

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Matematica

**Sui modelli topologici
per la biochimica**

Tesi di Laurea in Topologia Algebrica

Relatore:
Prof.ssa Alessia Cattabriga
Correlatore:
Prof.ssa Maria Carla Tesi

Presentata da:
Alice Farneti

Anno Accademico 2016/2017

Introduzione

Overview

La *teoria dei nodi* è lo studio degli aspetti geometrici legati alla forma di un nodo. Nodi e trecce sono stati uno strumento prezioso per l'esistenza e lo sviluppo dell'uomo fin dalla sua comparsa sulla Terra. Basti pensare alla costruzione delle prime armi e utensili in epoca preistorica, come ad esempio asce formate da una pietra legata ad un bastone, reti da pesca ecc. Nonostante gli uomini abbiano fatto uso dei nodi fin dall'alba della loro esistenza, l'attuale studio matematico di questi oggetti è relativamente recente, più vicino ad avere 100 anni che 1000. Pare che uno dei primi matematici a mostrare interesse per questo argomento sia stato Gauss (1777-1855), ma ancor di più il suo allievo Listing al quale è stato intitolato il "nodo di Listing" o nodo ad otto (Figura 1).

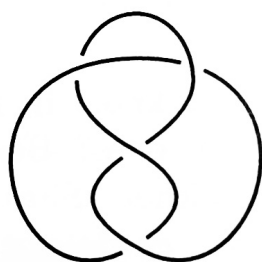


Figura 1: Nodo di Listing.

In seguito il matematico J.W. Alexander (1888-1971) scoprì che la teoria dei nodi è fondamentale per lo studio della topologia tridimensionale, fatto

sottolineato anche dai lavori di H. Seifert. Dopo la II Guerra Mondiale la ricerca in questo campo si spostò dall'Europa negli USA ed in Giappone, dove continua tuttora. Ad oggi, non solo la teoria dei nodi si è sviluppata ed è cresciuta per suo proprio diritto, ma si è scoperto che la matematica utilizzata da questa teoria ha applicazioni in varie branche della scienza tra cui la fisica, la biologia molecolare e la chimica.

Nel campo della biochimica, la teoria dei nodi è stata applicata con successo allo studio della struttura tridimensionale di alcune molecole. Tra queste, ci sono due delle molecole più importanti che stanno a fondamento degli organismi viventi: il DNA e le proteine. Queste due macromolecole, insieme all'RNA sono fondamentali per i processi vitali degli organismi. Possiamo riassumere il loro ruolo con le parole di Crick, premio Nobel nel 1962 insieme a Watson e Wilkins per la scoperta della struttura a doppia elica del DNA: *“Il DNA fa l'RNA, l'RNA fa le proteine e le proteine fanno noi”*.

Tali molecole presentano una struttura complessa, a più livelli, e sono coinvolte in un gran numero di processi che ne mutano la struttura e che permettono loro di compiere la propria funzione all'interno dell'organismo. Per questo motivo, nonostante DNA e proteine vengano studiati da molti decenni, ci troviamo solo agli albori della comprensione del loro funzionamento. Infatti le tecnologie e gli strumenti usati in biologia, pur essendo in continuo avanzamento, spesso non bastano. È proprio in questo contesto che ha trovato spazio un nuovo tipo di ricerca, di carattere matematico, che si basa sull'idea di modellizzare la struttura tridimensionale di DNA e proteine mediante nodi e link. Questa modellizzazione ha permesso di trasportare in ambiente matematico un gran numero di problemi biologici e, grazie ad una combinazione tra ragionamento matematico e dati basati sull'evidenza empirica, di addentrarsi in territori inesplorati e giungere a nuove scoperte.

In questa tesi, dopo aver fornito gli strumenti di base della teoria dei nodi e una breve panoramica su struttura e funzioni del DNA e delle proteine, abbiamo voluto analizzare diversi esempi di modellizzazione di tali molecole che fanno uso della teoria dei nodi. Ci siamo resi conto che gli approcci utilizzati

dai ricercatori in questo campo, che può sembrare molto specifico, sono in realtà innumerevoli e molto diversificati, tanto che spesso danno l'impressione di non comunicare tra loro. Il nostro lavoro si propone di dare un'idea di diversi aspetti del panorama attuale riguardante tale argomento e fornire le conoscenze di base a chi, per la prima volta, decidesse di interessarsi a questo ambito. Oltre ad essere uno stimolante campo di applicazione della teoria dei nodi, lo studio della struttura tridimensionale di DNA e proteine è di fondamentale importanza per lo sviluppo tecnologico in campo biomedico oltre che per la realizzazione di nuove terapie nel caso di svariate patologie negli animali e nell'uomo. È ormai appurato, infatti, che la struttura tridimensionale delle molecole quali le proteine e il DNA ne influenza il corretto funzionamento e, recentemente, sono emerse numerose ricerche che evidenziano una stretta correlazione tra mutazioni di tale struttura e malattie specifiche. La topologia e la matematica in generale possono quindi rivelarsi uno strumento prezioso per progredire in tal senso e giungere ad una comprensione sempre più completa dei costituenti fondamentali nel nostro stesso organismo.

Struttura della tesi

Nel Capitolo 1 introdurremo i concetti fondamentali della teoria dei nodi: definiremo i nodi, i link e studieremo il significato di equivalenza tra nodi. Presenteremo il concetto di diagramma e le cosiddette mosse di Reidemeister che ci permetteranno di convertire lo studio dei nodi, intesi come entità tridimensionali, nello studio di loro particolari proiezioni sul piano. Per ridurre ulteriormente la complessità del problema, introdurremo alcuni invarianti numerici ed una relazione fondamentale tra di essi. Presenteremo infine i concetti di tangle e di nodo razionale. In questa trattazione ci limiteremo a presentare i concetti della teoria dei nodi utili alla modellizzazione e allo studio del DNA e di alcune proteine di nostro interesse. Per approfondimenti si vedano [1, 10, 22].

Nel Capitolo 2 introdurremo alcune nozioni di base della biochimica. In particolare parleremo delle proteine, delle loro funzioni e della loro strut-

tura su vari livelli, introdurremo il concetto di folding proteico e vedremo che la comprensione di questo meccanismo assume grande importanza dal punto di vista biologico e medico. In seguito introdurremo la struttura e la funzione del DNA, soffermandoci sull'importanza di quello che viene chiamato supercoiling della molecola di DNA, per poi concentrarci sull'azione di due particolari enzimi che ne modificano la topologia, ovvero gli enzimi topoisomerasi I e II. Per approfondimenti si veda [3, 5, 11, 23]

Nel Capitolo 3 vedremo alcuni esempi pratici di come la teoria dei nodi si possa applicare alla biochimica e, in particolare, allo studio del DNA. Prenderemo come riferimento i testi [1, 22] ed un articolo del 2009 di Dorothy Buck dal titolo *DNA topology* [5] per proporre un modello del DNA circolare prima sotto forma di anello topologico e successivamente di link a due componenti. Dopo di che ne individueremo le proprietà geometriche e topologiche e discuteremo degli effetti delle topoisomerasi (gli enzimi che regolano la topologia della molecola) dal punto di vista matematico. Ci concentreremo poi su un altro processo che coinvolge il DNA e che, come effetto secondario, ne modifica la topologia: la ricombinazione sito specifica. Per studiare questo processo proporrò un modello basato sul concetto di tangle e, per concludere, mostreremo alcuni esempi del suo utilizzo.

Nel Capitolo 4 riporteremo un altro esempio di applicazione della teoria dei nodi, questa volta allo studio di una specifica proteina, la leptina, un ormone che gioca un ruolo chiave nel regolare l'assorbimento e il consumo di energia, il metabolismo e l'ipertensione. A questo scopo analizzeremo l'articolo *The unique cysteine knot regulates the pleiotropic hormone leptin*¹ mettendone in luce prevalentemente gli aspetti di modellizzazione matematica (per approfondimenti si veda [14]).

Nel Capitolo 5 mostreremo che la conformazione annodata è molto frequente nei *polimeri*. Con il termine polimero si intende una macromolecola che, come il DNA e le proteine, è costituita da una lunga catena di gruppi molecolari uniti mediante la ripetizione dello stesso tipo di legame. Queste

¹“L'unico nodo di cisteina regola l'ormone pleiotropico leptina”.

lunghe catene polimeriche si trovano normalmente in uno stato altamente aggrovigliato o sono aggrovigliate ad altri polimeri. Inoltre, le caratteristiche di tali “grovigli”, detti *entanglement*, sono fortemente influenzate dai vincoli geometrici dovuti al fatto che le molecole si trovano spesso confinate in uno spazio ristretto. A questo scopo analizzeremo l’articolo *Topological entanglement complexity of polymer chains in confined geometries*² [29], pubblicato nel 1998, che si occupa proprio di investigare gli effetti di differenti tipi di vincoli geometrici sulla complessità degli entanglement di polimeri, calcolando le probabilità di formazione di nodi e link nelle catene polimeriche soggette a vincoli geometrici.

Infine nell’Appendice A riporteremo alcune nozioni di topologia algebrica, introducendo in particolare le categorie Top Diff e PL, al fine di dare un punto di riferimento per la terminologia usata nel corso di tutta la trattazione e di renderla il più possibile completa e autoconsistente.

Considerazioni conclusive e questioni aperte

Gli articoli [5, 14, 29] analizzati rispettivamente nei Capitoli 3, 4 e 5 hanno prodotto dei risultati importanti in vari campi della ricerca relativa al DNA, alle proteine ed ai polimeri in generale; ma lasciano anche spazio a ulteriori approfondimenti in tantissime direzioni. Riportiamo quindi un breve resoconto dei risultati ottenuti nei tre articoli e delle questioni aperte su cui è possibile indirizzare eventuali ricerche future.

Topologia del DNA e ricombinazione sito-specifica: Attraverso l’esplorazione topologica, la comprensione della ricombinazione sito-specifica è progredita drammaticamente. Le soluzioni in termini di tangles sono state classificate completamente sia per una generica ricombinasi in tirosina che per molte ricombinasi in serina. Il modello a tangle ha illuminato diverse fasi di questo processo, ma ci sono molti passaggi bui nei quali i tipi di nodo o link di particolari prodotti sono ancora sconosciuti e il modello a tangle

²“Complessità di intricatezza topologica di catene polimeriche in geometrie confinate”.

non può essere sfruttato. Una nuova e potente strategia combina un modello predittivo (come quello presentato in Sezione 3.5.3) con altri modelli a tangle (come quelli analizzati nelle Sezioni 3.5.1, 3.5.2) per investigare ulteriormente il meccanismo di ricombinazione. In ogni caso, risulta chiaro che la topologia del DNA ha profonde conseguenze biologiche, ad esempio per quanto riguarda i meccanismi di replicazione e ricombinazione, e che le tecniche puramente biologiche non sono ancora in grado di determinare l'esatta topologia del DNA. Inoltre il meccanismo delle topoisomerasi di tipo II, che regolano il tipo di nodo o link di DNA, non è ancora stato spiegato con precisione ed esistono altri processi enzimatici la cui comprensione è solamente agli albori. Per questo motivo la ricerca in questo campo, in varie direzioni, è attualmente in fase di sviluppo.

Topologia e mutazioni dell'ormone leptina: Mentre la visione corrente è che le mutazioni della leptina che causano malattie siano associate ad una scorretta formazione del ponte disolfuro, in [14] si propone una nuova interpretazione collegata piuttosto alla mancata formazione di una struttura a cappio. D'altra parte, questo lavoro suggerisce che il ponte disolfuro abbia un ruolo regolatorio più importante di quanto si pensi, infatti esso sembra mediare fortemente le funzioni biologiche. Inoltre, basandosi sui dati raccolti, si suggerisce che queste mutazioni nella struttura della leptina possano interessare tre aspetti principali:

- la distruzione delle interazioni responsabili di trainare parte della molecola e di infilarla nella struttura a cappio;
- la riduzione della grandezza del cappio;
- la compromissione dei siti di ancoraggio, che si traduce nell'impossibilità per la leptina di legarsi ai recettori.

Quindi la questione principale diventa esplorare il folding della leptina concentrandosi sulle mutazioni della *struttura a cappio*. Tale scoperta apre molte possibilità per esplorare le funzioni della leptina e di conseguenza per lo

sviluppo di terapie di inibizione e attivazione della leptina, che si dimostrerebbero di fondamentale importanza per la cura di patologie anche gravi.

Probabilità di annodamento e linking dei polimeri, influenza dei vincoli geometrici: In [29] si è mostrato, tramite risultati teorici, che la probabilità di annodamento di tutti i polimeri, tranne un numero esponenzialmente piccolo, confinati in una slab (“lastra”) o un prisma tende ad 1 in maniera esponenziale quando la lunghezza n del polimero tende a infinito. Si è anche mostrato che entrambe le componenti di un link sono annodate con probabilità molto alta e si è caratterizzata la dipendenza da L (cioè dalla grandezza del vincolo geometrico: L -slab o (L, L) -prisma) dell’andamento esponenziale di tale probabilità, mostrando che se $L \rightarrow \infty$ la probabilità di annodamento/linking tenderà alla stessa probabilità nel caso non vincolato. Dopo di che è stato investigato l’andamento della probabilità per n piccoli, attraverso metodi numerici, e sono state osservate caratteristiche simili per le probabilità di annodamento e linking. Più precisamente, entrambe mostrano un massimo quando vengono rappresentate come funzioni di L (in un L -slab o (L, L) -prisma) per valori fissi di n . Ciò potrebbe riflettere la competizione tra una probabilità di essere aggrovigliati (annodati, linked o entrambi), più alta per anelli vincolati in quanto più compatti, e la difficoltà di formazione di nodi e link in geometrie confinate dovuta all’ingombro sterico tra le catene (ovvero il fenomeno prodotto dalla repulsione reciproca tra gli atomi che formano una molecola che impedisce di avvicinare eccessivamente due parti della catena che compone un polimero). Tra le questioni aperte c’è, ad esempio, l’influenza reciproca dei vincoli geometrici e della struttura topologica dei polimeri in casi diversi dalle catene lineari e dai poligoni. Oppure ci si potrebbe interessare al caso di vincoli geometrici meno regolari di quelli qui considerati.

Indice

Introduzione	i
1 Richiami di teoria dei nodi	1
1.1 Concetti fondamentali	1
1.1.1 Equivalenza tra nodi (I)	3
1.1.2 Equivalenza tra nodi (II)	6
1.1.3 Equivalenza tra nodi (III)	7
1.1.4 I link	11
1.2 Diagrammi e mosse di Reidemeister	12
1.2.1 Diagrammi regolari	13
1.2.2 Mosse di Reidemeister	15
1.3 Invarianti di un nodo (o link)	16
1.3.1 Numero di incroci	16
1.3.2 Numero di ponti	17
1.3.3 Linking number	18
1.3.4 Nastri o anelli topologici	20
1.4 Tangles e link razionali	22
1.4.1 Tangles	23
1.4.2 Link a due ponti (o razionali)	30
1.5 Somma di nodi	32
2 DNA e proteine	35
2.1 Le proteine	35
2.1.1 Il folding proteico	37

2.2	Il DNA	41
2.2.1	Il supercoiling	44
2.2.2	Azione degli enzimi topoisomerasi I e II.	48
3	Topologia del DNA	53
3.1	Il DNA come nastro topologico	53
3.2	Il modello di nastro topologico applicato allo studio del supercoiling	56
3.3	Regolazione della topologia del DNA	58
3.4	La ricombinazione sito-specifica	59
3.5	Modello a tangle per la ricombinasi	63
3.5.1	Modello a tangle per una ricombinasi in serina	67
3.5.2	Modello a tangle per una ricombinasi in tirosina	70
3.5.3	Un modello predittivo	74
4	Topologia di una proteina:	
	il caso della leptina	81
4.1	Funzioni della leptina e sue mutazioni	81
4.2	Topologia della leptina	82
4.3	Ruolo della topologia annodata nei processi di folding e unfolding	85
4.4	Conclusioni dell'articolo	89
5	Topologia dei polimeri in geometrie confinate	91
5.1	Introduzione all'articolo	92
5.2	Cammini e poligoni senza autointersezioni	92
5.3	Probabilità di annodamento e linking: risultati teorici	94
5.3.1	Probabilità di annodamento per poligoni non vincolati	94
5.3.2	Probabilità di annodamento per poligoni in ambiente vincolato	97
5.3.3	Probabilità di linking per coppie di poligoni non vincolati	100
5.3.4	Probabilità di linking per coppie di poligoni in ambiente vincolato	102

5.4	Probabilità di annodamento e linking: risultati numerici . . .	102
5.4.1	Probabilità di annodamento per poligoni in geometrie confinata	103
5.4.2	Probabilità di linking	106
5.4.3	Coppie di poligoni contenute in un cubo	107
5.4.4	Coppie di poligoni contenute in slabs	109
5.5	Conclusioni dell'articolo	110
A	Categorie Top, Diff, PL	113
	Bibliografia	121

Elenco delle figure

1	Nodo di Listing.	i
1.1	Nodo banale.	2
1.2	Tavola di nodi primi da 0 a 7 incroci.	2
1.3	Una diversa rappresentazione, dovuta al matematico Ochiai, del nodo banale.	3
1.4	Nodo trifoglio sinistro e nodo trifoglio destro.	3
1.5	Nodo come poligonale chiusa.	4
1.6	Esempio di wild knot.	4
1.7	Movimenti elementari.	5
1.8	Le due possibili orientazioni del nodo trifoglio sinistro.	6
1.9	Esempio di movimento elementare orientato.	6
1.10	Isotopia ambientale.	9
1.11	Traslazione di K	10
1.12	Esempio di link.	11
1.13	Due link equivalenti: questo link viene chiamato Whitehead link.	12
1.14	Diagramma del nodo trifoglio.	13
1.15	Proiezione nodo trifoglio sul piano.	13
1.16	Punti multipli inaccettabili.	14
1.17	Diagrammi e proiezione.	15
1.18	Mosse di Reidemeister.	15
1.19	Esempi di diagrammi e numero di ponti.	17
1.20	Configurazioni di un incrocio.	18

1.21	Esempi di calcolo del linking number.	19
1.22	Esempi di anello.	20
1.23	Diagramma di un anello.	21
1.24	Esempi di diagrammi di (2,2)-tangles.	23
1.25	Sfera con 4 punti fissati.	24
1.26	Diagramma di un tangle.	24
1.27	Numeratore e denominatore di un tangle.	25
1.28	Somma di tangles.	26
1.29	Esempio di tangle localmente annodato.	27
1.30	Tangles eccezionali.	27
1.31	Esempio di tangle banale (o razionale).	27
1.32	Torsione destra e sinistra.	28
1.33	Esempio di costruzione di un tangle banale.	29
1.34	Esempio di somma di tangles razionali che non è razionale.	30
1.35	Esempio di tangle primo.	30
1.36	Esempio di diagramma standard di un link a due ponti.	31
1.37	Somma di nodi.	33
2.1	Struttura generica di un amminoacido, R rappresenta un gruppo laterale specifico per ogni amminoacido.	36
2.2	Struttura di una proteina (in particolare dell' <i>Emoglobina</i>) su vari livelli.	38
2.3	Esempio di folding proteico: simulazione di una via di ripiegamento di parte di una proteina chiamata villina.	39
2.4	Processo di folding dal punto di vista termodinamico.	40
2.5	Struttura del DNA su vari livelli.	42
2.6	Supercoiling di una molecola circolare di DNA.	45
2.7	Supercoiling prodotto dalla separazione delle due catene di una struttura a doppia elica.	46
2.8	Azione degli enzimi topoisomerasi di tipo I.	49
2.9	Azione degli enzimi topoisomerasi di tipo II.	50

3.1	Modello di DNA circolare sotto forma di anello topologico. . .	54
3.2	Esempi di variazione del linking number, twisting number e writhe di una molecola di DNA.	57
3.3	Possibili azioni della ricombinasi sito-specifica.	60
3.4	I siti di ricombinazione in una molecola circolare di DNA possono avere orientazione diretta (immagine a sinistra) o inversa (immagine a destra).	61
3.5	Ricombinazione sito-specifica di una molecola circolare di DNA.	62
3.6	Esempio di ricombinazione vista come sostituzione di un tangle T con un altro R	65
3.7	Prodotto della ricombinazione visto come numeratore di una somma di tangles.	65
3.8	Substrato visto come numeratore di una somma di tangles. . .	66
3.9	Prodotti della ricombinazione processiva della Tn3 resolvasi su un nodo banale superavvolto.	67
3.10	Ricombinazione processiva della Tn3 resolvasi, vista come somma di tangles.	68
3.11	Possibili coppie di tangles (O, R)	69
3.12	Prima ricombinazione effettuata dalla Tn3 resolvasi, quando $E = T(0)$	70
3.13	Modello per il substrato e prodotto di una ricombinazione in tirosina regolata dall'enzima Flp.	71
3.14	Toro con un'orientazione fissata.	72
3.15	Substrati e prodotti della ricombinazione Flp.	73
3.16	Possibili conformazioni del complesso di ricombinazione e possibili substrati: nodi o link torici $t(2, m)$, nodo banale, link banale.	75
3.17	Il link di Hopf è il link torico $t(2, 2)$ e borda un anello topologico.	76
3.18	Possibili conformazioni di $B \cap J$ quando vale l'assunzione 1. . .	77
3.19	Esempi di substrati J e superfici D bordate da J nel caso in cui valga l'assunzione 2.	77

3.20	Esempi di substrati J e superfici D bordate da J nel caso in cui non valga l'assunzione 2.	78
3.21	Famiglia $F(p, q, r, s)$ dei prodotti della ricombinazione sito-specifica.	78
4.1	Struttura nativa (attiva) della leptina.	84
4.2	Rappresentazione della leptina in diversi stati conformazionali.	85
4.3	Confronto tra la forma ossidata e ridotta dello stato denaturato e nativo della leptina.	87
4.4	Folding della leptina ossidata e ridotta.	88
5.1	Esempio di pattern in \mathbb{Z}^2 che non è un Kesten pattern.	95
5.2	Esempio di arco annodato in \mathbb{Z}^3	96
5.3	Probabilità di annodamento $P(n, L)$ rispetto ad L per poligoni confinati in una slab di ampiezza L per diversi valori di n fissati.	104
5.4	Grafico di $\ln P^0(n, L)$ rispetto ad n per poligoni confinati in una slab, per tre valori di L fissati.	105
5.5	Grafico del raggio di girazione quadratico medio rispetto ad L per diversi valori di n	106
5.6	Probabilità per le coppie di poligoni di essere omologicamente linked in funzione di n e per diversi valori di L	108
5.7	Probabilità di linking omologico di una coppia di poligoni contenuta in un cubo, in funzione del rapporto $n/L^{1/\nu}$	109
5.8	Probabilità di linking topologico dato che è presente il linking omologico in funzione del rapporto $n/L^{1/\nu}$	110
A.1	Equivalenza di triangolazioni.	117
A.2	Relazione tra le categorie Top , $Diff$ e PL nelle varie dimensioni.	119

Capitolo 1

Richiami di teoria dei nodi

In questo capitolo introdurremo i concetti fondamentali della teoria dei nodi: definiremo i nodi, i link e studieremo il significato di equivalenza tra nodi. In seguito introdurremo il concetto di diagramma e le cosiddette mosse di Reidemeister che ci permetteranno di convertire lo studio dei nodi, intesi come entità tridimensionali, nello studio di loro particolari proiezioni sul piano. Per ridurre ulteriormente la complessità del problema, introdurremo alcuni invarianti numerici ed una relazione fondamentale tra di essi. Presenteremo infine i concetti di tangle e di nodo razionale. In questa trattazione ci siamo limitati a presentare i concetti della teoria dei nodi utili alla modellizzazione e allo studio del DNA e di alcune proteine di nostro interesse. Per approfondimenti si vedano [1, 10, 22].

1.1 Concetti fondamentali

Immaginiamo di prendere un pezzo di corda, annodarlo e poi incollare tra loro le due estremità libere, quello che otteniamo in ambito matematico viene chiamato *nodo*. Non dobbiamo preoccuparci della lunghezza o dello spessore della corda, ma solamente della sua forma. Per essere più precisi:

Definizione 1.1. Un *nodo* è una curva semplice chiusa nello spazio.

Il nodo che otteniamo, a partire da una corda, congiungendo semplicemente le due estremità senza annodarla è detto *nodo banale*. Più formalmente:

Definizione 1.2. Un *nodo banale* è un nodo che borda un disco (vedi Figura 1.1).

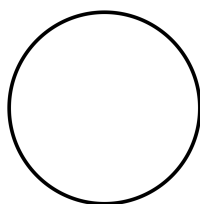


Figura 1.1: Nodo banale.

Utilizzando alcune caratteristiche che vedremo più avanti, si possono costruire delle *tavole di nodi* come quella riportata in Figura 1.2, che rappresenta i *nodi primi* (si veda Definizione 1.25) e che utilizza il *numero di incroci* (si veda Sezione 1.3.1 per la definizione di numero di incroci).

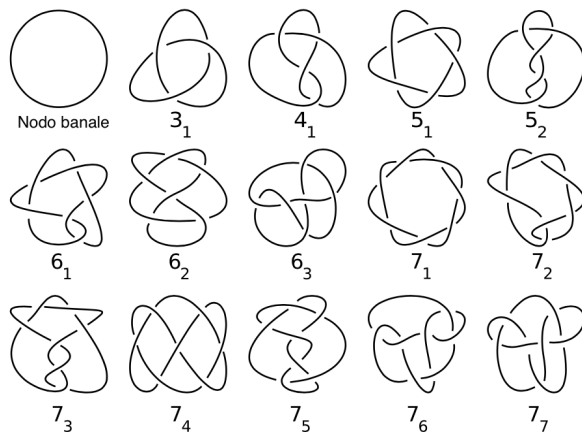


Figura 1.2: Tavola di nodi primi da 0 a 7 incroci.

Per ottenere una classificazione è necessario determinare quanti nodi distinti con n incroci esistono e dunque quando due nodi dati siano in realtà

lo stesso nodo, o meglio equivalenti tra loro (si veda la sezione successiva). Consideriamo ad esempio il nodo rappresentato in Figura 1.3: può sembrare complicato, ma in realtà non è altro che una diversa rappresentazione del nodo banale.

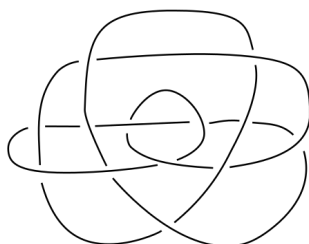


Figura 1.3: Una diversa rappresentazione, dovuta al matematico Ochiai, del nodo banale.

Al contrario i nodi rappresentati in Figura 1.4 possono sembrare molto simili tra loro, ma guardando più attentamente si può notare che essi differiscono per diversi aspetti, infatti non sono equivalenti tra loro.

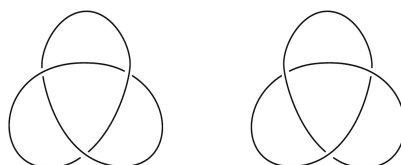


Figura 1.4: Nodo trifoglio sinistro e nodo trifoglio destro.

In generale dati due nodi, li consideriamo *equivalenti* se è possibile deformare con continuità uno di essi, senza mai tagliarlo, fino ad ottenere l'altro. Ma come possiamo definire in maniera più precisa questo concetto?

1.1.1 Equivalenza tra nodi (I)

Il problema dell'equivalenza è stato uno dei principali motori della teoria dei nodi dal XIX secolo ad oggi, vediamo i diversi approcci con cui è stato

affrontato. Finora abbiamo visto i nodi come delle curve chiuse, ma per comodità possiamo immaginarli d'ora in poi come delle *poligonali chiuse* nello spazio tridimensionale (Figura 1.5) così da escludere i cosiddetti “wild knot”, ovvero casi limite che renderebbero necessaria l'introduzione di condizioni specifiche e che non risultano di alcun interesse per la nostra trattazione (vedi Figura 1.6 per un esempio di wild knot).

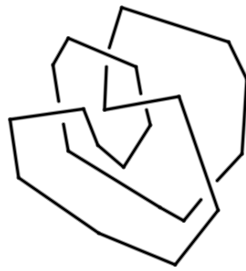


Figura 1.5: Nodo come poligonale chiusa.

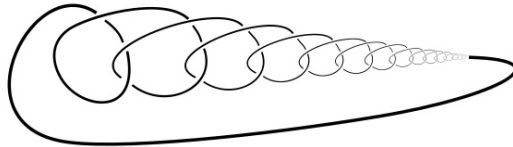


Figura 1.6: Esempio di wild knot.

La prima definizione di equivalenza tra nodi fa uso del concetto di *movimento elementare*:

Definizione 1.3. Su un nodo K possiamo compiere le seguenti due operazioni chiamate *movimenti elementari* e rappresentate in Figura 1.7 :

1. Dividere un suo lato AB in due lati AC e CB , individuando un punto C su di esso e viceversa.

2. Dato un lato AB ed un punto C non appartenente a K , se il triangolo ABC non interseca K in altri punti oltre ad AB , allora possiamo rimuovere AB ed aggiungere gli altri due lati del triangolo AC e BC e viceversa.

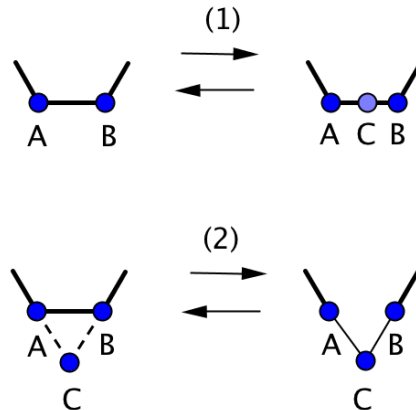


Figura 1.7: Movimenti elementari.

Definizione 1.4. Prima definizione di equivalenza.

Due nodi K e K' che possono essere trasformati l'uno nell'altro attraverso un numero finito di movimenti elementari sono detti *equivalenti*. Denoteremo tale equivalenza con: $K \approx K'$.

Osservazione 1. Intuitivamente si può dire che due nodi sono equivalenti se, nello spazio, possiamo deformare uno di essi in modo continuo, senza causare autointersezioni, fino a trasformarlo nell'altro.

Essendo un nodo una curva semplice chiusa possiamo definire su di esso un'*orientazione*. Come si può immaginare, le orientazioni possibili sono due e possiamo indicarle con una freccia sulla curva (come in Figura 1.8). Possiamo quindi rivedere il concetto di equivalenza per i *nodi orientati* nel modo seguente.

Definizione 1.5. Prima definizione di equivalenza (con orientazione). Due nodi *orientati* K e K' che possono essere deformati l'uno nell'altro

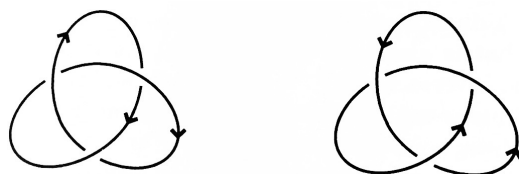


Figura 1.8: Le due possibili orientazioni del nodo trifoglio sinistro.

attraverso un numero finito di movimenti elementari *orientati* (Figura 1.9) sono detti *equivalenti (con orientazione)*.

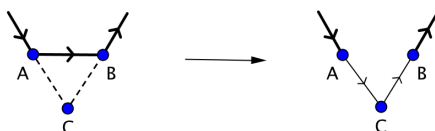


Figura 1.9: Esempio di movimento elementare orientato.

Osservazione 2. Chiaramente due nodi non orientati equivalenti tra loro non rimangono necessariamente equivalenti (con orientazione) quando assegnamo loro un'orientazione. D'altra parte è possibile scegliere un'orientazione tale che lo siano.

1.1.2 Equivalenza tra nodi (II)

I concetti di equivalenza visti finora hanno a che fare con i movimenti elementari, ovvero con trasformazioni di carattere locale. Noi invece siamo interessati ad una definizione di equivalenza tra nodi in termini di *trasformazioni globali*.

Definizione 1.6. Seconda definizione di equivalenza.

Diciamo che due nodi K e K' sono *equivalenti* se esiste un omeomorfismo PL da \mathbb{R}^3 in se stesso (si veda Appendice A) che conserva una fissata orientazione di \mathbb{R}^3 e che manda K in K' .

Definizione 1.7. Seconda definizione di equivalenza (con orientazione). Consideriamo due nodi K e K' con orientazioni assegnate. Se è possibile mandare K in K' attraverso un omeomorfismo PL di \mathbb{R}^3 che preserva una fissata orientazione su \mathbb{R}^3 e che manda l'orientazione di K in quella di K' , allora i due nodi K e K' sono detti *equivalenti con orientazione*.

Il vantaggio di questa seconda definizione di equivalenza, rispetto alla prima, sta nel fatto che essa ci permette di vedere in maniera più diretta e intuitiva un certo numero di equivalenze tra nodi.

1.1.3 Equivalenza tra nodi (III)

Presentiamo ora un terzo e ultimo modo di definire l'equivalenza tra nodi, utilizzando l'*isotopia*.

