



**UNIVERSITÁ DEGLI STUDI DI NAPOLI “FEDERICO II”**

FACOLTA’ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

TESI DI DOTTORATO IN SCIENZE CHIMICHE

XXIII CICLO

**MACROMOLECOLE SECRETE E DI  
MEMBRANA COINVOLTE NELLA RISPOSTA  
IMMUNITARIA DEI TELEOSTEI ANTARTICI**

Candidata  
Dott.ssa Sonia Varriale

Tutore  
Prof Lelio Mazzarella

Relatore  
Prof. Piero Pucci

Cotutore  
Dott. Antonello Merlino

Cotutore  
Dott. Umberto Oreste

Coordinatore Prof. Lucio Previtera

## ABBREVIAZIONI

CDR	Regioni che Determinano la Complementarietà
CH 1-4	I-IV dominio costante della catena pesante
DEPC	dietilpirocarbonato
EDTA	Acido etilen diammino tetracetico
EMPD	Dominio extracellulare prossimale alla membrana
FR	Regione cornice
Ig	Immunoglobulina
IgH	Catena pesante delle immunoglobuline
IgL	Catena leggera delle immunoglobuline
IgM	Immunoglobulina di classe M
LB	Terreno di cultura Luria- Bertani
MD	Dinamica Molecolare
POPC	Palmitoil-oleil-fosfatidilcolina
PCR	Reazione di polimerizzazione a catena
SDS	Sodio dodecil solfato
TM 1-2	I-II dominio transmembrana delle Ig di membrana
X-Gal	5-bromo-4-cloro-3-indolil- $\beta$ -D-galattopiranoside

<b>1.Introduzione</b>	<b>1</b>
1.1 Le immunoglobuline	2
1.2 Il dominio immunoglobulinico	3
1.3 Le catene, pesante e leggera, delle immunoglobuline	7
1.4 Siti di glicosilazione delle Ig	10
1.5 Flessibilità della molecola Ig	10
1.6 Le Ig di membrana	10
1.7 <i>Splicing</i> alternativo per la sintesi della forma secreta e di membrana della catena pesante	13 14
1.8 Trasduzione del segnale del BCR	17
1.9 Analisi strutturale delle Ig	20
1.10 Le Ig dei teleostei	24
1.11 Le Ig dei teleostei antartici	27
1.12 Contesto della tesi	32
<b>2. Materiali e metodi</b>	<b>36</b>
2.1 Metodi di biologia molecolare	37
2.1.1 Campioni biologici	37
2.1.2 Preparazione dell'RNA da milza	38
2.1.3 Amplificazione mediante PCR	39
2.1.4 Estrazione di DNA genomico da tessuto	41
2.1.5 Amplificazione di frammenti genomici mediante	42

PCR	
2.1.6 Elettroforesi del DNA su gel d'agarosio	44
2.1.7 Estrazione del DNA da gel d'agarosio	44
2.1.8 Clonaggio e trasformazione	44
2.1.9 Preparazione in piccola scala del DNA plasmidico	45
2.1.10 Verifica della lunghezza dei frammenti clonati e determinazione della sequenza nucleotidica	46
2.1.11 <i>Southern blotting</i>	47
2.1.12 Trascrizione <i>in vitro</i>	48
2.1.13 Espressione di un peptide ricombinante	49
2.2 Metodi di biochimica	53
2.2.1 Rivelazione del frammento ricombinante	53
2.3 Metodi computazionali	54
2.3.1 Analisi di sequenza	54
2.3.2 Costruzione di modelli molecolari	56
2.3.3 Dinamica Molecolare	57
2.3.4 Dataset	59
2.4 Metodi chimico-fisici	60
2.4.1 Dicroismo circolare	60
2.4.2 Microscopia a forza atomica	61
<b>3 Risultati</b>	<b>63</b>

3A	Identificazione dell'isotipo L2 di catena leggera in <i>Trematomus bernacchii</i>	66
3A.1	Cloni di cDNA di isotipo L2 ottenuti per PCR	66
3A.2	Analisi delle sequenze amminoacidi che dedotte	67
3A.3	Confronto con gli isotipi delle altre specie	72
3A.4	Modelli molecolari relativi agli isotipi del dominio costante della catena L	75
3B	Studio delle interazioni del dominio trans membrana delle Ig di <i>Chionodraco hamatus</i>	78
3B.1	Costruzione dei modelli e studio di dinamica molecolare dell'elica trans membrana delle Ig	78
3B.2	Associazione elica-elica delle IgTM e loro stabilità	79
3B.3	Confronto della regione TM delle Ig di <i>Chionodraco hamatus</i> con quella di altre specie	83
3C	Studio dello <i>splicing</i> atipico delle Ig di membrana di <i>Chionodraco hamatus</i>	92
3C.1	Confronto della sequenza della catena pesante delle Ig dei teleostei	92
3C.2	Espressione di un peptide ricombinante comprendente le regioni CH2 e EMPD delle Ig di <i>Chionodraco hamatus</i>	96
3C.3	Analisi della regione genomica compresa tra gli esoni CH2 e TM in <i>Chionodraco hamatus</i>	98

3C.4 Analisi comparativa delle sequenze antiparallele e delle relative controparti	101
3C.5 Studi di frammenti di RNA complementari ad una porzione del locus delle Ig in <i>Chionodraco</i> <i>hamatus</i> e <i>Bovicthus diacanthus</i>	104
3C5.1 Predizione di struttura secondaria	105
3C5.2 Dicroismo circolare	112
3C5.3 Microscopia a forza atomica	118
<b>4 Conclusione e discussione</b>	<b>122</b>
<b>5 Bibliografia</b>	<b>126</b>
<b>6 Appendice</b>	<b>137</b>

## **1. INTRODUZIONE**

## 1.1. Le Immunoglobuline

In tutti gli animali, il sistema immunitario è in grado di difendere l'organismo dai patogeni (virus, batteri e parassiti). Nei vertebrati i vari meccanismi della risposta immunitaria possono essere distinti in due categorie: quelli afferenti all'immunità innata che coinvolgono molecole con sequenza costante, determinata a livello genomico, e quelli afferenti all'immunità adattativa, che coinvolgono molecole con sequenza variabile, determinata a livello somatico. La molecola chiave della risposta adattativa è l'Immunoglobulina (Ig) (Fig. 1). Essa è presente in due forme: una secreta nel sangue e l'altra inserita nella membrana cellulare dei linfociti B dove, insieme all'eterodimero  $Ig\alpha$ - $Ig\beta$  concorre alla formazione del complesso recettoriale denominato BCR (Recettore delle cellule B).

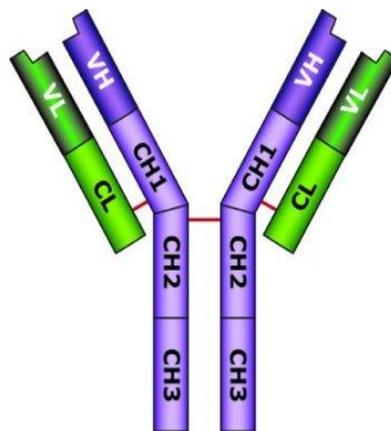


Figura 1. La molecola di Immunoglobulina.

Le Ig fanno parte di un'ampia famiglia di proteine denominata Superfamiglia delle Immunoglobuline (IgSF). I diversi membri di tale famiglia presentano similarità strutturali e di sequenza e condividono la stessa organizzazione genica; tra loro sono presenti varie proteine

di membrana cellulare coinvolte nella risposta immunitaria come i recettori delle cellule T (TCR), i componenti del complesso CD3, i recettori Fc, i marcatori di sottopopolazioni linfocitarie CD4 e CD8. Tale famiglia include anche mediatori del riconoscimento cellulare quali CD28 e CD80 (Fig. 2).

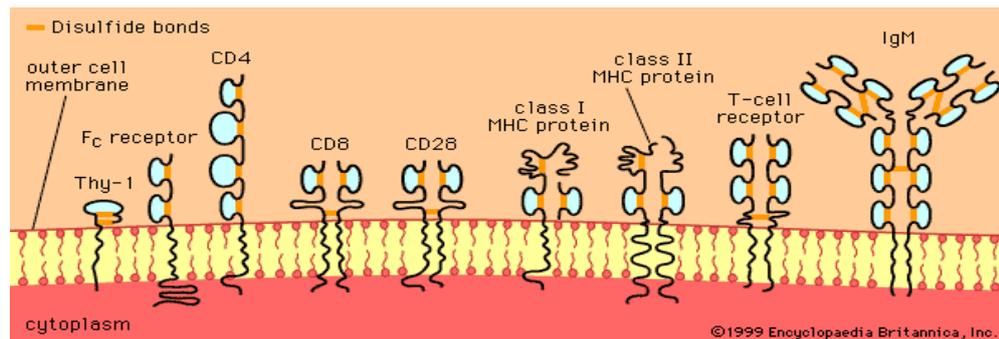


Figura 2. Esempi di molecole appartenenti alla Superfamiglia delle Ig.

## 1.2. Il dominio immunoglobulinico

Le molecole dell'IgSF sono caratterizzate da un elemento strutturale comune denominato dominio Ig, consistente in una coppia di foglietti  $\beta$  tenuti insieme da un ponte disolfurico tra due residui conservati di cisteina. Ciascun foglietto è formato da  $\beta$ -strand antiparalleli, contenente da 5 a 10 residui amminoacidici. In ciascun  $\beta$ -strand si alternano residui idrofilici rivolti verso l'esterno del dominio e residui idrofobici rivolti verso l'interno, che ulteriormente stabilizzano la struttura complessiva del dominio. I residui amminoacidici presenti nei  $\beta$ -strand sono più conservati di quelli presenti nei *loop* che connettono i  $\beta$ -strand. I domini Ig si distinguono da altri domini simili, a doppio foglietto  $\beta$  per il numero dei residui tra le due cisteine che formano il ponte; se tale numero è inferiore a 40 il

dominio non è definibile come dominio Ig. A seconda del numero di  $\beta$ -strand che formano ciascun foglietto si distinguono quattro tipi di dominio Ig. Domini non Ig possono presentare entrambi i foglietti  $\beta$

- composti da soli tre  $\beta$ -strand.

Nel dominio di tipo V un foglietto è costituito da quattro  $\beta$ -strand (A, B, E e D), l'altro da cinque (C, C', C'', F e G); le due cisteine che formano il ponte disolfurico tra strand B e F sono separati da circa 65-75 residui amminoacidici (Fig. 3).

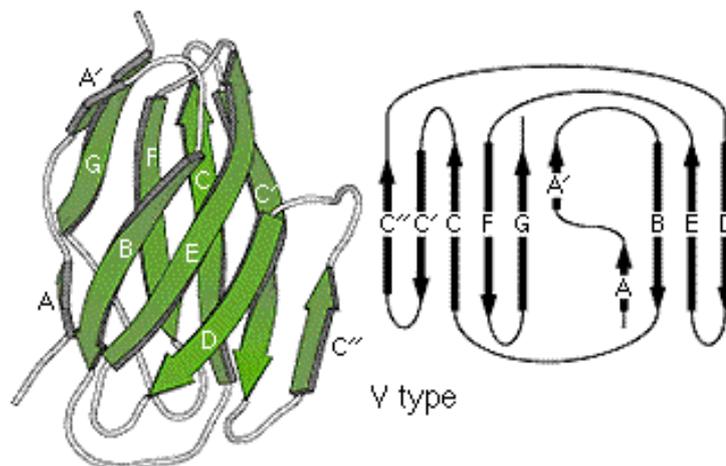


Figura 3. Rappresentazione di un dominio Ig di tipo V

Nel dominio di tipo C1 un foglietto è costituito da quattro  $\beta$ -strand (A, B, E e D), l'altro da tre (C, F e G); il numero di residui amminoacidici fra le due cisteine del ponte è variabile tra 55 e 60 (Fig. 4).

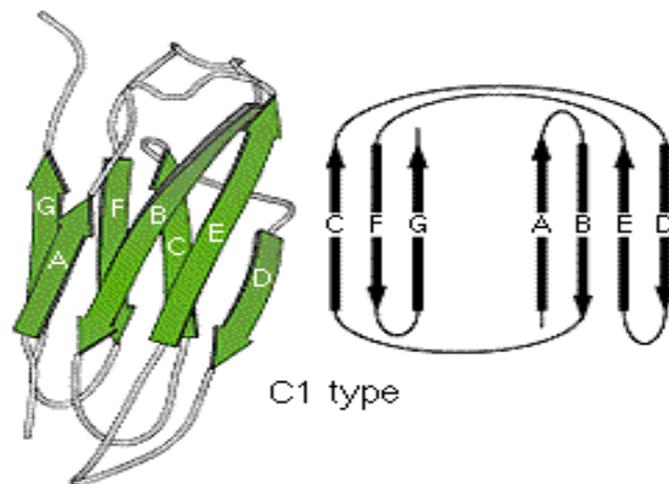


Figura 4. Rappresentazione di un dominio Ig di tipo C1.

Nella struttura del dominio di tipo C2 il numero di  $\beta$ -strand è, come per il tipo C1, tre (A, B ed E) e quattro (C, C', F e G), ma manca il  $\beta$ -strand D ed i *loop* sono in genere più corti (Williams AF and Barclay AN, 1988) (fig. 5).

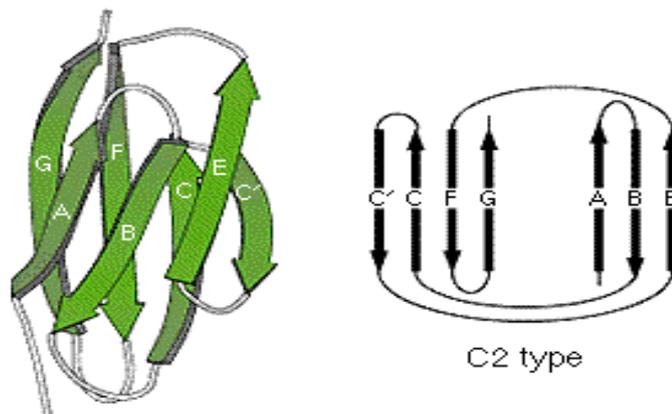


Figura 5. Rappresentazione di un dominio Ig di tipo C2.

Nel 1994 è stato descritto un nuovo tipo di dominio Ig nella proteina *Telokin*, una proteina della muscolatura liscia simile alla chinasi MYLK (Harpaz Y and Chothia C, 1994). Tale dominio,

denominato I (intermedio), rappresenta una versione troncata del dominio di tipo V che manca del  $\beta$ -strand C'' (Fig. 6).

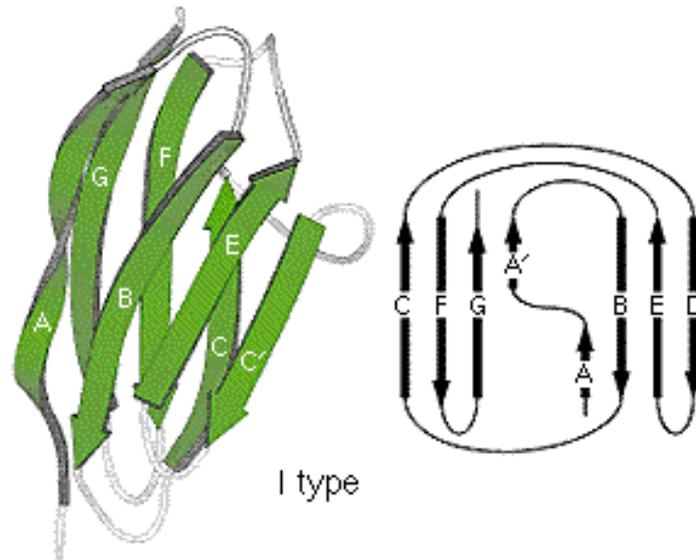


Figura 6. Rappresentazione di un dominio Ig di tipo I.

