 638-644
- ⁴ M. Kolbe, H. B.-O. (2000). Structure of a Light-Driven Chloride Pump at 1.8 Å Resolution. *Science* , 288, 1390-1396.
- ⁵ Essen L., S. R. (1998). Lipid patches in membrane protein oligomers: crystal structure of the bacteriorhodopsin-lipid complex. *Proc. Nat. Acad. Sci. USA* , 77, 338-342.
- Hinsen. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins* , 33, 417.
- ⁶ Morris, M. B. (2009). Rhodopsin: structure, signal transduction and oligomerisation. *The international journal of biochemistry & cell biology* , 41, 721-724.
- ⁷ Luecke, H. S.-t.-p. (1999). Structure of Bacteriorhodopsin at 1 . 55 Å. *Journal of Molecular Biology* , 291, 899-911.
- ⁸ Huber, T. B. (2004). Membrane Model for the G-Protein-Coupled Receptor Rhodopsin : Hydrophobic Interface and Dynamical Structure. 86 , 2078-2100.
- ⁹ Miyano, M. (2000). Atomic structure of bovine rhodopsin: a seven transmembrane receptor. *Science* , 7-9.
- ¹⁰ J. C. Kendrew et al. (1958). "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis". *Nature* 181, 662–666.
- ¹¹ David C. Teller, Tetsuji Okada, Craig A. Behnke, K. P. and R. E. S. (2006). Advances in Determination of a High-Resolution Three- Dimensional Structure of Rhodopsin, a Model of G-Protein- Coupled Receptors (GPCRs). *Biochemistry* 40, 7761–7772.
- ¹² http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1988/michel.html
- ¹³ Subramaniam, S. H. (2000). Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature* , 406, 653-657.
- ¹⁴ Krebs, A. E. (2003). The three-dimensional structure of bovine rhodopsin determined by electron cryomicroscopy. *The Journal of biological chemistry* , 278, 50217-50225.
- ¹⁵ Salomon D., L. D. (2006). Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *PNAS* , 103, 16123-16128.
- ¹⁶ Kpegba, K., Murtha, M. & Nesnas, N. (2006). Cyclodextrin retinylidene: a biomimetic kinetic trap model for rhodopsin. *Bioorganic & medicinal chemistry letters* 16, 1523–1526.
- ¹⁷ D. Oesterhelt and L. Schuhmann (1974). Reconstitution of bacteriorhodopsin. *FEBS Letters*, 44, 262-265.
- ¹⁸ Cooper, A., Dixon, S. F., Nutley, M. A. & Robb, J. L. (1987). Mechanism of Retinal Schiff Base Formation and Hydrolysis in Relation to Visual Pigment

Photolysis and Regeneration: Resonance Raman Spectroscopy of a Tetrahedral Carbinolamine Intermediate and Oxygen- 1 8 Labeling of Retinal at the Metarhodopsin Stage in. *Journal of the American Chemical Society* 109, 1254–1263.

¹⁹ Bayburt T. H., L. A. (2007). Transducin activation by nanoscale lipid bilayers containing one and two rhodopsins. *Journal of Biological Chemistry* , 282, 14875-14881.

²⁰ Ernst O. P., G. V. (2007). Monomeric G protein-coupled receptor rhodopsin in solution activates its G protein transducin at the diffusion limit. . *Proc. Natl. Acad. Sci USA* , 104, 10859-10864.

²¹ Kota P, R. P. (2006). Opsin is present as dimers in COS1 cells: identification of amino acids at the dimeric interface. *Proc. Natl. Acad. Sci. USA* , 103, 3054-3059.

²² Poland H. j., F. M. (1986). Early picoseconds events in the photocycle of bacteriorhodopsin. *Biophysical journal* , 49, 651-662.

²³ Hermone, A. & Kuczera, K. (2010). Free-Energy Simulations of the Retinal Cis f Trans Isomerization in. *Biochemistry*, 2960, 2843–2853.

²⁴ Lozier, R. H., Xie, a, Hofrichter, J. & Clore, G. M. (1992). Reversible steps in the bacteriorhodopsin photocycle. *Proceedings of the National Academy of Sciences of the United States of America* 89, 3610–3614.

²⁵ Onufriev, A., Smondyrev, A. & Bashford, D. (2003). Proton Affinity Changes Driving Unidirectional Proton Transport in the Bacteriorhodopsin Photocycle. *Journal of Molecular Biology*, 332, 1183–1193.

²⁶ Edman, K. N.-P. (1999). High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature* , 401, 822-826.

²⁷ Lanyi J. K. (2006). Proton transfers in bacteriorhodopsin photocycle. *Biochimica et Biophysica Acta* , 1757, 1012-1018.