Infatti l'idea intuitiva di “deformare” un nodo nell'altro corrisponde proprio ad un'isotopia fra la funzione identità su \mathbb{R}^3 ed un omeomorfismo che porta il primo nodo nel secondo. Possiamo quindi riformulare la definizione di equivalenza nel modo seguente.

Definizione 1.8. Terza definizione di equivalenza.

Due nodi K e K' sono *equivalenti* se esiste un'isotopia ambientale PL, ovvero una funzione PL data da $H : \mathbb{R}^3 \times I \longrightarrow \mathbb{R}^3 \times I$ con $H(x, t) = (h_t(x), t)$ tale che $h_t(x)$ è un omeomorfismo PL per ogni $t \in I$, e tale che in particolare:

- $H(\cdot, 0) = id_{\mathbb{R}^3}$
- $H(K, 1) = K'$

Si può dimostrare che le tre definizioni di equivalenza sono tutte equivalenti tra loro. Per far ciò osserviamo che la sfera tridimensionale S^3 è spesso pensata come compattificazione 1-punto di \mathbb{R}^3 , cioè come \mathbb{R}^3 con un punto (∞) aggiunto all'infinito. In alcune occasioni è conveniente pensare ad un nodo come giacente su S^3 , invece che su \mathbb{R}^3 , assumendo ovviamente che il nodo non contenga il punto all'infinito. Con questa premessa possiamo riformulare le tre definizioni di equivalenza e dimostrare il teorema seguente.

Teorema 1.1.1 (Equivalenza delle equivalenze). *Siano K e K' due nodi in S^3 . Le seguenti affermazioni sono equivalenti:*

1. *Esiste un omeomorfismo PL che preserva l'orientazione $f : S^3 \rightarrow S^3$ che manda K in K' .*
2. *K e K' sono equivalenti tramite un'isotopia ambientale da $S^3 \times I$ in se stesso.*
3. *K e K' sono equivalenti in senso combinatorio ovvero attraverso i movimenti elementari.*

Prima di dimostrare il teorema richiamiamo due risultati che ci saranno utili. Inoltre per le nozioni topologiche non precedentemente definite si rimanda all'Appendice A.

Lemma 1.1.2 (Teorema di Alexander–Schoenflies). *Sia $i : S^2 \rightarrow S^3$ un embedding (o immersione iniettiva) PL. Allora:*

$$S^3 = B_1 \cup B_2, \quad i(S^2) = B_1 \cap B_2 = \partial B_1 = \partial B_2,$$

dove B_i per $i = 1, 2$, è una palla tridimensionale combinatoria (cioè è il corpo di un 3-simplesso).

Per la dimostrazione si vedano [2, 4].

Lemma 1.1.3 (Alexander–Tietze). *Ogni omeomorfismo PL f di una palla n -dimensionale B^n , che è il corpo di un n -simplesso, che mantiene fisso il bordo è isotopo all'identità attraverso un'isotopia ambientale PL che mantiene fisso il bordo.*

Per una prova del Lemma 1.1.3 e per maggiori approfondimenti si veda [10].

Dimostrazione del Teorema 1.1.1. (1) \Rightarrow (2): Per prima cosa mostriamo che esiste un'isotopia ambientale $H(x, t) = (h_t(x), t)$ di S^3 tale che $h_1 f$ fissa un 3-simplesso $[P_0, P_1, P_2, P_3]$. Se $f : S^3 \rightarrow S^3$ ha un punto fisso, lo scegliamo

come P_0 ; altrimenti, prendiamo come P_0 un qualsiasi punto interno di un 3-simplesso $[s^3]$ di S^3 . Esiste un'isotopia ambientale di S^3 che lascia fisso $\overline{S^3 - [s^3]}$ e che manda P_0 in un qualsiasi altro punto interno di $[s^3]$. Se $[s'^3]$ è un altro 3-simplesso di S^3 avente una faccia in comune con $[s^3]$, si può costruire facilmente un'isotopia ambientale che manda un punto interno P_0 di $[s^3]$ in un punto interno P'_0 di $[s'^3]$ e che è l'identità al di fuori di $[s^3] \cup [s'^3]$ (si veda la Figura 1.10).

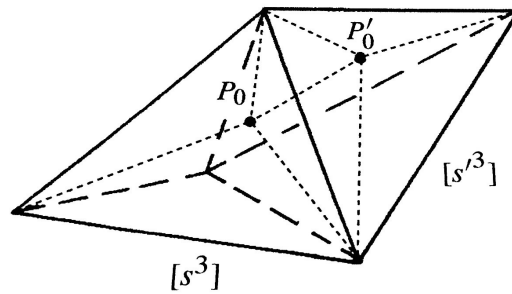


Figura 1.10: Isotopia ambientale (fonte dell'immagine [10]).

Dunque esiste un'isotopia ambientale H^0 con $h_1^0 f(P_0) = P_0$, in quanto ogni coppia di 3-simplessi può essere connessa da una catena di 3-simplessi adiacenti. Ora scegliamo un punto $P_1 \neq P_0$ nella stella di P_0 , e con un ragionamento analogo costruiamo un'isotopia ambientale H^1 tale che $h_1^1 h_1^0 f$ lasci fisso l'1-simplesso $[P_0, P_1]$. Al passo successivo otterremo $h_1^2 h_1^1 h_1^0 f$ che lascia fisso il 2-simplesso $[P_0, P_1, P_2]$. A questo punto assume importanza l'ipotesi che f debba preservare l'orientazione. Infatti grazie a questa ipotesi un punto $P_3 \notin [P_0, P_1, P_2]$, ma che appartiene alla stella di $[P_0, P_1, P_2]$, sarà mandato da $h_1^2 h_1^1 h_1^0 f$ in un punto P'_3 nello stesso semispazio rispetto al piano individuato da P_0, P_1, P_2 . Questo assicura l'esistenza di un'ultima isotopia ambientale H^3 tale che $h_1^3 h_1^2 h_1^1 h_1^0 f$ mantenga fisso $[P_0, P_1, P_2, P_3]$. A questo punto è sufficiente osservare che $H = H^3 H^2 H^1 H^0$ è un'isotopia ambientale, $H(x, t) = (h_t(x), t)$. Inoltre il Lemma 1.1.2 ci assicura che il complementare di $[P_0, P_1, P_2, P_3]$ è una palla tridimensionale combinatoria ed il Lemma 1.1.3 che esiste un'isotopia ambientale tra $h_1 f$ e l'identità su S^3 .

(2) \Rightarrow (1) segue dalla definizione di isotopia ambientale.

(1) \Rightarrow (3): Sia $h : S^3 \rightarrow S^3$ un omeomorfismo che preserva l'orientazione e tale che $K' = h(K)$. Per quanto già dimostrato, esiste un altro omeomorfismo che preserva l'orientazione $g : S^3 \rightarrow S^3$, con $g(K) = K$, tale che hg lascia fisso un qualche 3-simplesso $[s^3]$ che dovrà essere scelto al di fuori di un intorno regolare di K e K' . Preso un punto interno P di $[s^3]$ consideriamo $S^3 - P$ come lo spazio euclideo \mathbb{R}^3 , allora esiste una traslazione τ di \mathbb{R}^3 , tale che $\tau(K) \subseteq [s^3] - P$. È facile mostrare che K e $\tau(K)$ sono equivalenti tramite i movimenti elementari (vedi Figura 1.11).

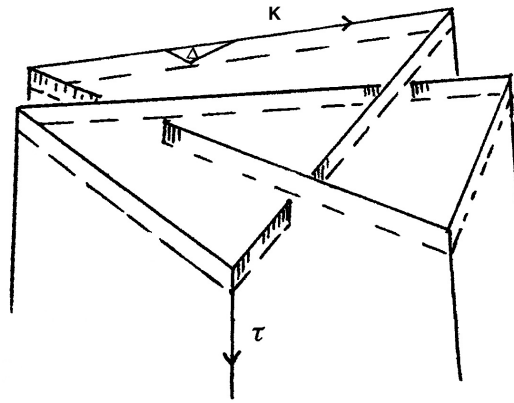


Figura 1.11: Traslazione di K (fonte dell'immagine [10]).

Inoltre vogliamo affermare che $K' = hg(K)$ e $hg\tau(K) = \tau(K)$ sono anch'essi equivalenti tramite movimenti elementari, cosa che concluderebbe la dimostrazione. Infatti, scegliamo una suddivisione della triangolazione di S^3 tale che i triangoli usati nell'equivalenza (tramite movimenti elementari) tra K e $\tau(K)$ formino un sottocomplesso di S^3 , allora esisterà un'equivalenza (tramite movimenti elementari) $K \rightarrow \tau(K)$ definita sui triangoli della suddivisione, ed $hg : S^3 \rightarrow S^3$ manderà il sottocomplesso in un altro sottocomplesso trasportando anche l'equivalenza tramite movimenti elementari (si veda [13]).

(3) \Rightarrow (1): Non è difficile costruire un omeomorfismo di S^3 in se stesso che realizzi uno dei movimenti elementari lasciando fisso il resto del nodo.

Scegliamo un intorno regolare U del 2-simplesso che definisce il movimento elementare, avente il bordo che interseca il nodo in due punti. Estendendo linearmente si può ottenere un omeomorfismo che produca il movimento elementare in U lasciando fisso $S^3 - U$. \square

1.1.4 I link

Un insieme disgiunto di nodi è detto *link* (vedi Figura 1.12). Più formalmente:

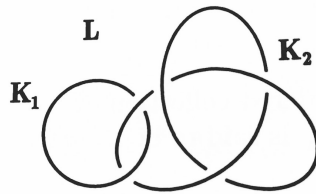


Figura 1.12: Esempio di link (fonte dell'immagine [22]).

Definizione 1.9. Un *link* L è una collezione finita di nodi K_1, \dots, K_m che non si intersecano tra loro:

$$L = K_1 \sqcup K_2 \sqcup \dots \sqcup K_m.$$

Ogni nodo K_i è detto *componente* del link.

Come ci si può aspettare, è possibile estendere il concetto di nodo banale al caso dei link. Dunque il *link banale* a n componenti sarà semplicemente una collezione di n nodi banali ciascuno dei quali borda un disco, e tali che tutti i dischi siano a due a due disgiunti.

Definizione 1.10. Due link $L = K_1 \sqcup K_2 \sqcup \dots \sqcup K_m$ e $L' = K'_1 \sqcup K'_2 \sqcup \dots \sqcup K'_n$ sono *equivalenti* se vale:

- $m = n$, cioè se hanno lo stesso numero di componenti;

- e se possiamo mandare L in L' con un numero finito di movimenti elementari. Più precisamente, a meno di riordinare le componenti, mandando K_i in K'_i per $i = 1, \dots, m$.

Per un esempio di link equivalenti si veda la Figura 1.13

Osservazione 3. Alla luce del Teorema 1.1.1, la seconda condizione può essere sostituita da una qualsiasi delle altre due nozioni di equivalenza tra nodi.

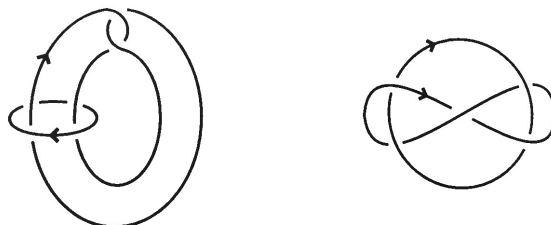


Figura 1.13: Due link equivalenti: questo link viene chiamato Whitehead link (fonte dell'immagine [22]).

Se ogni componente del link è orientata, allora le definizioni di equivalenza sono solo un'estensione di quelle date nel caso dei nodi. In generale molte delle proposizioni che valgono per i nodi possono essere estese anche ai link.

1.2 Diagrammi e mosse di Reidemeister

Come abbiamo visto, un nodo (o link) è un oggetto nello spazio tridimensionale. Ciononostante spesso è utile rappresentarlo su di un piano, seguendo alcune semplici regole che ci permettono di mantenere le informazioni importanti della sua struttura tridimensionale. Questo tipo di rappresentazione, del quale troviamo esempio in tutte le figure di nodi finora presentate, viene chiamato *diagramma* di un nodo, o link (si veda Figura 1.14). Vediamo ora i criteri con cui questi diagrammi sono costruiti.

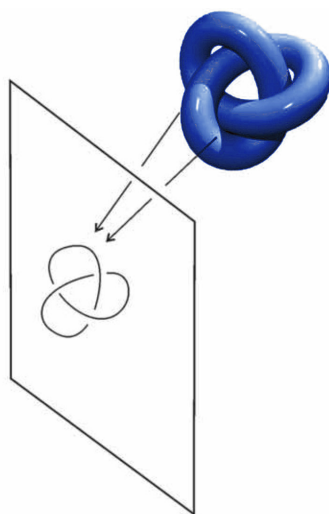


Figura 1.14: Diagramma del nodo trifoglio.

1.2.1 Diagrammi regolari

Consideriamo un link L in \mathbb{R}^3 , in cui abbiamo fissato un sistema di riferimento ortogonale $Oxyz$, e l'applicazione p che proietta ortogonalmente ogni punto di \mathbb{R}^3 sul piano xy . Chiamiamo quindi $\hat{L} = p(L)$ la proiezione di L sul piano xy (Esempio in Figura 1.15).

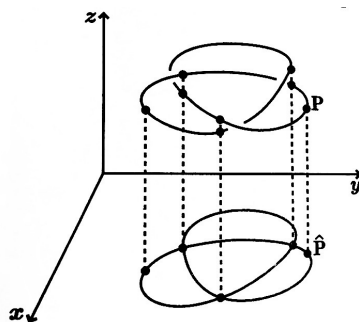


Figura 1.15: Proiezione del nodo trifoglio sul piano (fonte dell'immagine [22]).

Se L ha un'orientazione assegnata, \hat{L} eredita la sua orientazione in modo naturale. Consideriamo ora \hat{L} , esso non è in generale una curva semplice

chiusa sul piano, in quanto possiederà diversi punti di autointersezione. Ma, a meno di effettuare un numero finito di movimenti elementari su L , possiamo imporre le condizioni seguenti:

- \hat{L} ha un numero finito di punti di autointersezione.
- Se Q è un punto di autointersezione di \hat{L} , allora la sua immagine inversa in L : $p^{-1}(Q) \cap L$ è composta da esattamente due punti. Ovvero Q è un punto doppio di \hat{L} .
- Un vertice di L (considerato come nodo poligonale) non è mai mandato da p in un punto doppio di \hat{L} .

In pratica non sono ammessi i casi rappresentati in Figura 1.16.

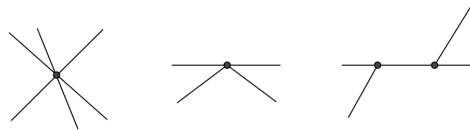


Figura 1.16: Punti multipli inaccettabili.

Una proiezione \hat{L} che soddisfa tali condizioni è detta *proiezione regolare*. Rimane però il problema della perdita di informazione, ovvero in ogni punto doppio non sappiamo quale dei due punti della controimmagine stia sopra e quale sotto. Per rimuovere questa ambiguità possiamo disegnare la proiezione come se fosse stata “tagliata” nei punti di intersezione; in questo modo si ottiene un effetto ottico che dà la sensazione della tridimensionalità dell’oggetto. Questo tipo di rappresentazione viene chiamato *diagramma regolare*, o semplicemente *diagramma*. Nella Figura 1.17 possiamo osservare come due nodi non equivalenti vengano rappresentati da due diagrammi distinti (le prime due immagini a sinistra), mentre se avessimo avuto a disposizione solo la loro proiezione (immagine a destra) sarebbero stati indistinguibili. In un diagramma, quindi, le intersezioni sono rappresentate tracciando interamente la curva che passa “sopra” al resto del nodo, mentre un piccolo

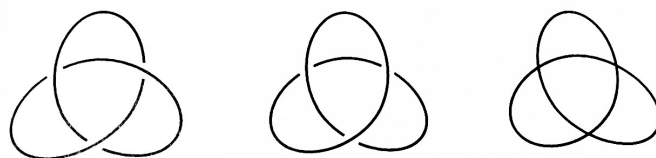


Figura 1.17: Diagrammi e proiezione (fonte dell'immagine [22]).

tratto della parte di nodo che passa “sotto” viene omessa per ottenere l'effetto visivo desiderato.

1.2.2 Mosse di Reidemeister

Come si può immaginare, un unico movimento elementare applicato ad un link produce diversi cambiamenti nel suo diagramma regolare. È però possibile ricondurre questi cambiamenti a solamente 3 tipi di movimenti del diagramma, chiamati *mosse di Reidemeister* e rappresentati in Figura 1.18. Si osservi che ognuno di questi movimenti corrisponde ad una delle tre situazioni non permesse in una proiezione regolare.

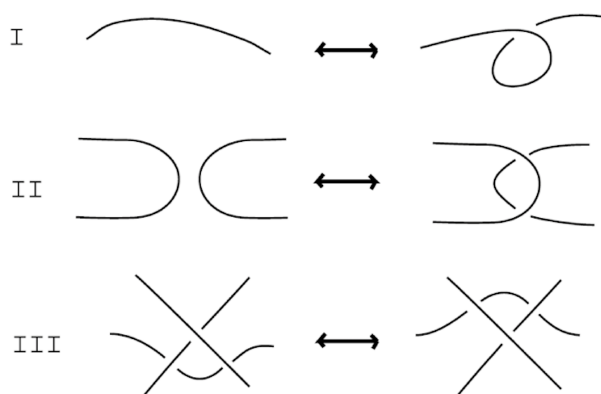


Figura 1.18: Mosse di Reidemeister.

Definizione 1.11. Se modifichiamo un diagramma regolare D attraverso un numero finito di mosse di Reidemeister fino ad ottenerne un altro D' allora

possiamo dire che D e D' sono *equivalenti*. Denotiamo tale equivalenza con: $D \approx D'$. Il simbolo utilizzato è lo stesso usato per indicare l'equivalenza tra link. Tale abuso di notazione è giustificato dal risultato seguente.

Teorema 1.2.1. *Siano D e D' diagrammi regolari dei nodi (o link) L e L' rispettivamente. Allora:*

$$L \approx L' \Leftrightarrow D \approx D'.$$

Per una dimostrazione del Teorema 1.2.1 si veda *Theorem 4.1.1.* in [22]. Possiamo quindi concludere che il problema dell'*equivalenza tra nodi (o link)* è lo stesso dell'*equivalenza tra diagrammi*.

I diagrammi dunque, non sono solamente utili per una più chiara rappresentazione grafica dei nodi e dei link, ma ci permettono di trasferire un complicato problema di topologia tridimensionale ad un problema bidimensionale, che quindi sarà molto più facile da affrontare.

1.3 Invarianti di un nodo (o link)

Un ulteriore step necessario alla semplificazione del problema di studiare e classificare nodi e link è quello di definire degli invarianti. In questo caso un *invariante* è una quantità caratteristica del nodo (o link) che non varia quando applichiamo ad esso un qualsiasi movimento elementare. Come conseguenza del Teorema 1.2.1, un invariante può anche essere pensato come una quantità associata ad un diagramma che rimane inalterata quando applichiamo una qualsiasi delle mosse di Reidemeister al diagramma regolare del nodo (o link) in esame.

1.3.1 Numero di incroci

Il *numero di incroci (MCN)* di un nodo, o link, è semplicemente il numero minimo di incroci di un diagramma del nodo/link. Vale la proprietà che per ogni numero di incroci n fissato, esiste un numero finito di nodi con n incroci.

Questo permette di costruire delle tavole di nodi come quella in Figura 1.2, che ordina i nodi proprio sulla base del loro numero di incroci.

1.3.2 Numero di ponti

Consideriamo un link L e immaginiamo di “schiacciarlo” il più possibile su di un piano orizzontale, se ora osserviamo il nodo ponendoci di taglio rispetto al piano su cui è schiacciato, tutto quello che potremo osservare sono solo alcuni archi che fuoriescono da tale piano. Questi archi corrispondono alle porzioni di filo che negli incroci passano sopra ad almeno un'altra porzione di filo (ma anche più di una), e vengono chiamati *ponti*.

Un altro approccio è quello di definire i *ponti* direttamente sul diagramma di un nodo o link. Infatti, utilizzando la convenzione di rappresentare come tagliata la porzione di nodo che passa sotto in un incrocio, possiamo chiamare ponte *ogni arco connesso del diagramma che passa sopra ad almeno un incrocio*. Indichiamo con $br(D)$ il numero di ponti di un diagramma D . Ad esempio, per i diagrammi rappresentati in Figura 1.19, otterremo:

$$br(D_1) = 3, \quad br(D_2) = 2, \quad br(D_3) = 2.$$

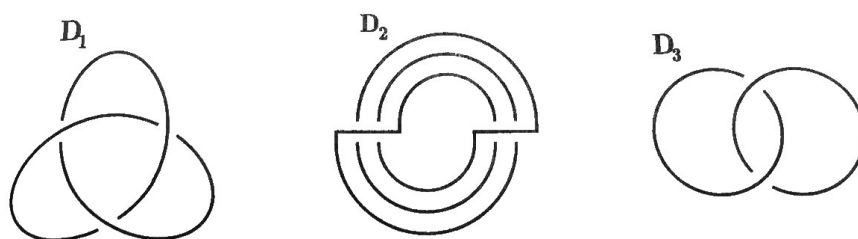


Figura 1.19: Esempi di diagrammi e numero di ponti (fonte dell'immagine [22]).

Definizione 1.12. A questo punto possiamo definire il *numero di ponti* di un link L come:

$$br(L) = \min_{D \in \mathcal{D}} br(D),$$

dove \mathcal{D} è l'insieme di tutti i diagrammi regolari di L .

Teorema 1.3.1. *Il numero di ponti $br(L)$ di un link è un invariante per L .*

1.3.3 Linking number

Il numero di ponti non dipende dall'orientazione del nodo o del link esaminato. In questa sezione, invece, introdurremo un invariante molto importante per i *link orientati*, il cosiddetto *linking number*. Se consideriamo un link costituito da due nodi nello spazio, intuitivamente il *linking number* rappresenta il numero di volte in cui un nodo si avvolge attorno all'altro. Come prima, però, ci conviene dare una definizione più precisa che si fondi sui diagrammi.

Per prima cosa possiamo assegnare un valore $+1$ o -1 ad ogni incrocio del diagramma regolare di un nodo (o link) orientato. In particolare:

Definizione 1.13. Un incrocio c di un diagramma regolare orientato può assumere due configurazioni (Figura 1.20).

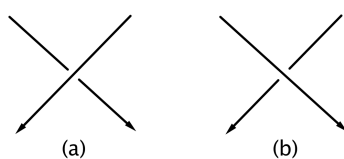


Figura 1.20: Configurazioni di un incrocio.

- se l'incrocio è di tipo (a) lo chiamiamo *positivo* e gli assegnamo il valore $sign(c) = +1$;
- se l'incrocio è di tipo (b) lo chiamiamo *negativo* e gli assegnamo il valore $sign(c) = -1$;

Definizione 1.14. Supponiamo che D sia un diagramma regolare orientato di un link a due componenti $L = K_1 \sqcup K_2$, e chiamiamo c_1, c_2, \dots, c_m gli incroci di D in cui si intersecano le proiezioni di K_1 e K_2 (non consideriamo le intersezioni di ciascun nodo con se stesso). Allora il *linking number* di K_1 e K_2 è:

$$lk(K_1, K_2) = \frac{1}{2} \{sign(c_1) + sign(c_2) + \dots + sign(c_m)\}$$

In Figura 1.21 sono rappresentati alcuni link con i relativi linking number.

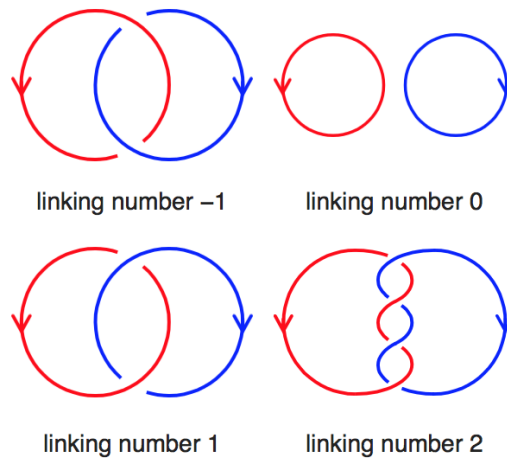


Figura 1.21: Esempi di calcolo del linking number.

Teorema 1.3.2. *Il linking number è un invariante di un link orientato a due componenti.*

Osservazione 4. Cambiando l'orientazione di una delle due componenti del link, il linking number cambia di segno.

Generalizziamo ora il concetto ad un link di n componenti.

Definizione 1.15. Sia L un link con n componenti, K_1, K_2, \dots, K_n . Allora:

$$lk(L) := \sum_{1 \leq i < j \leq n} lk(K_i, K_j),$$

è chiamato il *linking number totale* di L .

Teorema 1.3.3. *Anche il linking number totale è un invariante di un link orientato.*

1.3.4 Nastri o anelli topologici

In topologia un *anello* è una superficie immersa iniettivamente nello spazio, omeomorfa ad una corona circolare, ovvero a $S^1 \times [0, 1]$ (si veda Figura 1.22). Si dice che due anelli sono equivalenti se è possibile trasformare l'uno nell'altro attraverso deformazioni continue. Più precisamente:

Definizione 1.16. Due anelli, immersi iniettivamente nello spazio \mathbb{R}^3 (o S^3), sono considerati *equivalenti* se esiste un omeomorfismo PL di \mathbb{R}^3 (o S^3) che preserva l'orientazione e che manda un anello nell'altro.



Figura 1.22: Esempi di anello.

Immaginiamo un nastro (o anello) A e consideriamo solamente le curve che rappresentano i bordi del nastro, che chiameremo C_1 e C_2 e che corrispondono a $S^1 \times \{0\}$ e $S^1 \times \{1\}$. Abbiamo ottenuto così un modello del nastro sotto forma di **link** a due componenti. Inoltre, possiamo fissare un'orientazione sull'asse C , corrispondente a $S^1 \times \{\frac{1}{2}\}$, e considerare le orientazioni di C_1 e C_2 ereditate da C , otterremo così un **link orientato** come quello in Figura 1.23.

Abbiamo visto che il *linking number* è un invariante per i link orientati. Se ci riferiamo ad un modello di nastro come quello che abbiamo appena presentato, in cui consideriamo le curve C_1 e C_2 come le due componenti del

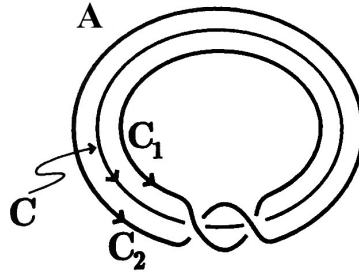


Figura 1.23: Diagramma di un anello (fonte dell'immagine [22]).

link orientato, il linking number può essere definito in una varietà di modi equivalenti. Una di queste, equivalente a quella data in Sezione 1.3.3, è la sua descrizione in termini dell'integrale di Gauss:

$$Lk(A) = \frac{1}{4\pi} \int_{C_1} \int_{C_2} \frac{(x' - x)(dydz' - dzdy') + (y - y')(dzdx' - dxdz') + (z' - z)(dxdy' - dydx')}{[(x' - x)^2 + (y' - y)^2 + (z' - z)^2]^{3/2}}$$

dove (x, y, z) varia su C_1 e (x', y', z') varia su C_2 .

È importante osservare che qui l'invariante è calcolato direttamente sul link e non su un suo diagramma.

Altri due invarianti che si possono definire in maniera analitica per gli anelli sono il *writhe* e il *twisting number*. Questi ultimi non sono invarianti del link ottenuto dalla modellizzazione dell'anello, ma sono invarianti del nastro stesso pensato come superficie nello spazio, inoltre essi non dipendono dall'orientazione di C , C_1 e C_2 .

Il *writhe* indica quanto l'asse C dell'anello sia contorto nello spazio. Se l'asse C è un nodo banale, allora $Wr(A) = 0$, si può quindi affermare che il *writhe* ci dà una misura della non-planarità di A . Il *writhe* può essere definito rigorosamente come:

$$Wr(A) = \frac{1}{4\pi} \int_C \int_C \frac{(x' - x)(dydz' - dzdy') + (y - y')(dzdx' - dxdz') + (z' - z)(dxdy' - dydx')}{[(x' - x)^2 + (y' - y)^2 + (z' - z)^2]^{3/2}}$$

dove (x, y, z) e (x', y', z') variano entrambi sull'asse C del nastro.

Ora, invece, concentriamoci su quante torsioni l'anello A compie rispetto al proprio asse centrale C , ovvero quanto strettamente l'anello si avvolge su di esso. Il *twisting number* di A , che si denota con $Tw(A)$, è un indice proprio del numero di queste torsioni. Anche per il *twisting number* possiamo dare una definizione rigorosa, che fa uso di un sistema di riferimento di Frenet.

In particolare, consideriamo un sistema di riferimento di Frenet (T, N, B) sulla curva C (asse dell'anello), con T vettore tangente, N vettore normale e B binormale. Consideriamo inoltre le equazioni di Serret-Frenet: $\dot{T} = \kappa N$, $\dot{N} = -\kappa T + \tau B$, $\dot{B} = -\tau N$. Allora possiamo definire il *twisting number* dell'anello A come la torsione totale della curva C :

$$Tw(A) = \frac{1}{2\pi} \int_C \langle B, \dot{N} \rangle ds = \frac{1}{2\pi} \int_C \tau ds.$$

Osserviamo infatti che $\langle B, \dot{N} \rangle$ è la proiezione di \dot{N} nella direzione di B che, per le equazioni di Serret-Frenet, è uguale a τ .

Si può dimostrare che sia il *writhe* che il *twisting number* non sono necessariamente numeri interi.

Il seguente teorema spiega il legame tra writhe, twisting number e linking number.

Teorema 1.3.4. [21] (*Călugăreanu-White-Fuller*). *Sia A un anello in \mathbb{R}^3 , si ha:*

$$Lk(A) = Tw(A) + Wr(A).$$

Come vedremo nel Capitolo 3, questa relazione fondamentale tra proprietà topologiche (linking number) e geometriche (writhe e twisting number) di un anello si è rivelata di enorme importanza per lo studio del DNA.

1.4 Tangles e link razionali

Un altro concetto utile per la nostra trattazione è quello di *tangle*, si tratta di una variazione del concetto di nodo introdotta da Conway a cavallo tra gli anni '60 e '70 che ha permesso di definire una nuova classe di link: i link algebrici. Dopo aver presentato i tangles e le loro proprietà, investigheremo un particolare tipo di link algebrici, ovvero i *link a due ponti*, anche detti *link razionali* per il loro legame con i numeri razionali.

1.4.1 Tangles

Definizione 1.17. Consideriamo la sfera S^2 , frontiera della palla B^3 , e fissiamo su di essa $2n$ punti. Un (n, n) -*tangle* T è formato da n curve semplici, aventi i $2n$ punti come estremi, che non si intersechino tra di loro e che siano tutte contenute all'interno di B^3 .

Più precisamente un *tangle* è la coppia ordinata (B^3, T) , ma per comodità spesso lo si denota semplicemente con T .

Analogamente a quanto fatto per i link anche per i tangles si può definire il concetto di diagramma regolare; vediamone alcuni esempi in Figura 1.24.

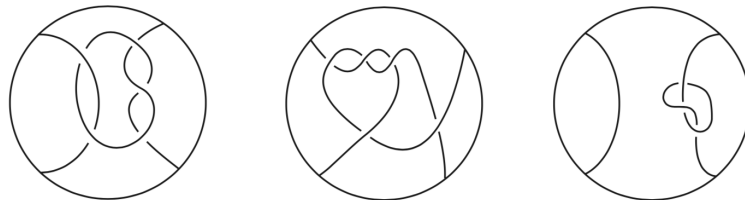


Figura 1.24: Esempi di diagrammi di $(2,2)$ -tangles.

Osservazione 5. Siccome ci occuperemo maggiormente di $(2,2)$ -*tangles* d'ora in poi ci riferiremo ad essi semplicemente come *tangles*.

Legame tra tangles e link

Consideriamo ora la sfera S^2 e quattro punti su di essa, che chiameremo NE , NW , SE e SW (si veda Figura 1.25).

A partire da questi punti costruiamo poi un tangle che possiamo rappresentare attraverso il suo diagramma regolare.

Definizione 1.18. Sia T un tangle avente per estremi NW , NE , SW , SE , come rappresentato in Figura 1.26.

- Il nodo (o link a due componenti) ottenuto connettendo NW con NE ed SW con SE attraverso curve semplici e non annodate tra loro né

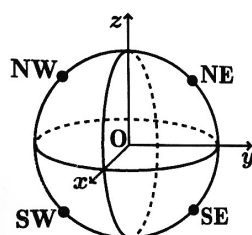


Figura 1.25: Sfera con 4 punti fissati (fonte dell'immagine [22]).