Caratteristica dei domini Ig è la capacità di appaiarsi mediante interazioni idrofobiche che coinvolgono la stessa faccia di ciascun dominio. I *loop* viceversa, data la loro flessibilità, hanno la possibilità di interagire con un elevato numero di *partners* diversi, come nel caso del dominio V delle Immunoglobuline i cui *loop* sono preposti al riconoscimento degli antigeni.

Un esame della struttura di domini Ig, ottenuta mediante diffrazione ai raggi X, condotto da Lesk e Chothia (1982), ha mostrato che i residui di cisteine che formano il ponte disolfurico e il residuo conservato di triptofano del foglietto C, spazialmente adiacente, formano un motivo strutturale nel core del dominio: Il triptofano copre il ponte disolfurico mentre altri sei residui bloccano l'intorno. Tale motivo è denominato "pin".

### 1.3. Le catene, pesante e leggera, delle Immunoglobuline

Le Immunoglobuline sono costituite da due catene pesanti (H) e due catene leggere (L) tenute insieme da ponti disolfurici.

Le catene H e L sono composte di una serie di domini Ig, ciascuno di circa 110 residui amminoacidi (Fig. 7).

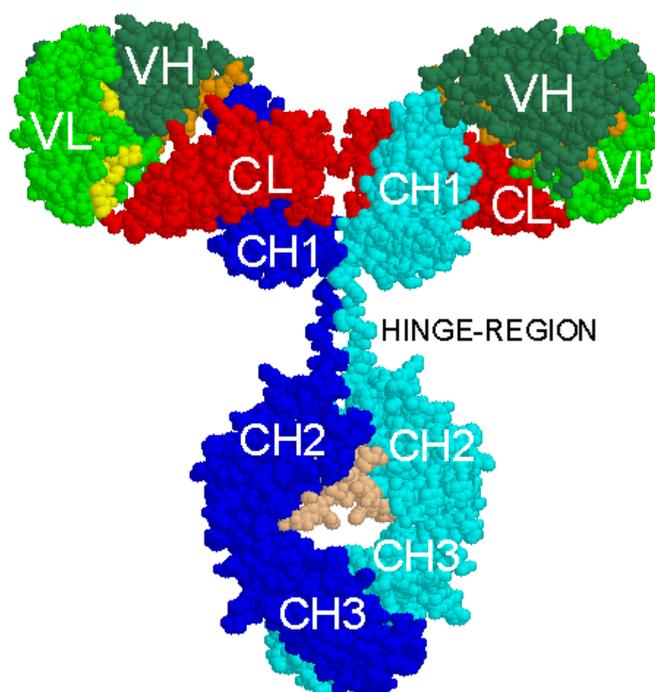


Figura 7. Rappresentazione schematica di un'immunoglobulina secreta.

Le catene leggere sono costituite da un dominio V ammino-terminale ed un dominio C1 carbossi-terminale; nei mammiferi si distinguono due isotipi di catene leggere ( $\kappa$  and  $\lambda$ ). Il dominio C1 è essenzialmente identico in tutte le catene leggere di ciascun isotipo ed è denominato regione costante della catena leggera (CL). Il dominio V è diverso a seconda del clone cellulare che lo sintetizza e contribuisce al riconoscimento dell'antigene. A causa della sua

variabilità è denominato regione variabile delle catene leggere (VL). La variabilità della regione VL è essenzialmente localizzata in tre segmenti chiamati regioni ipervariabili o regioni che determinano la complementarità (CDR). Le CDR contengono i residui amminoacidici che forniscono il contributo della catena leggera al riconoscimento dell'antigene. Le CDR sono intervallate da regioni più conservate denominate regioni cornice o *framework*; tali regioni sono quattro e sono individuate come FR1-4.

Nei mammiferi esistono cinque isotipi di catena pesante denominati  $\mu$ ,  $\delta$ ,  $\gamma$ ,  $\alpha$ , ed  $\epsilon$ . Gli isotipi  $\gamma$  ed  $\alpha$  sono ulteriormente distinti in sottoisotipi. Secondo l'isotipo di catena pesante utilizzato, l'Ig appartiene alla classe delle IgM, IgD, IgG, IgA, o IgE.

Ciascuna catena pesante consiste di un dominio ammino-terminale VH e di tre o quattro domini C1 che insieme costituiscono la regione costante CH. I domini C1 sono numerati a partire dall'ammino-terminale come CH1, CH2, CH3 e CH4. La regione VH come quella VL contiene tre CDR, che insieme alle corrispondenti regioni delle catene leggere, concorrono a determinare la struttura del sito anticorpale, e quattro FR. Nella catena pesante è anche presente una regione spaziatrice tra il CH1 ed il CH2, denominata "*hinge*" che conferisce flessibilità alla molecola e permette che i due siti anticorpali si posizionino con una certa libertà su epitopi uguali di antigeni multivalenti. Nelle IgM tale struttura è localizzata tra il CH2 ed il CH3. I domini VH e VL sono tenuti insieme da interazioni non-covalenti tra le facce omologhe dei due domini. Anche ciascuna coppia di domini costanti è tenuta insieme da interazioni non covalenti e uno o più ponti disolfurici sono diversamente localizzati nella catena H a seconda della classe.

Mediante idrolisi con papaina, la molecola Ig dà tre frammenti: due simili, ciascuno comprendente la catena leggera e i domini VH e

CH1 della catena pesante (detti Fab) e l'altro comprendente i restanti domini della catena pesante (detto Fc). L'idrolisi con pepsina dà luogo al frammento (Fab')<sub>2</sub> ed a quattro frammenti minori. I frammenti Fab contengono i due siti che legano l'antigene, sono specifici per ciascun clone cellulare che sintetizza la molecola e riconoscono un unico epitopo. Inoltre, il Fab può essere diviso in regioni dette Fv composte dai due domini variabili VH e VL, e regioni Fc composte dai domini CL e CH1.

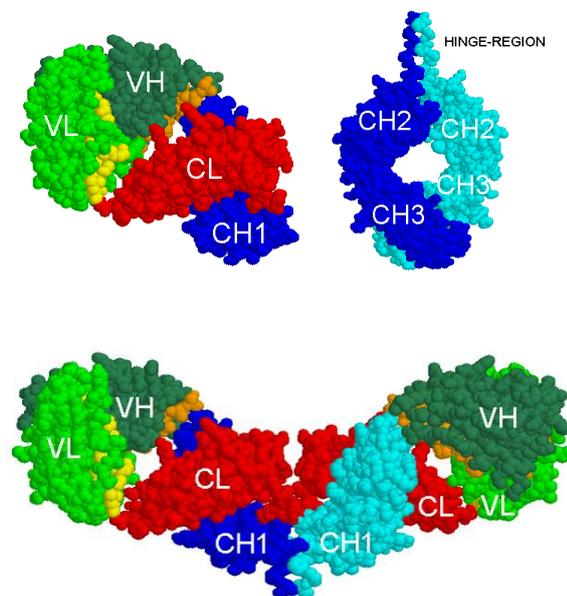


Figura 8. La struttura di un Fab (in alto a sinistra), di un frammento Fc (in alto a destra) e di un (Fab')<sub>2</sub> (in basso).

Mentre le IgG, IgD e IgE sono costituite da una singola unità strutturale H<sub>2</sub>L<sub>2</sub>, le IgA possono essere anche dimeriche (H<sub>2</sub>L<sub>2</sub>)<sub>2</sub>, e le IgM pentameriche (H<sub>2</sub>L<sub>2</sub>)<sub>5</sub>.

Le catene pesanti delle Ig di membrana presentano al carbossi-terminale una regione idrofobica che attraversa la membrana cellulare e una corta coda citoplasmatica altamente conservata, codificate dai due esoni supplementari, non trascritti nella forma secreta.

#### **1.4. Siti di glicosilazione delle Ig**

Tutte le Ig contengono catene oligosaccaridiche con legame all'asparagina; crescenti evidenze dimostrano che tali catene rivestono un ruolo strutturale e funzionale. Sebbene il tipo ed il numero delle catene oligosaccaridiche vari notevolmente tra le classi di Ig, una catena oligosaccaridica legata all'asparagina 297 nel dominio CH2 è conservata in tutte le IgG dei mammiferi e nelle corrispondenti posizioni del dominio CH2 delle IgA, IgD, e del dominio CH3 delle IgM ed IgE. La presenza dei carboidrati nell'Fc è stata dimostrata essere importante; infatti, essi determinano, a livello strutturale, il modo di appaiarsi dei domini che li contengono; proteggono siti di attacco proteolitico, modulando l'emivita della molecola; sono importanti nel meccanismo di secrezione cellulare; determinano alcune funzioni effettrici (Wawrzynczak EJ *et al.*, 1992; Arnold JN *et al.*, 2007). Inoltre, nelle IgG, la glicosilazione dei domini V influenza positivamente o negativamente il legame con l'antigene.

#### **1.5. Flessibilità della molecola Ig**

Sia i domini V che quelli CH tendono a formare dimeri compatti VH-VL e CH-CH. Tuttavia, la molecola nel suo complesso, essendo costituita da una sequenza di quattro o più di tali motivi dimerici legati come in una collana di perle, può essere vista come un paradigma della flessibilità molecolare.

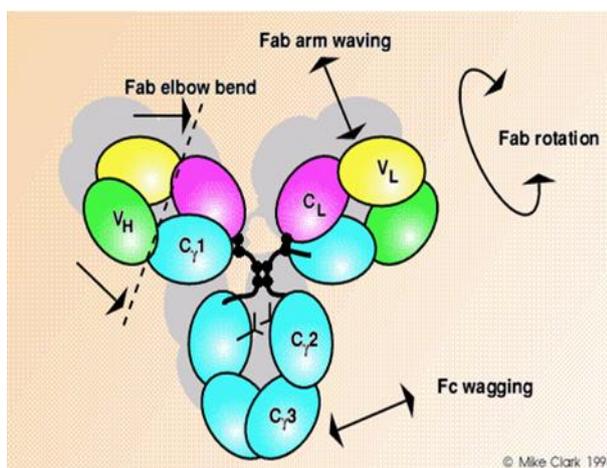


Figure 10. Rappresentazione della flessibilità della molecola immunoglobulinica e dei possibili movimenti. La flessibilità assiale e quella segmentale dipendono dalla “hinge”.

Sono stati descritti molteplici movimenti potenziali dei domini dovuti a flessibilità localizzata in diverse posizioni (Nezlin R, 1990). Alta flessibilità è localizzata tra le regioni Fv ed Fb del Fab che è denominata “*elbow bend*” (Fig. 11).

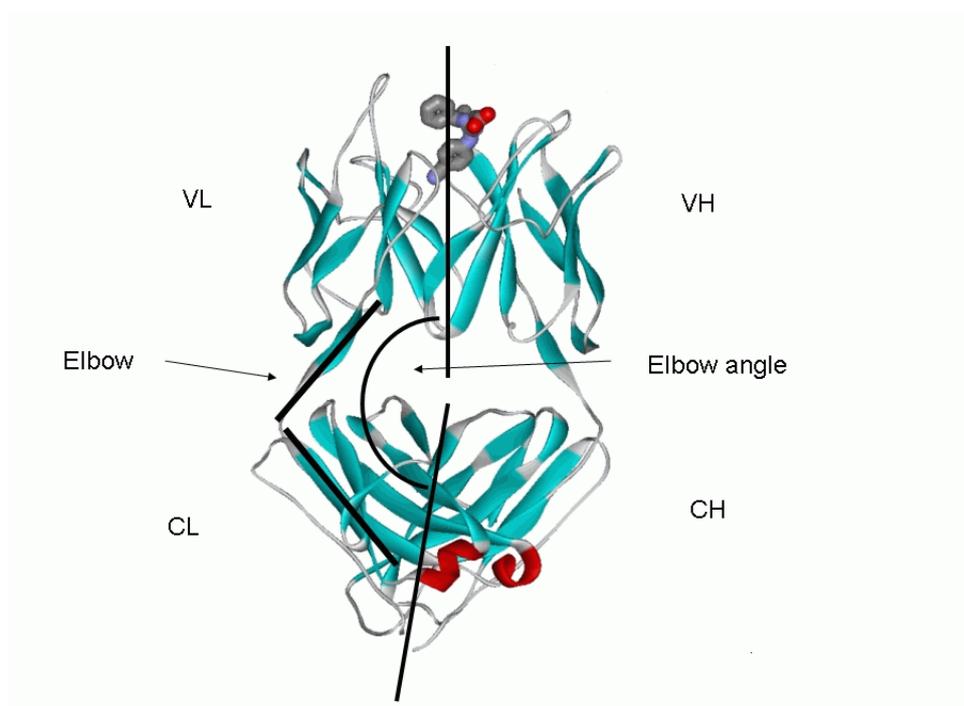


Figura 11. Curvatura del dimero VH-VL rispetto al dimero CH1-CL definita “*Elbow angle*”.

Un'interazione dinamica denominata “*ball and socket*” coinvolge 5 residui del FR1 del VH (*socket*) ed tre residui del CH1 (*ball*). Analoga situazione sussiste tra i domini VL e CL. Gli angoli “*elbow bend*” determinati da strutture cristallografiche di frammenti Fab monoclonali variano da 130° a 180° (Lesk AM and Chothia C, 1988).

Le catene leggere di isotipo  $\lambda$  sembrano permettere angoli più ampi di quelli possibili con le catene leggere di isotipo  $\kappa$  (Stanfield RL *et al.*, 2006).

Le catene pesanti delle Ig contengono un elemento strutturale che non è relazionato ai domini IgSF. Tale elemento viene chiamato “*hinge*” ed è localizzato tra il carbossi-terminale del CH1 e l'ammino-terminale del CH2. Nelle IgM che presentano quattro domini costanti la “*hinge*” è localizzata tra il CH2 ed il CH3. Le varie classi di Ig e le sottoclassi di IgG presentano “*hinge*” diverse in lunghezza e struttura. La “*hinge*” permette a ciascun Fab un movimento angolare relativo sia all'altro Fab che alla restante parte della molecola (Fc) (Fig. 12).

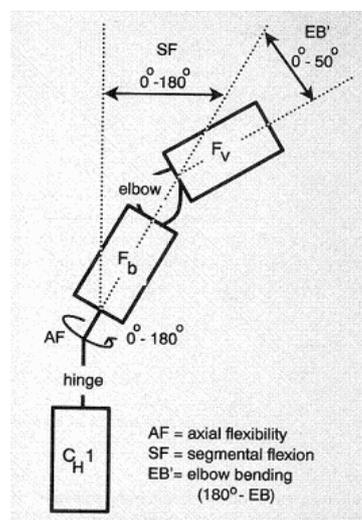


Figura 12. Flessibilità potenziale (SF) del frammento Fab, di cui sono indicati i domini Fv ed Fb, rispetto al frammento Fc, di cui è indicato solo il dominio CH1.

Ciò dà maggiore libertà ai due Fab di posizionarsi a distanze variabili tra di loro. L'ampiezza del movimento è determinato dal tipo

di “*hinge*”. Questa flessibilità ha una grande implicazione nella funzionalità degli anticorpi poiché permette ad un anticorpo bivalente di legare epitopi antigenici in una varietà di relative sistemazioni spaziali.

Sebbene non sia stato dimostrato, possiamo ipotizzare che la variazione della geometria Fab-Fc contribuisca ad ottimizzare l’attivazione di funzioni effettrici quando l’antigene è legato; ad esempio, la necessità di conservare il legame con l’antigene quando l’Ig interagisce con un recettore Fc sulla membrana di alcuni tipi cellulari, potrebbe dipendere dalla flessibilità della “*hinge*”.

Un altro ruolo chiave della “*hinge*” è la stabilizzazione dell’interazioni CH2-CH2 mediante ponti disolfurici promossi dai residui cisteina ivi presenti.