²⁸ Hermone, A. & Kuczera, K. (2010). Free-Energy Simulations of the Retinal Cis f Trans Isomerization in. *Biochemistry*, 2960, 2843–2853.

²⁹ Birge, R. R., Cooper, T. M., Lawrence, A. F., Masthay, M. B. & Zhang, C. (1991). Revised assignment of energy storage in the primary photochemical event in bacteriorhodopsin. *Journal of the American Chemical Society* 113, 4327–4328.

³⁰ Ludmann, K., Gergely, C. & Váró, G. (1998). Kinetic and thermodynamic study of the bacteriorhodopsin photocycle over a wide pH range. *Biophysical journal* 75, 3110–3119.

³¹ Onufriev, A., Smondyrev, A. & Bashford, D. (2003). Proton Affinity Changes Driving Unidirectional Proton Transport in the Bacteriorhodopsin Photocycle. *Journal of Molecular Biology* 332, 1183–1193.

³² Bolognani. (1990). *Compendio di bioenergetica*. Piccin Editore.

³³ Tozzini, V. (2010). Multiscale modeling of proteins. *Accounts of chemical research* , 43, 220-230.

³⁴ Frenkel, D. and Smit, B. (2001). *Understanding Molecular Simulation*. Academic Press

³⁵ Chang, C.-E., Shen, T., Trylska, J., Tozzini, V. & McCammon, J. A. (2006). Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse-grained model. *Biophysical journal* 90, 3880–3885.

-
- ³⁶Lindorff-Larsen, K. et al. (2012). Systematic validation of protein force fields against experimental data. *PloS one* 7, 1–6.
- ³⁷Tozzini V. (2005). Coarse-grained models for proteins. *Current Opinion in Structural Biology*, 15:144–150.
- ³⁸Trovato F. and Tozzini V. (2012). Minimalist models for biopolymers: Open problems, latest advances and perspectives. *AIP Conference Proceedings* 1456, 187.
- ³⁹H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81, 3684-3690.
- ⁴⁰P.H. Huñenberger (2005), Thermostat algorithms for molecular dynamics simulations, *Adv. Polymer. Sci.*, 173, 105-149.
- ⁴¹Nose, S (1984). A unified formulation of the constant temperature molecular-dynamics methods. *Journal of chemical physics* 81, 511–519.
- ⁴²Andersen, H. C. (1980) Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72, 2384-2393.
- ⁴³Langevin, P. (1908). On the Theory of Brownian Motion. *C. R. Acad. Sci.* 146: 530–533.
- ⁴⁴Ermak, D. L., and J. A. McCammon. (1978). Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.*, 69,1352–1360.
- ⁴⁵R. Salomon-Ferrer, D.A. Case, R.C.Walker. (2013). An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* 3, 198-210.
- ⁴⁶Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4, 187–217.
- ⁴⁷Oostenbrink, C., Villa, A., Mark, A. E. and Van Gunsteren, W. F. (2004), A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25, 1656–1676.
- ⁴⁸Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E., Berendsen H. J. (2005). "GROMACS: fast, flexible, and free". *Journal Computational Chemistry*, 26, 1701–1718.
- ⁴⁹Frontera A., Damm W., Jorgensen W. L., Maxwell D. S. and Tirado-Rives J. (1997). OPLS All-Atom Force Field for Carbohydrates. *Journal of Computational Chemistry*, 18, 1955-1970.
- ⁵⁰Weiner, S. J., Kollman, P. a., Nguyen, D. T. & Case, D. a. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry* 7, 230–252.
- ⁵¹Tirion, M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical review letters* , 77, 1905-1908.
- ⁵²Hinsen. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins* , 33, 417.
- ⁵³Haliloglu, T., Bahar, I. & Erman, B. (1997). Gaussian Dynamics of Folded Proteins. *Physical Review Letters* 79, 3090–3093.
- ⁵⁴Atilgan, A. R. (2001). Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal* , 80, 505-515.
- ⁵⁵Doruker P, Atilgan AR, Bahar I. (2000). Dynamics of proteins predicted by

molecular dynamics simulations and analytical approaches: Application to α -amylase inhibitor. *Proteins*, 40,512-524.

⁵⁶ D. M. F. Van Aalten, B. L. De Groot, J. B. C. Findlay, H. J. C. Berendsen, and A. Amadei. (1997). Comparison of Techniques for Calculating Protein Essential Dynamics. *Journal of Computational Chemistry*, 18, 169–181.

⁵⁷ Hamacher, K. and McCammon, J. A. (2006). Computing the amino acid specificity of fluctuations in biomolecular systems. *J. Chem. Theory Comput.* 2, 873–878.