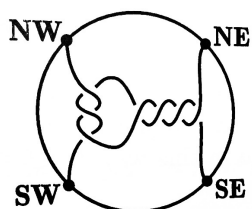


Figura 1.26: Diagramma di un tangle (fonte dell'immagine [22]).

autoannodate, fuori da B^3 , è detto il *numeratore* di T e si denota con $N(T)$ (Figura 1.27: immagine a sinistra).

- Quello ottenuto connettendo NW con SW ed NE con SE è detto il *denominatore* di T e si denota con $D(T)$ (Figura 1.27: immagine a destra).

Questo processo ci permette di costruire, a partire da un tangle, due differenti nodi (o links). Siccome un tangle può essere pensato come una “parte” di link, con i dovuti aggiustamenti possiamo estendere a questa categoria di oggetti diverse definizioni, come quelle di equivalenza.

Definizione 1.19. Due tangles T_1 e T_2 sono *equivalenti* se esiste un omeomorfismo PL che preserva l'orientazione $\varphi : B^3 \rightarrow B^3$ tale che:

- φ mantiene fisso S^2 ;
- $\varphi(T_1) = T_2$.

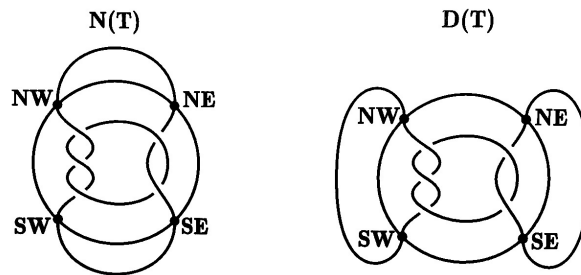


Figura 1.27: Numeratore e denominatore di un tangle (fonte dell'immagine [22]).

Definizione 1.20. Due tangles T_1 e T_2 in B^3 sono *equivalenti* quando è possibile collegare due loro diagrammi attraverso un numero finito di movimenti elementari, mantenendo fissi i quattro punti: NE , NW , SE , SW .

Intuitivamente, se possiamo muovere con continuità l'uno nell'altro senza causare intersezioni e tenendo fissi gli estremi. Come per i nodi (e link), le due definizioni sono equivalenti.

Somma di tangles

Per definire un'operazione di somma sull'insieme dei tangles l'idea è quella di connettere due dei punti "estremi" di un tangle a due punti "estremi" di un altro.

Nello specifico, prendiamo due tangles T_1 in una palla B_1^3 e T_2 in un'altra palla B_2^3 . Ora possiamo pensare di posizzarli parallelamente e di incollare l'emisfero est di B_1^3 all'emisfero ovest di B_2^3 in modo da far combaciare i punti:

- NE di T_1 con NW di T_2 ;
- SE di T_1 con SW di T_2 .

Infine rimuoviamo gli emisferi che sono stati incollati insieme, così da ottenere un unico tangle $T_1 + T_2$ nella nuova palla B_0^3 (si veda Figura 1.28). Il

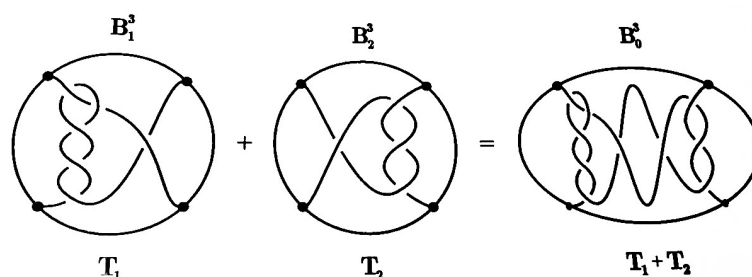


Figura 1.28: Somma di tangles (fonte dell'immagine [22]).

tangle $T(0)$, rappresentato in Figura 1.30, è l'elemento neutro della somma di tangles.

Tipi di tangle

I tangles possono essere divisi in tre famiglie disgiunte:

- tangles *localmente annodati*;
- tangles banali, o *razionali*;
- tangles *primi*.

Vediamo in dettaglio come sono definite le tre famiglie:

Definizione 1.21. Un tangle T è *localmente annodato* se esiste almeno una palla \hat{B} , interna a B^3 , tale che \hat{B} interseca T in due soli punti e l' $(1, 1)$ -tangle formato da $(\hat{B}, \hat{B} \cap T)$ non è banale (e quindi è annodato). Esempio in Figura 1.29. Un tangle che non è localmente annodato si dice *localmente banale*.

Se ci limitiamo agli $(1,1)$ -tangles, è ovvio quale debba essere l' $(1,1)$ -tangle *banale*; passando ai $(2,2)$ -tangles, invece, non è più così facile. Consideriamo ad esempio i diagrammi in Figura 1.30, essi rappresentano i quattro tipi più semplici di $(2,2)$ -tangle e sono chiamati tangles *eccezionali*. Nell'ordine li possiamo chiamare tangle di tipo (0) , $(0, 0)$, (-1) , (1) e denotarli con $T(0)$, $T(0, 0)$, $T(-1)$ e $T(1)$. Tutti questi sono tangles banali, ma non sono gli unici.

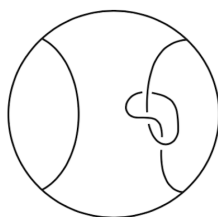


Figura 1.29: Esempio di tangle localmente annodato.

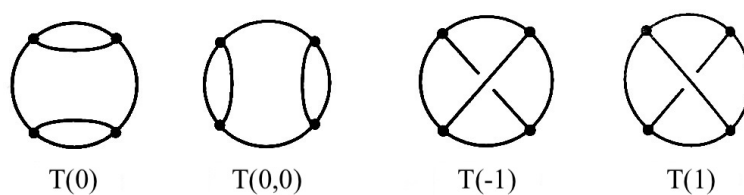


Figura 1.30: Tangles eccezionali (fonte dell'immagine [22]).

Definizione 1.22. Sia f un omeomorfismo che manda B^3 in se stessa e l'insieme dei punti $\{NW, NE, SW, SE\}$ in se stesso. Allora un tangle è *banale* (o *razionale*) se è l'immagine di $T(0,0)$ attraverso tale f . Esempio in Figura 1.31.

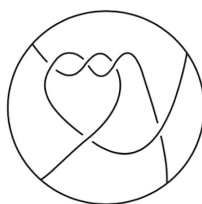


Figura 1.31: Esempio di tangle banale (o razionale).

Gli omeomorfismi di questo tipo sono infiniti, di conseguenza lo sono anche i tangles banali.

Ad esempio possiamo considerare la sfera S^2 e far ruotare l'emisfero sud attorno all'asse z , lasciando fisso l'emisfero nord. In questo modo i punti

SW e SE si scambiano posizione e otteniamo una torsione verticale (*vertical twist*) del tangle al suo interno. Analogamente ruotando l'emisfero est della sfera otteniamo una torsione orizzontale (*horizontal twist*). Inoltre, possiamo assegnare a queste torsioni un'orientazione:

- Nel caso verticale, la torsione considerata *positiva* è una *torsione destra*;
- Nel caso orizzontale, la torsione considerata *positiva* è una *torsione sinistra* (si veda Figura 1.32).



Figura 1.32: Torsione destra e sinistra.

Proposizione 1.4.1. *Un tangle banale può essere ottenuto con una sequenza finita di torsioni verticali e/o orizzontali a partire da $T(0)$ o $T(0,0)$, come si vede in Figura 1.33.*

Da questa proposizione segue che possiamo determinare completamente i tangles banali attraverso la sequenza di torsioni da applicare per ottenerli. La notazione per indicare uno di questi tangles sarà della forma:

$$T(a_1, a_2, \dots, a_n).$$

Ciò significa che per ottenere questo tangle dobbiamo applicare le seguenti procedure:

- Se n è dispari: partire da $T(0)$ ed applicare a_1 torsioni orizzontali, a_2 torsioni verticali, poi di nuovo a_3 orizzontali e così via fino ad a_n .
- Se n è pari: partire da $T(0,0)$ ed applicare a_1 torsioni verticali, a_2 torsioni orizzontali, poi di nuovo a_3 verticali e così via fino ad a_n .

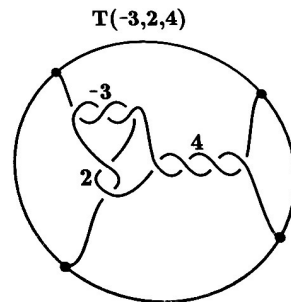


Figura 1.33: Esempio di costruzione di un tangle banale (fonte dell'immagine [22]).

Chiaramente i segni degli a_i indicano se la torsione da compiere è positiva o negativa.

I tangles banali vengono chiamati anche *razionali* in quanto possono essere messi in corrispondenza biunivoca con i numeri razionali estesi $\mathbb{Q} \cup \{\infty\}$. Questa corrispondenza, individuata per la prima volta dal matematico Conway, fa uso dell'espansione in termini di frazione continua dei numeri razionali. In pratica, il tangle banale $T(a_1, a_2, \dots, a_n)$ corrisponde alla frazione $\frac{p}{q}$ che ha un'espansione in termini di frazione continua data da:

$$[a_n, a_{n-1}, \dots, a_1] = \frac{p}{q}.$$

Il numero razionale $\frac{p}{q}$ è chiamato, in questo caso, *frazione del tangle*.

Esempio 1.1. La frazione di $T(0)$ è 0, e quella di $T(0,0)$ è $0 + \frac{1}{0}$, che denotiamo con ∞ .

Teorema 1.4.2. *Esiste una corrispondenza biunivoca tra l'insieme dei numeri razionali estesi $\mathbb{Q} \cup \{\infty\}$ e le classi di equivalenza dei tangles banali. In altre parole due tangles banali sono equivalenti se e solo se le loro frazioni coincidono.*

Osservazione 6. La somma di due tangles razionali non è necessariamente razionale, come si può osservare nell'esempio in Figura 1.34, dove T_1 e T_2 sono razionali mentre $T_1 + T_2$ non lo è.

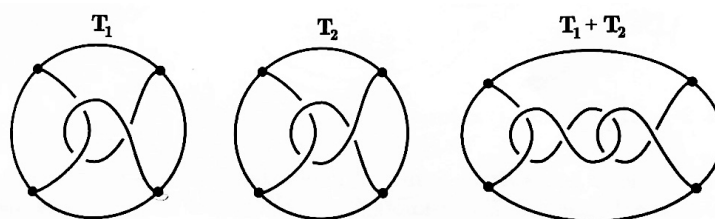


Figura 1.34: Esempio di somma di tangles razionali che non è razionale (fonte dell'immagine [22]).

Si chiama *algebrico* un tangle che può essere espresso come una somma finita di tangles banali.

Definizione 1.23. Tutti i tangles che non sono localmente annodati e che non sono banali (o razionali) vengono detti *tangles primi*. Un esempio è quello disegnato in Figura 1.35.

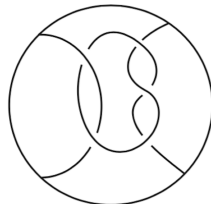


Figura 1.35: Esempio di tangle primo.

1.4.2 Link a due ponti (o razionali)

I *link a due ponti* costituiscono una famiglia di link strettamente legata ai tangles razionali, per questo vengono spesso chiamati *link razionali*. Come suggerisce il nome stesso, si tratta di link che hanno un numero di ponti uguale a 2 (si veda Sezione 1.3.2). Essi presentano al massimo due componenti, inoltre sono stati completamente classificati, ma alcune loro caratteristiche sono ancora argomento di ricerca.

Possiamo pensare al diagramma regolare standard di un link a due ponti come a quello in Figura 1.36.

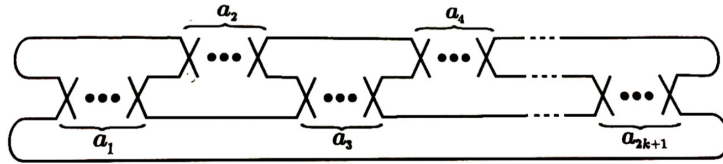


Figura 1.36: Esempio di diagramma standard di un link a due ponti (fonte dell'immagine [22]).

Teorema 1.4.3. *Un link a due ponti è sempre il denominatore di un tangle razionale e, viceversa, il denominatore di un tangle razionale è sempre un link a due ponti.*

Il numero di torsioni è denotato da degli interi a_i , e possiamo definire i segni di a_i come segue: se i è dispari allora la torsione destra è quella positiva, mentre se i è pari la torsione sinistra è quella positiva. Se la sequenza di interi (a_1, \dots, a_{2k+1}) individua il numero di torsioni presenti nel diagramma standard, diremo che il link a due ponti è di tipo (a_1, \dots, a_{2k+1}) e lo denoteremo con:

$$C(a_1, \dots, a_{2k+1}).$$

Inoltre si ha:

$$C(a_1, \dots, a_{2k+1}) = D(T(a_1, \dots, a_{2k+1}, 0)).$$

Abbiamo visto che il denominatore di un tangle razionale è un link a due ponti, ma lo è anche il numeratore, infatti:

Proposizione 1.4.4. *Per ogni intero b vale:*

$$\begin{aligned}
N(T(a_1, \dots, a_{2k+1})) &\approx N(T(a_1, \dots, a_{2k+1}, b, 0)) \\
&\approx D(T(-a_1, \dots, -a_{2k+1}, b)) \\
&\approx C(-a_1, \dots, -a_{2k+1}). \\
N(T(a_1, \dots, a_{2k})) &\approx D(T(-a_1, \dots, -a_{2k}, b)) \\
&\approx C(a_1, \dots, a_{2k} - 1, 1). \\
D(T(a_1, \dots, a_{2k})) &\approx D(T(a_1, \dots, a_{2k-1}, 0)) \\
&\approx C(a_1, \dots, a_{2k-1}). \\
D(T(a_1, \dots, a_{2k+1})) &\approx D(T(a_1, \dots, a_{2k}, 0)) \\
&\approx C(1, a_1 - 1, a_2, \dots, a_{2k}).
\end{aligned}$$

Per comodità, d'ora in poi, consideriamo solo i link a due ponti ottenuti da tangles razionali non eccezionali. Come per i tangles razionali, possiamo associare ogni link a due ponti $C(a_1, \dots, a_{2k+1})$ con $a_i \neq 0$ ad un numero razionale:

$$\frac{p}{q} = [a_1, \dots, a_{2k+1}], \quad MCD(p, q) = 1.$$

Diciamo allora che $C(a_1, \dots, a_{2k+1})$ è di tipo (p, q) e lo denotiamo con $b(p, q)$, dove $-p < q < p$ e $p > 0$.

Teorema 1.4.5. *Dati due link a due ponti $L = b(p, q)$ e $L' = b(p', q')$, essi sono equivalenti se e solo se vale:*

- $p = p'$ e $q \equiv q' \pmod{p}$
- oppure $p = p'$ e $q^{\pm 1} \equiv \pm q' \pmod{p}$.

1.5 Somma di nodi

Vediamo ora come è possibile ottenere un singolo nodo (orientato) a partire da due nodi (orientati) dati, ovvero come è definita la *somma di nodi*.

Consideriamo un punto P su di un nodo orientato K in S^3 . Possiamo pensare a P come al centro di una palla B^3 , di raggio molto piccolo, e che possieda le seguenti proprietà (si veda Figura 1.37 (a) e (b)):

1. K interseca la sfera che borda B^3 esattamente in due punti;
2. nell'interno di B^3 , l' $(1, 1)$ -tangle α , che si ottiene da K è l' $(1, 1)$ -tangle banale (si veda Sezione 1.4.1 per la definizione di $(1, 1)$ -tangle).

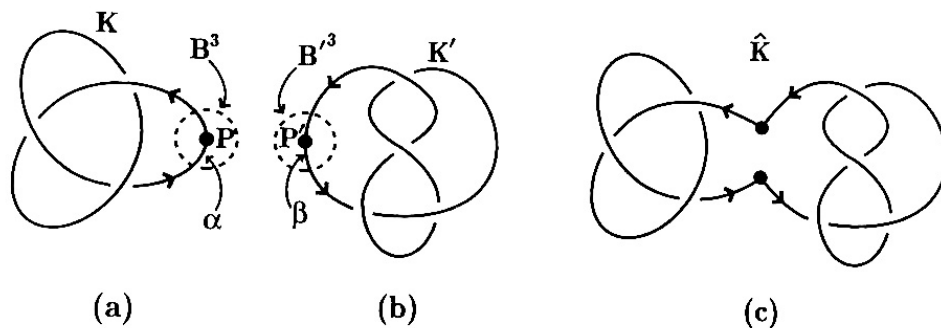


Figura 1.37: Somma di nodi (fonte dell'immagine [22]).

Ora consideriamo un altro nodo K' in un'altra sfera tridimensionale S^3 , analogamente prendiamo un punto P' su K' , centro di una palla B'^3 , in modo da ottenere un $(1, 1)$ -tangle banale β (come illustrato in Figura 1.37 (b)). Possiamo assegnare in modo naturale le orientazioni da K e K' rispettivamente, agli $(1, 1)$ -tangles α e β .

Sia \tilde{B}^3 la palla che si ottiene rimuovendo da S^3 i punti dell'interno di B^3 . Analogamente sia \tilde{B}'^3 la palla che si ottiene rimuovendo da S^3 i punti dell'interno di B'^3 . La superficie di ciascuna di queste palle, cioè di \tilde{B}^3 e \tilde{B}'^3 , è una sfera bidimensionale. Se ora incolliamo queste due palle lungo tale sfera, tramite un omeomorfismo che rovescia l'orientazione della sfera di una di queste palle, otterremo una sfera tridimensionale S^3 . Nel processo di incollamento vengono uniti:

- l'estremo iniziale di α con l'estremo finale di β ;

- l'estremo finale di α con l'estremo iniziale di β .

Dunque nella nuova sfera tridimensionale S^3 viene a formarsi un nuovo singolo nodo orientato \hat{K} (si veda Figura 1.37 (c)). Per costruzione, l'orientazione del nuovo nodo \hat{K} è compatibile con le orientazioni originali di K e K' .

Definizione 1.24. Il nodo \hat{K} , formato grazie al procedimento appena illustrato, è detto *somma* (o *somma connessa*) di K e K' e si denota con $K\#K'$.

Inoltre, la somma $K\#K'$ è indipendente dai punti P e P' scelti e si può dimostrare che essa dipende unicamente dai nodi K e K' .

Definizione 1.25. Un nodo che si può ottenere come somma di nodi entrambi non banali è detto *nodo composto*. Se, invece, ciò non è possibile si parla di *nodo primo*.

Osservazione 7. La somma di nodi è commutativa e il nodo banale è l'elemento neutro.

Capitolo 2

DNA e proteine

In questo capitolo introdurremo alcuni concetti fondamentali di biochimica che, come vedremo, possono essere modellizzati attraverso la teoria dei nodi. In particolare parleremo delle proteine, delle loro funzioni e della loro struttura su vari livelli, introdurremo il concetto di folding proteico e vedremo che la comprensione di questo meccanismo assume grande importanza dal punto di vista biologico e medico. In seguito introdurremo la struttura e la funzione del DNA, soffermandoci sull'importanza di quello che viene chiamato supercoiling della molecola di DNA, per poi concentrarci sull'azione di due particolari enzimi che ne modificano la topologia, ovvero gli enzimi topoisomerasi I e II. Per approfondimenti si veda [3, 5, 11, 23]

2.1 Le proteine

Le cellule degli organismi viventi sono composte per gran parte da lunghe e complesse molecole organiche, le **proteine**, che svolgono tutte le varie funzioni biologiche. Esse formano i tessuti per il sostegno e il movimento, producono energia, trasportano sostanze ai vari organi, fuori e dentro le cellule; inoltre, sotto il nome di **enzimi**, catalizzano le reazioni chimiche necessarie alla vita. Anche se alcune proteine sono assunte dall'esterno, la maggior parte viene prodotta dalle cellule in un processo detto **sintesi delle**

proteine. Questo processo è guidato dai cosiddetti **geni**, che fungono da “istruzioni biochimiche” per la composizione delle proteine. I geni, a loro volta, sono contenuti nei cromosomi all’interno di lunghi filamenti di **DNA**. I cromosomi infine vengono trasmessi da una generazione all’altra, determinando le caratteristiche ereditarie.

Ora che abbiamo presentato il quadro generale, possiamo concentrarci sulla descrizione delle *proteine*. Esse sono lunghe catene di molecole più semplici, gli **amminoacidi primari**, che si legano tra loro mediante un opportuno legame chimico detto **legame peptidico**. Per questo motivo spesso le proteine sono chiamate anche “polipeptidi”. Gli amminoacidi primari sono solamente 20 e quelli usati negli organismi viventi hanno una struttura variabile composta da 7 a 24 atomi, circa la metà dei quali sono atomi di Idrogeno (H). Gli unici altri elementi presenti sono Carbonio (C), Azoto (N), Ossigeno (O) e Zolfo (S). Gli amminoacidi primari assumono la struttura generica mostrata in Figura 2.1, dove *R* viene chiamato **gruppo laterale**.

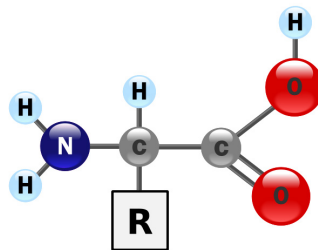


Figura 2.1: Struttura generica di un amminoacido, *R* rappresenta un gruppo laterale specifico per ogni amminoacido.

Il gruppo laterale *R* ha una struttura diversa in ciascuno dei 20 amminoacidi primari, queste differenze conferiscono agli amminoacidi proprietà chimiche distinte quando si trovano in una soluzione acquosa neutra.

Le proteine sono catene che possono andare da qualche decina a molte migliaia di questi amminoacidi; inoltre esistono proteine composte da molte catene polipeptidiche non legate covalentemente, chiamate subunità. I potenziali

elettrochimici tra l'ambiente e gli amminoacidi, e tra gli amminoacidi stessi, determinano il ripiegamento della proteina, chiamato processo di **fold**ing. La forma tridimensionale che ne risulta è specifica per ogni proteina ed è essenziale per lo svolgimento della sua funzione.

Abbiamo visto che le proteine hanno una struttura complessa, dunque per studiarle conviene suddividere tale struttura in vari livelli. Vediamo in Figura 2.2 i quattro livelli strutturali di una proteina ed analizziamoli più nello specifico:

- la **struttura primaria** consiste nella sequenza degli amminoacidi, anche detti *residui amminoacidici*, uniti da legami covalenti (soprattutto legami peptidici e ponti disolfuro);
- la **struttura secondaria** si riferisce a disposizioni particolarmente stabili dei residui amminoacidici che si organizzano in strutture ricorrenti, come ad esempio le α -*eliche* o le *strutture β* (*foglietto β*);
- la **struttura terziaria** descrive tutti gli aspetti di ripiegamento tridimensionale di un polipeptide;
- quando una proteina è costituita da due o più catene (o subunità) polipeptidiche, la loro disposizione reciproca nello spazio è detta **struttura quaternaria**.

2.1.1 Il folding proteico

Tutte le cellule sintetizzano le proteine sotto forma di lunghe catene lineari di amminoacidi. Due amminoacidi possono essere collegati unendo il gruppo amminico (NH_2 in Figura 2.1) di uno al gruppo carbossilico ($COOH$ in Figura 2.1) dell'altro con il rilascio di una molecola d'acqua, questo tipo di legame è il cosiddetto **legame peptidico**. Una catena di amminoacidi, dunque, terminerà da un lato con un gruppo amminico esposto e dall'altro con un gruppo carbossilico; questo fa sì che i polimeri (ovvero le catene di amminoacidi) possano essere sintetizzati in un ordine specifico. Dopo la sintesi,

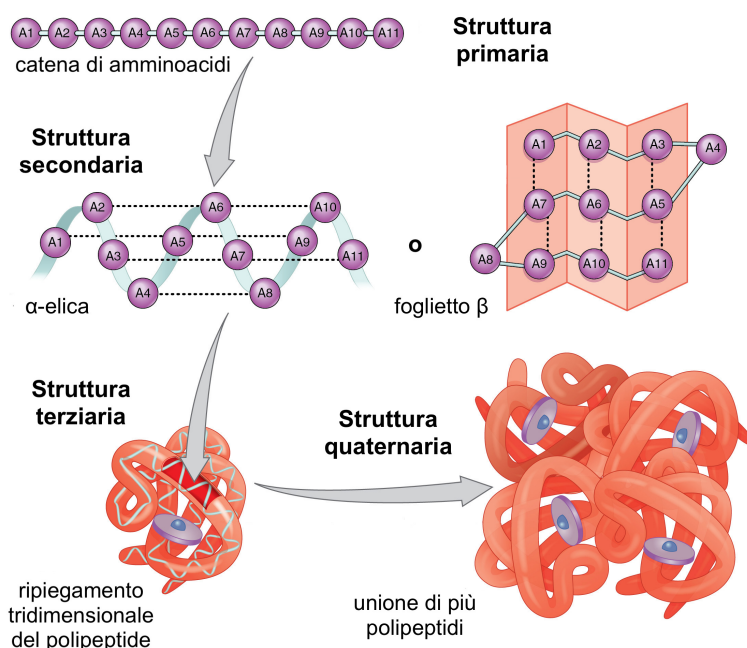


Figura 2.2: Struttura di una proteina (in particolare dell' *Emoglobina*) su vari livelli.

questa lunga catena polimerica è immersa in un solvente (l'acqua), che inizia ad interagire con i gruppi laterali in accordo con le loro proprietà specifiche. Questa interazione non coinvolge la creazione o rottura di legami covalenti, ma solo la creazione o rottura di legami molto più deboli, tra cui i legami idrogeno. Ciò che segue è un complesso meccanismo di ripiegamento, detto **folding**, nel quale la catena proteica assume una forma compatta e stabile, spingendo tutte le molecole d'acqua verso l'esterno. Questa conformazione è chiamata **conformazione nativa**, o attiva, della proteina ed è quella in cui la proteina svolge correttamente le proprie funzioni biologiche. Infatti la forma tridimensionale della proteina determina se essa sarà capace, o meno, di svolgere le proprie funzioni correttamente. Al contrario la conformazione casuale che la proteina presenta prima del folding viene chiamata **stato denaturato**. Inoltre chiameremo **unfolding** di una proteina il processo inverso del folding, ovvero il dispiegamento della proteina a partire da uno

stato ripiegato. In termini più specifici l'unfolding è il processo che porta la proteina dal suo stato nativo a quello denaturato. La via di ripiegamento (o folding) di una catena polipeptidica è complicata e non tutti i principi che guidano questo processo sono stati chiariti, ma gli studi effettuati fino ad ora hanno consentito di sviluppare diversi modelli plausibili. In uno di questi, il processo è visto come una serie di tappe con una gerarchia ben precisa, nella Figura 2.3 possiamo osservarne un esempio attraverso una simulazione di una via di ripiegamento di parte di una proteina chiamata villina.

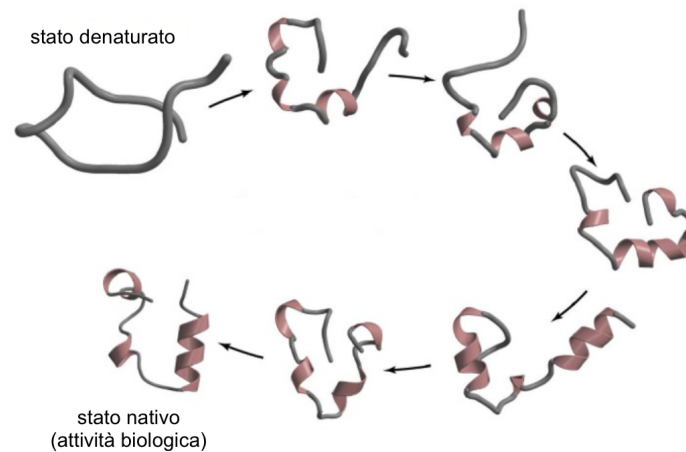


Figura 2.3: Esempio di folding proteico: simulazione di una via di ripiegamento di parte di una proteina chiamata villina.

A livello termodinamico, il processo di ripiegamento può essere visto come una specie di imbuto di energia libera. Il **paesaggio energetico** (o paesaggio di energia libera) è la mappatura di tutte le conformazioni possibili di un'entità molecolare. Ad esempio, una proteina può assumere teoricamente un numero di conformazioni praticamente infinito, nella realtà però essa si ripiega in una struttura (secondaria e terziaria) che minimizza l'energia libera. Per maggiori chiarimenti si veda l'ipotesi del "folding funnel" in [17]. In Figura 2.4 possiamo osservare come varia il paesaggio energetico durante il processo di folding di una proteina. Gli stati non ripiegati (o denaturati)

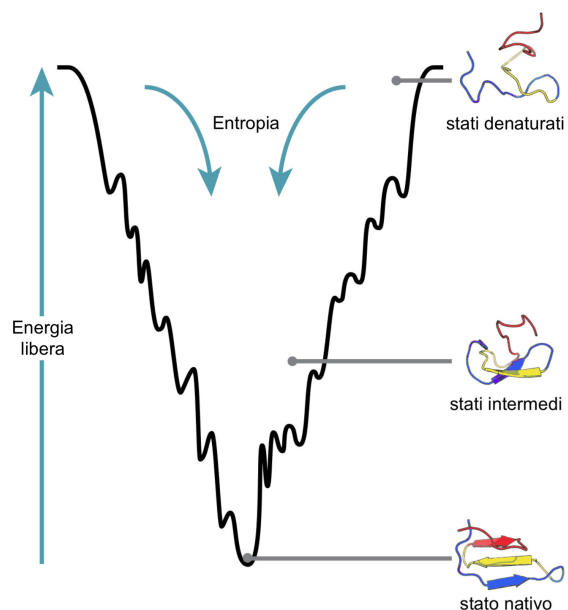


Figura 2.4: Processo di folding dal punto di vista termodinamico.

possiedono un elevato grado di **entropia** conformazionale ed elevati livelli di **energia libera**. Quando il processo di ripiegamento comincia il numero di conformazioni possibili diminuisce: tale diminuzione è rappresentata dalla forma a imbuto, mentre le piccole depressioni ai lati dell'imbuto indicano stati intermedi semistabili che rallentano leggermente la velocità del processo. Alla fine dell'imbuto le conformazioni presenti si riducono al solo stato nativo (oppure ad un piccolo insieme di conformazioni native). È importante precisare che, nonostante si parli per ogni proteina di una specifica conformazione nativa, in realtà molte proteine subiscono cambiamenti conformazionali controllati anche dopo che il processo di folding è terminato. Questi cambiamenti sono parte integrante delle loro attività, inoltre esse possono presentare regioni con una struttura casuale o amorfa. I difetti nel processo di folding delle proteine possono costituire le basi molecolari per un certo numero di malattie genetiche nell'uomo tra cui, ad esempio, la fibrosi cistica.

In generale possiamo pensare alle proteine come a piccole “macchine” che si auto-assemblano e che processano altre molecole in maniera specifica, ripeti-

bile e controllabile. La **forma tridimensionale** della proteina assemblata è determinata direttamente dalla sua sequenza amminoacidica; quest'ultima, a sua volta, è codificata dalla sequenza di DNA del gene per tale proteina. Dunque, se la sequenza di DNA viene mutata, verrà mutata anche la proteina risultante. Molte malattie sono causate proprio da proteine che non svolgono correttamente il proprio lavoro; ad esempio una proteina mal ripiegata sembra essere la causa di un certo numero di malattie rare nei mammiferi caratterizzate da degenerazione del tessuto nervoso. Tra queste la meglio conosciuta è forse quella detta della “mucca pazza” (encefalopatia bovina spongiforme), o altre patologie della famiglia delle encefalopatie spongiformi, malattie che causano demenza e perdita del coordinamento e che portano alla morte. Allo stesso tempo la maggior parte dei farmaci è costituita da piccole molecole che si attaccano alle proteine e le attivano/disattivano, o ne cambiano in qualche modo la funzione. Questi farmaci presentano sempre degli effetti collaterali, in quanto influiscono non solo sulle proteine difettose, ma anche su alcune proteine sane. Se arriveremo a comprendere appieno la forma esatta della proteina difettosa, potremo creare farmaci che colpiscano solo queste ultime, ovvero farmaci molto efficaci e senza effetti collaterali. Per questa ragione, comprendere il processo di folding ed essere in grado di predire la conformazione nativa di una proteina a partire dalla sua sequenza amminoacidica è un problema di grande rilevanza dal punto di vista sia biologico che medico.