## **1.6. Le Ig di membrana**

Le Ig possono essere espresse anche con una forma diversa di catena pesante. Le Ig di membrana, diversamente da quelle secrete, contengono una regione idrofobica che attraversa la membrana ed una coda citoplasmatica (Fig. 13).

La regione che attraversa la membrana cellulare è una tipica  $\alpha$ -elica consistente di 23 residui amminoacidici. La porzione citoplasmatica generalmente consiste di tre residui, solo quella delle IgE è più lunga (30 residui).

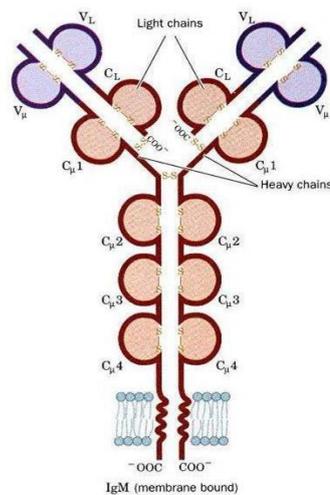


Figura 13. Rappresentazione schematica di un'immunoglobulina di membrana.

Essendo la porzione citoplasmatica generalmente corta, essa non è in grado di trasdurre il segnale del legame con l'antigene all'interno della cellula. La funzione di segnale viene assunta dai peptidi  $Ig\alpha$  ed  $Ig\beta$  che sono associati alla Ig nel contesto del BCR (Blum JH *et al.*, 1993). Tali peptidi contengono un dominio C1 nella regione extracellulare, un'elica transmembrana ed una lunga regione citoplasmatica contenente i motivi ITAM che fosforilandosi, innescano la cascata degli eventi del *signaling*. Ad ogni molecola di Ig è associato un singolo eterodimero  $Ig\alpha$ - $Ig\beta$ . Le interazioni che stabilizzano il complesso del BCR sono tuttora poco chiare: sono stati ipotizzati contatti extracellulari ed all'interno della membrana cellulare (Pleiman CM *et al.*, 1994).

### 1.7. *Splicing* alternativo per la sintesi della forma secreta o di membrana della catena pesante

I meccanismi dello *splicing* alternativo del trascritto primario (pre-mRNA) permettono di sintetizzare da un unico gene prodotti

multipli con funzionalità diversa; essi rappresentano un modo di regolare l'espressione di molti geni coinvolti nella risposta immunitaria (Lynch KW, 2004). Tra questi i geni del MHC (Ishitani A and Geraghty DE, 1992) e le forme di diversa lunghezza della catena  $\zeta$  del CD3 (Atkinson TP *et al.*, 2003).

La regione del locus IgH codificante la porzione costante contiene i quattro esoni costanti, seguiti al 3' da un segnale di poliadenilazione (s-pA) e due esoni (TM1 e TM2) codificanti la regione specifica della forma di membrana a circa 2 kb, seguiti da un secondo segnale di poliadenilazione (m-pA). L'ultima parte del dominio CH4 contiene un sito criptico di *splicing* che viene utilizzato nella sintesi della forma di membrana; la porzione al 5' di tale sito costituisce il dominio Ig mentre la porzione al 3' costituisce il cosiddetto frammento secretorio presente esclusivamente nella forma secreta ed importante nella polimerizzazione dell'IgM.

Sarà espressa la forma di membrana o quella solubile a seconda della diversa disponibilità dei segnali di poliadenilazione (Peterson ML, 2007). Nella cellula B il segnale s-pA è impegnato da una proteina regolatrice ed, in conseguenza, viene utilizzato il segnale m-pA con conseguente sintesi della forma di membrana. Quando la cellula si trasforma in plasmacellula secernente Ig, dal sito s-pA viene staccata la proteina regolatrice e viene sintetizzata la forma secreta (Fig. 14).

Nelle IgM dei mammiferi, gli esoni TM codificano 41 residui amminoacidici, dei quali 15 costituiscono il dominio extracellulare prossimale alla membrana (EMPD), 23 attraversano il doppio strato lipidico della membrana cellulare e 3 costituiscono la regione citoplasmatica.

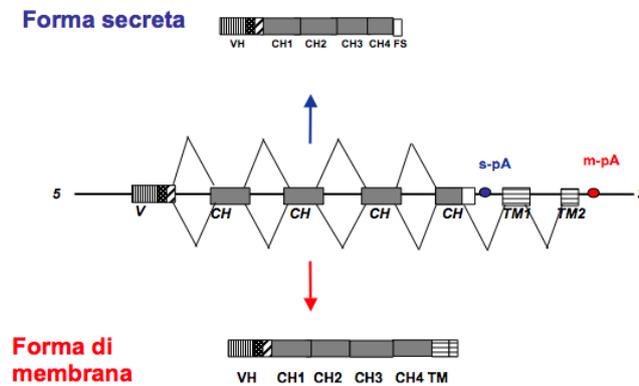


Figura 14. *Splicing* alternativo del trascritto primario dei mammiferi.

Predizioni di struttura secondaria suggeriscono che la regione che attraversa la membrana si strutturano ad  $\alpha$ -elica e che le EMPD delle due catene pesanti interagiscono tra loro (Major JG *et al.*, 1996). In particolare, studi sulle EMPD delle IgM umane hanno dimostrato che i residui di cisteina ivi presenti possono dar luogo ad un ponte disolfurico (Bestagno M *et al.*, 2001).

In genere l'EMPD è poco conservato evolutivamente in termini di sequenza e lunghezza, tuttavia, la maggior parte delle sequenze disponibili mostra alcuni elementi comuni come una sequenza prossimale alla membrana, ricca di residui carichi negativamente; inoltre, un alto contenuto di cisteine e proline è una caratteristica condivisa da altre regioni spaziatrici delle molecole immunitarie.

## 1.8. Trasduzione del segnale del BCR

Le Ig di membrana esplicano la loro funzionalità nell'ambito del BCR, complesso recettoriale dei linfociti B (Reth M, 1992). Il BCR è costituito da un'immunoglobulina con catena pesante di membrana (mIg) associata non covalentemente ad un eterodimero costituito dai due peptidi CD79A e CD79B, detti anche  $Ig\alpha$  ed  $Ig\beta$ , legati l'un l'altro da un ponte disolfurico. La stechiometria del complesso è stata recentemente dimostrata: ad un'Ig corrisponde un etero dimero.

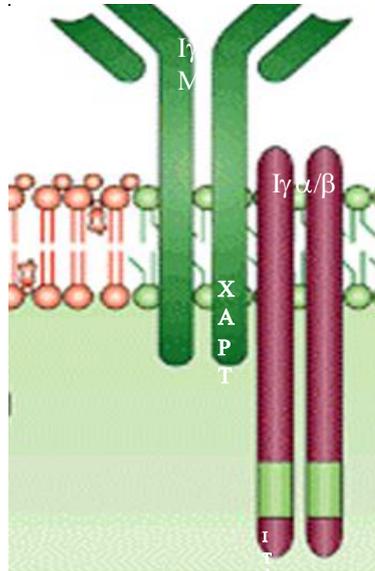


Figura 15. Rappresentazione schematica di un complesso BCR.

A seguito del legame con l'antigene, sono innescati due importanti meccanismi: la trasduzione del segnale, con una cascata di eventi biochimici, che induce processi di attivazione cellulare, e l'internalizzazione del complesso antigene-BCR. Nel complesso del BCR, il ruolo di riconoscimento e di legame con l'antigene è espletato dalla molecola mIg. Tuttavia, la coda citoplasmatica dell'mIg è troppo corta per interagire direttamente con l'apparato

della trasduzione del segnale. Tale funzione è, invece, a carico delle code citoplasmatiche dell'eterodimero  $Ig\alpha-Ig\beta$  che contengono i motivi ITAM dell'attivazione, mediata dalle interazioni con tirosin-chinasi specifiche. In assenza di antigene, cioè in condizione di riposo, i BCR appaiono dispersi sulla superficie delle cellule B. Essi, tuttavia, sembrano avere un'intrinseca propensione all'oligomerizzazione; si suppone, inoltre, che tale oligomerizzazione possa giocare un ruolo, ancora non ben definito, sull'inizio della trasduzione del segnale (Reth M, 2001).

In seguito al legame con l'antigene, i BCR formano *clusters* microscopici che reclutano la prima chinasi della trasduzione, Lyn, una chinasi della famiglia Src. Lyn fosforila le tirosine del motivo ITAM, contenuto nelle regioni citoplasmatiche dell'eterodimero  $Ig\alpha-Ig\beta$ , che a loro volta reclutano ai *cluster* di BCR altri componenti dei meccanismi di internalizzazione e trasduzione del segnale (Dal Porto JM *et al.*, 2004). Finora non è stato chiarito come l'informazione del legame con l'antigene nella regione extracellulare del BCR sia trasferita, attraverso la membrana cellulare, alla regione citoplasmatica di  $Ig\alpha-Ig\beta$ .

La funzione di riconoscimento dell'antigene da parte del BCR non si verifica solo con antigeni solubili, ma anche con antigeni sulla superficie cellulare delle cosiddette cellule presentanti l'antigene (APC). Il contatto tra le due cellule, APC e il linfocita B, rappresenta un momento fondamentale nell'attivazione. Batista *et al.* hanno evidenziato che sui linfociti B che vengono a contatto con l'antigene sulle APC, si osservano sinapsi immunologiche molto ordinate, con il caratteristico *pattern* di aggregazione dei BCR ad "occhio di bue" incorniciato da una corona di molecole di adesione quali LFA1 (Antigene associato a Funzione Leucocitaria 1) ed ICAM1 (Molecola di Adesione InterCellulare 1) (Batista F *et al.*, 2001).

Più recentemente, è stato osservato che i BCR che si aggregano al punto iniziale di contatto della APC con il linfocita B, inducono un'estesa adesione delle superfici cellulari. Successivamente, altri BCR incontrano l'insieme di antigeni e si spostano lungo le fibre di actina verso il centro dell'area di contatto delle cellule dove concentrandosi, formano la ben organizzata sinapsi immunitaria (Carrasco YR *et al.*, 2004; Fleire SJ *et al.*, 2006; Weber M *et al.*, 2008).

Studi più recenti hanno dimostrato che i microdomini di membrana biologica chiamati "*lipid raft*" sono coinvolti nell'attivazione del BCR ed agiscono da piattaforma per la trasduzione del segnale. I *lipid rafts* sono microdomini della membrana cellulare ricchi in colesterolo e di alcuni sfingolipidi, tra i quali la sfingomieline. Molti sfingolipidi presentano catene aciliche completamente sature che conferiscono al microdominio uno spessore maggiore ed un impacchettamento migliore rispetto al resto della membrana (McIntosh TJ *et al.*, 2003). Il colesterolo funziona da collante che tiene unito il microdominio. Viceversa i fosfolipidi, ricchi di catene insature, tendono ad impaccarsi meno efficientemente formando una fase disordinata, più fluida; ciò permette un rapido movimento laterale dei *lipid raft* con conseguente separazione di fase nella membrana. Tolar *et al.* hanno proposto che l'oligomerizzazione dei BCR porta ad una loro associazione con microdomini *raft* (Tolar P *et al.*, 2009). La condensazione dei *raft* intorno ai BCR potrebbe facilitare l'associazione con la chinasi Lyn o, alternativamente, l'ambiente lipidico dei *raft* potrebbe indurre una variazione conformazionale delle regioni citoplasmatiche del BCR che le renderebbe stericamente più disponibili all'interazione con Lyn (Fig. 16).

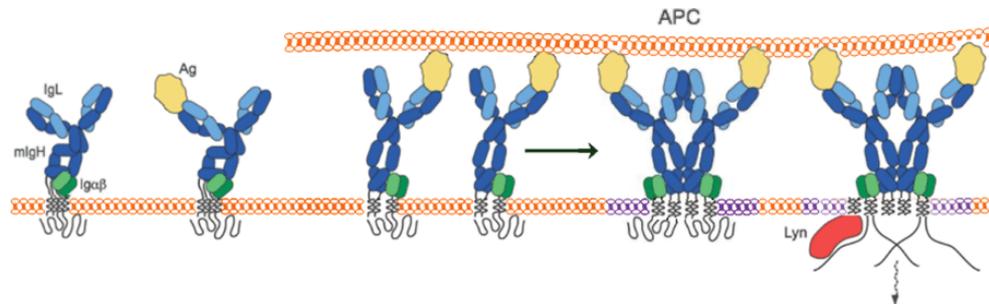


Figura 16. Modello di induzione della trasduzione del segnale dipendente dall'oligomerizzazione del BCR, conseguente al legame dell'antigene: l'interazione cellulare provocherebbe una variazione conformazionale degli oligomeri con conseguente apertura delle code citoplasmatiche dell'eterodimero, rese così disponibili all'interazione con Lyn.

## 1.9. Analisi strutturali delle Ig

Negli ultimi anni l'interesse per la molecola Ig è cresciuto enormemente e per analizzare il rapporto struttura-funzione a livello molecolare, diversi metodi di analisi strutturale sono stati proposti.

Centinaia di strutture cristallografiche comprendenti interi anticorpi monoclonali, anticorpi ingegnerizzati, frammenti Fab da soli o in complesso con l'antigene relativo, sono state risolte. Mentre le strutture dei singoli domini o frammenti sono facilmente risolvibili, la risoluzione dell'intera molecola comporta delle difficoltà a causa della sua elevata flessibilità.

Nel 2002 Saphire *et al.* (Saphire EO *et al.*, 2002) hanno presentato la prima struttura cristallografica di una IgG umana completa della "hinge": questa struttura evidenzia un'estrema asimmetria, indicativa della straordinaria flessibilità tra i domini (Fig. 17). Inoltre, la struttura evidenzia l'asimmetria nella composizione carboidratica di ciascun dominio CH2.

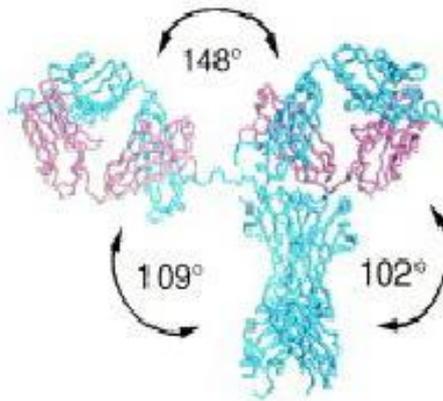


Figura 17. Struttura cristallografica di IgG (1HZH.pdb). Si può osservare il dominio Fc completo ed il possibile movimento relativo dei Fab.

In aggiunta all'approccio classico della cristallografia e dell'NMR, nuovi metodi sono stati sviluppati per studiare la struttura delle Ig. Uno di questi è la crio-elettrotomografia (cryo-ET), una tecnica di microscopia elettronica a temperature criogeniche che usa la tomografia per ottenere immagini tridimensionali. Tale tecnica è usata per studiare cellule e macromolecole di grandi dimensioni e, nel 2004, è stata usata per analizzare una IgG monoclonale (Sandin S *et al.*, 2004). I tomogrammi hanno rivelato molecole di IgG con la classica forma ad Y: due ellissoidi corrispondenti ai Fab ed una forma a cuore corrispondente allo stelo Fc (Fig. 18).

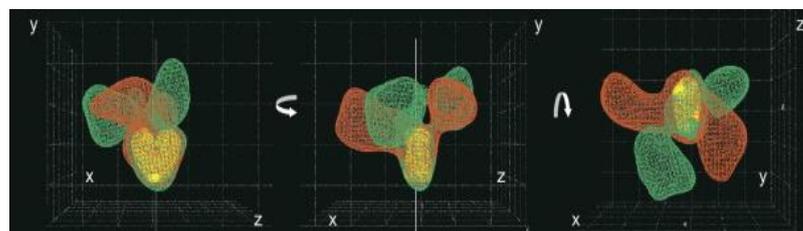


Figura 18. Confronto tra singole molecole di IgG in soluzione. I tomogrammi sono sovrapposti in base allo stelo Fc (in giallo). Le braccia dei Fab di due molecole sono riportate in arancione e verde. La visione in tre dimensioni è mostrata relativamente a rotazioni di 90° rispetto all'asse delle y e delle x.