⁵⁸ Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and ‘en-route’ intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298, 937–953.

⁵⁹ Jeong, J. I., Jang, Y. & Kim, M. K. (2005). A connection rule for α -carbon coarse-grained elastic network models using chemical bond information. *J. Mol. Graph Model*, 24, 296–306.

⁶⁰ Zheng, W., Brooks, B. R. & Hummer, G. (2007). Protein conformational transitions explored by mixed elastic network models. *Proteins* 69, 43–57.

⁶¹ Chu, J.-W. V. (2007). Coarse-grained free energy functions for studying protein conformational changes: a double-well network model. *Biophysical journal*, 93, 3860-3871.

⁶² Maragakis, P. K. (2005). Large Amplitude Conformational Change in Proteins Explored with a Plastic Network Model : Adenylate Kinase. *Journal of Molecular Biology*, 352, 807-822.

⁶³ Tekpinar, M. Z. (2010). Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. *Proteins*, 78, 2469-2481.

⁶⁴ Gō N. and Scheraga H. A. (1976). On the Use of Classical Statistical Mechanics in the Treatment of Polymer Chain Conformation. *Macromolecules*, 9, 535–542.

⁶⁵ Tozzini, V; McCammon, J. A. (2011). A coarse grained model for the dynamics of flap opening in HIV-1 protease. *Chemical Physics Letters*, 413, 123-128.

⁶⁶ Di Fenza, A., Rocchia, W. & Tozzini, V. (2009). Complexes of HIV-1 integrase with HAT proteins: Multiscale models, dynamics, and hypotheses on allosteric sites of inhibition. *Proteins* 76, 946–958.

⁶⁷ "The DL_POLY Classic User Manual", by W. Smith, T.R. Forester and I.T.Todorov, which is published by Daresbury Laboratory, United Kingdom

⁶⁸ Yong C. W. (2011). "The DL_FIELD User Manual".STFC, Daresbury Laboratory.

⁶⁹ Jang, H., Crozier, P. S., Stevens, M. J. & Woolf, T. B. (2004). How environment supports a state: molecular dynamics simulations of two states in bacteriorhodopsin suggest lipid and water compensation. *Biophysical Journal*, 87, 129–145.

⁷⁰ Schobert, B., Cupp-Vickery, J., Hornak, V., Smith, S. O. & Lanyi, J. K. Crystallographic Structure of the K Intermediate of Bacteriorhodopsin : Conservation of Free Energy after Photoisomerization of the Retinal. *Journal of Molecular Biology* 2836, 715–726 (2002).

-
- ⁷¹ Edman, K. et al. Deformation of helix C in the low temperature L-intermediate of bacteriorhodopsin. *The Journal of biological chemistry* 279, 2147–58 (2004).
- ⁷² Schobert, B., Brown, L. S. & Lanyi, J. K. Crystallographic Structures of the M and N Intermediates of Bacteriorhodopsin: Assembly of a Hydrogen-bonded Chain of Water Molecules Between Asp-96 and the Retinal Schiff Base. *Journal of Molecular Biology* 330, 553–570 (2003).
- ⁷³ Chen, D. & Lanyi, J. K. Structural changes in the N and N' states of the bacteriorhodopsin photocycle. *Biophysical journal* 96, 2779–88 (2009).
- ⁷⁴ Zhang, J. et al. Crystal structure of the O intermediate of the Leu93→Ala mutant of bacteriorhodopsin. *Proteins* 80, 2384–96 (2012).
- ⁷⁵ <http://www.rcsb.org/pdb/home/home.do>
- ⁷⁶ Rouhani, S. et al.(2001). Crystal structure of the D85S mutant of bacteriorhodopsin: model of an O-like photocycle intermediate. *Journal of molecular biology* 313, 615–28.
- ⁷⁷ Okumura, H., Murakami, M. & Kouyama, T. (2005). Crystal structures of acid blue and alkaline purple forms of bacteriorhodopsin. *Journal of molecular biology*, 351, 481–95.
- ⁷⁸ <http://www.stfc.ac.uk/>
- ⁷⁹ www.ks.uiuc.edu/Research/vmd/
- ⁸⁰ Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", *J. Molec. Graphics*, 1996, vol. 14, pp. 33-38.
- ⁸¹ Bakan A, Meireles LM, Bahar I. (2011). ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*, 27, 1575-1577.
- ⁸² Williams T. and Kelley C. (2012) "Gnuplot 4.6. An Interactive Plotting Program".
- ⁸³ Murphy M. (1997). Octave: A Free, High-Level Language for Mathematics. *Linux Journal*.