2.2 Il DNA

Il *DNA*, Acido Desossiribonucleico, è la molecola (o gruppo di molecole) responsabile della codifica di tutta l'informazione genetica negli organismi viventi. L'informazione contenuta nel DNA determina tutti i tratti ereditari, ad esempio il colore degli occhi o la predisposizione alle malattie ereditarie. La struttura del DNA è stata notoriamente scoperta negli anni '50 da Francis Crick e James Watson, sulla base delle immagini a raggi X di Ro-

salind Franklin. Come possiamo vedere in Figura 2.5, il DNA è composto da coppie di unità ripetute chiamate **nucleotidi**. Ogni nucleotide è formato da un gruppo fosfato, uno zucchero (desossiribosio) ed una delle quattro basi: Adenina, Timina, Citosina, Guanina. Questi nucleotidi formano una struttura simile ad una scala a pioli attorcigliata, la famosa **doppia elica**. Più precisamente il DNA è costituito da due **catene** elicoidali, avvolte attorno allo stesso asse per formare una doppia elica destrorsa. Lo scheletro

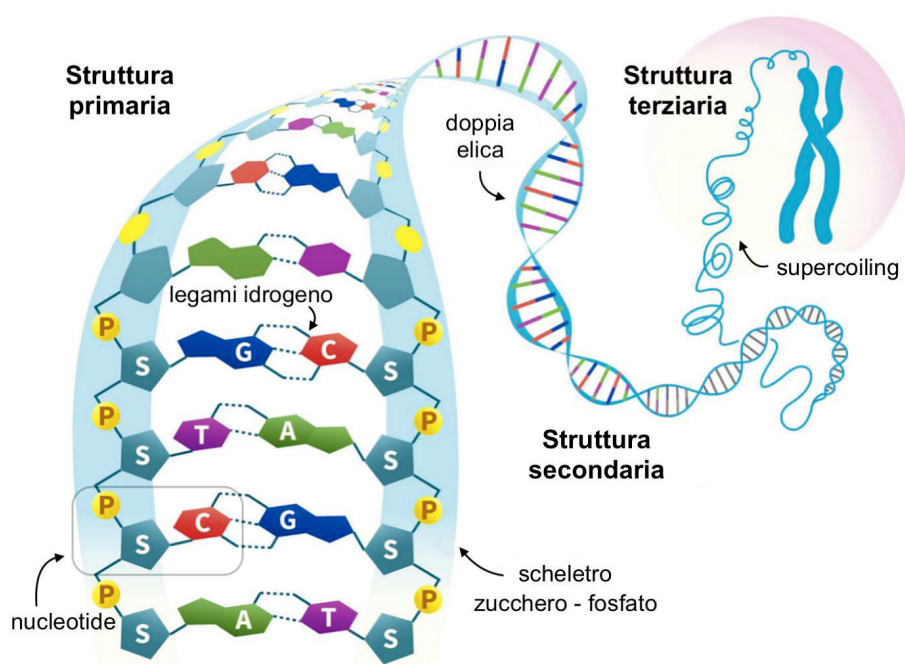


Figura 2.5: Struttura del DNA su vari livelli: P rappresenta il gruppo fosfato, S lo zucchero desossiribosio, A T C G rappresentano rispettivamente le quattro basi Adenina, Timina, Citosina, Guanina.

(ovvero le due catene che possiamo immaginare come i montanti della scala) è formato dall'alternarsi di molecole di zucchero e gruppi fosfati, e si trova all'esterno della doppia elica, cioè a contatto con l'ambiente circostante. I "pioli", invece, sono formati da **coppie di basi** che seguono una regola precisa: A si accoppia solo con T e C solo con G. Un "piolo" di tipo AT è formato da due legami idrogeno tra A e T, mentre un "piolo" di tipo CG è

più forte in quanto consiste in tre legami idrogeno tra C e G. La lunghezza del DNA è misurata in termini di coppie di basi (bp); ad esempio il DNA del genoma umano è composto approssimativamente da 3 miliardi di coppie di basi, mentre quello di un batterio comune, come l'*Escherichia coli*, è di circa 4.4 milioni di coppie di basi. Ogni nucleotide ha una lunghezza di circa 3 Angstroms ($= 3 \times 10^{-10}m$) e l'ampiezza media della doppia elica è di circa 25 Angstroms. Lo zucchero presente nel DNA (il desossiribosio) è uno zucchero pentoso, cioè a cinque atomi di carbonio. In ogni zucchero, il carbonio in posizione 3' si lega ad uno dei gruppi fosfato adiacenti e il carbonio in posizione 5' all'altro gruppo fosfato adiacente. Questi legami conferiscono ad ogni catena un'orientazione e si può osservare che i due filamenti sono allineati antiparallelemente uno rispetto all'altro: il cosiddetto "filamento di Crick" va da 5' a 3', mentre il "filamento di Watson" va da 3' a 5'. Di solito le sequenze di nucleotidi si scrivono seguendo il filamento di Crick, ovvero la direzione da 5' a 3'. Queste sequenze, ovvero la **struttura primaria** del DNA, codificano tutta l'informazione ereditaria, così come le istruzioni per tutti i processi cellulari. Per quanto riguarda la **struttura secondaria**, sia *in vitro* che *in vivo* compare una collezione di forme sorprendentemente diversificate tra loro. Ai fini della nostra trattazione considereremo la struttura prevalente: la doppia elica descritta in precedenza, nella quale ogni torsione completa presenta circa 10.5 coppie di basi. Anche la struttura secondaria del DNA, ovvero la sua geometria elicoidale, ha profonde conseguenze dal punto di vista biologico. Ad esempio è più difficile accedere alle basi in una struttura a doppia elica rispetto ad una struttura a scala lineare, tale difficoltà si traduce in un minor numero di modifiche accidentali della sequenza del DNA e quindi in una maggiore protezione del codice genetico. Circa dieci anni dopo la scoperta della doppia elica, furono individuate le caratteristiche fondamentali della **struttura terziaria** del DNA; si scoprì che anche l'asse del DNA può essere contorto nello spazio, dando vita al cosiddetto DNA superavvolto, o **supercoiled**. Spesso, come nel caso del genoma umano, i filamenti di DNA sono lineari e sono attaccati ad un'impalcatura proteica

che ne vincola la topologia; ma è anche frequente imbattersi in molecole di DNA con asse **circolare**, ovvero nel quale le due estremità della doppia elica sono legate covalentemente. È il caso, ad esempio, del DNA genomico dei batteri, del DNA dei cloroplasti e di quello dei mitocondri umani. Può accadere, inoltre, che l'asse centrale del DNA si presenti sotto forma di **nodo** o **link**. L'informazione portata dal DNA, sotto forma di sequenze di coppie di nucleotidi base, non è influenzata da come esso è "aggrovigliato"; tuttavia è stato scoperto recentemente che la tipologia di nodo formato dal DNA ha un effetto importante sulla **funzione** della molecola di DNA nella cellula (si veda ad esempio [20]).

Alcuni link di DNA, chiamati in letteratura biologica **catenani**, compaiono ad esempio durante la **replicazione** (meccanismo attraverso cui viene prodotta una copia del DNA cellulare) e la **ricombinazione** (riarrangiamento del DNA che sta alla base della variabilità genetica degli organismi). Osserviamo che il vincolo dell'asse circolare, abbinato ai due livelli di avvolgimento (doppia elica e supercoiling) è sufficiente a produrre una topologia non banale del DNA.

2.2.1 Il supercoiling

Analizziamo più nello specifico che cosa si intende con il termine *supercoiling*, o superavvolgimento. Esso sta ad indicare "l'avvolgimento di qualcosa già avvolto": in particolare il DNA è avvolto in una forma di doppia elica nella quale le due catene ruotano attorno ad un asse centrale, un ulteriore ripiegamento o una torsione di tale asse su se stesso determinano un superavvolgimento (o **supercoiling**) del DNA (si veda la Figura 2.6). Al contrario, se non vi è alcun ripiegamento dell'asse su se stesso, si dice che il DNA è in uno stato **rilassato**.

Il supercoiling rappresentato in Figura 2.6, in cui l'asse del DNA si avvolge su se stesso in maniera elicoidale, viene chiamato **supercoiling plectonemico**. In seguito ci riferiremo principalmente a questo tipo di superavvolgi-

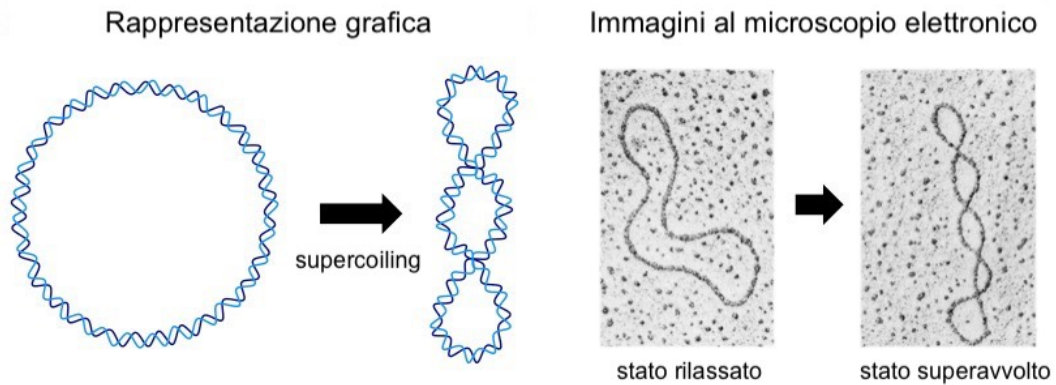


Figura 2.6: Supercoiling di una molecola circolare di DNA.

mento che quindi chiameremo semplicemente supercoiling, anche se in realtà esiste un altro tipo di supercoiling, chiamato **supercoiling a solenoide**, che si manifesta principalmente quando il DNA si avvolge attorno a strutture proteiche.

Osserviamo brevemente in quali casi si presenta il DNA superavvolto e perchè. Innanzitutto, il DNA cellulare è composto da filamenti lunghissimi che devono essere immagazzinati in uno spazio ristretto; per questo motivo le molecole di DNA devono essere strettamente compatte. Come si può immaginare, il DNA che si trova in questo stato altamente impacchettato è superavvolto, ma si è osservato che molte molecole di DNA circolare rimangono superavvolte anche dopo che sono state separate dagli altri componenti cellulari. Ciò indica che il supercoiling è un aspetto **intrinseco** della struttura terziaria del DNA. Infatti si tratta, in genere, della manifestazione di una **tensione strutturale** della molecola. Per quanto riguarda il DNA cellulare, il supercoiling non solo è frequente, ma onnipresente; esso rappresenta lo stato nativo delle molecole di DNA praticamente in tutte le cellule ed è rigidamente regolato all'interno della cellula stessa. Si è osservato che *in vivo* il suo orientamento favorito è negativo, cioè generalmente l'elica formata dall'asse del DNA si avvolge in senso opposto rispetto alla doppia elica originale. Il supercoiling

si ritrova in tutti gli organismi viventi per tre ragioni fondamentali: esso facilita la **compattificazione** delle lunghe molecole di DNA all'interno delle cellule, agevola l'**accesso alle coppie di basi** durante i processi metabolici a cui è soggetto costantemente il DNA ed infine permette di immagazzinare **energia libera**, che servirà come motore per i processi metabolici stessi.

Vediamo ora alcuni esempi di questi processi che coinvolgono il DNA e che producono, o modificano, il supercoiling della molecola. Consideriamo, in particolare: la **replicazione** e la **trascrizione** (processo mediante il quale le informazioni contenute nel DNA vengono trascritte in una molecola complementare di RNA). Durante questi processi risulta necessaria una temporanea separazione dei due filamenti che compongono l'elica di DNA, resa difficile dal fatto che essi sono avvolti l'uno sull'altro. Per questo motivo la separazione causa una tensione strutturale che, riversandosi sulla parte restante della molecola, produce un superavvolgimento della stessa. Vediamo una simulazione di questo meccanismo in Figura 2.7.



Figura 2.7: Supercoiling prodotto dalla separazione delle due catene di una struttura a doppia elica (fonte dell'immagine [23]).

In generale, il superavvolgimento avviene quando il DNA è sottoposto ad una sorta di deformazione strutturale. In quasi tutti i casi la deformazione è il risultato di un **disavvolgimento** del DNA circolare a doppia elica. In altre

parole, quando il DNA presenta un numero di giri di elica *minore* rispetto a quelli che ci si aspetterebbe nella sua forma rilassata, esso viene a trovarsi in una forma meno stabile e svantaggiata dal punto di vista termodinamico. La deformazione, allora, può essere riequilibrata tramite l'avvolgimento dell'asse del DNA su se stesso, ovvero tramite la formazione di un superavvolgimento. Questa situazione si presenta molto più spesso di quanto si creda, infatti la cellula disavvolge attivamente il suo DNA attraverso dei processi regolati da enzimi. La deformazione del DNA così ottenuta rappresenta una forma di immagazzinamento di energia all'interno della cellula. Inoltre le cellule mantengono il DNA in uno stato parzialmente disavvolto per facilitarne il compattamento e per agevolare la separazione delle due catene e l'accesso alle coppie di basi da parte di alcuni enzimi. Ovviamente lo stato disavvolto può essere mantenuto solo se il DNA è di forma circolare chiusa o se è vincolato dall'interazione con proteine in modo che le catene non siano libere di ruotare su se stesse. In caso contrario la rotazione libera provocherebbe un ritorno spontaneo allo stato rilassato del DNA. *In un DNA circolare chiuso, invece, il numero di giri di elica è fisso e non può essere modificato, a meno che una delle catene di DNA non sia interrotta.* Il numero di giri di elica di una molecola di DNA è quantificabile e fornisce perciò una descrizione precisa del superavvolgimento.

Le proprietà quantificabili del superavvolgimento, che sono state definite negli ultimi decenni, hanno fornito molti chiarimenti sulla struttura e sulla funzione del DNA. Come vedremo, lo studio di questi argomenti si avvale abbondantemente della **topologia** e della **teoria dei nodi**. Nei capitoli successivi tradurremo dunque il problema dello studio del DNA in termini topologici. Per il momento teniamo presente che, nel caso del DNA, le *deformazioni continue* comprendono modificazioni conformazionali dovute a movimenti termici o a interazioni con proteine o altre molecole; mentre *deformazioni discontinue* richiedono la rottura delle catene del DNA. Dunque, per le molecole di DNA circolare, una proprietà topologica è una proprietà che rimane inalterata quando avviene una deformazione, senza rotture, delle

catene del DNA. Due forme di DNA circolare che differiscono solo per una proprietà topologica sono chiamati **topoisomeri**.

2.2.2 Azione degli enzimi topoisomerasi I e II.

Il superavvolgimento è un processo altamente regolato che influenza numerosi aspetti del metabolismo del DNA. In ogni cellula vi sono enzimi il cui unico compito è quello di regolare la quantità di supercoiling del DNA. Questi enzimi, scoperti all'inizio degli anni '70, sono chiamati **topoisomerasi** e svolgono un ruolo importante specialmente in processi quali la replicazione e l'impacchettamento delle molecole di DNA. Esistono due famiglie di topoisomerasi chiamate di tipo I e II; in questo capitolo sorvoleremo su molti dettagli, in particolare sulle differenze minori tra i vari membri di ciascuna famiglia, per maggiori approfondimenti si veda [23].

Le **topoisomerasi di tipo I** agiscono nel modo seguente: si legano alla molecola di DNA, rompono in maniera transitoria una delle due catene della doppia elica e fanno passare l'altra catena attraverso il taglio, infine richiudono il taglio e rilasciano la molecola. Attraverso questo processo, illustrato in Figura 2.8, le topoisomerasi di tipo I **rilassano** i superavvolgimenti negativi nei procarioti e tutti i tipi di superavvolgimento negli eucarioti.

Una delle situazioni in cui le topoisomerasi di tipo I entrano in gioco è quando la doppia elica del DNA viene aperta, cioè quando vengono separate le due catene. Ciò avviene, ad esempio, durante il processo di **replicazione** del DNA. Come abbiamo osservato nella Sezione 2.2.1 sul supercoiling, se in una molecola di DNA con estremi vincolati vengono separate le due catene, allora nella regione rimanente si produrrà un aumento del supercoiling (si veda la Figura 2.7). In queste condizioni, se un enzima topoisomerasi di tipo I non rilassasse il superavvolgimento, la forza di torsione potrebbe diventare insostenibile e la molecola di DNA potrebbe rompersi. Probabilmente è per questo motivo che le topoisomerasi di tipo I sono state trovate in tutti gli organismi studiati finora, infatti la mancanza di questi enzimi sarebbe letale per la cellula.

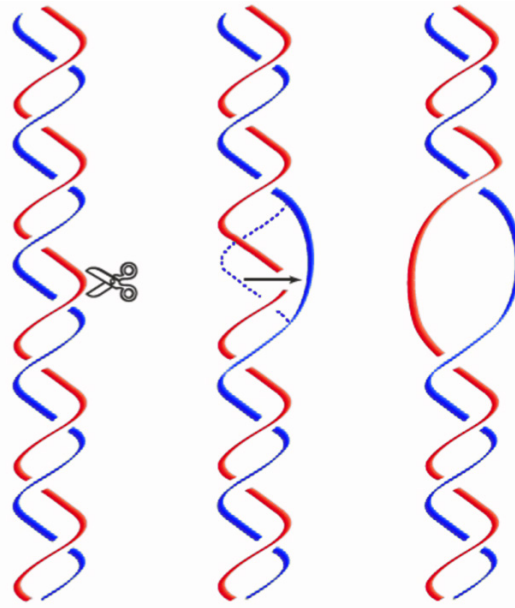


Figura 2.8: Azione degli enzimi topoisomerasi di tipo I (fonte dell'immagine [5]).

Le **topoisomerasi di tipo II**, invece, rompono entrambe le catene del DNA e consentono il passaggio di un'altra parte di doppia elica attraverso la rottura, esaminiamo questo processo più nel dettaglio. Un generico enzima della famiglia delle topoisomerasi di tipo II si lega alla molecola di DNA in una regione in cui sono giustapposte due doppie eliche (ad esempio in una regione superavvolta o in un incrocio di un nodo o link di DNA). La prima doppia elica, generalmente chiamata *segmento G* (che sta per "Gate" ovvero "cancello"), viene tagliata per permettere il passaggio dell'altra doppia elica, chiamata *segmento T* (che sta per "Transported" ovvero "trasportato"). Tale passaggio avviene in maniera unidirezionale e al suo termine il segmento G viene richiuso e la molecola rilasciata (si veda la Figura 2.9).

Anche le topoisomerasi di tipo II si trovano in tutti gli organismi e sono essenziali alla vita delle cellule. La loro funzione primaria è quella di **cambiare il tipo di nodo o link** formato dal DNA, ma esse possono anche **diminuire o incrementare il supercoiling** della molecola. Questi en-

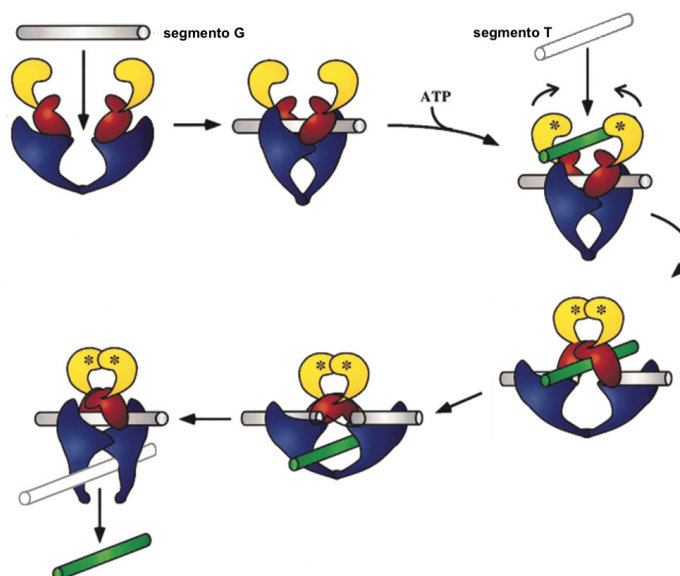


Figura 2.9: Azione degli enzimi topoisomerasi di tipo II (fonte dell'immagine [5]).

zimi entrano in gioco, ad esempio, durante il processo di **replicazione** di molecole di DNA circolari. Infatti questo processo ha come risultato finale due molecole circolari di DNA concatenate, e quindi vincolate l'una all'altra; se un enzima topoisomerasi di tipo II non dividesse le due molecole risultanti, la cellula non potrebbe dividersi propriamente nelle due cellule figlie e morirebbe. A causa del loro ruolo cruciale in diversi processi cellulari, le topoisomerasi di tipo II sono considerate ottimi bersagli per i farmaci sia contro il cancro che contro le malattie infettive negli esseri umani. Ad esempio, siccome le topoisomerasi di tipo II presenti nei batteri sono diverse da quelle umane, molti antibiotici funzionano inibendo tali enzimi, e uccidono i batteri proprio impedendo la divisione delle molecole di DNA circolare durante la replicazione. In quest'ottica, la comprensione dei meccanismi d'azione delle topoisomerasi di tipo II diventa un problema di grande importanza nel campo biochimico e farmaceutico. In generale, quindi, *il grado di superavvolgimento del DNA ed il tipo di nodo, o link, che esso forma sono regolati dall'attività*

delle topoisomerasi I e II.

Capitolo 3

Topologia del DNA

In questo capitolo vedremo alcuni esempi pratici di come la teoria dei nodi si possa applicare alla biochimica e, in particolare, allo studio del DNA. Prenderemo come riferimento i testi [1, 22] ed un articolo del 2009 di Dorothy Buck dal titolo: “DNA topology” [5] per proporre un **modello del DNA** circolare prima sotto forma di anello topologico e successivamente di link a due componenti. Dopo di che ne individueremo le proprietà geometriche e topologiche e discuteremo degli **effetti delle topoisomerasi** (gli enzimi che regolano la topologia della molecola) dal punto di vista matematico. Ci concentreremo poi su un altro processo che coinvolge il DNA e che, come effetto secondario, ne modifica la topologia: la **ricombinazione sito specifica**. Per studiare questo processo proporrò un modello basato sul concetto di tangle e, per concludere, mostreremo alcuni esempi del suo utilizzo.

3.1 Il DNA come nastro topologico

Consideriamo una molecola di DNA *con asse circolare*, ovvero una molecola in cui le estremità delle due catene che formano la doppia elica sono legate covalentemente. Questo genere di molecola è molto frequente in natura: il DNA genomico dei batteri, il DNA dei cloroplasti e quello dei mitocondri umani sono esempi di DNA con asse circolare. La struttura del DNA

circolare può essere modellizzata attraverso un **nastro** (o anello) topologico, come vediamo in Figura 3.1

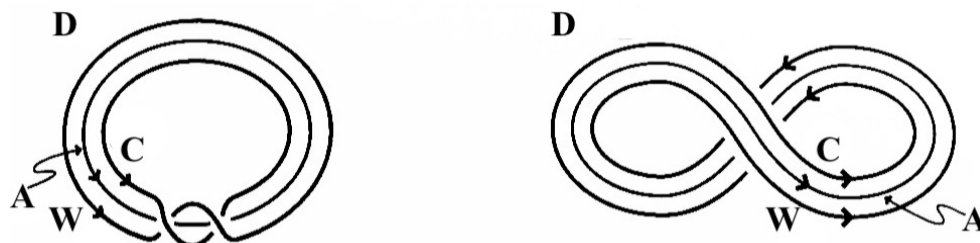


Figura 3.1: Modello di DNA circolare sotto forma di anello topologico (fonte dell'immagine [22]).

Ad ognuna delle due catene del DNA, associamo una curva nello spazio (cioè uno dei due bordi del nastro), che chiameremo rispettivamente C per il filamento di Crick e W per il filamento di Watson. Anche se nella struttura chimica i due filamenti sono orientati antiparallelemente, nel modello è conveniente assegnare a C e W orientazioni parallele tra loro, che ereditano dall'orientazione dell'asse A (scelta a piacere). Questo causa una variazione controllata del linking number, che cambierà di segno (si veda l'Osservazione 4). Si può pensare alle due immagini in Figura 3.1 come a rappresentazioni della stessa molecola di DNA in due momenti consecutivi. In entrambi i casi, il nastro ottenuto dalla loro modellizzazione può essere pensato come un link a due componenti $D = C \sqcup W$:

- nel primo caso (immagine a sinistra) otterremo:
 $Lk(D) = Lk(C, W) = 1, Wr(D) = 0, Tw(D) = 1;$
- nel secondo caso (immagine a destra) otterremo:
 $Lk(D) = Lk(C, W) = -1, Wr(D) = -1, Tw(D) = 0.$

Come previsto dal Teorema 1.3.4, si può osservare che linking number, twisting number e writhe rispettano la relazione $Lk = Tw + Wr$. Ciò significa che al variare di una di queste quantità le altre due saranno costrette a riadattarsi per compensare il cambiamento; in altre parole, se si modifica una

delle caratteristiche geometriche della molecola, si ottiene un effetto rilevante su tutta la sua struttura. Vediamo ora quali caratteristiche della struttura biologica del DNA sono individuate da questi tre invarianti.

Il **linking number**, come abbiamo visto, è un invariante topologico del nastro e del link, per questo motivo esso varia solo in presenza di una deformazione discontinua, ovvero di una rottura di una o entrambe le catene della doppia elica. Quando la molecola è in stato rilassato, il linking number è semplicemente il numero di coppie di basi diviso per il numero di coppie di basi per giro di elica, che normalmente è vicino a 10.5. Una variazione del linking number (vedremo che in genere si tratta di una sua riduzione) può essere causata dall'azione delle topoisomerasi che agiscono tagliando una o entrambe le catene della doppia elica.

Il **twisting number** è un indicatore del *numero di giri di elica* della molecola di DNA circolare, nello stato rilassato dunque coincide con il linking number, ma essi si differenziano in presenza di supercoiling.

Il **writhe**, invece, rappresenta proprio questo *supercoiling* (positivo se ha segno + e negativo se ha segno -). Infatti, come abbiamo già osservato, esso ci dà una misura di quanto l'asse del nastro topologico sia contorto nello spazio. Riesaminiamo in quest'ottica la Figura 3.1, tenendo presente che, grazie alla relazione $Lk = Tw + Wr$, possiamo pensare al linking number come se fosse composto da due componenti strutturali. Dunque, quando si verifica un cambiamento del linking number, parte della tensione risultante è solitamente compensata da modificazioni del writhe e parte dal twisting number. Nel caso rappresentato in Figura 3.1, nel passaggio dalla prima conformazione (immagine a sinistra) alla seconda (immagine a destra), il linking number diminuisce passando da 1 a -1 , quindi possiamo ipotizzare che una topoisomerasi abbia agito sulla molecola, come descritto nella Sezione 2.2.2. A questo punto writhe e twisting number sono costretti a riadattarsi alla nuova situazione: il writhe passa da 0 a -1 causando una torsione dell'asse del nastro e quindi un superavvolgimento e, di conseguenza, il twisting number diminuisce passando da 1 a 0. Nell'esempio proposto è presente un solo giro

di elica, rappresentato dal twisting number, che scompare nel passaggio dalla prima conformazione alla seconda. Normalmente in una molecola di DNA il numero di giri di elica è molto maggiore, dunque quando il twisting number diminuisce quello che accade è un disavvolgimento parziale della doppia elica. Il processo appena descritto rappresenta un esempio di un fenomeno alquanto comune: infatti è ormai appurato che diminuire il linking number di un doppio filamento di DNA produce un aumento del supercoiling (positivo o negativo) e che il supercoiling, a sua volta, può permettere un parziale disavvolgimento della struttura a doppia elica. Ma può avvenire anche il processo inverso: in una molecola circolare o con estremi vincolati accade spesso che i due filamenti della doppia elica debbano essere disavvolti e separati per permettere meccanismi quali la replicazione, o in generale per permettere l'accesso alle coppie di basi. Come abbiamo visto nella Sezione 2.2.1, dedicata al supercoiling, questa separazione produce un incremento del supercoiling nella parte restante della molecola (si veda la Figura 2.7). Ancora una volta il Teorema 1.3.4 si adatta perfettamente a descrivere la situazione. Infatti, dal punto di vista della teoria dei nodi, quello che avviene è semplicemente una diminuzione del twisting number (disavvolgimento), senza possibilità di modificare il linking number (poiché le catene non vengono rotte), questo porta necessariamente ad un incremento dei termini di writhe (superavvolgimento). Riportiamo in Figura 3.2 qualche altro esempio che ci permetta di visualizzare la relazione tra linking number, twisting number e writhe.

3.2 Il modello di nastro topologico applicato allo studio del supercoiling

Consideriamo le molecole di DNA circolare, il caso più comune in cui si osserva un writhe non banale è quello del *supercoiling plectonemico*, in cui l'asse della molecola si avvolge su se stesso in maniera elicoidale. Le molecole di DNA estratto dalle cellule di una grande varietà di organismi, tra cui quello umano, presentano tutte approssimativamente la stessa **densità**

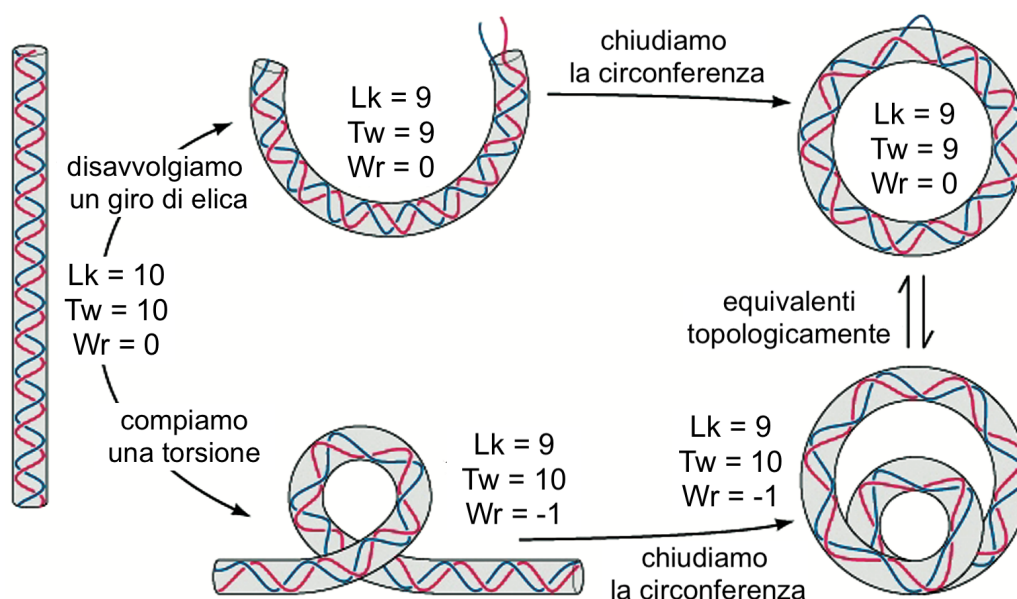


Figura 3.2: Esempi di variazione del linking number, twisting number e writhe di una molecola di DNA (fonte dell'immagine [5]).

di **supercoiling** (plectonemico). Questa densità può essere quantificata attraverso la differenza di linking number tra lo stato superavvolto e lo stato rilassato del DNA. Sia N il numero di coppie di basi della molecola circolare di DNA ed h_0 il numero di coppie di basi per ogni giro di elica (nello stato rilassato), allora il linking number che la molecola avrebbe nel suo stato rilassato è dato da $Lk_0 := N/h_0$. Dunque possiamo definire la densità di supercoiling σ come:

$$\sigma = \frac{Lk - Lk_0}{Lk_0}.$$

Il modello di nastro topologico permette di spiegare anche alcune delle funzioni del supercoiling esaminate nella Sezione 2.2.1. Ad esempio, esso chiarisce in che modo il supercoiling facilita l'**accesso alle coppie di basi** da parte di alcuni enzimi. Infatti, se la molecola di DNA rimane chiusa covalentemente durante il supercoiling, il suo linking number rimane necessariamente costante. Allora, per la relazione $Lk = Tw + Wr$, la variazione del writhe dovrà essere compensata da una variazione di twisting number, cioè

da una diminuzione del numero di giri di elica. Questo disavvolgimento, a sua volta, causa una leggera apertura della doppia elica che permette un più facile accesso alle coppie di basi, agevolando una grande varietà di processi biologici.

Un'altra delle funzioni biologiche del supercoiling è l'immagazzinamento di **energia libera**. Infatti molte delle reazioni che coinvolgono il DNA, come ad esempio la replicazione, richiedono di separare le due catene della doppia elica, anche se si tratta di un processo difficile e costoso dal punto di vista energetico. Fortunatamente il Teorema 1.3.4 implica che il linking number del DNA superavvolto negativamente (cioè nel suo stato naturale all'interno delle cellule) sia minore di quello del DNA rilassato ed è proprio questo 'deficit' in termini di linking number che fornisce l'energia necessaria alla separazione locale dei due filamenti. Infatti si è osservato sperimentalmente che l'energia libera ΔG del supercoiling, per DNA circolare di lunghezza superiore a 2000 bp, rispetta la relazione:

$$\Delta G = K(Lk - Lk_0)^2$$

dove K è una costante di proporzionalità che dipende dal numero di coppie di basi N . Per N molto grandi si è visto sperimentalmente che il prodotto NK è indipendente dalla lunghezza della molecola.