Tale studio ha permesso di analizzare l'alta flessibilità della molecola poiché le posizioni relative dei Fab rispetto all'Fc differivano ampiamente tra molecola e molecola. La distribuzione delle strutture individuali ha fornito un'analisi quantitativa della dinamica del sistema.

Per ottenere nuovi dati riguardo alla struttura della molecola Ig in soluzione è stato proposto l'approccio della Cristalloidrodinamica. Dati sperimentali idrodinamici di soluzioni diluite di IgG sono applicati, tramite l'algoritmo SOLpro, ad ellipsoidi costruiti su dati cristallografici; la tecnica descrive i tempi medi di orientamento degli ellipsoidi costituenti i Fab e l'Fc in soluzione diluita (Carrasco B *et al.*, 2001; Lu Y *et al.*, 2007). Un'immagine dei risultati ottenuti è riportata nella figura 19.

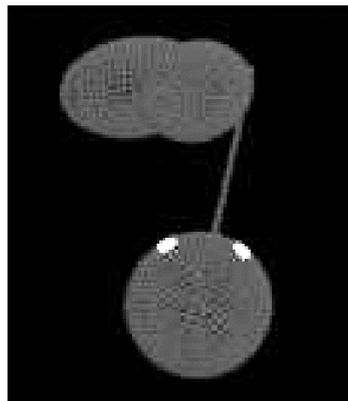


Figura 19. Modelli di “best fit” per un'IgG3 umana ottenuti mediante Cristalloidrodinamica. Il dominio Fc (ellissoide oblato) è in basso ed è congiunto, mediante la “hinge”, ai due domini Fab (ellissoidi prolata) parzialmente sovrapposti.

Anche metodi biochimici sono stati utili per ottenere informazioni strutturali: riduzione limitata con ditiotreitolo (DTT) dei ponti disolfurici delle IgG è stata usata da Hong *et al.* (Hong J *et al.*, 2009). Mediante tale tecnica si è visto che il ponte disolfurico tra la catena leggera e la pesante viene idrolizzato più facilmente di quello tra le due catene pesanti.

Più recentemente, metodi computazionali quale la Dinamica Molecolare (MD) hanno fornito un notevole contributo a questi studi. Questa si basa sulla risoluzione dell'equazione del moto di Newton. Integrando l'equazione del moto è possibile ottenere per ciascun atomo del sistema la sua posizione e velocità in funzione del tempo; l'insieme dei dati definisce una traiettoria del modello che identifica gli stati di minore energia potenziale e, quindi, di equilibrio più stabile. Mediante MD è stato possibile studiare sistemi biologici complessi di un numero elevatissimo di atomi. Nel 2005 una simulazione di MD è stata condotta su un sistema antigene-anticorpo di elevata affinità per analizzare la flessibilità del sito anticorpale, e per interpretare la dinamica della reazione con l'antigene. Essa risultava ascrivibile al meccanismo in due passi del tipo "adattamento indotto", cioè aggiustamento dell'interazione primaria piuttosto che al meccanismo in un unico passo del tipo "chiave-serratura". Inoltre, con questo metodo è stato possibile capire il significato funzionale della selezione dei residui amminoacidici carichi e di quelli polari nelle regioni più o meno rigide. (Sinha N and Smith-Gill SJ, 2005).

Poiché è molto complesso ottenere dati strutturali a livello atomico di un'intera Ig, Brandt JP *et al.* (Brandt JP *et al.*, 2010) hanno costruito un modello molecolare da strutture cristallografiche di frammenti di una IgG monoclonale (Trastuzumab); tale modello è stato sottoposto a simulazioni MD e sulla dimensione e forma della molecola intera sono stati raccolti dati idrodinamici sperimentali; per capire quale delle traiettorie ottenute in una serie di simulazioni fosse quella più vicina alla realtà è stata usata la tecnica del "*continuum hydrodynamic modeling*" che calcola le proprietà idrodinamiche sulla dimensione e topologia della sua superficie accessibile al solvente. Il paragone dei risultati ottenuti sperimentalmente con quelli calcolati,

ha determinato il prodotto di quale simulazione fosse quello più vicino alla realtà (Fig. 20).

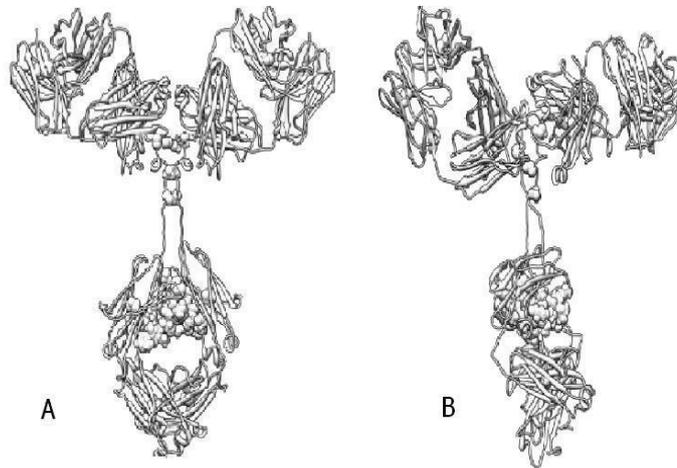


Figura 20. Anticorpo monoclonale IgG1 prima (A) e dopo una simulazione di 7.4 ns.

### 1.10. Le Ig dei teleostei

Le Ig dei teleostei sono più simili alle IgM dei mammiferi: esse consistono di due catene pesanti, ognuna di 75 kDa, e di due catene leggere, ciascuna di 22-25 kDa. Questi anticorpi non sono presenti solo nel sangue ma anche nelle secrezioni come muco cutaneo e bile.

Le Ig dei teleostei sono state isolate da molte specie e sono state ben caratterizzate: si tratta di molecole simili alle IgM, di circa 800 kDa (Wilson MR and Warr GW, 1992). Nei teleostei, la forma polimerica delle Ig predominanti nel plasma è quella tetrameric, costituita dall'assemblaggio di quattro unità di base, ma non è chiaro se esse siano tenute insieme da un peptide ortologo di J dei mammiferi; soltanto per alcune specie di teleostei è stato possibile dimostrare la presenza di un peptide simile a J (Warr GW, 1995). La

forma pentamerica dei condrostei, invece, presenta un peptide J per mantenere unite le cinque unità  $H_2L_2$ .

Sebbene la forma secreta delle IgM dei teleosti presenti alcune caratteristiche strutturali e funzionali simili a quelle delle IgM dei mammiferi, una differenza sostanziale tra le due consiste nel fatto che le IgM di alcuni teleostei si associano tra loro in modo diverso fino a formare una miscela eterogenea semplicemente variando il numero di ponti disolfurici fra i monomeri. Questa eterogeneità nella struttura di base è collegabile alle varie forme redox ed ha implicazioni significative nel processo di maturazione dell'affinità dell'anticorpo (Ye J *et al.*, 2010) e potrebbe rendere conto dell'assenza del peptide J.

Nel 1990 Wilson *et al.* hanno individuato un *pattern* di *splicing* alternativo insolito per le IgM di membrana del teleosteo *Ictalurus punctatus*. Anche in questa specie, entrambe le forme di catena pesante vengono sintetizzate dallo stesso gene ma, la forma di membrana risulta più corta di quella secreta. Successivamente, questo dato è stato confermato anche per altre specie di teleostei analizzate. In questi pesci è stata riscontrata l'assenza del sito criptico di *splicing* del CH4 che permette in tutte le altre specie di agganciare gli esoni TM al CH4. Nei teleostei, il sito donatore di *splicing* utile per agganciare gli esoni TM è quello al 3' dell'esone CH3 (Warr GW *et al.*, 1992). In conseguenza di ciò, la proteina sintetizzata risulta mancante del dominio CH4 e quindi più corta (Fig. 21).

Da un'analisi comparativa dei trascritti codificanti le catene pesanti delle IgM di teleostei (Artero S and Lefranc MP, 2000) risulta che i due esoni, TM1 e TM2, codificano 45-48 residui amminoacidici; i grafici di idropaticità indicano che 17-22 residui appartengono all'EMPD, i successivi 23 attraversano il doppio strato lipidico e gli

ultimi tre sono attribuibili alla coda citoplasmatica. Questi ultimi mediano l'interazione con i filamenti di actina del citoscheletro nei mammiferi (Park JI and Jongstra-Bilen J, 1997).

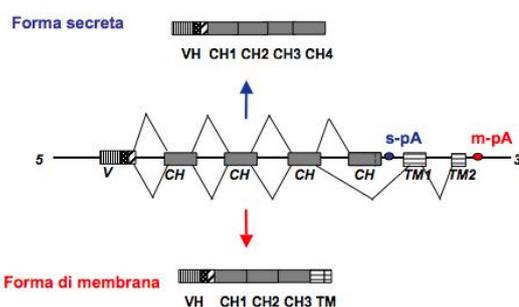


Figura 21. *Splicing* alternativo del trascritto primario IgH dei teleostei.

L'isotipo M non è, però, l'unico presente nei teleostei: infatti, recentemente è stato descritto un nuovo isotipo definito come IgD in base all'omologia di alcuni suoi domini costanti con quelli delle IgD dei mammiferi, in specie quali *Ictalurus punctatus* (Rycyzyn MA *et al.*, 1996), *Paralichthys olivaceus* (Hirono I *et al.*, 2003), *Gadus morhua* (Stenvik J *et al.*, 2001), *Salmo salar* (Hordvik I *et al.*, 2002). Inoltre è stato identificato anche un terzo isotipo IgH in *Danio rerio* e in *Oncorhynchus mykiss*: tale Ig è stata chiamata IgZ in *D. rerio* (Danilova N *et al.*, 2005) e IgT in *O. mykiss* (Hansen JD *et al.*, 2005).

Un'ulteriore differenza tra le Ig dei teleostei e quelle dei mammiferi riguarda la struttura genica codificante le catene leggere.

Infatti per tutti gli isotipi di catene leggere dei teleostei esistono molteplici *clusters* VJC come mostrato in figura 22.

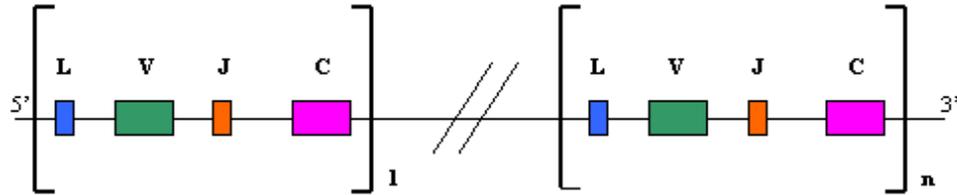


Figura 22. Organizzazione genica a *cluster* della catena leggera nei teleostei.

### 1.11. Le Ig dei teleostei antartici

I Nototenioidi sono un gruppo di teleostei che vivono nei mari che circondano il continente antartico (Fig. 23).



Figura 23. Il continente antartico.

Essi si sono evoluti a partire da circa 40 milioni di anni fa, da progenitori dell'ordine dei Perciformi che popolavano le acque costiere temperate dell'antico supercontinente Gondwana. La fauna ittica costiera originaria si diversificò e si adattò progressivamente

alle sempre più estreme condizioni ambientali durante la deriva del continente antartico verso il Polo Sud geografico.

Nell'evoluzione dei Nototenioidei ha avuto grande rilievo la formazione della Convergenza Antartica, un sistema di correnti marine circum-antartiche che costituisce una barriera termica tra le gelide acque vicine al continente e le masse oceaniche circostanti più temperate.

I pesci antartici hanno sviluppato meccanismi di adattamento fisiologico e biochimico che li rende unici al mondo. Essi dimostrano che, sotto la spinta di una forte pressione ambientale, gli organismi sono in grado di dare risposte strategiche variamente diversificate ed adeguate. Attualmente l'ittiofauna antartica è dominata da un singolo gruppo di teleostei, ordine Perciforme, sottordine Nototenioidei. Quest'ultimo comprende 120 delle 272 specie che popolano l'oceano meridionale finora conosciute; di queste, 174 si trovano nella regione antartica, delle quali 95 sono Nototenioidei. La dominanza di questo gruppo tassonomico è dimostrata dal fatto che il 45% delle specie ittiche che vivono nella regione polare appartengono ad esso (Eastman JT, 2005). Questi pesci, definiti altamente stenotermi, ovvero, in grado di vivere solo in un ristretto intervallo di temperatura, si sono adattati a vivere alla temperatura di  $-1,87\text{ }^{\circ}\text{C}$ , costante tutto l'anno. In particolare, il loro sistema enzimatico è specializzato per funzionare alle basse temperature ed inattivarsi al di sopra dello zero (Hazel JR, 1995).

Tra le forme di specializzazione fisiologica che li caratterizzano come organismi unici vi è senza dubbio la sintesi di particolari glicoproteine anticongelamento che impedisce il congelamento dei loro fluidi corporei a temperature più basse dello zero (Chen L *et al.*, 1997). Un altro elemento distintivo dell'ittiofauna degli ambienti polari è l'assenza di vescica natatoria, compensata dalla modificazione del

tessuto osseo divenuto simile a quello cartilagineo e dall'accumulo di depositi di grasso: ciò rende il loro peso nell'acqua quasi nullo, con una forte diminuzione del consumo di energia durante gli spostamenti. A favorire un minor consumo da parte dell'organismo compete anche il ridotto numero di globuli rossi con conseguente diminuzione della viscosità del sangue altrimenti troppo alta a temperature prossime allo zero.

Le Ig dei teleostei antartici sono studiate quasi esclusivamente dal gruppo di ricerca presso il quale è stato prodotto questo lavoro di tesi. In 15 specie sono state determinate le sequenze complete della regione costante della catena pesante delle Ig sia come forma secreta che di membrana; di altre 5 specie sono state determinate sequenze parziali; di 9 specie sono state ottenute sequenze genomiche; di due specie sono state ottenute sequenze di catene leggere.

I principali dati fin qui ottenuti sono stati recentemente raccolti in una review (Coscia MR *et al.*, 2010); ho partecipato personalmente alla stesura di questo lavoro, che è riportato integralmente in appendice. In esso sono parzialmente riportati anche dati riportati nel capitolo dei risultati di questa tesi. Di seguito riepilogo i risultati ottenuti precedentemente all'inizio della tesi.

Ig di *Trematomus bernacchii* (Fig. 24) sono state purificate dal siero e dalla bile (Pucci B *et al.*, 2003): la massa molecolare relativa della forma polimerica  $(H_2L_2)_4$  risulta di 830 kDa, quelle della catena H ed L rispettivamente 78 e 25 kDa. Il valore di massa molecolare relativa risulta leggermente più alto che nelle specie temperate dove risulta di 70 kDa (Wilson Mr and Warr GW, 1992); la differenza tra tali valori può essere attribuita ad un più alto contenuto di carboidrati delle specie polari. Infatti, il contenuto di carboidrati dell'intera molecola di Ig di *T. bernacchii* è stato determinato come 12.8 % della

massa totale; è da sottolineare che solo le catene pesanti sono state trovate glicosilate. Il pI della molecola intera varia da 4.4 a 6.0 a seconda della regione variabile presente, quello delle catene H isolate risulta tra 4.0 e 6.0. La sequenza amminoacidica della catena pesante di *T. bernacchii* contiene 7 siti putativi di N-glicosilazione; esso è il numero più alto tra le sequenze di tutte le specie di vertebrati. Sperimentalmente, mediante N-deglicosilazione limitata, sono stati evidenziati quattro catene carboidratiche di massa simile. L'elevato grado di glicosilazione può essere correlato all'alta solubilità di queste molecole anche a bassa temperatura.