3.3 Regolazione della topologia del DNA

In generale, come visto nel Capitolo 2, si chiamano **topoisomeri** le molecole di DNA che differiscono solamente per una proprietà topologica; d'ora in avanti useremo questo termine per indicare molecole circolari di DNA che differiscono tra loro solo per il **linking number**, dato che, come osservato sopra, questo è l'invariante che controlla la topologia del DNA. Dalla termodinamica standard e da quanto detto finora, si ottiene che la concentrazione di DNA con linking number Lk_i è data da:

$$[Lk_i] := \frac{1}{Z} \exp\left(\frac{-\Delta G}{RT}\right) = \frac{1}{Z} \exp\left(\frac{-K(Lk_i - Lk_0)^2}{RT}\right),$$

dove Z è una costante di normalizzazione, T è la temperatura in Kelvin e R è la costante di Boltzmann. In altre parole, per una popolazione in stato di equilibrio di grandi anelli di DNA che differiscono solo per il linking number (e quindi topoisomeri), i valori del linking number Lk_i si distribuiscono con distribuzione normale centrata in Lk_0 con deviazione standard $\sqrt{\frac{RT}{2K}}$. Per mantenere questo equilibrio nella popolazione, si sono evolute le due famiglie di enzimi: le **topoisomerasi di tipo I e II**, che hanno come funzione primaria quella di trasformare il linking number delle molecole di DNA, ovvero di convertire la molecola in un suo topoisomero. Infatti, come abbiamo già osservato (si vedano le Figure 2.8 e 2.9), le **topoisomerasi di tipo I** tagliano una delle catene del DNA e fanno passare l'altra attraverso il taglio, per poi richiuderlo. Le **topoisomerasi di tipo II**, invece, tagliano entrambe le catene della doppia elica di DNA e fanno passare attraverso il taglio un'altra doppia elica, per poi richiuderlo. In entrambi i casi, l'azione dell'enzima causa un *cambiamento del linking number* e, come già osservato, questo cambiamento produrrà una variazione anche del twisting number e del writhe (e quindi del livello di supercoiling).

Si potrebbe pensare che le topoisomerasi di tipo II, essendo molecole di piccole dimensioni, agiscano indipendentemente dalla topologia globale della molecola. Sorprendentemente, invece, l'azione di questi enzimi segue una direzione preferenziale: essi prediligono la semplificazione o lo scioglimento di nodi e link, piuttosto che la loro creazione. Questo fenomeno è conosciuto in letteratura come **semplificazione topologica**. Non solo le topoisomerasi di tipo II semplificano la topologia del DNA, ma lo fanno in maniera straordinariamente efficiente; tant'è che i meccanismi che utilizzano per raggiungere questo scopo sono stati, e sono ancora, ampiamente dibattuti.

3.4 La ricombinazione sito-specifica

Come abbiamo accennato nel Capitolo 2, i nodi e i link di DNA compaiono in un gran numero di processi cellulari; ad esempio possono risultare come

prodotto della replicazione, della ricombinazione e trasformarsi attraverso l'azione degli enzimi. Oltre alle topoisomerasi, esistono altre due famiglie di enzimi che danno vita e che modellano i nodi e i link di DNA: le *ricombinasi* e le *trasposasi*. Concentriamoci ora su un particolare tipo di ricombinasi: le **ricombinasi sito-specifiche**, una famiglia di proteine che si occupa della *ricombinazione sito-specifica*, ovvero del rimescolamento della sequenza genetica. Il risultato dell'azione di questo enzima è l'eliminazione, l'inserimento o l'inversione di segmenti di DNA (si veda la Figura 3.3).

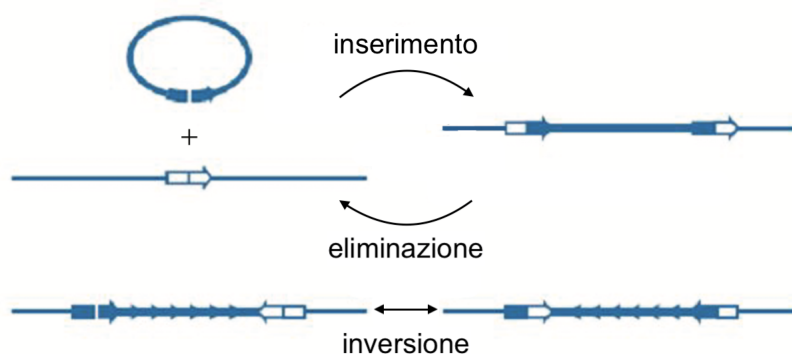


Figura 3.3: Possibili azioni della ricombinasi sito-specifica (fonte dell'immagine [5]).

La ricombinazione sito-specifica ha un ruolo chiave in numerosi processi biologici, inclusi alcuni passaggi cruciali nelle infezioni virali. Inoltre molte industrie farmaceutiche e agricole mostrano un interesse sempre crescente nella modifica genetica degli organismi o nel testare se un particolare gene può causare una malattia e uno strumento per manipolare in modo preciso il DNA sono proprio le ricombinasi sito-specifiche. Il cambiamento della topologia del DNA non è la funzione primaria di questi enzimi, ma può essere un prodotto secondario della loro azione. Infatti se la molecola di DNA è superavvolta, i superavvolgimenti possono essere trasformati in nodi o link durante il processo di ricombinazione. Normalmente un sistema di ricombinazione sito-specifica è costituito dalle ricombinasi e da due brevi

segmenti di DNA (30-50 coppie di basi) chiamati *siti di ricombinazione*, che possono appartenere alla stessa molecola o a due diverse molecole di DNA. Nel caso in cui entrambi appartengano alla stessa molecola circolare possono trovarsi tra loro in *orientazione diretta* o *inversa* (si veda la Figura 3.4).

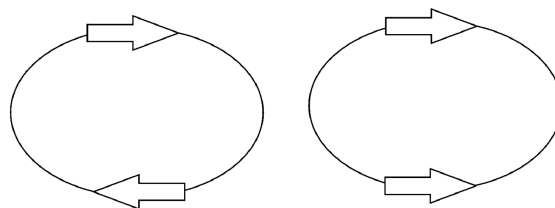


Figura 3.4: I siti di ricombinazione in una molecola circolare di DNA possono avere orientazione diretta (immagine a sinistra) o inversa (immagine a destra).

Le molecole di DNA coinvolte nel processo di ricombinazione, prima che questa avvenga, sono chiamate *substrato*. Consideriamo, ad esempio, come substrato una sola molecola di DNA circolare e analizziamo brevemente come avviene il processo di ricombinazione, che si può suddividere in tre fasi principali (si veda la Figura 3.5):

1. per prima cosa le ricombinasi si legano ai due siti di ricombinazione e li avvicinano, dando vita al cosiddetto *complesso di ricombinazione* definito come la più piccola regione convessa che contiene le ricombinasi e i due siti di ricombinazione. Chiamiamo B il complesso di ricombinazione e osserviamo che è possibile modellarlo con una **palla topologica**;
2. in seguito le ricombinasi tagliano una o entrambe le catene del DNA all'interno dei siti di ricombinazione e le richiudono riarrangiandole in vari modi (dipendenti dalla sottofamiglia a cui appartengono le ricombinasi coinvolte);
3. infine il DNA riarrangiato, chiamato *prodotto*, viene rilasciato.

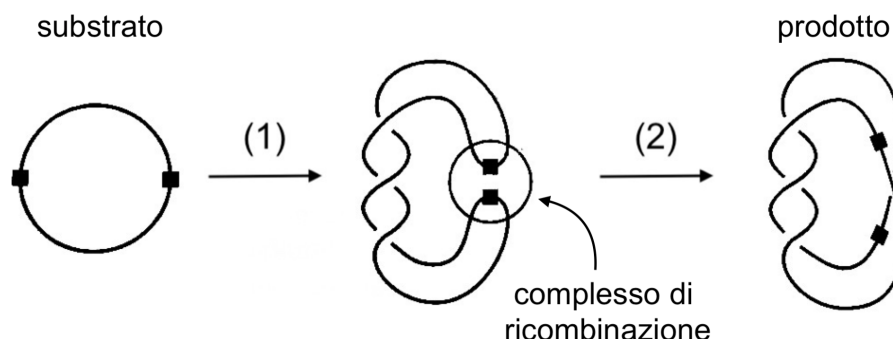


Figura 3.5: Ricombinazione sito-specifica di una molecola circolare di DNA (fonte dell'immagine [22]).

In generale:

- se il substrato è un *nodo* di DNA e i siti di ricombinazione hanno orientazione diretta allora il prodotto sarà un link a due componenti, se l'orientazione è inversa il prodotto sarà di nuovo un nodo;
- se invece il substrato è un *link* a due componenti con un sito di ricombinazione su ogni componente, allora il prodotto sarà un singolo nodo di DNA.

Le ricombinasi sito-specifiche si dividono in due famiglie: le *ricombinasi in serina* (anche dette *resolvasi*) e quelle *in tirosina* (anche dette *integrasi*), che differiscono per i meccanismi con cui tagliano e riarrangiano il DNA. Le ricombinasi in serina possono avere bisogno di uno specifico numero di superavvolgimenti prima di iniziare la ricombinazione; ad esempio, un tipo di ricombinasi in serina chiamata *Tn3 resolvasi* necessita di tre superavvolgimenti negativi per dare inizio al processo. Anche le ricombinasi in tirosina hanno bisogno della presenza di superavvolgimenti, ma la maggior parte di esse tollera una certa variazione del loro numero. Questi superavvolgimenti, che si trovano al di fuori del complesso di ricombinazione, insieme al complesso di ricombinazione stesso vengono chiamati *complesso sinaptico*.

È importante notare che i due tipi di ricombinasi possono agire più volte sulla stessa molecola, ma lo fanno in modo leggermente diverso tra loro. Le ricombinasi in serina, dopo essersi attaccate al DNA, sono in grado di effettuare diversi scambi (o riarrangiamenti) consecutivi, rilasciando la molecola solamente alla fine (ricombinazione *processiva*). Invece le ricombinasi in tirosina tagliano, scambiano i filamenti e rilasciano il DNA; il complesso sinaptico rimane poi in una sorta di configurazione intermedia finché l'enzima non ricomincia l'intero ciclo (ricombinazione *distributiva*).

La ricombinazione sito-specifica può trasformare il DNA circolare superavvolto in un nodo o link. Infatti, dato un substrato in cui i siti di ricombinazione presentano un'orientazione specifica, un singolo ciclo di ricombinazione mediato dalla ricombinasi in serina produce un *unico* prodotto topologico. Ad esempio la Tn3 resolvasi che agisce sul nodo banale con siti di ricombinazione posti in orientazione diretta produce univocamente il *link di Hopf* (si veda Figura 3.9). Se poi si continua con nuovi cicli di ricombinazione processiva si ottengono: il *nodo a otto*, il *link di Whitehead* ed infine un *nodo a 6 incroci* (sempre in Figura 3.9). Al contrario, un singolo ciclo di ricombinazione mediato da una generica ricombinasi in tirosina conduce ad un ampio spettro di nodi e link di DNA. La distribuzione di nodi e link così ottenuti riflette la densità di supercoiling del substrato. Comprendere con precisione quali nodi e link compaiono durante la ricombinazione sito-specifica può fornire un grande aiuto nell'analisi di questo processo biologico. Ci sono diverse modelli e tecniche topologiche aventi questo scopo che utilizzano il concetto di tangle richiamato nella Sezione 1.4.1, nel seguito ne mostreremo alcune.

3.5 Modello a tangle per la ricombinasi

Supponiamo di conoscere la topologia del substrato e del prodotto per una data ricombinasi sito-specifica e di voler determinare i dettagli del percorso che va da uno all'altro. Per fare ciò possiamo costruire un modello che si basa sul concetto di **tangle**; in particolare possiamo rappresentare le diver-

se regioni della molecola di DNA con dei tangles ed interpretare l'azione di una ricombinasi come un cambiamento di uno di essi. Date la topologia del substrato e del prodotto si individuano tutte le combinazioni di tangles che possono spiegare l'azione della ricombinasi presa in esame, alcune di queste soluzioni matematicamente possibili vengono poi escluse grazie a considerazioni biologiche. Quello che si ottiene è una lista di soluzioni delle cosiddette **equazioni di tangles**, che conducono ad un corrispondente insieme di possibilità per il meccanismo di ricombinazione.

Iniziamo presentando due assunzioni che sono state confermate dall'evidenza empirica e che ci aiuteranno nella costruzione del modello.

Proposizione 3.5.1. (Proposition 13.3.1. in [22]) *Si è osservato sperimentalmente che:*

1. *praticamente tutti i prodotti ottenuti dalla ricombinazione sito-specifica, prendendo come substrato il nodo banale, sono **nodi o link razionali**, cioè nodi o link a due ponti;*
2. *il complesso di ricombinazione, ovvero la parte del complesso sinaptico su cui agisce la ricombinasi, è un **tangle** (ogni arco del tangle rappresenta un segmento di doppia elica di DNA).*

Dunque la ricombinazione consiste nel rimpiazzare un tangle con un altro tangle (si veda la Figura 3.6).

Ancora più precisamente, questo processo può essere descritto attraverso la definizione che abbiamo dato di **somma di tangles**. Infatti, dal punto di vista topologico, è perfettamente ragionevole considerare anche O , ovvero la parte del substrato esterna a T , come un altro tangle. In questo modo si arriva a descrivere il **prodotto** semplicemente come il numeratore della somma di O ed R , Figura 3.7.

Dunque vale la seguente “equazione”:

$$N(O + R) = \text{prodotto}.$$

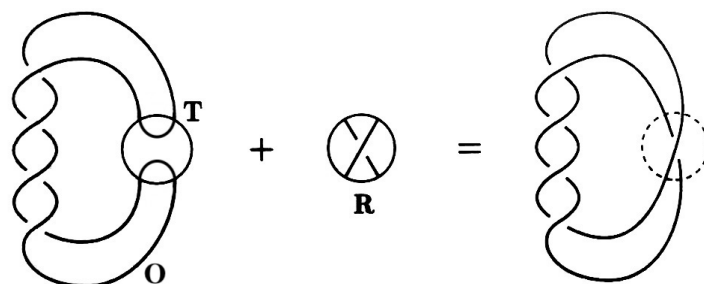


Figura 3.6: Esempio di ricombinazione vista come sostituzione di un tangle T con un altro R (fonte dell'immagine [22]).

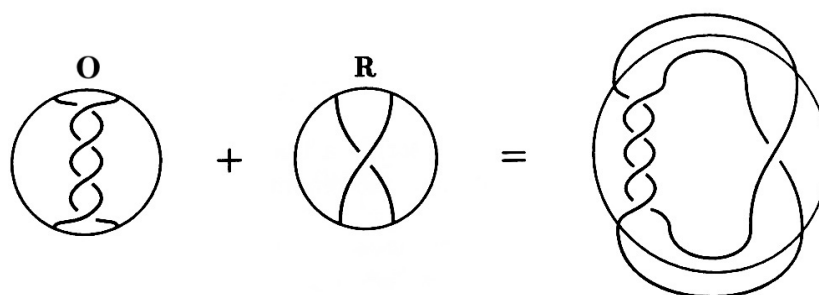


Figura 3.7: Prodotto della ricombinazione visto come numeratore di una somma di tangles (fonte dell'immagine [22]).

A questo punto ci conviene dividere anche il substrato nel tangle esterno O e interno E , in questo modo il **substrato** non è altro che il numeratore della somma di O ed E , come rappresentato in Figura 3.8.

Si ottiene dunque, anche per il substrato, la seguente “equazione”:

$$N(O + E) = \text{substrato}.$$

Quindi, se conosciamo il substrato ed il prodotto, l'ideale sarebbe poter determinare O , E e R dalle due equazioni di tangles trovate. Senza fare ulteriori assunzioni questo non è possibile in quanto sono presenti tre incognite e solamente due “equazioni”, ma grazie ai dati sperimentali possiamo ottenere

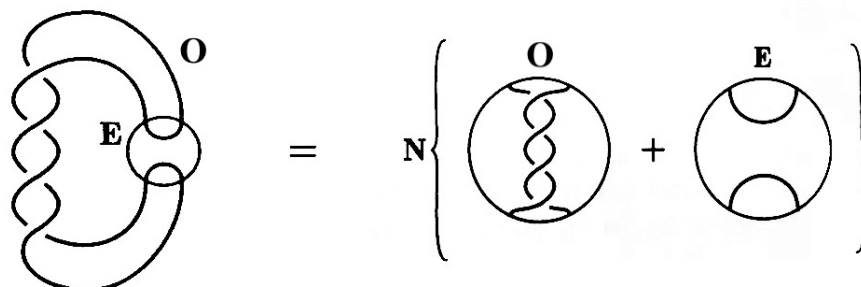


Figura 3.8: Substrato visto come numeratore di una somma di tangles (fonte dell'immagine [22]).

altre informazioni utili. Ad esempio, in seguito all'osservazione dei risultati sperimentali si è formulata la seguente congettura.

Congettura 3.5.2. (Supposition 13.3.3. in [22]) *L'effetto di un enzima, ovvero il cambiamento dal tangle E al tangle R , dipende solamente dall'enzima originale. Dunque questo processo è indipendente da forma, posizione e grandezza del substrato.*

Ciò significa che, per ogni ricombinazione successiva che eseguiremo sulla stessa molecola, otterremo una nuova “equazione” che però non contiene nuove incognite. In questo modo è possibile ottenere un sistema di tre (o più) “equazioni” in tre incognite, e quindi potenzialmente risolvibile. Un'altra osservazione utile, che emerge dai dati sperimentali, è la seguente.

Proposizione 3.5.3. *Il prodotto di una serie di ricombinazioni sito-specifiche si può esprimere come il numeratore di una somma di tangles, cioè nella forma:*

$$N(O + R + R + \cdots + R).$$

Accettando queste due assunzioni (Proposizione 3.5.2 e 3.5.3) è possibile costruire dei modelli potenzialmente risolvibili sia per la ricombinazione sito-specifica mediata dalle ricombinasi in serina che dalle ricombinasi in tirosina.

3.5.1 Modello a tangle per una ricombinasi in serina

Consideriamo un enzima specifico della famiglia delle ricombinasi in serina: la **Tn3 resolvasi**. Questo enzima catalizza la ricombinazione sito-specifica per substrati di DNA circolare in cui i siti di ricombinazione hanno orientazione diretta. Come già accennato, affinché il processo inizi è necessario un numero preciso di superavvolgimenti; quando tale condizione è soddisfatta inizia il processo di riarrangiamento del DNA, che può ripetersi diverse volte prima del rilascio della molecola. Quando il substrato è formato dal nodo banale, il prodotto dell'azione della Tn3 resolvasi è un link a due componenti: il *link di Hopf* (inoltre sappiamo che se si considera l'orientazione della molecola di DNA, il link di Hopf ottenuto dalla ricombinazione avrà linking number -1). Se poi l'enzima effettua ulteriori cicli di ricombinazione processiva, si è osservato sperimentalmente che si ottengono i seguenti prodotti: dopo il secondo ciclo il *nodo a otto*, dopo il terzo ciclo il *link di Whitehead* e dopo il quarto ciclo uno specifico *nodo a 6 incroci* (si veda la Figura 3.9).

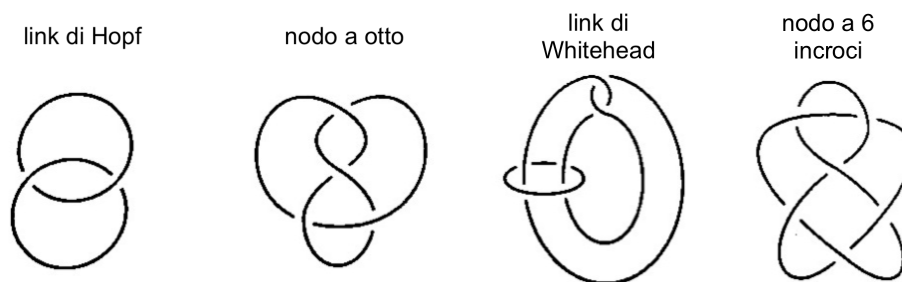


Figura 3.9: Prodotti della ricombinazione processiva della Tn3 resolvasi su un nodo banale superavvolto (fonte dell'immagine [22]).

Per ricavare i tangles O , E ed R non è necessario il quarto ciclo di ricombinazione, ma lo useremo come una verifica per il funzionamento del modello. Da quanto detto finora otteniamo la serie di diagrammi in Figura 3.10.

Siccome tutti i nodi e link ottenuti come prodotti sono nodi o link razionali, possiamo rappresentarli con la notazione introdotta nella Sezione 1.4.2,

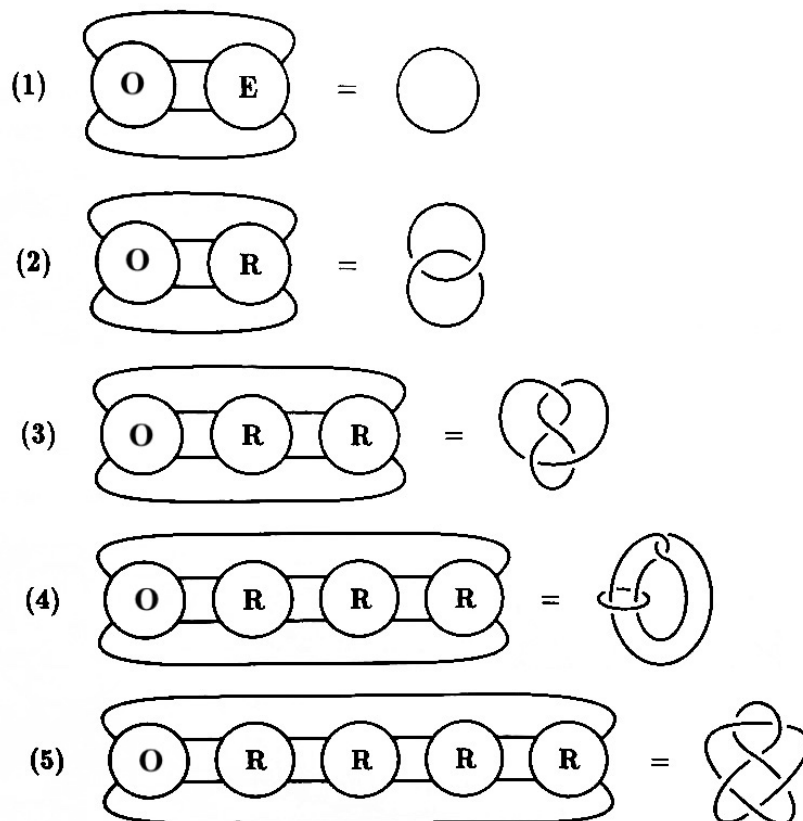


Figura 3.10: Ricombinazione processiva della Tn3 resolvasi, vista come somma di tangles (fonte dell'immagine [22]).

otterremo così le seguenti **equazioni di tangles**:

1. $N(O + E) = C(1)$
2. $N(O + R) = C(2)$
3. $N(O + R + R) = C(2, 1, 1)$
4. $N(O + R + R + R) = C(1, 1, 1, 1, 1)$
5. $N(O + R + R + R + R) = C(1, 1, 1, 2, 1)$.

Grazie a queste equazioni è possibile determinare O ed R , come afferma il seguente teorema.

Teorema 3.5.4. *Otteniamo che:*

1. le possibili coppie di tangle (O, R) che soddisfano le prime tre equazioni sono le quattro coppie in Figura 3.11;
2. inoltre, affinché valga anche la quarta equazione O ed R devono essere come in Figura 3.11 (a); ovvero $O = T(-3, 0)$ e $R = T(1)$.

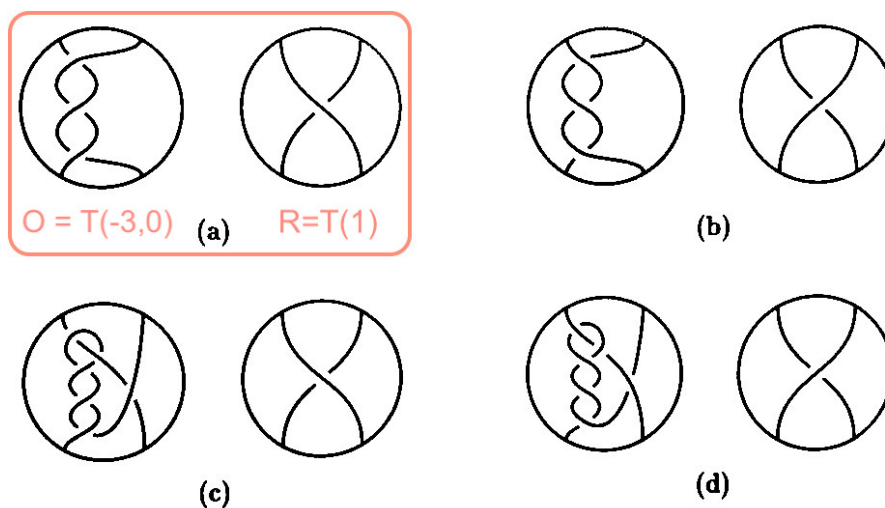


Figura 3.11: Possibili coppie di tangles (O, R) (fonte dell'immagine [22]).

Per una dimostrazione del Teorema 3.5.4 si veda *Theorem 13.4.1.* in [22]. Con queste informazioni non possiamo ancora determinare E , ma se assumiamo che esso sia il tangle $T(0)$ possiamo pensare alla prima ricombinazione effettuata dalla Tn3 resolvasi come al processo in Figura 3.12.

La coppia di tangles (O, R) ottenuta grazie al Teorema 3.5.4 soddisfa anche la quinta equazione e ci permette di prevedere i prodotti di ulteriori ricombinazioni. D'altra parte questo teorema non getta alcuna luce sulla struttura di E , in effetti non è ancora possibile determinarne la forma con certezza. Solo in alcuni casi specifici la struttura di E può essere intravista grazie al teorema seguente.

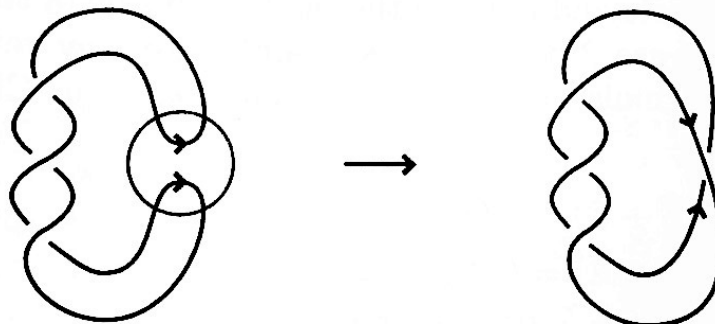


Figura 3.12: Prima ricombinazione effettuata dalla Tn3 resolvasi, quando $E = T(0)$ (fonte dell'immagine [22]).

Teorema 3.5.5. *Se E è un tangle **razionale** e soddisfa le due condizioni:*

1. $N(T(-3, 0) + E) = C(1)$
2. $N(T(-3, 0) + T(1)) = C(2)$ con linking number -1

allora E è della forma $T(2x, 3, 0)$, con $x \in \mathbb{Z}$.

Per una dimostrazione del Teorema 3.5.5 si veda *Theorem 13.4.2.* in [22].

3.5.2 Modello a tangle per una ricombinasi in tirosina

Le ricombinasi in tirosina non necessitano di un numero fisso di superavvolgimenti per dare il via alla ricombinazione; per questo motivo nel caso della ricombinazione sito-specifica dovuta a enzimi di questa famiglia le cose si complicano ulteriormente. Consideriamo ad esempio l'enzima **Flp** (pronuncia 'flip') della famiglia delle ricombinasi in tirosina, tenendo presente che esistono diverse altre proteine i cui prodotti risultano topologicamente identici a quelli ottenuti con Flp. Questa volta modellizziamo ognuno dei substrati e dei prodotti come il numeratore di una somma di **tre** tangles, in particolare quello che finora abbiamo chiamato O lo riscriviamo come somma di due tangles:

- O_v che indica la parte di DNA che sta fuori dal complesso di ricombinazione E , ma che è fisicamente vincolata dall'enzima Flp;
- O_l^k che indica la parte di DNA che sta fuori dal complesso di ricombinazione E e che è libera dai vincoli dell'enzima, l'indice $k \in \{0, 1, 2, 3\}$ rappresenta il grado di superavvolgimento del substrato presente prima dell'azione di Flp.

In una prima versione del modello, proposto da Ernst e Summers [12], si suppone che il processo di rottura e riarrangiamento del DNA sia completamente localizzato; ovvero che due dei tre tangles (O_l^k e O_v) rimangano immutati sotto l'azione dell'enzima, che agirebbe semplicemente sostituendo il tangle E con un altro tangle R , come in Figura 3.13.

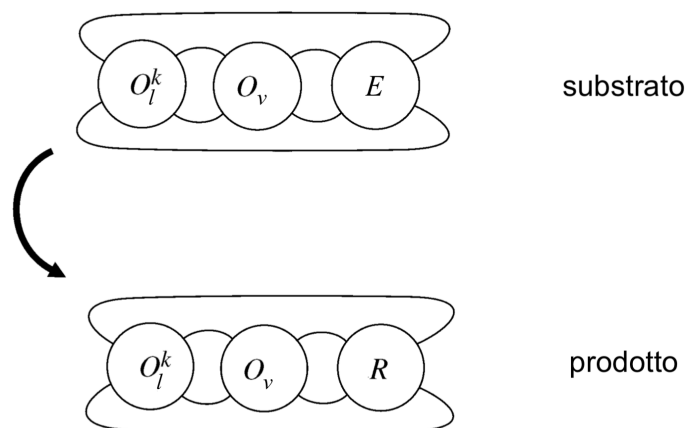


Figura 3.13: Modello per il substrato e prodotto di una ricombinazione in tirosina regolata dall'enzima Flp (fonte dell'immagine [5]).

Dunque possiamo scrivere le seguenti equazioni:

1. $N(O_l^k + O_v + E) = \text{substrato}$,
2. $N(O_l^k + O_v + R) = \text{prodotto}$.

Al variare del numero di superavvolgimenti k presenti nel substrato, l'azione di Flp sul DNA produce vari **link torici**. Un nodo (o link) è detto

torico se può essere immerso iniettivamente (ossia disegnato senza autointersezioni) sulla superficie di un toro standard in \mathbb{R}^3 (si veda Figura 3.14). Se si definisce su di un toro un sistema di riferimento (come in Figura 3.14), in cui l viene chiamata *longitudine* (rappresentata in blu) e m viene chiamato *meridiano* (rappresentato in rosso), allora un link torico è definito come:

$$t(l, m) := ll + mm,$$

dove l rappresenta il numero di giri che il link compie lungo la longitudine ed m in numero di giri che compie lungo il meridiano. La coppia (l, m) identifica

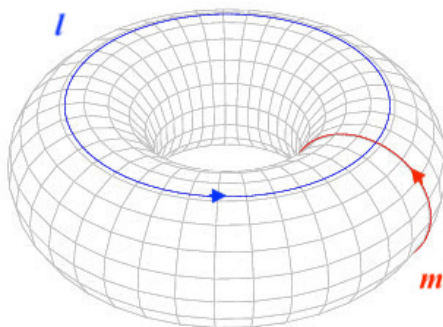


Figura 3.14: Toro con un'orientazione fissata.

univocamente un link torico, permettendo così una classificazione di questo genere di link. Inoltre:

1. il numero di componenti di un link torico è dato da $\text{MCD}(l, m)$;
2. il numero di ponti di un link torico è dato dal $\min(l, m)$.

Ma torniamo ora a concentrarci sui prodotti della ricombinazione dovuta all'azione di Flp: gli esperimenti hanno mostrato che quando il substrato presenta siti di ricombinazione con orientazioni inverse si ottengono dei nodi torici, tra cui il nodo banale (con una diversa sequenza di DNA) o nodi che presentano fino a 11 incroci. Grazie all'uso di microscopi elettronici sono state ottenute le immagini dei prodotti più semplici, e si è mostrato che sono i nodi torici: $t(1, 1)$ (nodo banale), $t(3, 1)$ e $t(5, 1)$.

Osservazione 8. Riferendoci alla Figura 3.15, questi dati sperimentali indicano che:

1. Flp agisce su un nodo banale $t(1, 1)$ di DNA con siti di ricombinazione posti in orientazione inversa convertendolo in un **nodo torico** della forma $t(\pm(2k + 1), 1)$, dove $k \in \{0, 1, 2, 3\}$;
2. quando invece le orientazioni sono dirette si ottiene un **link torico** della forma: $t(\pm 2k, 1)$, dove $k \in \{0, 1, 2, 3\}$.

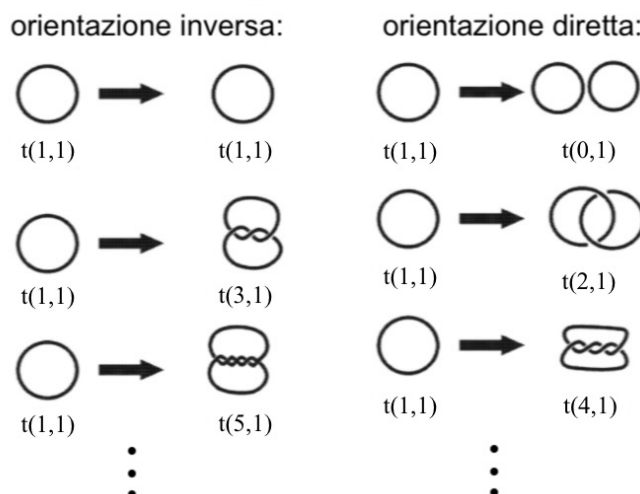


Figura 3.15: Substrati e prodotti della ricombinazione Flp (fonte dell'immagine [5]).

Ad esempio, per siti con orientazione inversa le equazioni che modellizzano l'azione di Flp saranno:

1. $N(O_l^k + O_v + E) = t(1, 1) = \text{nodo banale}$, con $k \in \{0, 1, 2, 3\}$
2. $N(O_l^0 + O_v + R) = t(1, 1) = \text{nodo banale}$,
3. $N(O_l^k + O_v + R) = t(\pm(2k + 1), 1) = \text{nodo torico}$, per $k \in \{0, 1, 2, 3\}$.

Dove la prima equazione indica il substrato e le seguenti indicano i prodotti della ricombinazione.

Osserviamo che, in generale, le equazioni di tangles per la ricombinazione in serina sono più facilmente trattabili rispetto a quelle per la ricombinazione in tirosina, perché in queste ultime è necessario tenere in considerazione anche il numero di superavvolgimenti k del substrato, che invece è fisso nel caso precedente.

Le strategie utilizzate nella risoluzione delle equazioni di tangles comprendono varie tecniche di rappresentazione delle 3-varietà come i *rivestimenti ramificati*, le *decomposizioni di Heegard* e la *chirurgia sui tangles* (in particolare *chirurgia di Dehn*). Per approfondire queste nozioni si può consultare [25], per comprendere che ruolo ricoprono nello studio della ricombinazione sito-specifica si veda [5] e per un esempio del loro utilizzo nello studio di un membro generico della famiglia delle ricombinasi in tirosina si vedano [8, 9]. Grazie a queste tecniche, le soluzioni in termini di tangles sono state classificate completamente sia per una generica ricombinasi in tirosina che per molte ricombinasi in serina.

3.5.3 Un modello predittivo

Gli studiosi Buck e Flapan [6, 7] hanno sviluppato un **modello predittivo** per nodi e link di DNA che risultano come prodotti della ricombinazione sito-specifica. Questo modello non si focalizza su un tipo specifico di ricombinasi, come hanno fatto molti degli studi precedenti, al contrario esso fornisce una previsione dei nodi e link prodotti dalla ricombinazione sito-specifica *in generale*. L'approccio è quello di descrivere topologicamente il processo di formazione dei nodi e link di DNA che compaiono come risultato di uno o più cicli di ricombinazione processiva, a partire da diversi tipi di substrato. I tipi di substrato considerati sono (si veda la Figura 3.16)¹:

¹Ogni curva in Figura 3.16 rappresenta l'asse di una molecola (a doppia elica) di DNA.

- il nodo banale o il link banale superavvolti plectonemicamente (si veda la Sezione 3.2 per la definizione di supercoiling plectonemico);
- i nodi o link torici $t(2, m)$.

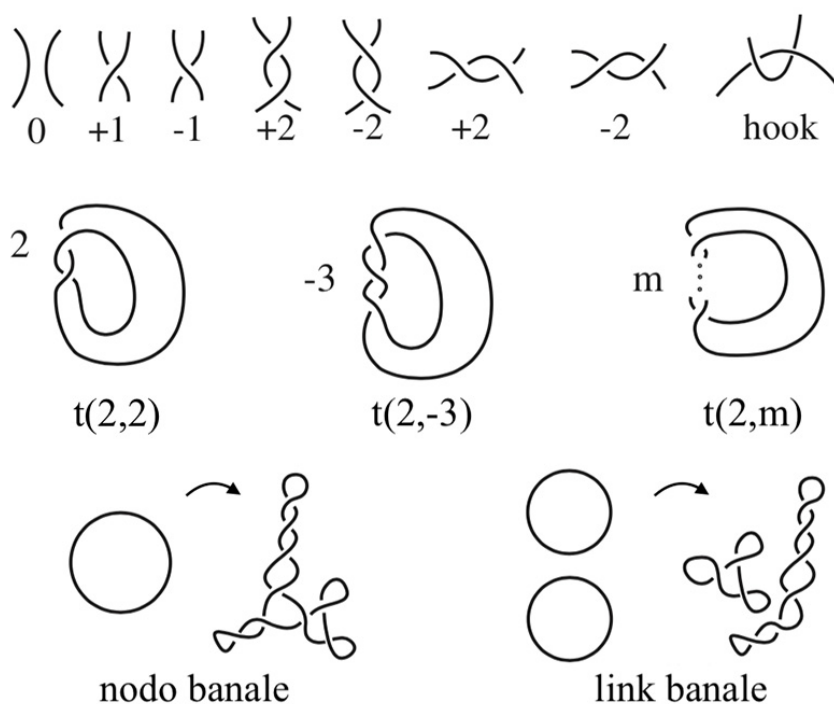


Figura 3.16: *Sopra*: possibili conformazioni del complesso di ricombinazione. *Sotto*: possibili substrati: nodi o link torici $t(2, m)$, nodo banale, link banale (fonte dell'immagine [5]).

L'idea generale è quella di considerare una palla B che contenga l'inviluppo convesso delle molecole di ricombinasi, in pratica il corrispettivo topologico del complesso di ricombinazione, e una superficie D tale che l'asse del substrato di DNA sia il bordo di tale superficie. Per il nodo banale D sarà un disco, per il link banale sarà costituito da due dischi disgiunti e per i nodi o link torici sarà un anello (si veda la Figura 3.17 per il caso del link di Hopf

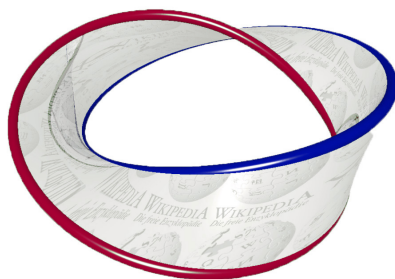


Figura 3.17: Il link di Hopf è il link torico $t(2, 2)$ e borda un anello topologico.

$t(2, 2)$). Consideriamo $D \cap B$ prima e dopo la ricombinazione e, analogamente, caratterizziamo $D \cap \text{cl}(S^3 \setminus B)$ per ogni substrato. Dopo aver ristretto le possibilità per entrambi questi oggetti con ragionamenti topologici, incolliamo ognuna delle forme di $D \cap B$ ottenute dopo la ricombinazione con ogni forma possibile di $D \cap \text{cl}(S^3 \setminus B)$, avremo così ottenuto **tutti i possibili nodi e link prodotto della ricombinazione**.

Questo modello è indipendente dalla grandezza del substrato, dal segno dei superavvolgimenti e dall'orientazione dei siti di ricombinazione. Esso si basa su tre assunzioni, ognuna delle quali è supportata da evidenze biologiche.

Se chiamiamo J il link che modella il substrato (che può essere un nodo o link torico $t(2, m)$, il nodo banale o il link banale) e $C := \text{cl}(S^3 \setminus B)$, allora le tre assunzioni si possono esprimere in termini matematici come segue.

1. $B \cap J$ contiene due archi ed esiste una sua proiezione che ha al massimo un incrocio tra i due archi e nessun incrocio di ciascun arco con se stesso. Dunque, possiamo fissare una proiezione di J tale che $B \cap J$ abbia una delle forme illustrate in Figura 3.18.
2. Esiste una superficie D tale che: (i) J sia il bordo di D (ii) $D \cap \partial B$ sia formato da due archi e (iii) $D \cap C$ sia *non annodato rel ∂B* . Diciamo che $D \cap C$ è *non annodato rel ∂B* , se c'è un'isotopia ambientale di C che fissa puntualmente il bordo ∂B e che manda $D \cap C$ in una superficie



Figura 3.18: Possibili conformazioni di $B \cap J$ quando vale l'assunzione 1 (fonte dell'immagine [6]).

planare con torsioni, dove i punti finali degli archi che definiscono le torsioni sono disgiunti da ∂B . Ad esempio $D \cap C$ è non annodata nel ∂B per ciascuno dei casi rappresentati in Figura 3.19 mentre lo è per quelli rappresentati in Figura 3.20.

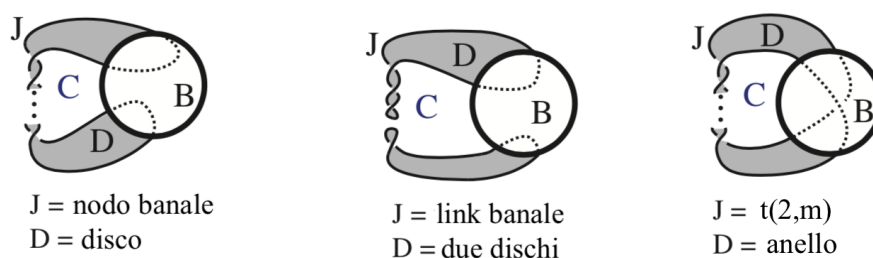


Figura 3.19: Esempi di substrati J e superfici D bordate da J nel caso in cui valga l'assunzione 2 (fonte dell'immagine [6]).

3. Le *ricombinasi in serina* tagliano ciascuno dei siti di ricombinazione e aggiungono un incrocio all'interno di B tra gli archi tagliati su siti diversi, quindi richiudono. Inoltre, nella ricombinazione processiva, tutti gli eventi di ricombinazione sono identici. Dopo una ricombinazione mediata da una *ricombinasi tirosina*, c'è una proiezione dei siti di ricombinazione che ha al massimo un incrocio.

Osserviamo che la terza assunzione riguarda il meccanismo di ricombinazione, perciò essa si differenzia a seconda della famiglia a cui appartiene la

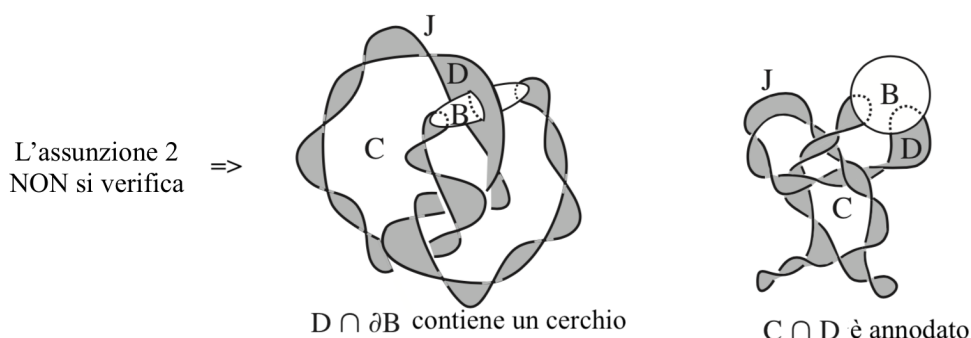


Figura 3.20: Esempi di substrati J e superfici D bordate da J nel caso in cui non valga l'assunzione 2 (fonte dell'immagine [6]).

ricombinasi in azione. Si veda [6] per approfondimenti sulle tre assunzioni e [7] per la loro giustificazione biologica.

Una volta accettate le assunzioni, il risultato previsto da questo modello è che i prodotti che si ottengono dalla ricombinazione sito-specifica fanno parte di un'unica famiglia di prodotti (illustrata in Figura 3.21) che chiameremo $F(p, q, r, s)$. Si tratta di una particolare famiglia di nodi e link che

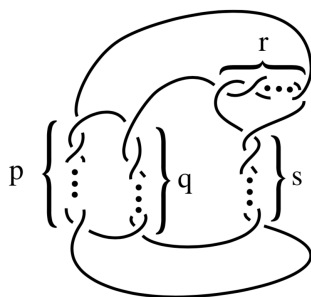


Figura 3.21: Famiglia $F(p, q, r, s)$ dei prodotti della ricombinazione sito-specifica (fonte dell'immagine [5]).

sono il numeratore del tangle ottenuto dalla somma di tre tangles razionali corrispondenti a $\frac{1}{p}$, $\frac{1}{q}$ e $\frac{r}{rs+1}$, con $p, q, r, s \in \mathbb{Z}$ (si veda la Sezione 1.4.1). Tale famiglia è una sottofamiglia dei *link di Montesinos* ossia dei link che sono il

numeratore della somma di un numero finito di tangles razionali. Fortunatamente questa famiglia è di piccole dimensioni e la biologia sperimentale può aiutarci a restringere ulteriormente il campo. Infatti il numero di incroci, o MCN (si veda Sezione 1.3.1), di un nodo o link di DNA può essere determinato sperimentalmente. Osserviamo che per n piccoli il numero di nodi e link con $MCN = n$ è piuttosto piccolo, ma esso cresce esponenzialmente in funzione di n (ad esempio ci sono 1.701.936 nodi con $MCN \leq 16$), quindi ovviamente conoscere l' MCN di un nodo o link non è sufficiente a determinarlo. D'altra parte, però, si può dimostrare [6] che il numero totale di nodi e link della famiglia $F(p, q, r, s)$ cresce linearmente come n^3 . Quindi, se consideriamo i nodi e link di $F(p, q, r, s)$ con $MCN = n$, rispetto a tutti i nodi e link con $MCN = n$ il rapporto decrescerà esponenzialmente all'aumentare di n . Allora, conoscere l' MCN di un nodo o link e sapere che esso fa parte della famiglia dei prodotti $F(p, q, r, s)$ ci permette di restringere significativamente i tipi di nodo o link possibili.

Le applicazioni di questo modello sono diverse, in quanto:

- permette di prevedere i nodi e link prodotto in alcuni casi non ancora conosciuti;
- permette di determinare il percorso di ricombinazione, ovvero se si tratta di una ricombinazione di tipo processivo o distributivo: in particolare se un link non è della famiglia $F(p, q, r, s)$ dev'essere necessariamente un prodotto della ricombinazione distributiva;
- aiuta a determinare la sequenza dei prodotti di diversi cicli di ricombinazione nel caso della ricombinazione processiva.

Inoltre, una nuova e potente strategia combina questo modello con modelli del tipo presentato nelle Sezioni 3.5.1 e 3.5.2, per investigare ulteriormente il meccanismo di ricombinazione. Questo processo, infatti, continua a presentare molti passaggi bui che le tecniche puramente biologiche non sono ancora in grado di illuminare, per questo motivo la ricerca topologica in questo campo si sta sviluppando in una grande varietà di traiettorie.

Capitolo 4

Topologia di una proteina: il caso della leptina

Vediamo ora un altro esempio di applicazione della teoria dei nodi, questa volta allo studio di una specifica proteina, la *leptina*, un ormone che gioca un ruolo chiave nel regolare l'assorbimento e il consumo di energia, il metabolismo e l'ipertensione. A questo scopo analizziamo l'articolo *The unique cysteine knot regulates the pleiotropic hormone leptin*¹ mettendone in luce prevalentemente gli aspetti di modellizzazione matematica (per approfondimenti si veda [14]).

4.1 Funzioni della leptina e sue mutazioni

La leptina è un ormone che si occupa di trasmettere segnali dal tessuto adiposo al cervello. Essa fa parte del sistema complesso che regola il mantenimento del peso corporeo: quando si accumula troppo tessuto adiposo la leptina invia un segnale al cervello innescando così un meccanismo a catena che induce ad assumere meno calorie e a consumare più energia tramite l'attività fisica e la termoregolazione. Si è osservato che una mutazione nel gene che codifica la leptina può portare, quindi, ad un aumento incontrollato della

¹“L'unico nodo di cisteina regola l'ormone pleiotropico leptina”.

massa adiposa. L'individuo che presenta tale mutazione non sarà in grado di regolare la temperatura corporea e di controllare l'appetito, di conseguenza avrà grandi probabilità di sviluppare malattie quali l'obesità e il diabete. Dunque la carenza di leptina può causare l'obesità, ma spesso accade anche che individui obesi presentino livelli di leptina maggiori della norma. Infatti, anche se l'ormone viene prodotto, è possibile che il meccanismo di segnalazione condotto dalla leptina sia in qualche modo ostacolato; questo induce il corpo a produrre leptina in maggiori quantità come tentativo (fallimentare) di risolvere il problema. Risulta quindi necessario indagare più a fondo le cause del suo malfunzionamento, a questo scopo può risultare molto utile analizzare la topologia della leptina e i suoi meccanismi di folding/unfolding. Come abbiamo visto nel Capitolo 2, infatti, la struttura tridimensionale (o struttura terziaria) di una proteina, ne determina fortemente il funzionamento. Oltre ad avere un ruolo nello sviluppo dell'obesità e del diabete, questo ormone è coinvolto nella regolazione delle infiammazioni e della risposta immunitaria. Inoltre, un certo numero di mutazioni nel gene che codifica la leptina negli umani è stato collegato sia a modifiche nei processi di sviluppo, quali l'inizio della pubertà, che a malattie tra cui l'Alzheimer.

4.2 Topologia della leptina

Come tutte le proteine, la leptina viene sintetizzata a partire da un gene specifico. Dopo la sintesi, essa si ripiega, attraverso il processo che abbiamo chiamato "folding proteico" (si veda Sezione 2.1.1), in un fascio formato da **4 eliche**, come rappresentato in Figura 4.1, fino a raggiungere il suo stato nativo. A questo punto la leptina è pronta a svolgere le sue funzioni biologiche, legandosi ad alcuni recettori posti al di fuori della cellula e trasmettendo segnali alla cellula stessa. L'articolo che stiamo presentando ha rivelato una complessità nascosta nella topologia della leptina, si è scoperto infatti che durante il suo ripiegamento si può formare un cosiddetto **nodo di cisteina**. La cisteina è un amminoacido e due unità di questo amminoacido

possono legarsi creando un “ponte di zolfo”, o **ponte disolfuro**, costituito da due atomi di zolfo (S) legati. La presenza del ponte disolfuro conduce alla conformazione chiamata **nodo di cisteina**, cioè un annodamento che è realizzato tramite un legame chimico tra due amminoacidi (cisteine) della proteina. La presenza di un nodo di cisteina plasma in maniera decisiva le strutture terziaria e quaternaria della proteina dalle quali dipende la sua azione biologica. Il nodo di cisteina osservato nelle molecole di leptina non era mai stato caratterizzato prima, l'ipotesi di questo articolo è che questa topologia unica possa offrire nuovi meccanismi per regolare l'attività della proteina. È stata utilizzata quindi una combinazione di esperimenti *in vivo* e *in silico* per provare il ruolo della topologia annodata (indotta dal ponte disolfuro) sul folding e sul funzionamento della leptina. Dal momento che studieremo la leptina nella sua forma ossidata e ridotta, è utile ricordare che in generale una struttura chimica si dice ossidata quando possiede più elettroni della struttura originale e ridotta quando ne possiede meno. Nella leptina **ossidata** è presente il ponte disolfuro, che rimane intatto durante il processo di folding/unfolding della proteina, mentre nella leptina **ridotta** il ponte disolfuro non si forma mai. Sono state modellizzate anche forme intermedie in cui il ponte disolfuro si rompe e ricrea spontaneamente durante il folding/unfolding, ma in questi casi i meccanismi di folding e unfolding non possono essere osservati sperimentalmente. Dunque si parla di proteina **annodata** nel caso della leptina ossidata, a causa della presenza del ponte disolfuro che induce la formazione del nodo di cisteina, e di proteina **non annodata** nel caso della leptina ridotta. I risultati ottenuti da questi studi hanno mostrato sorprendentemente che il paesaggio di energia libera (si veda Sezione 2.1.1) è lo stesso per la proteina annodata e non annodata. La complessità aggiuntiva dovuta alla conformazione annodata, però, si è rivelata strutturalmente importante. Inoltre, le analisi dello stato nativo hanno condotto alla scoperta che il ponte disolfuro gioca un ruolo importante nella fase di attaccamento ai recettori e dunque media l'attività biologica.

In Figura 4.1 possiamo vedere la topologia complessiva della leptina, in cui

la parte gialla (in basso a sinistra) rappresenta il ponte disolfuro e la parte in blu (di cui fanno parte le eliche α_4 e α_5) costituisce il nodo di cisteina.

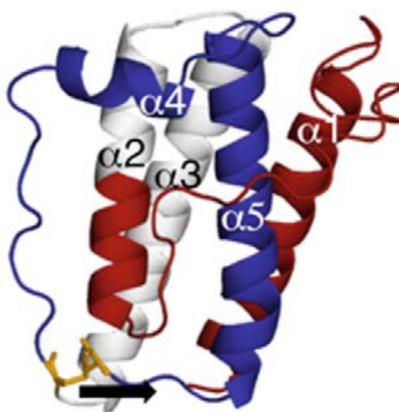


Figura 4.1: Struttura nativa (attiva) della leptina (fonte dell'immagine [14]).

Vediamo ora, in Figura 4.2, una modellizzazione dei diversi stati conformazionali della leptina in cui si enfatizza l'importanza del ponte disolfuro.

Gli stati denaturati (C e D) rappresentano le conformazioni iniziali, rispettivamente della forma ossidata e ridotta, dalle quali può iniziare il processo di ripiegamento (folding proteico) fino ad arrivare in entrambi i casi ad un unico stato nativo (A), cioè lo stato biologicamente attivo della proteina. Come si può osservare dalla sequenza di immagini in Figura 4.2 (scorrendo da destra verso sinistra), durante il processo di **folding** è necessario che gli N residui al termine della catena (rappresentati in rosso) più l'elica α_3 e metà dell'elica α_2 (rappresentate in bianco) si “infilino” all'interno del cappio creato dal nodo di cisteina (parte blu). Se invece analizziamo il processo contrario, **unfolding**, possiamo osservare che a partire dallo stato nativo la leptina può raggiungere uno stato denaturato anche senza rompere il ponte disolfuro (rappresentato in giallo).

In particolare la leptina ossidata si dispiega (processo di unfolding) lasciando intatto il ponte disolfuro, come si può osservare in Figura 4.2 C. La leptina ridotta, invece, durante il processo di unfolding rompe il ponte disolfuro e va

4.3 Ruolo della topologia annodata nei processi di folding e unfolding

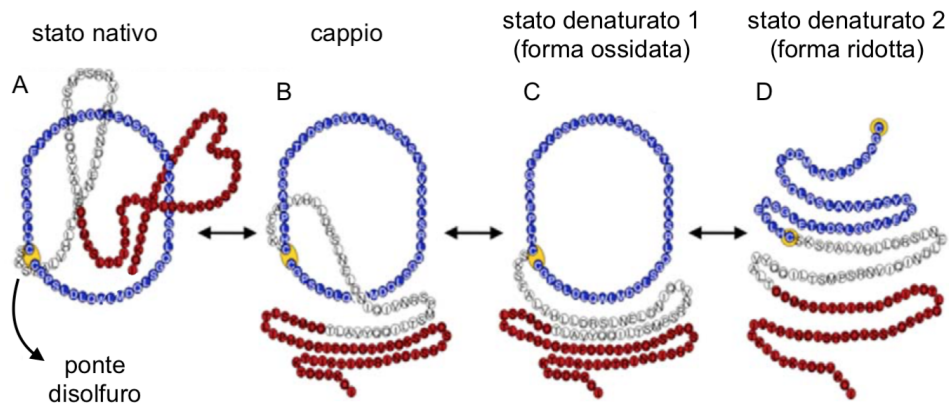


Figura 4.2: Rappresentazione della leptina in diversi stati conformazionali (fonte dell'immagine [14]).

a formare una catena lineare come quella in Figura 4.2 D. Gli autori dell'articolo hanno dimostrato che la leptina può ripiegarsi (folding), o compiere il processo inverso (unfolding), spontaneamente e con successo sia nella forma ossidata che ridotta.

4.3 Ruolo della topologia annodata nei processi di folding e unfolding

La letteratura biologica suggerisce che i fasci di eliche siano proteine abbastanza stabili. Questo articolo mostra che la stabilità della leptina è molto minore di quella osservata in un tipico fascio di eliche. Un'analisi della struttura ha rivelato che la formazione del ponte disolfuro nella leptina ossidata non solo stabilizza la proteina, ma conduce anche univocamente ad una struttura di **cisteina annodata** mai vista prima. Gli studiosi si sono chiesti perché la leptina presenti questa topologia unica quando gli altri fasci di 4 eliche sono sufficientemente stabili da esistere senza essere annodati o

presentare un ponte disolfuro. La scoperta di questa struttura ha indotto, quindi, a ipotizzare che questa univoca topologia annodata potesse offrire nuovi meccanismi di regolazione della leptina. Dunque le mutazioni che conducono a malattie potrebbero essere collegate alla formazione della struttura di nodo cisteina piuttosto che interessare solamente il ponte disolfuro. Ci si è chiesti inoltre se la leptina ossidata, che presenta questa topologia annodata, possa piegarsi spontaneamente dallo stato denaturato allo stato nativo e, se è possibile, con che velocità avviene questo processo rispetto alla leptina non annodata. Infatti, per investigare il ruolo delle mutazioni della leptina che conducono a malattie, è necessario comprendere il meccanismo di ripiegamento che porta alla conformazione di cisteina annodata e, in generale, alla formazione della proteina nel suo stato pienamente attivo. Per questo motivo, nell'articolo preso in esame, si studiano i meccanismi di folding della proteina lineare e della proteina annodata, comparandone i relativi cammini e tempi di ripiegamento. Gli autori dell'articolo hanno anche scoperto una correlazione tra i siti di attaccamento ai recettori e la topologia annodata, e suggeriscono che quest'ultima sia importante non solo nella trasmissione di segnali dal tessuto adiposo, ma anche nella risposta alla presenza di agenti esterni e quindi per la difesa immunitaria. Analizziamo ora in maniera più approfondita la topologia di questa proteina.

Come possiamo vedere in Figura 4.3, nella forma ossidata la presenza del ponte disolfuro crea un **vincolo geometrico** sia nello stato nativo che denaturato. Probabilmente è per questo motivo che, durante il processo di **unfolding**, la leptina ossidata presenta molti più tentativi falliti di passare dallo stato nativo allo stato denaturato; mentre nella forma ridotta non sono presenti particolari "trappole" durante il cammino di unfolding. Nonostante questa differenza, se si concede un tempo sufficiente, anche la leptina ossidata riesce a compiere il processo passando dallo stato nativo a quello denaturato. Per quanto riguarda il **folding**, esso conduce ad una conformazione nativa comune nel caso ossidato e ridotto. Come prima però la presenza del vincolo geometrico dovuto al ponte disolfuro crea delle differenze nei tempi di

4.3 Ruolo della topologia annodata nei processi di folding e unfolding

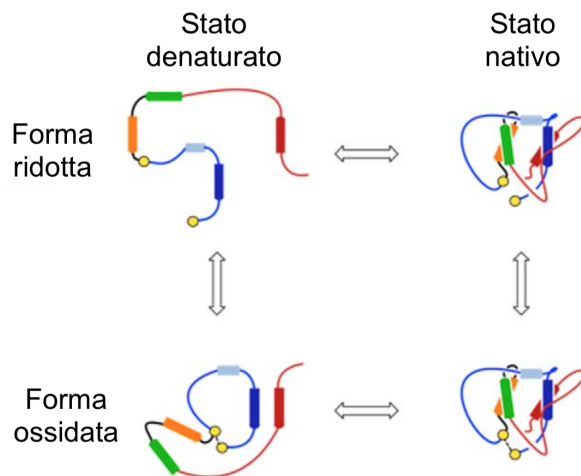


Figura 4.3: Confronto tra la forma ossidata e ridotta dello stato denaturato e nativo della leptina (fonte dell'immagine [14]).

ripiegamento. Infatti per la leptina ossidata alcune delle vie di ripiegamento possibili conducono a delle “trappole” che non permettono di raggiungere lo stato nativo e la percentuale di successo del folding si abbassa. Per questo motivo, nel caso della leptina ossidata i tempi di ripiegamento sono significativamente maggiori di quelli che compaiono nel caso ridotto (almeno il doppio). Questo indica un cammino principale di ripiegamento più complicato, ovvero i tempi di folding più lunghi sono il risultato di una topologia più complicata nella forma ossidata.

Nonostante queste differenze, sorprendentemente, il meccanismo di ripiegamento della leptina ossidata e ridotta è sostanzialmente lo stesso. Analizzando passo a passo le interazioni che si formano durante il folding, si osserva che nella forma ossidata la caratteristica struttura a cappio è la prima ad essere creata, più precisamente le due cisteine presenti nella catena si avvicinano e si legano formando il ponte disolfuro. Nella forma ridotta, invece, le due cisteine si avvicinano senza creare legami e la proteina assume una conformazione a ferro di cavallo. Nel primo caso il resto della catena si infila nel cappio, mentre questo ovviamente non avviene nel secondo caso,

la parte restante del processo di ripiegamento invece risulta praticamente la stessa nel caso ossidato e ridotto (si veda la Figura 4.4).

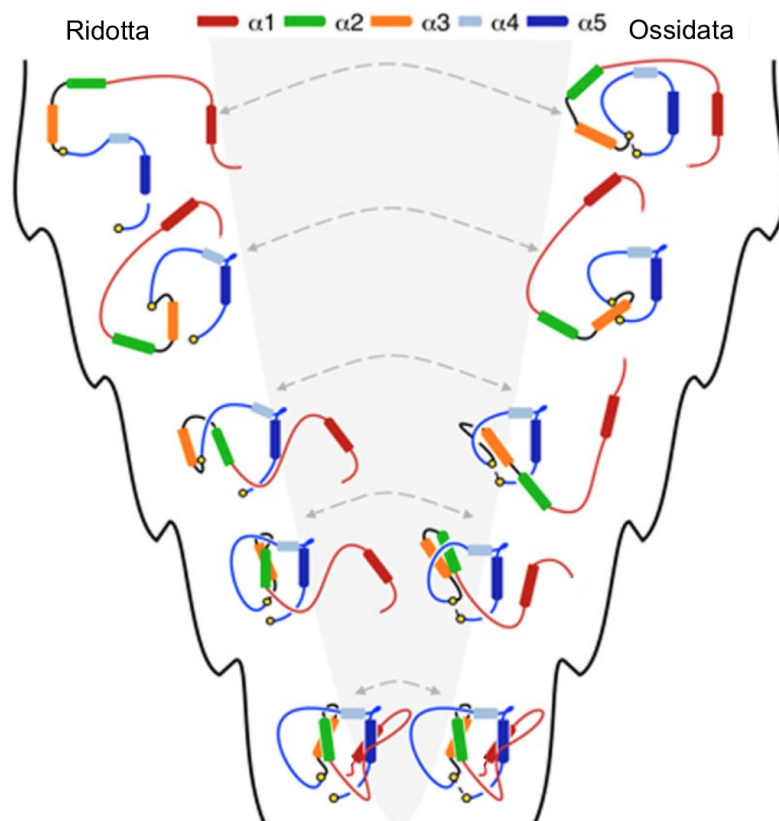


Figura 4.4: Folding della leptina ossidata e ridotta (fonte dell'immagine [14]).

Teoricamente numerose conformazioni geometriche sono possibili, ma la parte grigia al centro dell'imbuto non è accessibile tramite gli esperimenti. Questa figura descrive lo scenario del folding: in alto troviamo gli stati denaturati, che possono assumere un gran numero di conformazioni spaziali (alto grado di entropia conformazionale) e che presentano un basso numero di legami/contatti tra le varie parti della molecola. Scorrendo verso il basso si possono confrontare diversi stadi della molecola che sono simili per lo stato ossidato e ridotto, mostrando come i cammini di ripiegamento siano analoghi per le due forme di leptina. Ed infine come entrambi i cammini,

ossidato e ridotto, conducano ad un corretto stato nativo (ripiegato). Dalla somiglianza dei due cammini si deduce che i vincoli topologici imposti dal ponte disolfuro non determinano grandi variazioni nel paesaggio energetico associato al folding. Comunque, la rottura del ponte disolfuro gioca un ruolo centrale nell'efficienza del folding incrementandone la velocità. Ciò suggerisce una nuova interpretazione dei risultati sperimentali precedentemente pubblicati: il ripiegamento scorretto della leptina non è dovuto ad un fallimento nella formazione del ponte disolfuro, ma è causato da alterazioni dell'unico cammino di ripiegamento possibile per la forma annodata.