Figura 24. Esemplare di *Trematomus bernacchii*.

La sequenza dell'intera catena pesante di *T. bernacchii* è stata determinata; le differenze più rilevanti rispetto a quelle dei teleostei temperati non risiedono nella composizione amminoacidica, né nelle modifiche post-traduzionali, ma piuttosto nella lunghezza delle regioni spaziatrici tra i domini, regioni che assicurano una maggiore flessibilità delle molecole adattate al freddo. Infatti, un'inserzione di due residui è stata ritrovata a livello della regione di connessione tra i domini VH e CH1, che coinvolge il "ripiegamento a gomito" e determina la mobilità del frammento Fv, che lega l'antigene ed il dimero CH1-CL. Una seconda inserzione è stata trovata a livello

della cosiddetta “*hinge*”, la regione che determina molti dei movimenti dell'intera molecola (Coscia MR *et al.*, 2000).

Inoltre, una libreria di cDNA da milza è stata costruita per studiare come si genera la diversità anticorpale in questa specie: 43 cloni di cDNA codificanti altrettante regioni variabili sono stati sequenziati ed, in base all'omologia di sequenza, sono state definite due famiglie geniche *VH*. L'89% delle sequenze risultavano appartenere ad una famiglia definita *Trbe VH II*, mentre le rimanenti sequenze venivano attribuite ad un'altra famiglia denominata *Trbe VH I*. Il numero di famiglie è il più basso di tutte le specie di vertebrati (Coscia MR *et al.*, 2002). Alcune sequenze *VH* sono state identificate anche nel DNA genomico e le sequenze contenenti i segnali della ricombinazione (RSS) sono stati trovati altamente omologhi a quelli di tutti i vertebrati (Coscia MR and Oreste U, 2009). Sebbene *T. bernacchii* possieda solo due famiglie geniche, importanti indicazioni sono state evidenziate dall'analisi della lunghezza e della composizione delle CDR, in particolare i dati di entropia di Shannon relativi alla CDR2 sono più elevati rispetto a quelli delle altre CDR. Inoltre, le inserzioni/delezioni contribuiscono alla diversità più che in ogni altra specie (Oreste U and Coscia MR, 2002).

Duplicazioni geniche del locus codificante la catena pesante è un'altra caratteristica dei teleostei antartici. Infatti, in quattro specie, *Trematomus scotti*, *Trematomus newnesi*, *Trematomus pennellii*, and *Trematomus lepidorhinus*, mediante FISH su cromosomi in metafase, sono stati evidenziati segnali prodotti da una sonda IgH di *T. bernacchii* su due coppie di cromosomi ed in *T. newnesi* tre varianti alleliche sono state sequenziate (Pisano E *et al.*, 2007).

Ulteriori contributi sono stati forniti dal gruppo sulla specificità antigenica delle Ig dei teleostei antartici (Coscia MR and Oreste U,

1998; Coscia MR *et al.*, 2000) e su meccanismi fisiologici (Abelli L *et al.*, 2005).

Il gruppo di ricerca, dopo aver ottenuto i risultati descritti sopra ha cercato di estendere i dati ottenuti ad altre specie di Nototenioidi sia antartici che periantartici, collezionando sequenze di cDNA di catene L ed H e sequenze genomiche di catena H.

## 1. 12. Contesto della tesi

Nel 2006, al fine di svolgere la mia tesi di laurea, sono entrata nel gruppo di ricerca che aveva prodotto i risultati sopra descritti, inserendomi nei progetti in corso e sviluppando un tema specifico quello dell'analisi della regione transmembrana della catena pesante delle Ig del teleosteo antartico *Chionodraco hamatus* (Fig. 25).



Figura 25. Esemplare di *Chionodraco hamatus*.

L'interesse derivava da fatto che il dimero di tale regione, interagendo con l'eterodimero  $Ig\alpha-Ig\beta$  è probabilmente coinvolto nella trasduzione, all'interno della membrana, del segnale dell'avvenuto legame con l'antigene.

Ho, quindi, completato la determinazione della sequenza della forma di membrana della catena H delle Ig di *C. hamatus* e sulla sequenza ottenuta ho condotto analisi di Modeling e MD. I risultati ottenuti sono stati successivamente alla mia laurea, riportati in una pubblicazione (Merlino A *et al.*, 2008), riportati per esteso al termine di questa tesi e qui sintetizzati.

Mediante metodi computazionali sono state individuate le regioni EMPD, transmembrana e citoplasmatica della sequenza aminoacidica codificata dagli esoni TM1 e TM2. E' stato costruito un modello molecolare delle regioni transmembrana e citoplasmatica mediante Homology Modeling usando come template la struttura dell'elica H del centro della reazione fotosintetica di *Rhodobacter sphaeroides*. La struttura ottenuta è stata validata con PROCHECK ed usata come modello di partenza per simulazioni di MD in acqua. I modelli ottenuti consistevano di  $\alpha$ -eliche estese e vicine all'idealità (Fig. 26).

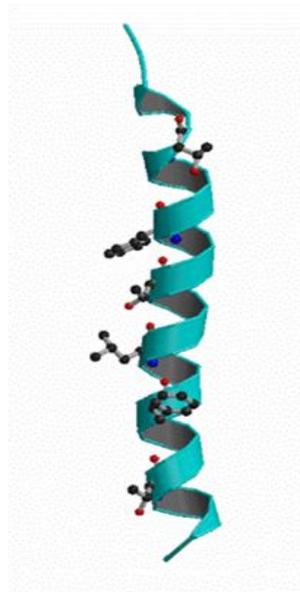


Figura 26. Rappresentazione del modello della regione transmembrana di *C. hamatus* ottenuto per omologia.

Successivamente, al fine di studiare la formazione spontanea dell'omodimero parallelo, un dimero con gli assi delle due eliche perfettamente allineati era generato ed usato per ottenere, mediante limitata rotazione di un'elica rispetto all'altra, altre tre conformazioni. Le quattro conformazioni sono state sottoposte a MD in acqua. Le strutture dimeriche ottenute mostravano impacchettamenti simili caratterizzati da un'elevata complementarità di superficie conseguenti ad interazioni idrofobiche tra due coppie di Fenilalanina e legami Idrogeno tra due coppie di Serina/Serina e Serina/Treonina (Fig. 27).

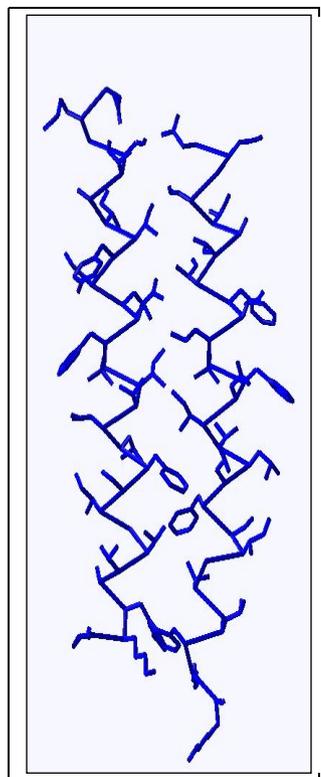


Figura 27: Rappresentazione di un dimero ottenuto mediante MD in ambiente acquoso.

Obiettivo primario del mio lavoro di tesi di dottorato è di estendere i dati finora ottenuti ad altre specie di Nototeniodi sia antartici che periantartici, collezionando sequenze di cDNA di catene

L ed H e sequenze genomiche di catena H. In particolare il tema centrale è stato quello di capire a livello chimico il meccanismo di *splicing* del trascritto primario codificante la forma di membrana della catena IgH di *C. hamatus*. Un altro obiettivo è stato quello di studiare la struttura del dimero transmembrana delle Ig di diverse specie in doppio strato lipidico, per verificare se il motivo responsabile nell'appaiamento delle eliche transmembrana nella sequenza dell'Ig di *C. hamatus* fosse presente anche in altre specie di vertebrati.

## **2. MATERIALI E METODI**

## 2.1. Metodi di biologia molecolare

### 2.1.1. Campioni biologici

I campioni biologici usati nella presente tesi sono stati ottenuti da teleostei pescati nei mari antartici da diversi ricercatori partecipanti a spedizioni scientifiche antartiche e subantartiche nell'ambito del Programma Nazionale di Ricerche in Antartide.

Le specie di teleostei utilizzate appartengono all'ordine *Perciformes*, sottordine Nototenioidi: *Bovichtus diacanthus* appartiene alla famiglia *Bovichtidae*; *Eleginops maclovinus* appartenente alla famiglia *Eleginopidae*; *Dissostichus mawsoni*, *Trematomus bernacchii*, *Gobionotothen gibberifrons* e *Notothenia coriiceps* appartenenti alla famiglia *Nototheniidae*; *Pogonophryne scotti* appartenente alla famiglia *Artedidraconidae*; *Gymnodraco acuticeps* e *Cygnodraco mawsoni* appartenenti alla famiglia *Bathydraconidae*; *Pagetopsis macropterus*, *Chaenocephalus aceratus* e *Champscephalus esox* appartenenti alla famiglia *Channichthyidae*.

Particolare attenzione è stata rivolta alla specie *Chionodraco hamatus* appartenente alla famiglia *Channichthyidae* considerata specie modello per lo studio dall'adattamento alle basse temperature.

## 2.1.2. Preparazione dell'RNA da milza

### *Estrazione e precipitazione dell'RNA*

L'RNA è stato estratto dalla milza dei pesci utilizzando il kit *SV Total RNA Isolation System* (PROMEGA). Sono stati utilizzati circa 160 mg di tessuto per ciascuna specie, tagliuzzati con un bisturi sterile e posti in un tubo da 2 ml; è stato aggiunto 1 ml di *SV RNA Lysis Buffer*, al quale era stato precedentemente aggiunto  $\beta$ -mercaptoetanolo. Il tessuto è stato omogeneizzato con un pestello fino ad ottenere una sospensione uniforme; da tale omogenato sono stati prelevati 175  $\mu$ l ai quali sono stati aggiunti 350  $\mu$ l di *SV Dilution Buffer*. L'omogenato è stato posto in un bagnetto termostatico a 70 °C per 3 min. Successivamente, la miscela è stata centrifugata per 10 min a 11.500 rpm. Il surnatante è stato trasferito in un tubo da 1,5 ml e ad esso sono stati aggiunti 200  $\mu$ l di etanolo al 96%. Il campione è stato trasferito in una *Spin Column Assembly* e centrifugato per 1 min a 11.500 rpm. Dopo aver allontanato il filtrato, sono stati aggiunti 600  $\mu$ l di *SV Wash Solution* alla *Spin Column Assembly*; successivamente, è stata eseguita una centrifugazione per 1 min a 11.500 rpm. Dopo aver nuovamente allontanato il filtrato, è stata aggiunta alla colonna una miscela costituita da 40  $\mu$ l di *Yellow core Buffer*, 5  $\mu$ l di  $MnCl_2$  0,09 M e 5  $\mu$ l dell'enzima Dnasi I. Il volume totale, così ottenuto, è stato attentamente applicato sulla superficie della membrana e la colonna è stata incubata in un bagnetto per 15 min a 24 °C. Terminata l'incubazione, sono stati aggiunti alla colonnina 200  $\mu$ l di *SV Dnase Stop Solution* alla colonnina che è stata poi centrifugata per 1 min a 11.500 rpm. Sono stati aggiunti altri 600  $\mu$ l di *SV Wash Solution* alla *Spin Column Assembly* ed è stata centrifugata nuovamente per 1 min a 11.500 rpm. Eliminato il filtrato,

sono stati aggiunti altri 250  $\mu$ l di *SV Wash Solution* e centrifugati al massimo della velocità per 2 min. La colonna è stata trasferita in un tubo sterile e ad essa sono stati aggiunti 100  $\mu$ l di acqua *Nuclease-Free*; è stata eseguita una centrifugazione per 1 min a 11.500 rpm e l'eluato contenente l'RNA è stato conservato a -80 °C.

#### *Elettroforesi dell'RNA su gel di agarosio e determinazione della densità ottica mediante lettura spettrofotometrica*

I campioni di RNA sono stati separati mediante una corsa elettroforetica a 100 V costanti su un gel di agarosio all'1%, precedentemente sciolto a caldo in una soluzione di Tris-acetato 0,04 M, pH 8, contenente EDTA 1 mM, DEPC 0,1% (v/v) e bromuro di etidio (0,5  $\mu$ g/ml). La densità ottica di ciascun campione è stata letta a 260 nm e 280 nm e la concentrazione di ciascun RNA è stata successivamente calcolata.

### **2.1.3. Amplificazione mediante PCR**

#### *Sintesi del cDNA*

Dalla preparazione di RNA totale sono stati prelevati 5  $\mu$ g ai quali sono stati aggiunti 5  $\mu$ l di oligo-dT+adattatore (50 pmol/ $\mu$ l), 5'-CGGAGATCTTCCAATGTGATGGGAAATTCT-3'.

Il campione è stato incubato a 70 °C per 10 min e immediatamente posto in ghiaccio per 5 min. Sono stati poi aggiunti 2  $\mu$ l di *RT buffer* 10X, 3  $\mu$ l di *dNTP mix* 10 mM (dATP, dCTP, dGTP, dTTP), 0,5  $\mu$ l di *Ribonuclease inhibitor* (40 U/ $\mu$ l) e 2  $\mu$ l *M-MULV RT* (100 U/ $\mu$ l). La miscela è stata equilibrata a temperatura ambiente per

5 min e poi incubata a 42 °C per un'ora. La trascrittasi inversa è stata inattivata ponendo il campione a 70 °C per 15 min.

### *Amplificazione del cDNA*

A 5 µl del cDNA sintetizzato sono stati aggiunti 27 µl di H<sub>2</sub>O distillata sterile, 5 µl di *PCR Buffer* 10X (100 mM Tris-HCl, pH 8,3, KCl 500 mM), 1,5 µl di MgCl<sub>2</sub> 50 mM, 1 µl di *dNTP mix* 10 mM, 5 µl del *primer* antisenso, (10 µM), 5µl del *primer* senso (10 µM). È stato scelto come *primer* oligonucleotidico antisenso *PK72*, complementare alla sequenza dell'adattatore utilizzato per la sintesi del cDNA. Come *primer* senso è stato scelto l'oligonucleotide opportuno. Le sequenze oligonucleotidiche dei *primers* e le rispettive temperature di *annealing* sono riportate nella tabella I.

Tabella I. Oligonucleotidi di innesco utilizzati per l'amplificazione del cDNA mediante PCR.

Nome	Sequenza	T <i>annealing</i> (°C)
<b>FWL2</b>	5' -TATGGTTCTGGTTTCTCATCTC- 3'	62,0
<b>R2</b>	3' -GTGTAAGTCTGAAGAACTACG-5'	60,0
<b>FR4</b>	5' -CAATGGTCACCGTCAGC- 3'	68,0
<b>PK72</b>	5'-CGGAGATCTTCCAATGTGATGGGAATTC-3'	65,1

Le reazioni di amplificazione dei frammenti di interesse sono state eseguite utilizzando 0,5 µl di *EuroTaq* DNA polimerasi (5 U/µl, AMBION). La miscela è stata sottoposta al seguente ciclo:

1° step: 94 °C per 1 min

2° step ripetuto per 30 volte:

94 °C per 1 min

T *annealing* per 1 min

72 °C per 1 min

3° step: 72 °C per 15 min

Il prodotto di PCR è stato poi sottoposto ad una seconda amplificazione, svolta nelle stesse condizioni della PCR primaria, allo scopo di aumentare la quantità del frammento di nostro interesse.