4.4 Conclusioni dell'articolo

Gli studi precedenti suggerivano che la forma ridotta della leptina non fosse del tutto attiva. I risultati riportati in questo articolo hanno invece rivelato che la leptina ridotta è comunque in grado di svolgere la sua funzione di trasmissione di segnali, anche se gli esperimenti indicano che essa è meno attiva della leptina ossidata. Questi risultati suggeriscono che la leptina ridotta abbia la conformazione corretta per potersi legare ai recettori e attivare l'effetto a cascata di segnali tipico dell'azione di questo ormone. La minore attività che si riscontra nella forma ridotta può essere un effetto delle diverse dinamiche presenti nello stato nativo e/o un effetto della minore stabilità dovuta all'assenza del ponte disolfuro. Questo può condurre sia ad una minore affinità della leptina ridotta con i recettori, che ad un minore effetto a cascata dei segnali una volta che il legame con i recettori è già avvenuto. Si è scoperto che il ponte disolfuro funziona come un punto di tensione che influenza i movimenti locali dei siti di ancoraggio al recettore. Rompere il ponte disolfuro aumenta i movimenti locali nella parte inferiore della struttura, ovvero quella che si lega al recettore, e quindi ne ostacola l'ancoraggio. Mentre la visione corrente è che le mutazioni della leptina che causano malattie siano associate ad una scorretta formazione del ponte disolfuro, in questo articolo si propone una nuova interpretazione collegata piuttosto alla

mancata formazione di una struttura a cappio. D'altra parte, questo lavoro suggerisce che il ponte disolfuro abbia un ruolo regolatorio più importante di quanto si pensi, infatti esso sembra mediare fortemente le funzioni biologiche. Inoltre, basandosi sui dati raccolti, si suggerisce che queste mutazioni nella struttura della leptina possano interessare tre aspetti principali:

- la distruzione delle interazioni responsabili di trainare parte della molecola e di infilarla nella struttura a cappio;
- la riduzione della grandezza del cappio;
- la compromissione dei siti di ancoraggio, che si traduce nell'impossibilità per la leptina di legarsi ai recettori.

Quindi la questione principale diventa esplorare il folding della leptina concentrandosi sulle mutazioni della **struttura a cappio**. Tali scoperte aprono molte possibilità per esplorare le funzioni della leptina e di conseguenza per lo sviluppo di terapie di inibizione e attivazione della leptina, che si dimostrerebbero di fondamentale importanza per la cura di patologie anche gravi.

Capitolo 5

Topologia dei polimeri in geometrie confinate

Un *polimero*, o catena polimerica, è una macromolecola costituita da un considerevole numero di gruppi molecolari uniti “a catena” mediante la ripetizione dello stesso tipo di legame. Sia le proteine, costituite da lunghe catene di amminoacidi, che il DNA, in quanto catena di coppie di nucleotidi, sono esempi di polimeri. In questo contesto, la leptina analizzata nel capitolo precedente è un esempio di polimero che presenta una topologia annodata. In questo capitolo vogliamo mostrare che questa non è affatto un’eccezione, al contrario è molto comune che le lunghe catene polimeriche si trovino in uno stato altamente aggrovigliato o che siano aggrovigliate ad altri polimeri. Le caratteristiche di tali “grovigli”, detti *entanglement*, sono fortemente influenzate dai **vincoli geometrici** dovuti al fatto che le molecole si trovano spesso confinate in uno spazio ristretto. A questo scopo analizzeremo l’articolo *Topological entanglement complexity of polymer chains in confined geometries*¹ [29], pubblicato nel 1998, che si occupa proprio di investigare gli effetti di differenti tipi di vincoli geometrici sulla complessità degli entanglement di polimeri, calcolando le **probabilità di formazione di nodi e link** nelle catene polimeriche soggette a vincoli geometrici.

¹“Complessità di intricatezza topologica di catene polimeriche in geometrie confinate”.

5.1 Introduzione all'articolo

Le catene polimeriche lineari sono molecole altamente flessibili e sono caratterizzate da un alto grado di libertà conformazionale. Esse si possono aggrovigliare su loro stesse o con altre molecole e la struttura che assumono influirà fortemente sulle loro proprietà. Se il polimero subisce poi una reazione che chiude ad anello la catena, l'entanglement che si è formato può rimanere intrappolato in un **nodo** o **link**. Dunque gli entanglement che si formano nella prima fase, ovvero nella catena lineare, possono essere studiati attraverso l'analisi dei nodi e link che si presentano nella seconda fase, quando la molecola è chiusa ad anello. I polimeri chiusi ad anello possono essere modellizzati da poligoni di n lati senza autointersezioni su un reticolo regolare, questo approccio è stato utilizzato da diversi studiosi per analizzare la formazione di nodi o link non banali da parte di tali polimeri. Si è mostrato, con argomentazioni teoriche, che poligoni sufficientemente "lunghi" sono annodati con probabilità uno. Questi risultati sono stati poi estesi al caso più generale di grafi immersi iniettivamente in \mathbb{Z}^3 e al problema della complessità di aggrovigliamento dei cosiddetti "cammini senza autointersezioni" (*self-avoiding walks*). Come abbiamo già accennato, i polimeri sono spesso confinati in spazi ristretti. Per studiare l'effetto che tale vincolo ha sulle proprietà geometriche e topologiche del polimero, in questo articolo sono stati studiati polimeri chiusi ad anello, modellizzati da poligoni nel reticolo cubico, prima nel caso libero da vincoli e poi nel caso in cui tali poligoni siano confinati in uno slab ("lastra") o in un prisma.

5.2 Cammini e poligoni senza autointersezioni

Per comprendere la maggior parte dei risultati ottenuti in questo articolo, è necessario fornire alcune conoscenze di base sui cammini senza autointersezioni. Per una review molto esaustiva sull'argomento si rimanda alla re-

ferenza [18]. Consideriamo il reticolo cubico \mathbb{Z}^3 : i vertici di questo reticolo sono i punti interi di \mathbb{R}^3 e i lati uniscono coppie di vertici che distano una unità l'uno dall'altro.

Definizione 5.1. Un *cammino di n lati* è una sequenza ordinata

$$\omega_0, \omega_1, \omega_2, \dots, \omega_n$$

di punti di \mathbb{Z}^3 tali che due punti successivi distano una unità uno dall'altro. Si chiamerà *cammino di n lati senza autointersezioni*, o più brevemente *n – SAW*, se gli $n + 1$ punti che lo compongono sono tutti distinti. Chiameremo i punti $\omega_0, \omega_1, \omega_2, \dots, \omega_n$ *vertici* del cammino e considereremo due cammini equivalenti quando uno è una traslazione dell'altro.

Dato che i cammini sono equivalenti a meno di traslazioni, per i nostri scopi conviene considerare solo i cammini con primo vertice fisso nell'origine, indicheremo con c_n il numero di tali cammini.

Definizione 5.2. Un *circuito di n lati senza autointersezioni* è un $(n - 1)$ – SAW in cui il primo e l'ultimo vertice, ω_0 e ω_{n-1} , distano una unità.

Proposizione 5.2.1. *Se $\omega_0, \omega_1, \omega_2, \dots, \omega_{n-1}$ sono i vertici successivi di un circuito di n lati senza autointersezioni, allora ogni permutazione ciclica o inversa di una permutazione ciclica dei vertici è ancora un circuito di n lati senza auto intersezioni.*

Definizione 5.3. Dunque la classe di equivalenza di $2n$ circuiti di n lati senza autointersezioni può essere considerata come un unico oggetto geometrico che chiameremo *poligono di n lati senza autointersezioni*, o più semplicemente *n – SAP*. Dove, come prima, due poligoni sono equivalenti se uno è una traslazione dell'altro; chiamiamo p_n il numero di classi di equivalenza di n – SAPS.

Un n – SAP si dice *rooted* (radicato) se contiene un vertice etichettato, chiamato *radice*, che coincide con l'origine e siccome la radice può essere scelta in n modi distinti il numero di n – SAPs rooted distinti sarà np_n .

Di seguito useremo il termine *poligono* sia per denotare gli $n - SAPs$ che gli $n - SAPs$ rooted.

Teorema 5.2.2. *Esiste una costante, detta costante di connettività, $\kappa > 0$ tale che, per n grandi:*

$$p_n = e^{\kappa n + o(n)}.$$

Lo stesso risultato vale per i cammini, con la stessa costante κ , ovvero:

$$c_n = e^{\kappa n + o(n)}.$$

5.3 Probabilità di annodamento e linking: risultati teorici

Indaghiamo ora la probabilità di un poligono di essere **annodato** e come essa varia in presenza di diversi tipi di vincoli geometrici. Dopo di che ci concentreremo sulla probabilità di **linking** di due poligoni e anche qui analizzeremo l'influenza di diversi tipi di ambienti che ne vincolano la geometria.

5.3.1 Probabilità di annodamento per poligoni non vincolati

Per prima cosa richiamiamo brevemente alcuni risultati riguardanti i poligoni in generale, in particolare riportiamo un breve resoconto di un risultato che ci dice che *tutti i poligoni sufficientemente "lunghi", tranne un numero esponenzialmente piccolo, sono annodati*.

L'idea è concettualmente molto semplice, e si divide in due parti: la prima parte consiste nell'utilizzare un teorema che riguarda i *pattern*, il **teorema di Kesten**, e la seconda si avvale del concetto di *arco annodato*.

Definizione 5.4. Un *pattern* è un qualsiasi cammino finito senza autointersezioni.

Teorema 5.3.1 (Teorema di Kesten). *Se esiste un cammino senza autointersezioni nel quale il pattern K compare almeno tre volte (come sottocammino), allora il pattern K comparirà in quasi tutti (a parte un numero esponenzialmente piccolo) i cammini senza autointersezioni sufficientemente lunghi.*

Definizione 5.5. Pattern di questo tipo, cioè che compaiono almeno tre volte in un cammino, vengono chiamati *Kesten patterns*.

Osserviamo che non tutti i pattern sono Kesten pattern, ad esempio il pattern in Figura 5.1 non è un Kesten pattern in \mathbb{Z}^2 in quanto si può vedere facilmente che esso può presentarsi due volte in un cammino senza autointersezioni in \mathbb{Z}^2 , ma non tre volte.

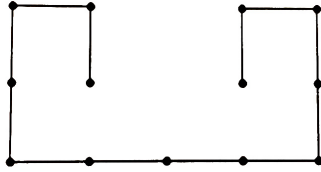


Figura 5.1: Esempio di pattern in \mathbb{Z}^2 che non è un Kesten pattern (fonte dell'immagine [18]).

Il secondo ingrediente della prova è il concetto di *arco annodato* (per la definizione rigorosa si veda [27]). L'idea è che possiamo catturare la parte importante di un nodo, ad esempio il nodo trifoglio, e stringerla così tanto sul reticolo che il resto del cammino non può più passare in un intorno di questo sottocammino e sciogliere il nodo. Un esempio di tale cammino, che è un Kesten pattern annodato, è dato in Figura 5.2. Per quanto visto finora, tale pattern dovrà comparire almeno una volta in quasi tutti i poligoni senza autointersezioni sufficientemente lunghi, ovvero *quasi tutti i poligoni sufficientemente lunghi contengono questi archi annodati e dunque sono annodati*.

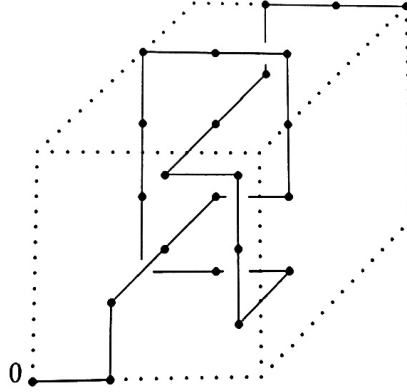


Figura 5.2: Esempio di arco annodato in \mathbb{Z}^3 (fonte dell'immagine [18]).

Teorema 5.3.2. *Se chiamiamo p_n^0 il numero di **poligoni non annodati**, il limite $\lim_{n \rightarrow \infty} n^{-1} \log p_n^0$ esiste ed è tale che:*

$$p_n^0 = e^{\kappa_0 n + o(n)},$$

inoltre il teorema di Kesten implica che $0 < \kappa_0 < \kappa$. In altre parole, la probabilità $P(n)$ che il poligono sia annodato tende ad 1 esponenzialmente:

$$P(n) = 1 - p_n^0/p_n = 1 - e^{-\alpha n + o(n)},$$

per una qualche costante positiva $\alpha = \kappa - \kappa_0$.

Si può quindi concludere che i poligoni non annodati costituiscono una frazione esponenzialmente piccola di tutti i poligoni quando il numero di lati n tende a infinito. In realtà il teorema dice qualcosa di più generale, infatti, dato che ogni nodo può essere associato ad un Kesten pattern, il teorema implica che ogni nodo deve apparire da qualche parte in un poligono sufficientemente lungo e, inoltre, dovrà apparire diverse volte. Ciò significa che quasi tutti i nodi presenti in lunghi poligoni sono *nodi composti* (cioè somma di nodi non banali, si veda Definizione 1.25) e che la complessità dell'entanglement è molto alta. Risultati simili sono validi in \mathbb{R}^3 per varie versioni continue di questi entanglement lineari a tratti.

5.3.2 Probabilità di annodamento per poligoni in ambiente vincolato

I risultati presentati finora sono validi per poligoni in un ambiente non vincolato, ora vogliamo affrontare questioni simili nel caso di poligoni confinati in diversi sottoinsiemi del reticolo cubico \mathbb{Z}^3 .

Definizione 5.6. Definiamo un L – slab come il grafo della sezione di \mathbb{Z}^3 tale che:

- l'insieme dei suoi vertici è l'insieme dei vertici di \mathbb{Z}^3 con coordinata z appartenente all'intervallo $0 \leq z \leq L$;
- e l'insieme dei suoi lati è l'insieme dei lati di \mathbb{Z}^3 che passano per due di questi vertici (ricordando che due vertici consecutivi distano 1 unità l'uno dall'altro).

Definizione 5.7. Similmente un (L_1, L_2) – prisma è il grafo della sezione di \mathbb{Z}^3 con vertici aventi le coordinate y e z che soddisfano rispettivamente $0 \leq y \leq L_1$ e $0 \leq z \leq L_2$, e con i lati di \mathbb{Z}^3 che passano per due di questi vertici.

Sia $c_n(L)$ il numero di cammini senza autointersezioni in un L – slab e $p_n(L)$ il numero di poligoni (non rooted) confinati in un L – slab. Considereremo due cammini o due poligoni, confinati in uno slab, distinti se non possono essere sovrapposti con una traslazione nelle direzioni x o y . Ora possiamo provare un risultato analogo al Teorema 5.3.2 per poligoni confinati in un L – slab. Per fare ciò, richiamiamo il seguente lemma.

Lemma 5.3.3 (Hammersley e Whittington 1985). *Se $c_n(L)$ è il numero di cammini senza autointersezioni di n lati in un L – slab, allora il limite*

$$\lim_{n \rightarrow \infty} n^{-1} \log c_n(L) = \kappa(L)$$

esiste per ogni L . Inoltre $\kappa(L)$ è monotona strettamente crescente in L e $\lim_{L \rightarrow \infty} \kappa(L) = \kappa$.

Ora dobbiamo generalizzare il teorema di Kesten (Teorema 5.3.1), che abbiamo visto per i cammini, ai cammini confinati in un L -slab.

Definizione 5.8. Per ogni $\alpha, \beta > 0$, definiamo $K_{\alpha, \beta}$ *pattern* un qualsiasi cammino senza autointersezioni ω tale che almeno tre copie disgiunte di ω compaiano in un qualche cammino ω^* , dove uno degli estremi di ω^* è l'origine, e l'altro è $(\alpha, \beta, 0)$. Inoltre, ω^* è completamente contenuto in $D_{\alpha, \beta} = \{(x, y, z) \in \mathbb{Z}^3 \mid 0 \leq x \leq \alpha, 0 \leq y \leq \beta, 0 \leq z \leq L\}$.

Lemma 5.3.4. Per ogni $\alpha, \beta > 0$, sia P un $K_{\alpha, \beta}$ *pattern*, allora:

$$\limsup_{n \rightarrow \infty} n^{-1} \log c_n(\bar{P}, L) = \kappa(\bar{P}, L) < \kappa(L),$$

dove $c_n(\bar{P}, L)$ è il numero di cammini di n lati in un L -slab che non contiene il *pattern* P .

I cammini e i poligoni hanno la stessa costante di connettività e questo continua a valere quando essi sono confinati in L -slabs, infatti vale l'affermazione seguente.

Lemma 5.3.5 (*Theorem 8.2.2.* in [18]). Se $p_n(L)$ è il numero di poligoni non rooted in un L -slab, allora:

$$\lim_{n \rightarrow \infty} n^{-1} \log p_n(L) = \kappa(L)$$

esiste per ogni L . Inoltre $\kappa(L)$ è monotona strettamente crescente in L e $\lim_{L \rightarrow \infty} \kappa(L) = \kappa$.

Dato che cancellare un lato non può creare un *pattern*, abbiamo:

$$p_n(\bar{P}, L) \leq c_{n-1}(\bar{P}, L).$$

Tale disuguaglianza, insieme ai due lemmi precedenti (Lemma 5.3.4 e 5.3.5), giustifica un ulteriore risultato.

Lemma 5.3.6. Se esiste un cammino senza autointersezioni in un L -slab in cui compare un Kesten *pattern* P , allora il numero di poligoni in un L -slab in cui P non compare mai è tale che:

$$\limsup_{n \rightarrow \infty} n^{-1} \log p_n(\bar{P}, L) \leq \kappa(\bar{P}, L) < \kappa(L).$$

Questo implica che i poligoni in un $L - slab$ che non contengono un dato pattern P sono esponenzialmente rari paragonati al numero totale di poligoni in un $L - slab$. Di conseguenza, otteniamo il seguente teorema, che estende il risultato del Teorema 5.3.2 al caso di poligoni senza autointersezioni in un $L - slab$.

Teorema 5.3.7. [28] *Se denotiamo con $p_n^0(L)$ il numero di $n - SAPs$ non annodati in un $L - slab$, allora la probabilità $P(n, L)$ che un $n - SAP$ nelle stesse condizioni sia annodato è:*

$$P(n, L) = 1 - p_n^0(L)/p_n(L) = 1 - e^{-\alpha(L)n+o(n)}$$

con $\alpha(L) > 0$.

Dimostrazione. Prendiamo come $K_{\alpha,\beta}$ pattern:

$$T = \{-j, -j, k, -j, -j, -k, i, i, j, j, j, -i, k, -i, -i, -k, -j, -j, i, i, k\},$$

dove i, j e k sono i vettori unitari nelle direzioni delle tre coordinate. Questa sequenza di lati è un arco annodato in un $1 - slab$ e la sua presenza in un poligono assicura che tale poligono sia annodato. \square

Una conseguenza immediata di questo teorema è il seguente corollario.

Corollario 5.3.8. *Tutti i poligoni sufficientemente lunghi confinati in un $L - slab$ di \mathbb{Z}^3 , tranne una frazione esponenzialmente piccola, sono **annodati**.*

Riguardo alle funzioni $\alpha(L)$ e $\kappa(L)$, si può mostrare che:

1. $\lim_{L \rightarrow \infty} \alpha(L) = \alpha$;
2. $\lim_{L \rightarrow \infty} \kappa(L) = \kappa$.

Inoltre, se consideriamo il numero di poligoni non annodati e non rooted in un $L - slab$, $p_n^0(L)$, il limite

$$\lim_{n \rightarrow \infty} n^{-1} \log p_n^0(L) = \kappa_0(L)$$

esiste per ogni L ed anche $\lim_{L \rightarrow \infty} \kappa_0(L) = \kappa_0$.

Lemma 5.3.9. *Le funzioni $\kappa(L)$ e $\kappa_0(L)$ sono funzioni concave di L .*

Un risultato analogo al Teorema 5.3.7 è stato provato anche per poligoni confinati in un **prisma rettangolare** (per approfondimenti si veda [26]). Si può quindi affermare che.

Proposizione 5.3.10. *Tutti i poligoni sufficientemente lunghi confinati in un prisma rettangolare (di misure (L, M) , con $L, M > 2$), tranne una frazione esponenzialmente piccola, sono **annodati**.*

5.3.3 Probabilità di linking per coppie di poligoni non vincolati

Finora abbiamo analizzato la probabilità di essere annodato di un solo poligono, ora invece vogliamo studiare la complessità di un entanglement formato da più poligoni, concentrandoci in particolare sui casi in cui i poligoni in questione sono confinati all'interno di uno slab o di un prisma. Per semplicità, restringiamoci ora al caso di due poligoni (e quindi di un link a due componenti). Chiamiamo *tipo di link* la classe di equivalenza di un link.

Definizione 5.9. Introduciamo due diverse definizioni di *linking*:

1. due curve semplici chiuse disgiunte C_1 e C_2 si dicono *topologicamente unlinked* se esiste un omeomorfismo di \mathbb{R}^3 in se stesso, $H : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, tale che le immagini $H(C_1)$ e $H(C_2)$ siano separate da un piano bidimensionale, in tal caso si dice anche che il link $C_1 \sqcup C_2$ è *split*;
2. C_1 e C_2 si dicono *omologicamente unlinked* se C_1 borda una superficie orientabile che è disgiunta da C_2 , o viceversa (la relazione è simmetrica).

Osserviamo che il **linking omologico** implica quello **topologico**.

Si può dimostrare che per capire quando due curve C_1 e C_2 sono omologicamente linked è sufficiente calcolare il linking number $Lk(C_1, C_2)$ e verificare che sia diverso da zero; individuare il linking topologico, invece, è un problema un po' più complicato. Per il momento quindi consideriamo il linking

nel senso omologico, vedremo in seguito in quale misura i risultati ottenuti ci danno informazioni anche sul linking topologico.

Per coppie di poligoni linked non soggetti a vincoli geometrici, è stato dimostrato che dati due poligoni che formano un tipo di link τ fissato (a eccezione del link banale), tale che i due poligoni siano non annodati, la costante di connettività dei poligoni è uguale a κ_0 indipendentemente dal tipo di link τ .

Teorema 5.3.11. [24] *Se $p_n^{(2)}(t)$ è il numero di immersioni iniettive in \mathbb{Z}^3 di due poligoni, ognuno di n lati, che formano un link di tipo τ , dove τ è un qualsiasi link con entrambe le componenti non annodate (eccetto il link banale), allora:*

$$\lim_{n \rightarrow \infty} \frac{\log p_n^{(2)}(\tau)}{2n} = \kappa_0.$$

Si veda [24] per la dimostrazione.

Si può anche pensare di aumentare il tipo di link τ considerando l'unione di tutti i tipi di link che possono essere ottenuti da τ prendendo la somma (si veda Sezione 1.5) di ogni componente unknotted (non annodata) con un nodo qualsiasi (compreso il nodo banale). Chiamiamo questa unione, e quindi questa classe di equivalenza aumentata, il *link aumentato* $\tau^\#$. Ovviamente, ogni componente del link aumentato può essere annodata e ci chiediamo con quale probabilità questo avvenga. La risposta a questa domanda è data dal teorema seguente:

Teorema 5.3.12. [24] *La probabilità $P^{(2)}(n, \tau^\#)$ che entrambe le componenti (con n lati) di un link aumentato $\tau^\#$ siano annodate **tende a uno** come:*

$$P^{(2)}(n, \tau^\#) = 1 - e^{-\alpha n + o(n)}$$

quando $n \rightarrow \infty$.

Ovvero ognuna delle due componenti del link aumentato sarà annodata per $n \rightarrow \infty$, escluso un numero esponenzialmente piccolo di eccezioni.

5.3.4 Probabilità di linking per coppie di poligoni in ambiente vincolato

Vediamo ora come i risultati raccolti finora possono essere estesi al caso di coppie di poligoni contenute in un L – slab. In particolare.

Teorema 5.3.13. *Se chiamiamo $p_n^{(2)}(L, \tau^\#)$ il numero di immersioni iniettive in \mathbb{Z}^3 di due poligoni ognuno di n lati, confinati in un slab di ampiezza L , e che formano un link aumentato $\tau^\#$, dove τ è un qualsiasi link non banale, allora:*

$$\lim_{n \rightarrow \infty} \frac{\log p_n^{(2)}(L, \tau^\#)}{2n} = \kappa(L).$$

Teorema 5.3.14. *La probabilità $P^{(2)}(n, L, \tau^\#)$ che entrambe le componenti (con n lati) di un link aumentato $\tau^\#$, contenuto in una slab di ampiezza L , siano annodate **tende a uno** come:*

$$P^{(2)}(n, L, \tau^\#) = 1 - e^{-\alpha(L)n + o(n)}$$

quando $n \rightarrow \infty$.

Per le dimostrazioni dei Teoremi 5.3.13 e 5.3.14 si veda [29]. Lo stesso tipo di analisi può essere fatta per coppie di poligoni in un (L_1, L_2) – prisma ottenendo il risultato seguente.

Teorema 5.3.15. *Se chiamiamo $p_n^{(2)}((L_1, L_2), \tau^\#)$ il numero di immersioni iniettive in \mathbb{Z}^3 di due poligoni ognuno di n lati, confinati in un prisma di ampiezza (L_1, L_2) , e che formano un link aumentato $\tau^\#$, dove τ è un qualsiasi link non banale, allora:*

$$\lim_{n \rightarrow \infty} \frac{\log p_n^{(2)}((L_1, L_2), \tau^\#)}{2n} = \kappa(L_1, L_2).$$

5.4 Probabilità di annodamento e linking: risultati numerici

I risultati mostrati finora riguardano il comportamento asintotico dei poligoni. Se si è interessati, invece, alla probabilità di annodamento o di linking

per valori finiti di n , l'unico approccio utile sino ad ora consiste nell'utilizzo di metodi numerici. Con tali metodi si è scoperto ad esempio che, nonostante la probabilità che un poligono di n lati sia annodato tenda ad 1 in maniera esponenziale per $n \rightarrow \infty$, la stessa probabilità è in realtà molto bassa per valori intermedi di n (fino a circa 2000 lati). Per quanto riguarda le probabilità di annodamento e di linking dei poligoni confinati in un L -slab o in un (L, L) -prisma, con alcuni metodi numerici essa si può calcolare in maniera abbastanza precisa. Insorgono però dei problemi quando L diventa molto piccolo, in particolare risulta utile analizzare in maniera isolata il caso di $L = 1$. Ad esempio, non è necessario calcolare numericamente la probabilità che un poligono contenuto in un (1,1)-prisma sia annodato, in quanto sia l'intuito che i risultati teorici mostrano che nessun poligono contenuto in un (1,1)-prisma può essere annodato.

5.4.1 Probabilità di annodamento per poligoni in geometrie confinate

È ragionevole pensare che la probabilità di essere annodato aumenti per un poligono contenuto in un L -slab al decrescere di L . Tuttavia, non è chiaro se questa tendenza continui quando L si avvicina ad 1. In Figura 5.3 viene mostrata la probabilità di annodamento in funzione di L per poligoni con diverso numero di lati: $n = 600, 800, 1200$ e 1600 confinati all'interno di slabs di diverse ampiezze L .

Per ogni n fissato la probabilità sembra raggiungere un massimo per un valore finito di $L = L(n)$, dopo di che la frazione di poligoni annodati inizia a decrescere (per $L > L(n)$). Per confermare l'esistenza del massimo raggiunto in $L = L(n)$, è stata calcolata la probabilità di annodamento anche per $L = 1$, che è risultata essere praticamente uguale a zero per ogni n . La presenza di un massimo nella probabilità potrebbe riflettere la competizione tra una maggiore probabilità di annodamento per poligoni più compatti da un lato e la difficoltà di annodamento in geometrie confinate dovuta al cosiddetto *ingombro sterico* dall'altro. L'ingombro sterico è un fenomeno prodotto

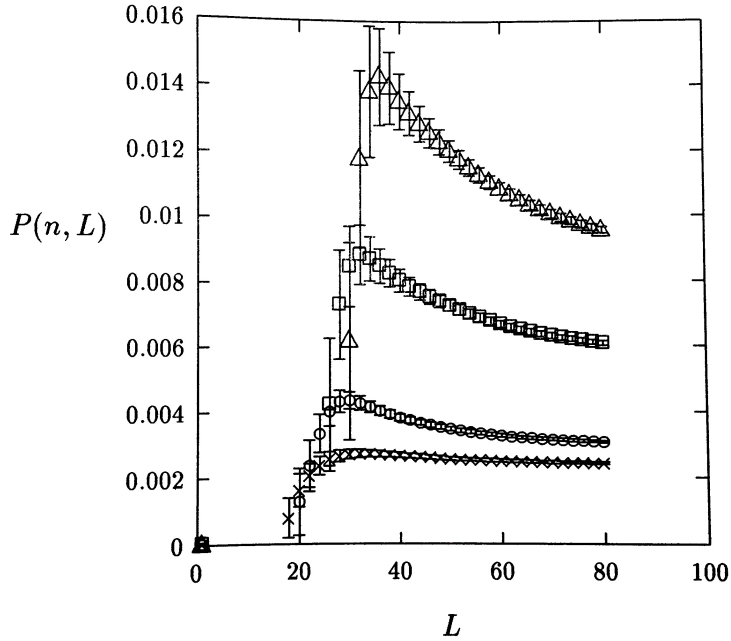


Figura 5.3: Probabilità di annodamento $P(n, L)$ rispetto ad L per poligoni confinati in una slab di ampiezza L per diversi valori di n fissati ($n = 600(\times)$, $n = 800(o)$, $n = 1200(\square)$, $n = 1600(\triangle)$) (fonte dell'immagine [29]).

dalla repulsione reciproca tra gli atomi che formano una molecola che quindi impedisce di avvicinare eccessivamente due parti della catena che compone un polimero. Inoltre, se il poligono viene schiacciato lungo un dato asse fino a raggiungere un'ampiezza L in quella direzione, allora esso si spalmerà lungo le altre direzioni, e ci si aspetterà una decrescita del numero di conformazioni annodate. Il Teorema 5.3.7 suggerisce che la frazione di conformazioni non annodate in un $L - slab$, $P^0(n, L)$ si comporti come:

$$P^0(n, L) \simeq \mathcal{A}e^{-\alpha(L)n}.$$

In Figura 5.4 viene riportato il grafico di $\ln P^0(n, L)$ in funzione di n per $L = 40, 50, 60$, ed il comportamento lineare che si osserva supporta proprio la precedente affermazione.

Inoltre da questa rappresentazione risulta chiaro che il valore di $\alpha(L)$ de-

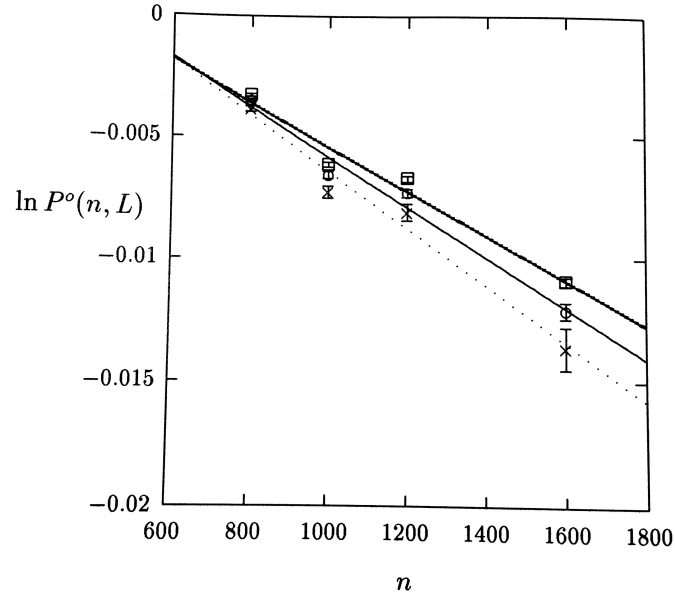


Figura 5.4: Grafico di $\ln P^o(n, L)$ rispetto ad n per poligoni confinati in una slab, per tre valori di L fissati ($L = 40(\times)$, $L = 50(o)$, $L = 60(\square)$) (fonte dell'immagine [29]).

crebbe al crescere di L .

Nella Figura 5.5 è invece rappresentato il grafico del raggio di girazione quadratico medio rispetto ad L per diversi valori di n . Il raggio di girazione (detto anche raggio d'inerzia) descrive come le componenti di un oggetto sono distribuite attorno al suo centro di massa e ci dà una misura di quanto l'oggetto in questione sia compatto. Osserviamo che, per n fissato, ogni curva ha un minimo approssimativamente in corrispondenza dello stesso $L = L(n)$ nel quale la probabilità di annodamento raggiunge il suo massimo. Questa osservazione supporta l'idea che l'annodamento dei poligoni sia strettamente correlato alla “taglia” della conformazione: ovvero più il poligono è “compatto” (e quindi il suo raggio di girazione è piccolo), più è probabile che sia annodato. Per di più, i poligoni confinati in un 1 – slab presentano raggi di girazione quadratici medi molto grandi ed infatti per questi poligoni si osserva una probabilità di annodamento molto bassa.

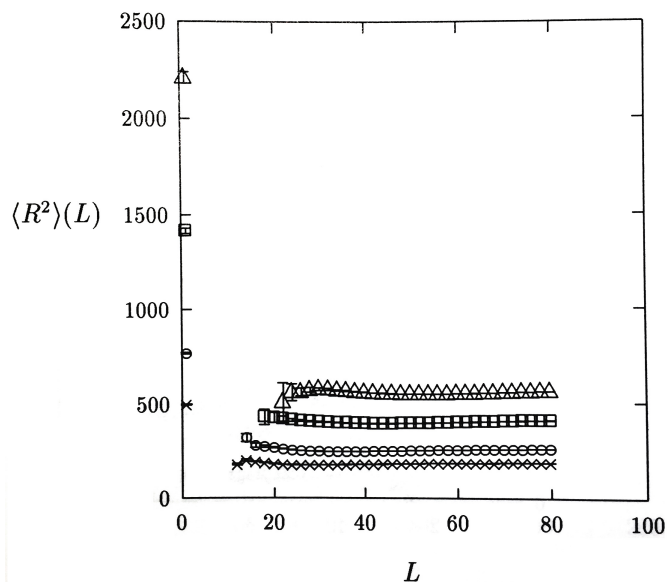


Figura 5.5: Grafico del raggio di girazione quadratico medio rispetto ad L per diversi valori di n ($n = 600(\times)$, $n = 800(o)$, $n = 1200(\square)$, $n = 1600(\triangle)$) (fonte dell'immagine [29]).

Per i poligoni confinati in un (L, L) -prisma è stato osservato esattamente lo stesso comportamento che abbiamo trovato nel caso di poligoni in una slab. L'unica differenza è il valore dei massimi, che sono considerevolmente più alti rispetto al caso delle slab; infatti all'interno di un prisma un poligono è più vincolato, e dunque più compatto, e per questo motivo ci aspettiamo una probabilità maggiore di annodamento. Inoltre, anche in questo caso si è osservato che il raggio di girazione quadratico medio presenta un minimo approssimativamente in corrispondenza del valore di L in cui la probabilità di annodamento raggiunge un massimo.

5.4.2 Probabilità di linking

Passiamo ora al caso di più poligoni, in particolare consideriamo come prima il caso di due poligoni e analizziamo con quale probabilità essi sono linked

tra loro. Siamo interessati a situazioni in cui i due poligoni non siano troppo distanti l'uno dall'altro, perchè in caso contrario essi saranno certamente unlinked. Per concentrare l'attenzione su questo genere di coppie di poligoni è stato utilizzato un opportuno metodo di campionamento. Per analizzare la dipendenza della probabilità di linking da un vincolo geometrico, come ad esempio un cubo di lato L , è stato scelto, per un L fissato, il sottoinsieme di coppie di poligoni che potevano essere contenute nel L - cubo. Per poter contare il numero di modi in cui una data coppia può essere contenuta in un L - cubo, due entanglement (formati dalla coppia di poligoni) verranno considerati distinti anche quando sono sovrapponibili per traslazione lungo le direzioni x , y o z all'interno del cubo.

5.4.3 Coppie di poligoni contenute in un cubo

I risultati ottenuti nelle Sezioni 5.3.3 e 5.3.4, in particolare il Teorema 5.3.11, riguardavano coppie di poligoni le cui configurazioni non erano influenzate dalla presenza di vincoli. Per due poligoni (con n lati ciascuno) confinati in un cubo di lato L , ci aspettiamo che tale comportamento continui finché $L \sim n^\nu$, dove ν rappresenta l'esponente che caratterizza la dipendenza da n del raggio di girazione quadratico medio. Per $L < n^\nu$, i poligoni saranno più compatti, e tale effetto aumenterà al decrescere di L (per un n fisso). Concentriamoci su questo andamento: in Figura 5.6 è riportata la probabilità per le coppie di poligoni di essere omologicamente linked (ovvero di avere linking number non nullo) in funzione di n e per diversi valori di L .

Le caratteristiche qualitative sono visibili: la probabilità di linking cala all'aumentare di L per n fissato. In effetti ciò che guida il comportamento di tale probabilità è il rapporto tra le due misure n^ν e L . Nella Figura 5.7 si può osservare proprio la probabilità di linking in funzione del rapporto $n/L^{1/\nu}$, ed è chiaro come tutti i dati, per differenti valori di n ed L , cadono su una singola curva. Inoltre si può osservare che la probabilità di linking è ancora molto piccola per $n/L^{1/\nu} = 1$.

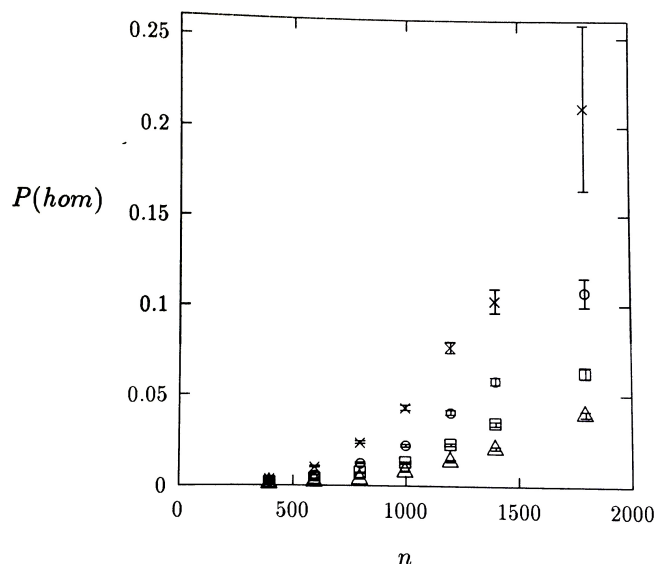


Figura 5.6: Probabilità per le coppie di poligoni di essere omologicamente linked in funzione di n e per diversi valori di L ($L = 72(\times)$, $L = 84(\circ)$, $L = 96(\square)$, $L = 108(\triangle)$) (fonte dell'immagine [29]).

Dato che il linking omologico è la forma più debole di linking, è interessante analizzare quanto esso sia efficace come indicatore del linking topologico. In Figura 5.8 è riportata la probabilità per una coppia di poligoni di essere omologicamente linked, dato che essi sono topologicamente linked, in funzione di $n/L^{1/\nu}$. È necessario precisare che gli indicatori usati per individuare il linking di tipo topologico non sono del tutto accurati, e qualche coppia di poligoni topologicamente linked potrebbe essere rimasta esclusa, ma si ritiene che per questi valori di n l'errore non sia rilevante ai fini dell'analisi.

Come possiamo vedere, ci sono due diversi tipi di andamento che corrispondono alla dominanza di uno dei due parametri (n^ν e L) sull'altro. In particolare per $n \ll L^{1/\nu}$, che corrisponde a un cubo grande, quasi tutte le coppie topologicamente linked sono omologicamente *unlinked*. D'altra parte, al crescere di $n/L^{1/\nu}$, la frazione di coppie di poligoni topologicamente linked che sono anche omologicamente linked cresce e tende all'unità per

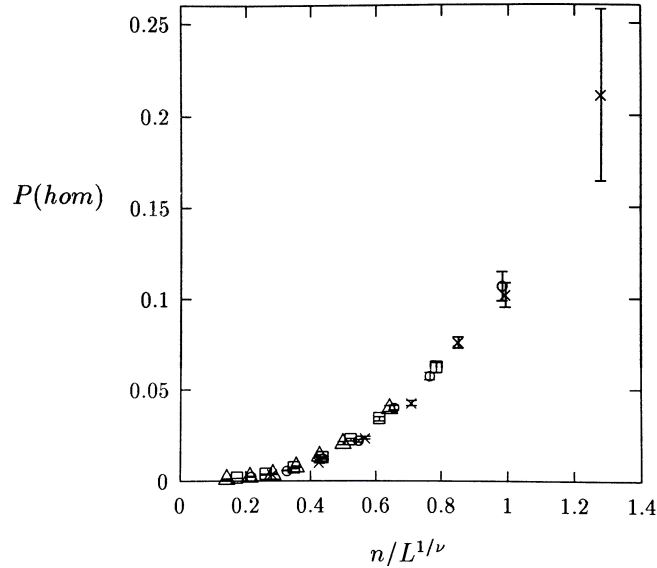


Figura 5.7: Probabilità di linking omologico di una coppia di poligoni contenuta in un cubo, in funzione del rapporto $n/L^{1/\nu}$ (fonte dell'immagine [29]).

$n \gg L^{1/\nu}$. Questo comportamento, che non è ovvio a priori, suggerisce che il linking omologico è un buon indicatore del linking topologico per configurazioni in cui i due poligoni si compenetrano fortemente, mentre diventa un indicatore meno efficace per configurazioni in cui le due componenti sono, in media, più lontane tra loro.

5.4.4 Coppie di poligoni contenute in slabs

Consideriamo ora coppie di poligoni (ciascuno di n lati) con distanza dei loro centri di massa non maggiore di d . Ancora una volta abbiamo due parametri, d ed n^ν , e ci aspettiamo che il rapporto tra di essi condizioni l'andamento della probabilità di linking. Anche in questo caso si può osservare la presenza di due tipi di regime: per $n \ll d^{1/\nu}$ i centri di massa dei due poligoni sono molto distanti tra loro e la probabilità di linking è quasi nulla. Al crescere di $n/d^{1/\nu}$, invece, i poligoni si avvicinano tra loro ed è

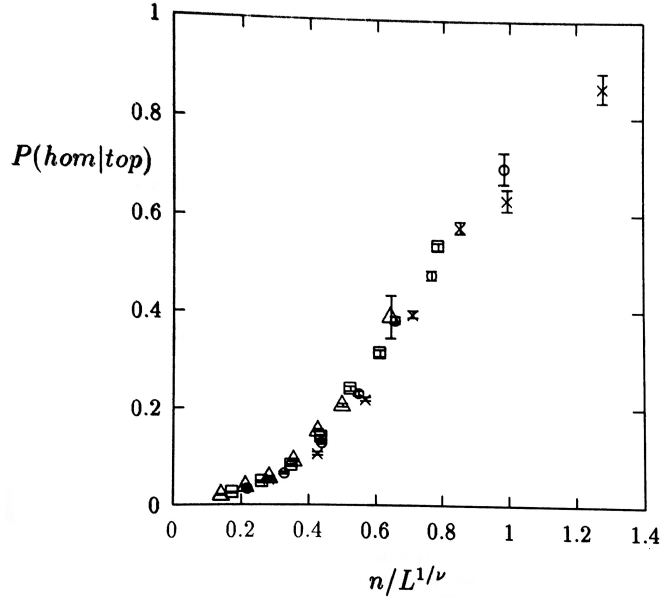


Figura 5.8: Probabilità di linking topologico dato che è presente il linking omologico in funzione del rapporto $n/L^{1/\nu}$ (fonte dell'immagine [29]).

più facile che siano linked. Le caratteristiche qualitative globali che abbiamo osservato per la probabilità di linking sono molto simili a quelle osservate per la probabilità di annodamento. In particolare, anche nel caso di coppie di poligoni contenute in uno slab o prisma, la probabilità di linking incontra un massimo quando è rappresentata in funzione di L , la cui posizione dipende da n . Questo indica che le probabilità di annodamento e linking di polimeri chiusi ad anello è influenzata essenzialmente nello stesso modo dalla presenza di vincoli geometrici, come slabs o prismi, e dunque che si tratta di due quantità che indicano essenzialmente caratteristiche topologiche simili.

5.5 Conclusioni dell'articolo

In questo articolo è stata analizzata l'influenza dei vincoli geometrici (come slabs e prismi) sulla probabilità di annodamento e linking di polimeri chiusi ad anello. Si è mostrato che la probabilità di annodamento di tutti i

polimeri, tranne un numero esponenzialmente piccolo, confinati in una slab o un prisma tende ad 1 in maniera esponenziale quando $n \rightarrow \infty$. Si è anche mostrato che entrambe le componenti di un link sono annodate con probabilità molto alta e si è caratterizzata la dipendenza da L del fattore esponenziale $\alpha(L)$, mostrando che $\lim_{L \rightarrow \infty} \alpha(L) = \alpha$. Dopo di che è stato investigato l'andamento della probabilità per n piccoli, attraverso metodi numerici, e sono state osservate caratteristiche simili per le probabilità di annodamento e linking. Più precisamente, entrambe mostrano un massimo quando vengono rappresentate come funzioni di L (in un L-slab o (L,L)-prisma) per valori fissi di n .

Ciò potrebbe riflettere la competizione tra una probabilità di essere aggrovigliati (annodati, linked o entrambi), più alta per anelli vincolati in quanto più compatti, e la difficoltà di formazione di nodi e link in geometrie confinate dovuta all'ingombro sterico tra le catene.

Ovviamente, ci sono ancora molte domande aperte, ad esempio riguardanti l'influenza reciproca dei vincoli geometrici e della struttura topologica dei polimeri in casi diversi dalle catene lineari e dai poligoni. Oppure ci si potrebbe interessare al caso di vincoli geometrici meno regolari di quelli qui considerati.

Appendice A

Categorie Top, Diff, PL

In questa appendice richiamiamo alcune notazioni e le principali nozioni sulle categorie *Top*, *Diff* e *PL*. Per ulteriori riferimenti si vedano [15, 16, 19]. Indichiamo con:

- \mathbb{N}^0 i numeri naturali (zero compreso), $\mathbb{N} = \mathbb{N}^0 - 0$;
- $\mathbb{N}_n^0 = \{i \in \mathbb{Z} \mid 0 \leq i \leq n\}$, $\mathbb{N}_n = \mathbb{N}_n^0 - 0$;
- \mathbb{E}^n lo spazio euclideo di dimensione n ;
- $B^n = \{x \in \mathbb{E}^n \mid \|x\| \leq 1\}$, $\text{int}(B^n) = \{x \in \mathbb{E}^n \mid \|x\| < 1\}$;
- $S^n = B^{n+1} - \text{int}(B^{n+1}) = \{x \in \mathbb{E}^{n+1} \mid \|x\| = 1\}$.

Definizione A.1. Siano $U \subset \mathbb{E}^k$, $V \subset \mathbb{E}^l$ insiemi aperti. Un'applicazione $f : U \rightarrow V$ si dice *di classe C^r* ($r \in \mathbb{N}$) se è continua e tutte le derivate parziali $\frac{\partial^h f}{\partial x_{i_1} \dots \partial x_{i_n}}$ ($h \in \mathbb{N}_r$) esistono e sono continue. Dati due insiemi arbitrari $X \subset \mathbb{E}^k$ e $Y \subset \mathbb{E}^l$, un'applicazione $f : X \rightarrow Y$ si dice *di classe C^r* se, per ogni $x \in X$ esistono un intorno aperto U di x ed un'applicazione $g : U \rightarrow \mathbb{E}^l$ di classe C^r tale che $f|_{U \cap X} = g|_{U \cap X}$. Si scriverà $f \in C^r(X, Y)$ o semplicemente $f \in C^r$; si definisce $C^\infty(X, Y) = \bigcap_{r \in \mathbb{N}} C^r(X, Y)$; infine, per $X \subset \mathbb{E}^k$ e $Y \subset \mathbb{E}^l$ qualunque, $C^0(X, Y)$ indicherà la classe delle applicazioni continue, rispetto alla topologia indotta da \mathbb{E}^k ed \mathbb{E}^l , da X a Y (anche dette *mappe*).

Definizione A.2. Un *diffeomorfismo di classe C^r* $f : X \rightarrow Y$ è un'applicazione biettiva di classe C^r , tale che $f^{-1} \in C^r(X, Y)$.

Definizione A.3. Un insieme $M \subset \mathbb{E}^k$ è detto *varietà di classe C^r* ($r \in \mathbb{N}^0$) di *dimensione n* , se per ogni $x \in M$ esiste un diffeomorfismo di classe C^r $f : W \rightarrow U$, dove W è un intorno di x in M (nella topologia indotta da \mathbb{E}^k) e U è un sottoinsieme aperto di \mathbb{E}^n .

Le varietà (di qualunque dimensione) di classe C^r , e le applicazioni di classe C^r fra di esse, formano una categoria, per r fissato. In particolare si indica con **Top** la categoria di varietà di classe C^0 (varietà *topologiche*) e delle applicazioni continue, e con **Diff** la categoria delle varietà e applicazioni di classe C^∞ (varietà e applicazioni *lisce*). Le due ultime categorie sono essenzialmente diverse, nel senso che esistono varietà equivalenti in *Top* (cioè omeomorfe) e non equivalenti in *Diff* (cioè non diffeomorfe).

Definizione A.4. Una *varietà con bordo* di classe C^r di dimensione n è un sottoinsieme $M \subset \mathbb{E}^k$ tale che, per ogni $x \in M$, esiste un diffeomorfismo $f : W \rightarrow U$ di classe C^r , dove W è un opportuno intorno di x in M e U è un sottoinsieme aperto del semispazio $\mathbb{E}_+^n = \{(x_1, \dots, x_n) \in \mathbb{E}^n \mid x_n \geq 0\}$. L'insieme $\text{int}(M)$ di punti di M che ammettono un intorno diffeomorfo a B^n forma una varietà di classe C^r di dimensione n nel senso precedente. Inoltre $\partial M = M - \text{int}(M)$ è una varietà di classe C^r di dimensione $(n-1)$ detta *bordo di M* . Le varietà, nel senso definito precedentemente, si possono considerare varietà con bordo vuoto; una varietà si dice *chiusa* se è compatta e con bordo vuoto. Spesso le varietà con bordo vuoto sono chiamate semplicemente *varietà*.

Definizione A.5. Dati, in uno spazio euclideo \mathbb{E}^m , $n+1$ punti affinementemente indipendenti v^0, \dots, v^n , la chiusura convessa di $\{v^0, \dots, v^n\}$, cioè

$$s^n = \{p \in \mathbb{E}^m \mid p = \sum_{i=0}^n \lambda_i v^i, \text{ con } \lambda_i \in \mathbb{R}, \lambda_i \geq 0, \sum_{i=0}^n \lambda_i = 1\},$$

si dirà *simplexso di dimensione n* (o *n -simplexso*) *generato da v^0, \dots, v^n* e si indicherà con il simbolo $\langle v^0, \dots, v^n \rangle$; v^0, \dots, v^n sono detti *vertici* di s^n . Con

un leggero abuso di linguaggio, uno 0-simplesso verrà sempre confuso con il proprio unico vertice. Gli 1-simplessi sono anche detti *lati* o *spigoli*. Dato un n -simplesso s^n , e dato un insieme di $q + 1$ suoi vertici ($q \geq 0$), il q -simplesso da essi generato si dice *faccia* di s^n ; una faccia s^q di s^n si dice *propria* se $s^q \neq s^n$.

Definizione A.6. Un *complesso simpliciale* (o, brevemente, *complesso*) K è un insieme di simplessi di uno stesso spazio \mathbb{E}^m , tale che:

1. se $s^n \in K$ e s^q è una faccia di s^n , allora $s^q \in K$;
2. se $s^n, s^p \in K$, allora $s^n \cap s^p$ o è vuoto, o è una faccia comune di s^n ed s^p ;
3. ogni simplesso di K è faccia di un insieme finito di simplessi di K ;
4. rispetto alla topologia indotta da \mathbb{E}^m , $A \subset \bigcup K$ è chiuso se e solo se, per ogni $s \in K$, $A \cap s$ lo è.

Si dice *dimensione* di un complesso K , e si indica con $\dim K$ la massima delle dimensioni dei suoi simplessi. Dati due complessi K ed L , un'*applicazione simpliciale* da K a L è un'applicazione $f : K \rightarrow L$ tale che:

1. per ogni vertice (cioè 0-simplesso) v di K , $f(v)$ è un vertice;
2. se $s^n = \langle v^0, \dots, v^n \rangle \in K$, allora $f(s^n)$ è un simplesso di L generato da $f(v^0), \dots, f(v^n)$.

Dato un complesso K , un insieme $L \subset K$ che sia un complesso si dice *sottocomplesso* di K .

Si ha che:

- Dato un simplesso s^n , l'insieme \bar{s}^n di tutte le facce di s^n (esso stesso compreso) è un complesso simpliciale; l'insieme \dot{s}^n di tutte le facce proprie di s^n costituisce un sottocomplesso di \bar{s}^n detto *bordo* di s^n .

- Dato un complesso K ed un semplice $s^q \in K$, la *stella di s^q in K* è l'insieme $st(s^q, K)$ formato da tutti i semplici di K contenenti s^q , e da tutte le loro facce; essa è un sottocomplesso di K . L'insieme $lk(s^q, K)$ formato dai semplici di $st(s^q, K)$ disgiunti da s^q è un sottocomplesso di $st(s^q, K)$ detto *cintura di s^q in K* .
- Dato un complesso K , l'insieme dei suoi semplici di dimensione $\leq r$ (dove $r \leq \dim K$) costituisce un suo sottocomplesso denotato da K^r e detto *r -scheletro di K* .

I complessi e le applicazioni simpliciali costituiscono rispettivamente gli oggetti e i morfismi di una categoria.

Ogni n -simpleso, con la topologia indotta da \mathbb{E}^m , è omeomorfo a B^n .

Definizione A.7. Si dice *corpo* di un complesso K di E^m lo spazio topologico $|K|$ il cui insieme di punti è l'unione dei semplici di K , e la cui topologia è quella indotta da \mathbb{E}^m . Per ogni punto $p \in |K|$, il semplice di minima dimensione fra quelli contenenti p è detto il *supporto* di p . Data una qualunque applicazione simpliciale $f : K \rightarrow L$, si noti che essa è completamente determinata dalla propria restrizione allo 0-scheletro K^0 . Ad f si associa una applicazione continua $|f| : |K| \rightarrow |L|$ come segue: per ogni $p \in |K|$, $p = \sum_{i=0}^q \lambda_i v^i$, dove $\{v^i\}_{i \in \mathbb{N}_q^0}$ è l'insieme dei vertici del supporto di p ; si definisce:

$$|f|(p) = \sum_{i=0}^q \lambda_i f(v^i).$$

“ $| \cdot |$ ” (che si legge “corpo”) è un funtore dalla categoria dei complessi e delle applicazioni simpliciali alla categoria degli spazi topologici e applicazioni continue. L'immagine di $| \cdot |$ è una sottocategoria di tale categoria: i suoi soggetti si dicono *poliedri* e i suoi morfismi *applicazioni lineari a tratti* (o *PL*).

Definizione A.8. Una *triangolazione* di uno spazio topologico X è una coppia (K, f) , dove K è un complesso e $f : |K| \rightarrow X$ è un omeomorfismo.

Talvolta si dirà che K triangola X o, con abuso, che è esso stesso una triangolazione di X . Si è già visto, per esempio, che ogni \bar{s}^n triangola B^n ; inoltre \dot{s}^n triangola S^{n-1} . Importante è il fatto che, se K triangola uno spazio compatto, K è finito. Due triangolazioni (K, f) , (L, g) di uno stesso spazio X si dicono *equivalenti* se esiste un omeomorfismo PL $h : |K| \rightarrow |L|$ che renda commutativo il diagramma in Figura A.1.

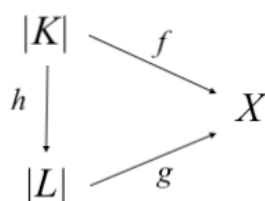


Figura A.1: Equivalenza di triangolazioni.

Definizione A.9. Una *suddivisione* di un complesso K è un complesso L tale che $|L| = |K|$ e ogni semplice di L è contenuto in un semplice di K .

Proposizione A.0.1. *Se due complessi K, L hanno lo stesso poliedro come corpo, essi ammettono una suddivisione comune.*

Definizione A.10. Un complesso n -dimensionale K si dice *varietà combinatoria n -dimensionale* se, per ogni vertice v di K , esiste un'applicazione simpliciale biettiva da un'opportuna suddivisione di $lk(v, K)$ ad un'opportuna suddivisione del bordo di un n -simpleso; detto altrimenti, $|lk(v, K)|$ è PL-omeomorfo al corpo del bordo di un n -simpleso. Il concetto di varietà combinatoria risulta, in effetti, invariante per suddivisione; pertanto è ben data la seguente definizione: un poliedro $M = |K|$ si dice *varietà PL* se K è una varietà combinatoria.

Le varietà PL e le applicazioni PL fra esse formano una categoria, indicata con **PL**.

Definizione A.11. Una *varietà combinatoria n -dimensionale con bordo* è definita come un complesso n -dimensionale K in cui, per ogni vertice $v \in K$, $|lk(v, K)|$ è PL-omeomorfo o (1) al corpo del bordo di un n -simpleso o (2) ad un $(n - 1)$ -simpleso. Per ogni vertice v vi è un omeomorfismo PL $f : |st(v, K)| \rightarrow |\bar{s}^n|$; per ogni vertice v di tipo (2), si considerino gli $(n - 1)$ -simplessi di $f^{-1}(|\dot{s}^n|)$ che non fanno parte di $|lk(v, K)|$; tali simplessi, con le loro facce, costituiscono un sottocomplesso di K detto *bordo* di K , che è in effetti una varietà combinatoria $(n - 1)$ -dimensionale. Una *varietà PL con bordo* è il corpo M di una varietà combinatoria con bordo; il *bordo* \dot{M} di M è il corpo del bordo di K .

Spesso si utilizza una triangolazione, nello studio di una varietà topologica, perché, in generale, calcoli e dimostrazioni sono più semplici nella categoria *PL* che nella categoria *Top*. Un problema essenziale è, quindi, sapere quali varietà siano triangolabili; un problema altrettanto importante è quello della validità della cosiddetta *Hauptvermutung* (o “congettura fondamentale della topologia”) che asserisce che due triangolazioni qualsiasi di una stessa varietà topologica hanno corpi PL-omeomorfi. Triangolabilità e *Hauptvermutung* sono banali in dimensione 1 e sono state dimostrate in dimensione 2 e 3; la *Hauptvermutung* non vale per dimensione $n \geq 5$ e, nelle stesse dimensioni, non tutte le varietà topologiche sono triangolabili mediante varietà combinatorie. Sono tuttora irrisolti i problemi della validità di *Hauptvermutung* in dimensione 4 e della triangolabilità (mediante complessi che eventualmente non siano varietà combinatorie) in dimensione $n \geq 4$. La non triangolabilità significa che, per $n \geq 5$, qualora si volesse sostituire la classificazione in *Top* (cioè di varietà topologiche a meno di omeomorfismi) con la classificazione in *PL* (cioè di varietà PL a meno di omeomorfismi PL), allora alcune varietà topologiche verrebbero trascurate; la falsità della *Hauptvermutung* significa che alcune varietà omeomorfe verrebbero distinte. Dato che, storicamente e per le applicazioni, la categoria più notevole è *Diff* è confortante sapere che le varietà lisce sono tutte triangolabili mediante varietà combinatorie. Invece esistono, per $n \geq 10$, n -varietà PL (e quindi

anche topologiche) che non ammettono alcuna struttura differenziabile; ma già in dimensione 4 esistono varietà topologiche che non ammettono struttura differenziabile né PL. Inoltre, una stessa n -varietà PL può ammettere diverse strutture differenziabili per $n \geq 7$. Un risultato positivo è l'esistenza e unicità di una struttura differenziabile su una 4-varietà PL, risultato che è stato poi esteso alle dimensioni 5 e 6. La coincidenza (indicata con =), o meno, delle equivalenze nelle varie categorie si può, dunque, riassumere nella tabella in Figura A.2.

Dimensione:	≤ 3	4	5, 6	≥ 7
	<i>Top</i> // \	<i>Top</i> ? ?	<i>Top</i> ✕ ✕	<i>Top</i> ✕ ✕
	<i>PL = Diff</i>	<i>PL = Diff</i>	<i>PL = Diff</i>	<i>PL ≠ Diff</i>

Figura A.2: Relazione tra le categorie *Top*, *Diff* e *PL* nelle varie dimensioni.

Un'analisi sistematica del confronto fra le tre categorie è contenuta in [15], si veda anche [19]. In ogni caso, ai fini di questa trattazione, è importante ricordare solamente il caso di dimensione ≤ 3 in cui le tre categorie: *Top*, *PL* e *Diff* coincidono.

Bibliografia

- [1] Adams C.C., *The knot book, an elementary introduction to the mathematical theory of knots*, W.H. Freeman and Company, New York (1994).
- [2] Alexander J.W., *An example of a simply connected surface bounding a region that is not simply connected*, Proc. Nat. Acad. Sci. USA 10 (1924), 8–10.
- [3] Benedetto D., Degli Esposti M., Maffei C., *Matematica per le scienze della vita*, Casa editrice Ambrosiana, Milano (2008).
- [4] Brown M., *Locally flat imbeddings of topological manifolds*, Ann. of Math., 75 (1962), 331–341.
- [5] Buck D., *DNA topology*, PROC SYM AP volume 66 (2009).
- [6] Buck D., Flapan E., *A topological characterization of knots and links arising from site-specific recombination*, J. Phys. A: Math. Theor. 40 (2007), 12377-12395.
- [7] Buck D., Flapan E., *Predicting knot or catenane type of site-specific recombination products*, J. Mol. Biol. 374 (2007), 1186-1199.
- [8] Buck D., Verjovsky Marcotte C., *Tangle solutions for a family of DNA-rearranging proteins*, Math. Proc. Cambridge Philos. Soc. 139 (2005), 59-80.

- [9] Buck D., Verjovsky Marcotte C., *Classification of tangle solutions for integrases, a protein family that changes DNA topology*, J. Knot Theory Ramifications 16 (2007), 969-995.
- [10] Burde G., Zieschang H., *Knots, second revised and extended edition*, Walter de Gruyter, Berlin (2002).
- [11] Dix D.B., *Mathematical models of protein folding*, preprint <http://people.math.sc.edu/dix/fold.pdf>.
- [12] Ernst C., Sumners D.W., *A calculus for rational tangles: applications to DNA recombination*, Math. Proc. Cambridge Philos. Soc. 108 (1990), 489-515.
- [13] Graeub W., *Die semilinearen Abbildungen*, Springer Berlin Heidelberg, Berlin (1950).
- [14] Haglund E., Sułkowska J.I., He Z., Feng G-S., Jennings P.A., Onuchic J.N., *The Unique Cysteine Knot Regulates the Pleiotropic Hormone Leptin*, Plos One 7 (2012), e45654.
- [15] Kirby R., Siebenmann L., *Foundational essays on topological manifolds, smoothings and triangulations*, Princeton Univ. Press, Princeton (1977).
- [16] Ferri M., Dispense di Topologia Algebrica, reperibili alla pagina web: <http://www.dm.unibo.it/ferri/hm/progtopalg.htm>.
- [17] Leopold P.E., Montal M., Onuchic J.N., *Protein folding funnels: a kinetic approach to the sequence-structure relationship*. Proc. Natl. Acad. Sci. USA 89 (1992), 8721-8725.
- [18] Madras N., Slade G., *The self-avoiding walk*, Birkhäuser, Boston (1993).
- [19] Mandelbaum R., *Four-dimensional topology: An introduction*, Bull. Amer. Math. Soc. 2 (1980), 1-159.

- [20] Maxwell A., Bates A.D., *DNA topology*, Oxford University Press, Oxford (2005).
- [21] Moffatt H.K., Ricca R.L., *Helicity and the calugareanu invariant*, Proc. L. Soc. London Ser. A, 439 (1906) 411–429, (1992).
- [22] Murasugi K., *Knot theory & it's applications*, Birkhäuser, Boston (1996).
- [23] Nelson D.L., Cox M.M., *I principi di biochimica di Lehninger*, Zanichelli, Bologna (2006).
- [24] Orlandini E., Janse Van Rensburg E.J., Tesi M.C., Whittington S.G., *Random linking of lattice polygons*, J. Phys. A: Math. Gen. 27 (1994), 335-45.
- [25] Prasolov V.V, Sossinsky A.B., *Knots, Links, Braids and 3-Manifolds: An Introduction to the New Invariants in Low-Dimensional Topology*, Translations of Mathematical Monographs volume 154, American Mathematical Society (1996).
- [26] Soteris C.E., *Knots and graphs in subsets of \mathbb{Z}^3* , in “Topology and geometry in ploymer science”, IMA volume 103, Springer, New York (1998).
- [27] Sumners D.W., Whittington S.G., *Knots in self-avoiding walks*, J. Phis. A: Math. Gen. 21 (1988), 1689-1694.
- [28] Tesi M.C. et al. in *Knot probability for lattice polygons in confined geometries*, J.Phys. A: Math. Gen. 27 (1994), 347-360.
- [29] Tesi M.C. et al. in *Topological entanglement complexity of polymer chains in confined geometries*, in “Topology and geometry in ploymer science”, IMA volume 103, Springer, New York (1998).

Ringraziamenti

Vorrei ringraziare infinitamente tutti coloro che mi hanno aiutata e mi sono stati vicini in questo percorso, a partire dalla prof. Alessia Cattabriga che mi ha seguita durante il percorso della tesi con professionalità e umanità, fornendomi tutto il supporto necessario e trasmettendomi interesse e passione. Ringrazio anche la prof. Maria Carla Tesi che ha mostrato grande disponibilità e mi ha aiutata a realizzare con serenità la mia tesi. Un pensiero va anche agli altri professori che ho incontrato durante questi anni e che hanno risvegliato, con il loro lavoro, la mia grande passione per la matematica. Un enorme grazie va inoltre a tutte le amiche e amici che mi hanno aiutata a staccare la spina quando ne avevo bisogno e che mi sono vicini ogni giorno. Infine un ringraziamento dovuto alla mia famiglia: i miei genitori, che mi hanno permesso di arrivare fino qui e mi hanno sempre sostenuta; ma anche mio fratello e sua moglie, la mia zia Liviana, le mie meravigliose nonne Valentina e Maria e tutti quanti i miei familiari.