#### **2.1.4. Estrazione di DNA genomico da tessuto**

Per ciascuna specie, il DNA genomico è stato estratto dai testicoli. Il tessuto prelevato è stato tagliato a pezzi con un bisturi sterile ed omogenato in ghiaccio con un tampone di estrazione (10 mM Tris, 0,1 M EDTA e 0,5% SDS) in un rapporto 0,1 g di tessuto:1 ml di tampone. L'omogenato è stato trasferito in un tubo di polipropilene ed incubato a 37 °C per un'ora dopo aver aggiunto RNasi A (Fermentas) (2µl/ml di omogenato). È stata aggiunta Proteinasi K (SIGMA) (10µl/ml) e la miscela è stata incubata a 50 °C per 2 ore. Al volume di omogenato è stato aggiunto un egual volume di fenolo/cloroformio/alcool isoamilico (25:24:1); il tubo è stato agitato dolcemente per inversione per 5 min e centrifugato a 6.000 rpm per 10 min. La fase superiore è stata prelevata, evitando il contatto con l'interfaccia delle due fasi al fine di evitare eventuali contaminazioni. Al volume prelevato sono stati aggiunti 2 volumi di etanolo al 96% ed 1/10 di volume di Acetato di sodio 3 M. È risultato visibile il caratteristico batuffolo di DNA che è stato prelevato con un'ansa di platino e posto in tubo sterile da 1,5 ml. Il DNA è stato lavato con 5 ml di etanolo al 70% per 5 min e centrifugato a 6.000 rpm per 10 min. Il surnatante è stato allontanato e si è ripetuto il lavaggio. Dopo aver

lasciato asciugare all'aria, il *pellet* di DNA è stato risospeso in 2 ml di acqua sterile per 16 ore in agitazione a temperatura ambiente.

La concentrazione del campione di DNA estratto è stata calcolata misurando la densità ottica a 260 nm, mentre un'eventuale contaminazione da proteine è stata stimata calcolando il rapporto  $A_{260}/A_{280}$ .

### **2.1.5. Amplificazione di frammenti genomici mediante PCR**

Per poter amplificare mediante PCR frammenti molto grandi (dell'ordine di 2-10 kb) è stato utilizzato il kit *Expand Long Template PCR System* (Roche). Il sistema utilizza una miscela di enzimi che contengono una Taq DNA polimerasi termostabile ed una Tgo DNA polimerasi: una DNA polimerasi termostabile con un'attività di riparo 3'-5' isolata da *Thermococcus gorgonarius*. Questa miscela di enzimi è stata progettata per avere un'alta resa di prodotti ottenuti per PCR da DNA genomico.

La miscela di reazione è composta da 200 ng di DNA genomico, 1,75 µl di dATP (10 mM), 1,75 µl di dCTP (10 mM), 1,75 µl di dGTP (10 mM), 1,75 µl di dTTP (10 mM), 2,5 µl di *primer reverse*, 2,5 µl di *primer forward*, 5 µl di Buffer 1 (ROCHE), 0,75 µl di enzima e portati a volume a 50 µl con acqua sterile. Il programma impostato al termociclatore Hybaid è stato il seguente:

1° step: 92 °C per 2 min

2° step ripetuto per 10 volte:

92 °C per 10 s

T annealing per 40 s

68 °C per 4 min  
 3° step ripetuto per 20 volte  
 92 °C per 15 s  
 T annealing per 30 s  
 68 °C per 4 min + 20 s per ogni ciclo  
 4° step  
 68 °C per 7 min.

I *primers forward* e *reverse* usati per l'amplificazione dei frammenti genomici di *C. hamatus* e *B. diacanthus* sono stati rispettivamente: *Hinge3* (complementare ad una sequenza nell'esone CH2) e *TManti2h* (complementare alla sequenza di parte dell'esone TM). La temperatura di *annealing* scelta per l'amplificazione era 60 °C.

Per la sintesi del frammento di DNA genomico comprendente l'esone CH2 e l'esone CH4 i *primers* scelti sono stati disegnati all'interno dell'esone CH2 (*sshamfw*) e all'inizio dell'introne CH4 (*sshamrev*) mentre la temperatura di *annealing* scelta è di 51,0°C.

Le sequenze di tutti i *primers* utilizzati sono riportate in tabella II.

Tabella II. Oligonucleotidi di innesco utilizzati per l'amplificazione di frammenti di DNA genomico mediante PCR.

Nome	Sequenza	T <sub>m</sub> (°C)
<b>Hinge3</b>	5'-TGTCTGTTTGAGGGGAAAGG-3'	57,3
<b>TManti2h</b>	5'-AACAGCAGAGTGATGAGGAAGAG3'	68,0
<b>sshamfw</b>	5'-TTGCAGCCATCTCTGTTT-3'	51,0
<b>sshamrev</b>	5'-ATTAGGGGAGCACTGTGAC-3'	57,0
<b>antiCH4endTb</b>	5'-CGGGGATGTTTCATGTTGAG-3'	56,7

### **2.1.6. Elettroforesi del DNA su gel d'agarosio**

I prodotti di PCR sono stati separati mediante una corsa elettroforetica su gel d'agarosio all'1% in Tris-acetato 0,04 M, pH 8, contenente EDTA 1 mM e bromuro di etidio (0,5 µg/ml). L'elettroforesi è stata effettuata a 100 V e le bande sono state poi osservate ponendo il gel su un transilluminatore UV. La dimensione di ciascun frammento è stata determinata riferendosi al marcatore *GeneRuler™ 100bp DNA Ladder Plus* (MBI FERMENTAS).

### **2.1.7. Estrazione del DNA da gel di agarosio**

La corsa elettroforetica è stata effettuata a 100 V su gel di agarosio all'1% in Tris-acetato 0,04 M, pH 8, contenente EDTA 1 mM e bromuro di etidio (0,5 µg/ml). Una volta individuate le bande di interesse, il gel è stato tagliato in corrispondenza di queste ed i frammenti sono stati eluiti dal gel utilizzando il kit *Gel Extraction* (QIAGEN).

### **2.1.8. Clonaggio e trasformazione**

La reazione di clonaggio è stata effettuata utilizzando il kit *StrataClone PCR Cloning* (STRATAGENE). Il vettore *pSC-A* consente di clonare facilmente i prodotti di PCR in quanto presenta all'estremità 3' una timina in grado di appaiarsi con l'adenina aggiunta dalla Taq polimerasi alle estremità 3' dei prodotti amplificati. Tale vettore possiede, accanto al sito di clonaggio, i siti di restrizione per l'enzima *EcoRI* permettendo la verifica della presenza e delle

dimensioni dell'inserito mediante una semplice digestione. Le cellule utilizzate per la trasformazione sono cellule competenti fornite dal kit commerciale. A due microlitri di ciascun frammento di DNA eluito dal gel (corrispondenti a circa 25 ng) sono stati aggiunti 1  $\mu$ l di *Strataclone Vector Mix* e 3  $\mu$ l di *StrataClone Cloning Buffer* e incubati a temperatura ambiente per 5 min.

Ad un'aliquota di 50  $\mu$ l di cellule competenti, è stato aggiunto 1  $\mu$ l della miscela di ligasi ed il tutto è stato tenuto in ghiaccio per 20 min. Successivamente, la reazione di clonaggio è stata sottoposta ad una trasformazione mediante shock termico, a 42 °C per 90 s. Dopo un'incubazione in ghiaccio per 2 min, al campione sono stati aggiunti 250  $\mu$ l di SOC medium (LB medium, Bacto triptone 2%, estratto di lievito 0,5%, NaCl 10 mM, KCl 2,5 mM, MgCl<sub>2</sub> 10 mM, MgSO<sub>4</sub> 10 mM, glucosio 20 mM) e la miscela è stata sottoposta ad agitazione a 37 °C per circa un'ora. Le cellule sono state piastrate nei volumi di 80 e 150  $\mu$ l su terreno solido LB Agar (NaCl 6 M, Bacto triptone 1%, estratto di lievito 0,5%, agar 2%, pH 7), contenente ampicillina (100  $\mu$ g/ml) e X-gal (20 mg/ml) ed incubate per la notte a 37 °C. E' stato così possibile distinguere le colonie bianche, contenenti l'inserito, da quelle blu, che ne erano prive.

### **2.1.9. Preparazione in piccola scala del DNA plasmidico**

Ogni singola colonia batterica, contenente il plasmide ricombinante, è stata incubata in 5 ml di *LB medium* (NaCl 17 mM, triptone 1%, estratto di lievito 0,5%, pH 7) contenente ampicillina (50  $\mu$ g/ml), per 16 ore a 37 °C. Dopo aver centrifugato 4 ml di ciascun

inoculo a 12.500 rpm per 5 min, il surnatante è stato aspirato e il precipitato è stato risospeso in 250 µl di *Buffer P1* del Kit *QIAprep Spin Miniprep* (QIAGEN). Sono stati aggiunti quindi altri 250 µl di *Buffer P2* e la soluzione è stata mescolata per inversione 4-6 volte. Successivamente, sono stati aggiunti 350 µl di *Buffer N3*, mescolati per inversione finché la soluzione non diventava opaca e centrifugati per 10 min a 13.000 rpm. Il surnatante è stato trasferito in una *QIAprep spin column* e ulteriormente centrifugato per 1 min a 13.000 rpm. Dopo aver allontanato l'eluato, sono stati aggiunti 500 µl di *Buffer PB* e la soluzione è stata centrifugata per 1 min a 13.000 rpm. Allontanato il filtrato, sono stati aggiunti 750 µl di *Buffer PE* e centrifugati nuovamente a 13.000 rpm. Dopo aver allontanato l'eluato, è stata effettuata un'ulteriore centrifugazione per 1 min alla massima velocità per eliminare le tracce di etanolo. La colonna è stata trasferita in un tubo da 1,5 ml e, quindi, dopo aver applicato 50 µl di *Buffer EB* (Tris-HCl 10mM, pH 8,5), la colonna è stata nuovamente centrifugata per 1 min a 13.000 rpm.

L'eluato contenente il DNA plasmidico è stato infine conservato a -20 °C.

#### **2.1.10. Verifica della lunghezza dei frammenti clonati e determinazione della sequenza nucleotidica**

Circa 100 ng di ciascun DNA plasmidico sono stati digeriti con 1 µl di *EcoRI* (BIOLabs, 10 U/µl) nella soluzione appropriata. Le miscele sono state incubate a 37 °C per un'ora ed i prodotti della digestione sono stati separati mediante corsa elettroforetica su gel di agarosio 1% in Tris-acetato 0,04 M, pH 8, EDTA 1 mM e bromuro di

etidio (0,5 µg/ml). L'elettroforesi è stata effettuata a 100 V e le bande sono state visualizzate ponendo il gel su un transilluminatore UV ed utilizzando come riferimento il marcatore *Gene Ruler™ DNA Ladder Mix* (FERMENTAS).

La sequenza nucleotidica è stata determinata con un sequenziatore automatico ABI PRISM 3100 (PRIMM, Napoli) utilizzando il sito di riconoscimento per il *primer* oligonucleotidico *M13 Forward* localizzato nel vettore.

### **2.1.11. Southern blotting**

#### *Sintesi di una sonda marcata*

Una sonda marcata con digossigenina, complementare alla regione comprendente gli esoni CH2, CH3 e CH4 del cDNA di *C. hamatus*, è stata sintetizzata usando il *PCR DIG Probe Synthesis* kit (ROCHE). Il frammento di interesse è stato amplificato mediante PCR utilizzando una miscela di nucleotidi contenente dTTP marcato con digossigenina. Il *primer forward* usato è stato disegnato all'interno dell'esone CH2 (*Hinge3*), quello *reverse* è stato disegnato alla fine dell'esone CH4 (*antiCH4endTb*); le sequenze e le temperature di *annealing* relative sono riportate in tabella II, la temperatura di *annealing* della reazione di PCR è stata di 57 °C. Il frammento ottenuto è stato clonato all'interno del vettore plasmidico *pSC-A STRATACLONE* (STRATAGENE) e sequenziato. Successivamente, il plasmide contenente la sonda è stato digerito con *EcoRI* e la sonda estratta dal gel d'agarosio.

### *Southern Blotting*

I campioni di DNA genomico sono stati caricati su gel d'agarosio allo 0,8% e la corsa elettroforetica è stata condotta a 50 V. Il gel è stato recuperato e trattato per 30 min con un tampone di denaturazione (1,5 M NaCl, 0,5M NaOH) in leggera agitazione. Dopo un lavaggio con acqua, il gel è stato trattato con un tampone di neutralizzazione (1,5 M NaCl, 0,5 M Tris a pH 7,5) per 30 min in agitazione. Successivamente, si è proceduto al trasferimento per capillarità su membrana di nylon (Amersham Hybond-N): è stato costruito un ponte di carta (Whatmann 3MM) le cui estremità pescavano in una soluzione di SSC 20X (0,3 M trisodio citrato, 3 M NaCl, pH 7,5) e su di esso è stato posto il gel d'agarosio. Il gel è stato poi coperto dalla membrana di nylon e da una pila di circa 5 cm di fogli di carta Whatmann 3 MM sulla cui sommità è stato posto un peso di circa 500 g. Il trasferimento è proseguito tutta la notte. Il giorno dopo, il filtro è stato lasciato asciugare all'aria per circa 30 min e posto in un incubatore a 80 °C per 2 ore. Una volta ritirato, si è proceduto al saggio seguendo il protocollo del *DIG High Prime DNA Labeling and Detection Starter kit II* (ROCHE).

E' seguita l'esposizione del filtro ad una lastra autoradiografica (Fuji X-Ray) per 1 min.

### **2.1.12. Trascrizione *in vitro***

La trascrizione *in vitro* è stata effettuata utilizzando il kit *RiboMAX™ Large Scale RNA Production Systems-SPA and T7* (Promega) che consente la sintesi in grande scala di RNA. Il template di DNA deve essere inserito all'interno di un plasmide

contenente il promotore dell'RNA polimerasi T7 ed essere linearizzato con un enzima che produca estremità *blunt*. È necessario inoltre che il filamento antisenso del frammento di DNA di interesse sia sotto il controllo del promotore T7.

La miscela di reazione conteneva 4  $\mu$ l di *T7 Transcription 5x Buffer* (Promega), 30  $\mu$ l di rNTP (rATP, rCTP, rGTP, rUTP) 25 mM, 5-10 g di DNA linearizzato e 10  $\mu$ l di *Enzime Mix T7* (Promega). Il volume finale è stato portato a 50  $\mu$ l con acqua DEPC. La reazione è stata incubata a 37 °C per 4 h.

È stata aggiunta *RQ1 RNase-Free DNase* (Promega) ad una concentrazione di 1U/ $\mu$ g di DNA template e la miscela è stata incubata a 37 °C per 15 min. L'RNA è stato estratto con un volume di fenolo saturato con citrato (pH 4.7), cloroformio:alcol isoamilico (125:24:1), mescolato su vortex per 1 min e centrifugato a 13.500 rpm per 2 min. Alla fase acquosa, trasferita in tubo pulito, sono stati aggiunti 2,5 volumi di etanolo al 96% e 0,1 volumi di sodio acetato 3 M (pH 5,2), è poi seguita un'incubazione in ghiaccio per 5 min. La miscela è stata centrifugata a 13.500 rpm per 10 min, aspirato gentilmente il surnatante e lavato il *pellet* con etanolo al 70%. L'RNA è stato nuovamente centrifugato a 13.500 rpm per 10 min, il surnatante aspirato e il *pellet* è stato asciugato all'aria. Successivamente è stato risospeso in 80  $\mu$ l di acqua DEPC e la concentrazione dell'RNA sintetizzato è stata calcolata leggendone l'assorbanza a 260 nm.

### **2.1.13. Espressione di un peptide ricombinante**

*Amplificazione del frammento e clonaggio*

Un frammento della regione extracellulare della catena pesante della forma di membrana delle Ig di *C. hamatus* comprendente il CH2 e l'EMPD è stato espresso come peptide ricombinante. La regione nucleotidica codificante tale regione è stata amplificata per PCR utilizzando come stampo il cDNA di *C. hamatus* e come *primers* *CH2Rec* (complementare ad una sequenza al 5' del dominio CH2) e *VerdeRec* (complementare alla sequenza al 3' dell'EMPD). In entrambe, è stata sostituita una base presente nello stampo per poter inserire il codone di inizio ed un codone di stop.

Il sistema di espressione scelto è pETBlue (Novagen) ed il vettore di espressione usato è *pETBlue-1 Perfectly Blunt Vector* (Novagen). Il sistema comprende due ceppi diversi di cellule: le *NovaBlue Singles Competent Cells*, utili per un clonaggio iniziale e per la verifica del costrutto, e le *Tuner(DE3)pLacI Competent Cells* che contengono una copia cromosomiale del gene della T7 RNA polimerasi e sono state designate per l'espressione inducibile con IPTG del gene target sotto il controllo del promotore *T7lac* nel vettore pETBlue-1.

Il frammento da esprimere è stato prima convertito in una forma *blunt* e fosforilata attraverso una reazione di conversione delle estremità: a 0,8 µg di frammento sono stati aggiunti 5 µl di *End Conversion Mix* (Novagen) e portati ad un volume finale di 10 µl con acqua *Nuclease-free*. La miscela di reazione è stata incubata a 22 °C per 15 min e la reazione è stata disattivata riscaldando a 75 °C per 5 min. La miscela di reazione è stata raffreddata in ghiaccio per 2 min e ad essa sono stati aggiunti 1 µl (pari a 50 ng) di *Blunt Vector* e 1 µl (pari a 4 U) di T4 DNA Ligasi. La reazione di ligasi è stata mescolata gentilmente e incubata a 22 °C per 15 min; 1 µl di tale reazione è stato usato per trasformare un'aliquota di cellule *NovaBlue Singles Competent Cells*, le cellule sono state poste in ghiaccio per 5 min,

hanno subito uno shock termico per 30 s a 42 °C e in ghiaccio per 2 min. Sono stati aggiunti 250 µl di SOC a temperatura ambiente e 30 µl di tale miscela sono stati distribuiti su terreno solido LB agar contenente 50 µg/ml carbenicillina, 12,5 µg/ml tetraciclina, 70 µg/ml X-gal e 80 µM IPTG, e incubati per la notte a 37 °C. È stato preparato il DNA plasmidico in piccola scala e sequenziato utilizzando gli oligonucleotidi universali presenti nel vettore, come già descritto.

Un microlitro di DNA plasmidico è stato usato per trasformare un'aliquota di cellule (DE3)pLacI mediante un'incubazione in ghiaccio per 5 min, a 42 °C per 30 s e in ghiaccio per 2 min. Le cellule sono state risospese in SOC medium e incubate a 37 °C per 1 ora in agitazione. Cinquanta microlitri di cellule trasformate sono stati distribuiti su LB agar contenente 50 µg/ml di carbenicillina, 34 µg/ml di cloramfenicolo e 1% glucosio. Le piastre sono state poste a 37 °C per la notte.

#### *Crescita delle cellule ed induzione dell'espressione della proteina ricombinante*

Singole colonie sono state fatte crescere in 3 ml di LB contenente 50 µg/ml di carbenicillina, 34 µg/ml cloramfenicolo e 1% glucosio. Sono state incubate a 37 °C a 250 rpm fino a che la densità ottica a 600 nm raggiunga un valore di 0,8. Le colture sono state conservate per la notte a +4 °C. Le cellule sono state pellettate e risospese in 3 ml di terreno di coltura fresco e sono state usate per inoculare 100 ml di terreno fresco. La coltura è fatta crescere in agitazione a 250 rpm fino ad un valore di 0,8 di densità ottica a 600 nm. Cinque millilitri di tale coltura sono stati prelevati e usati come controllo negativo dell'induzione mentre al resto della coltura è stato

aggiunto IPTG ad una concentrazione finale di 0,1 mM per la notte a +4 °C. Cinquanta microlitri delle due colture sono stati prelevati e conservati per analisi successive.

### *Separazione delle frazioni cellulari*

Le colture cellulari (indotte e non indotte con IPTG) sono state centrifugate a 5.000 rpm per 15 min. Il surnatante è stato eliminato e le cellule sono state pesate e risospese in un tampone PBS ( $\text{Na}_2\text{PO}_4$  20 mM pH 7,4, contenente 150 mM NaCl, 1% Triton ) in un rapporto 1 g di cellule:3 ml di PBS. Cinquanta microlitri di cellule sono stati prelevati e conservati. La rottura delle cellule è stata effettuata per shock termico: le cellule sono state poste per 5 min in una miscela di ghiaccio secco e alcool e per 5 min a 37 °C, per 3 volte. È stato poi aggiunto 1  $\mu\text{l}$  di Benzonase (Novagen) per g di cellule e pochi granelli di Lisozima (Fluka). Le cellule sono state tenute in agitazione a 250 rpm per un'ora a 37 °C e centrifugate a 13.500 rpm per 30 min. Il surnatante è stato recuperato e conservato mentre il precipitato di cellule è stato risospeso in un volume di PBS uguale a quello in cui erano state inizialmente risospese le cellule.

## 2.2. Metodi di biochimica

### 2.2.1. Rivelazione del frammento ricombinante

#### *SDS-PAGE*

L'elettroforesi in condizioni riducenti è stata effettuata seguendo il protocollo di Laemmli (1970).

I campioni sono stati denaturati per 5 min a 100 °C in una soluzione per campioni contenente SDS 2% (w/v),  $\beta$ -mercaptoetanololo 5% (v/v), blu di bromofenolo 0,001%, glicerolo 10% (v/v). I gel utilizzati sono stati preparati al 12% e al 15% di poliacrilammide utilizzando una soluzione acrilammide/bisacrilammide (37,5:1), mentre il tampone da corsa utilizzato è stato Tris-glicina 0,025 M, pH 8,3, contenente SDS 0,1% (w/v). La corsa è stata condotta per 30 min a 24 mA costanti e successivamente a 34 mA. La massa delle proteine separate mediante elettroforesi è stata stimata utilizzando la miscela di marcatori di peso molecolare noto *Broad Range* (MBI Fermentas). Le proteine sono state rilevate mediante colorazione con *Coomassie Brilliant Blue R-250* (Bio-Rad) allo 0,1% (w/v) in metanolo 40% (v/v) e acido acetico 10% (v/v), immergendo il gel nel colorante per circa 30 min e decolorando poi, con una soluzione di metanolo 40% (v/v) e acido acetico 10% (v/v), fino al rilevamento ottimale delle bande.

#### *Immunoblotting*

Dopo separazione mediante SDS-PAGE al 15 %, le proteine sono state trasferite dal gel su filtro di nitrocellulosa da 0,2  $\mu$ m

(Schleicher & Schuell BA 83) per 17 h a 24 mA e 80 V costanti, usando come tampone di trasferimento una soluzione di 0,025 M Tris, 0,2 M glicina, metanolo 20% (v/v), pH 8,3. I siti aspecifici del filtro sono stati saturati con BSA 5% in TBS-T 0,1% (0,02 M Tris-HCl, pH 7,6, 0,14 M NaCl, Tween-20 0,1% (v/v)) per un'ora a temperatura ambiente e sotto agitazione; sono seguiti diversi lavaggi del filtro in TBS-T 0,1%. Il filtro è stato incubato a temperatura ambiente in agitazione per un'ora con l'anticorpo primario (anticorpi di coniglio specifici per le Ig di *C. hamatus*), diluito 1:5.000 in TBS-T 0,1%, e lavato con TBS-T 0,5%. Il filtro è stato incubato poi con l'anticorpo secondario (anticorpi anti-IgG di coniglio) coniugato con perossidasi, diluito 1:4.000 in TBS-T 0,1%, per un'ora a temperatura ambiente, in agitazione. Dopo lavaggi con TBS-T 0,5%, le proteine sono state rivelate attraverso una reazione di chemiluminescenza catalizzata dalla perossidasi coniugata all'anticorpo secondario, utilizzando l'*ECL-Kit* (Amersham); è seguita l'esposizione a tempi diversi a lastre Fuji X-Ray Film.

## **2.3. Metodi computazionali**

### **2.3.1. Analisi di sequenze**

Gli allineamenti di sequenza sono stati condotti utilizzando il programma *ClustalX 1.83* (Thompson JD *et al.*, 1997) e sono stati analizzati con il *tool* disponibile in linea *WebLogo* (<http://weblogo.berkeley.edu/>). La variabilità di ciascuna posizione nell'allineamento multiplo delle sequenze amminoacidiche è stato

misurato con l'analisi dell'entropia di Shannon (Stewart JJ *et al.*, 1997) (<http://hcv.lanl.gov/content/sequence/ENTROPY/entropy.html>).

Il programma NetPhos 2.0 ha permesso di predire i diversi siti di fosforilazione specifici per le chinasi (Blom N *et al.*, 1999).

I domini transmembrana sono stati individuati utilizzando il programma: *TMpred* (Hofmann K and Stoffel W, 1993), disponibile on-line: [http://www.ch.embenet.org/software/TMPRED\\_form.html](http://www.ch.embenet.org/software/TMPRED_form.html); *HMMTOP* versione 2.0 (Tusnàdy GE and Simon I, 2001), disponibile on-line: <http://www.enzim.hu/hmmtop/html/documenti.html>.

I siti di N-glicosilazione sono stati predetti utilizzando il Server NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc>).

Gli alberi filogenetici sono stati costruiti con il metodo del *Neighbor-Joining* (NJ)(Saitou N and Nei M, 1987) utilizzando gli allineamenti generati con *ClustalX* e la matrice di peso di BLOSUM.

L'identità di sequenza di proteine e nucleotidi è stata determinata rispettivamente con i *tools SIM* e *LALIGN*, entrambi disponibili sul sito <http://www.expasy.ch>.

La velocità di sostituzioni sinonime e non sinonime è stata calcolata nell'allineamento nucleotidico di sequenze di cDNA come descritto in Nei e Gojobori (1986)

Il peso molecolare dei frammenti di RNA è stato predetto utilizzando il *tool online* <http://www.encorbio.com/protocols/Nuc-MW.htm> mentre il peso molecolare delle proteine è stato predetto con Compute pI/Mw disponibile sul sito <http://www.expasy.ch>.

## 2.3.2. Costruzioni di modelli molecolari

### *Catene leggere di T. bernacchii*

Modelli molecolari dei domini costanti delle catene leggere sono stati costruiti utilizzando le sequenze codificate dai cloni 1B11, 5141, 219 e L440, ognuno rappresentativo di un isotipo di IgL, tecniche di *modelling* per omologia (Morea V *et al.*, 1998) ed il programma SwissPdbViewer3.7. In tabella III sono riportati i parametri e i dettagli delle strutture scelte come templati per ciascun isotipo di catena leggera.

Tabella III. Strutture utilizzate come stampo per la costruzione dei modelli molecolari degli isotipi di catena leggera di *T. bernacchii*.

Clone	Isotipo	Templato CL	Specie	PDB code	Resol.Å	Identità di sequenza
1B11	1A	Parte di un anti Testosterone Fab 77	Topo	1L7T	2.10	38%
219	1B	Parte di un anticorpo KOL	Uomo	2IG2	3.00	39%
L440	II	Parte di un Fab	Uomo	1W72	2.15	44%
5141	III	Parte di un Fab SE1554	Topo	1MFE	2.00	41%

### *IgTM di C. hamatus*

Un modello molecolare per la sequenza composta da 29 residui amminoacidici, che contiene la regione transmembrana delle IgM di *C. hamatus*, è stato costruito utilizzando il programma *Homology* (MSI/Corporation). È stato scelto come templato la struttura cristallografica dell'elica H della proteina di membrana del centro di reazione fotosintetica di *Rhodobacter sphaeroides*, risolta a 2,1 Å. Allineamenti multipli tra sequenze amminoacidiche di proteine di

membrana, effettuati mediante il programma *3D-PSSM* (Kelley LA *et al.*, 2000), mostravano, infatti, che questa proteina presentava i più alti valori di identità di sequenza (32%) con la sequenza di *C. hamatus* in esame. La struttura è stata validata con PROCHECK (Laskowski RA *et al.*, 1993).

Due copie della stessa elica sono state utilizzate come modello di partenza per analizzare la formazione di possibili omodimeri, mediante un approccio di MD. Sono stati ottenuti dieci modelli disponendo le eliche a distanza minima dei centri di massa di circa 14 Å e mantenendo gli assi delle eliche paralleli.

### **2.3.3. Dinamica Molecolare**

Sono state condotte simulazioni di MD in acqua e POPC (palmitoil-oleil-fosfatidilcolina) per valutare la stabilità dei modelli. Il doppio strato lipidico di POPC può essere considerato adatto a mimare la membrana cellulare perché contiene sia una catena satura (C16) che una insatura (C18) di acidi grassi. Per confrontare ambienti diversi, le simulazioni sono state condotte anche in acqua.

Per studiare la formazione spontanea di un omodimero parallelo di IgTM, quattro configurazioni iniziali, che differivano per il diverso orientamento relativo delle due eliche, sono state costruite utilizzando il programma O (Jones TA *et al.*, 1990) e studiate mediante MD sia in acqua che in ambiente lipidico. Il primo dimero è stato generato con gli assi delle eliche perfettamente allineati; successivamente le altre tre configurazioni sono state generate ruotando un'elica attorno al suo asse di pochi gradi e traslando la stessa di pochi Å. In tutti i casi, le distanze tra i centri di massa delle

due eliche variavano tra i 14 e 16 Å. La minima distanza tra gli assi delle due eliche è di 5.2 Å.

Tutte le simulazioni sono state condotte utilizzando il pacchetto di programmi GROMACS versione 3.2 (Van Der Spoel D *et al.*, 2005) e il campo di forza GROMOS96. I modelli di partenza sono stati immersi in diverse box contenenti 3.500-16.000 molecole di acqua. Lo stato di ionizzazione dei residui carichi è stato scelto per mimare un ambiente a pH neutro. La carica totale del sistema è stata neutralizzata aggiungendo il numero appropriato di ioni ad almeno 7 Å di distanza dalla superficie della proteina. Nel primo passaggio del processo di equilibratura, il solvente è stato rilassato da una minimizzazione energetica seguita da 20 ps di MD a 300 K. L'intero sistema è stato poi minimizzato senza vincoli restrizionali tra le eliche prima della simulazione. Le lunghezze di legame sono state vincolate dall'algoritmo LINCS (Berendsen HJC *et al.*, 1981). Le equazioni del moto di Newton sono state integrate con un *time step* di 2 fs e le coordinate atomiche sono state salvate ogni 0,5 ps. Il valore della costante dielettrica scelto è 1. Per il trattamento delle interazioni elettrostatiche è stato usato un valore di *cut-off* pari a 14 Å. Tutte le simulazioni sono state condotte in un *ensemble* isobarico-isotermico (NPT) a 300 K, usando le condizioni periodiche al contorno nelle direzioni delle tre coordinate. La pressione è stata controllata utilizzando l'algoritmo di Berendsen a 1 bar mentre la temperatura è stata mantenuta costante usando un algoritmo di Berendsen (Berendsen HJC *et al.*, 1984).

I modelli di partenza sono stati posti in una box contenente una miscela di molecole di POPC organizzate in doppio strato lipidico e acqua. Prima dell'inserzione del modello, la box conteneva 128 molecole di lipidi e 3.755 molecole di acqua; dopo l'inserzione della proteina al centro della box, sono state eliminate tutte le molecole di

acqua aventi un atomo di ossigeno ad una distanza minore di 4 Å da un atomo non idrogeno della proteina e delle molecole lipidiche. La rimozione di tali molecole introduceva delle cavità all'interno del sistema che sono state facilmente rimosse durante il primo passaggio di equilibrizzazione del processo di 10 ps, durante il quale la proteina era vincolata alle posizioni iniziali. Il sistema risultante comprendeva 110 molecole di lipidi e 3.730 molecole di acqua.

Nel caso delle simulazioni in POPC, è stato usato per il trattamento delle interazioni elettrostatiche per atomi a distanza maggiore di 9 Å il metodo di Part Mesh Ewald (PME) (Darden T *et al.*, 1993); il campo di forza e le coordinate per il doppio strato lipidico di POPC sono state scaricate dal sito web del gruppo del prof. Tieleman (<http://mouse.bio.ucalgary.ca>).

La stabilità strutturale globale delle simulazioni è stata confermata analizzando diversi parametri (struttura secondaria, raggio di girazione, numero totale di legami idrogeno) in funzione del tempo.

#### **2.3.4. Dataset**

##### *IgTMD dataset*

Il dataset usato in questo studio è stato costruito cercando nei database sia sequenze di cDNA che di DNA genomico codificanti per IgH legate alla membrana. Nel caso di sequenze di cDNA, le regioni specifiche per la forma di membrana sono state identificate dal confronto con la corrispondente sequenza codificante la forma secreta delle IgH, in questo modo sono state raccolte sequenze appartenenti a tutti gli isotipi di tutte le classi di vertebrati. Le 68

sequenze selezionate in questo studio sono elencate nella Tabella S2 allegata in appendice.

### *PDBTM dataset*

È stato generato un dataset non ridondante comprendente le catene di proteine di membrana la cui struttura è nota sperimentalmente e presenti nel Data Bank di proteine transmembrana (PDBTM <http://pdbtm.enzim.hu/>) aggiornato al 10 marzo 2010. il criterio scelto per la selezione delle proteine è il seguente: la presenza di almeno un'elica transmembrana; un'identità di sequenza inferiore al 25% con le altre proteine presenti nel dataset; esclusione di proteine le cui strutture sono state dedotte tramite il metodo di *Homology Modelling*. I frammenti di eliche sono stati identificati mediante DSSP (Dictionary of Protein Secondary Structure) (Kabsch W and Sander C, 1983). Sono state selezionate per l'analisi soltanto sequenze con un numero di residui maggiore di sette, per poter garantire almeno due giri in tutte le eliche. Il dataset finale conteneva 236 catene di proteine e 24.886 residui amminoacidici.

## **2.4. Metodi chimico-fisici**

### **2.4.1. Dicroismo circolare**

Gli spettri di dicroismo circolare (CD) per i campioni di RNA sono stati registrati utilizzando uno spettropolarimetro *Jasco-J715* precedentemente calibrato con una soluzione acquosa di Tris 10

mM, EDTA 1 mM in acqua DEPC. L'apparecchio è dotato di un sistema *Peltier PTC-348WI* per il controllo della temperatura. Per registrare gli spettri nella regione lontano UV sono state utilizzate cuvette di quarzo con cammino ottico di 0,1 cm; la costante di tempo, la larghezza di banda e la velocità di scansione impostate sono state pari rispettivamente a 4 s, 1 nm e 2 nm/min. Valutando la variazione del segnale dicroico alla lunghezza d'onda di 267 nm all'aumentare della temperatura, sono stati registrati spettri CD di denaturazione termica. Il segnale dicroico è stato convertito in ellitticità per mole di residuo,  $[\theta]$ , espressa in grado  $\text{cm}^2 \text{dmol}^{-1}$ , utilizzando la seguente equazione:

$$[\theta] = [\theta]_{\text{oss}} - \frac{mrw}{100l}$$

Dove  $[\theta]_{\text{oss}}$  è l'ellitticità registrata in gradi,  $mrw$  è il peso molecolare medio di nucleotide,  $C$  la concentrazione dell'RNA in g/mL e  $l$  è il cammino ottico della cuvetta in cm.

#### 2.4.2. Microscopia a forza atomica

Una soluzione 1 nM di RNA di *C. hamatus* (4 mM HEPES pH 7.4, 10 mM NaCl, 2 mM MgCl<sub>2</sub>) è stato trattato in due modi diversi prima di essere analizzato alla microscopia a forza atomica: un'aliquota è stata riscaldata a 90 °C per 5 min e posta in ghiaccio; un'altra aliquota è stata riscaldata a 90 °C per 5 min e poi lasciata raffreddare per tutta la notte.

L'RNA di *B. diacanthus* è stato risospeso nello stesso buffer ad una concentrazione finale di 1 nM e riscaldato a 90 °C per 5 min e poi posto in ghiaccio.

Venti microlitri dei campioni da analizzare sono stati depositati su un disco di mica e lasciati incubare per 1-5 min. Il disco di mica è

stato poi lavato con pochi millilitri di acqua ultrapura e la superficie è stata asciugata con un debole flusso di azoto. Tutte le immagini raccolte sono state ottenute all'aria, a temperatura ambiente, con un microscopio Nanoscope III (Digital Instruments, Santa Barbara, CA) equipaggiato con uno scanner di tipo E (12 mm x 12 mm) ed in modalità "tapping".

I "cantilevers" usati sono di tipo commerciale *diving board-type silicon* (NSCS12, SiliconMDT Moscow). Le immagini (512 x 512 pixels) sono state raccolte ad una scansione di 2-4  $\mu\text{m}$  ad una velocità di scansione variabile tra le due e le quattro linee di scansione al secondo.

## **3. RISULTATI**

Durante la mia tesi di dottorato mi sono interessata prevalentemente a tre aspetti riguardanti le immunoglobuline di teleostei antartici che sono stati brevemente riassunti di seguito:

### **A. Studi delle catene leggere delle Ig del teleosteo antartico *T. bernacchii***

Nel laboratorio presso il quale ho svolto la mia attività di ricerca, era già stato avviato uno studio relativo alla ricerca e all'identificazione degli isotipi di catene leggere della specie *T. bernacchii*. Erano a disposizione le sequenze nucleotidiche di due isotipi: L1 (a sua volta distinguibile in due sottoisotipi L1A e L1B) ed L3. La prima parte dell'attività del mio dottorato ha riguardato la ricerca dell'isotipo di catena leggera, L2, già descritto in altre specie di teleostei.

### **B. Caratterizzazione del dominio transmembrana delle Ig del teleosteo antartico *C. hamatus***

La sequenza del dominio che attraversa la membrana dell'Ig di *C. hamatus* è stata già determinata. Utilizzando questa sequenza ed un approccio computazionale che prevede l'utilizzo combinato di *homology modelling* e MD è stato anche proposto un modello della struttura tridimensionale di questa regione. Nel corso della tesi di dottorato, con lo stesso approccio computazionale, ho voluto ampliare tale studio in due direzioni: ho verificato la formazione di omodimeri della regione IgTM di *C. hamatus* in un ambiente più simile a quello fisiologico, cioè in un doppio strato lipidico, al fine di evidenziare i residui amminoacidici responsabili della dimerizzazione. Inoltre, ho esteso tale studio alle regioni transmembrana di immunoglobuline di tutte le classi di vertebrati e appartenenti a isotipi

diversi, allo scopo di identificare eventuali caratteristiche comuni a tutte le specie antartiche.

### **C. Studio dello *splicing* atipico responsabile della sintesi delle Ig di membrana di *C. hamatus***

La scoperta in *C. hamatus* di una forma di membrana delle Ig più corta rispetto a quella di altri teleostei ha aperto la strada ad uno studio delle possibili cause a livello genomico e a livello trascrizionale. È stato pertanto interessante studiare la regione genomica codificante le Ig di tale specie ed effettuare studi di struttura sul rispettivo trascritto primario.

Con lo stesso approccio è stato studiato il gene e l'RNA delle Ig di una specie, ritenuta ancestrale, non antartica, *B. diacanthus*, che presenta uno *splicing* per la forma di membrana uguale a quella di altri teleostei allo scopo di evidenziarne le differenze.

In appendice, in tabella S1, sono riportati i numeri di accesso delle sequenze nucleotiche da me ottenute nel corso del mio dottorato di ricerca.

### 3A. Identificazione dell'isotipo L2 di catena leggera in *Trematomus bernacchii*

#### 3A.1. Cloni di cDNA di isotipo L2 ottenuti per PCR

E' stato costruito, con il programma ClustalX 1.8, un allineamento di sequenze nucleotidiche dell'isotipo L2 di alcune specie di teleostei quali: *Danio rerio*, *Cyprinus carpio*, *Salmo salar* e *Oncorhynchus mykiss* ed è stato possibile disegnare due oligonucleotidi nelle regioni conservate. I prodotti amplificati per PCR su campioni di cDNA di *T. bernacchii*, effettuate utilizzando un *primer* senso *FR2* ed un *primer* antisenso *R2*, sono stati clonati e 6 cloni sequenziati. Tutte le sequenze sono state identificate come appartenenti all'isotipo L2 con un'identità nucleotidica media rispettivamente del 51.6, 48.6 e 52.2 % con gli isotipi L1A, L1B ed L3 , già identificati in *T. bernacchii*. Le sequenze sono state depositate in banca dati (Tabella IV).

Tabella IV. Le sequenze identificate come appartenenti all'isotipo di catena leggera L2 in *T. bernacchii* e relativi numeri di accesso in banca dati.

Sequenza	N° accesso
L440	EF421437
L450	EF421438
L47	EF421436
L470	EF114789
L480	EF114790
L420	EF114785

La suddivisione dell'isotipo L1 in due sottoisotipi L1A e L1B è stata effettuata in base all'omologia di sequenza relativa e all'assenza/presenza di possibili inserzioni di DNA microsatellite. La distinzione in due sottoisotipi potrebbe suggerire che questo evento, probabilmente derivato da una duplicazione del locus genico sia datato circa 90 MYA, che è la data stimata per la divergenza degli Acantopterigi.

### **3A.2. Analisi delle sequenze amminoacidiche dedotte**

Confrontando le sequenze dell'isotipo L2 con quelle degli altri isotipi già noti, L1A, L1B ed L3, è stato possibile fare un'analisi approfondita delle caratteristiche sequenziali specifiche per il dominio CL di ciascun isotipo nella specie in esame (Fig. 28). Le sequenze a disposizione erano in tutto 28: 15 sono di isotipo L1A, 3 di isotipo L1B, 6 sono attribuite all'isotipo L2 e 4 all'isotipo L3.

Le percentuali di omologia di sequenza amminoacidica tra gli isotipi sono riportate nella tabella V.

Le sequenze sono state allineate con il programma ClustalX 1.8 ed è stato possibile distinguere le regioni C e quelle V; tutti i cloni, ad eccezione del 5141 ed L180, codificano il dominio VL solo parzialmente. In figura 28 è riportato l'allineamento multiplo delle sequenze amminoacidiche dedotte dedotte dei cloni isolati mediante PCR e relativi ai tre isotipi di CL di *T. bernacchii*.

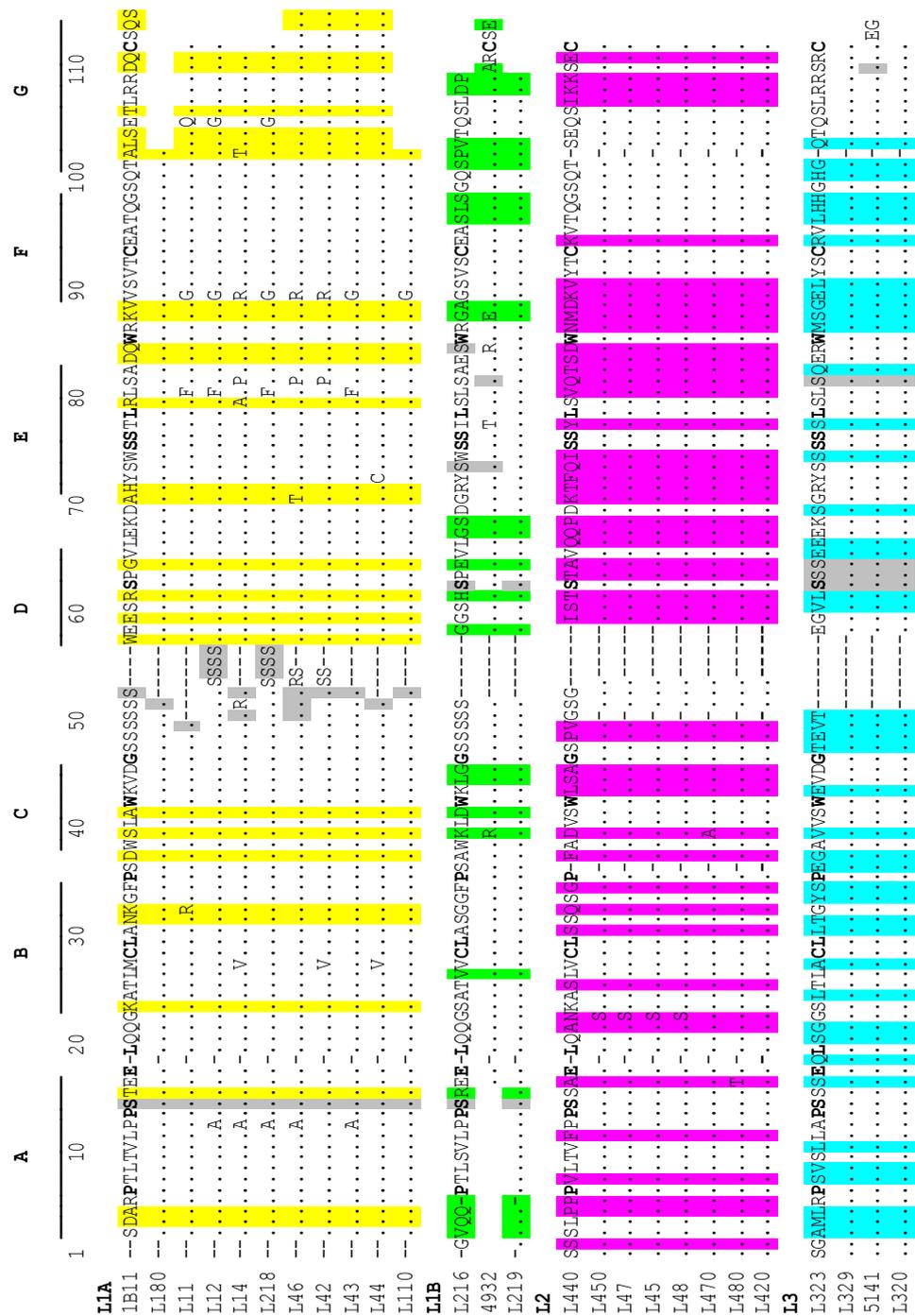


Figura 28. Allineamento multiple delle sequenze amminoacidiche dedotte dei cloni isolati mediante PCR e relativi ai tre isotipi di catene leggere di *T. bernacchii*. In grassetto sono state evidenziate le posizioni conservate mentre con colori diversi sono state evidenziate le posizioni specifiche per ciascun isotipo. In grigio sono stati evidenziate gli ipotetici siti di fosforilazione. Con le lettere A, B, C, D, E, F e G, sono indicati i foglietti. Il segno "=" indica un residuo identico a quello della prima linea di ciascun isotipo, mentre "-" indica una posizione mancante.

Tabella V. Percentuali medie di omologia di sequenza amminoacidica di ciascun isotipo di catena leggera con gli altri isotipi.

<b>L1A</b>	<b>L1B</b>	<b>L2</b>	<b>L3</b>	
91.7	57.9	37.2	34.0	<b>L1A</b>
	95.7	34.6	38.1	<b>L1B</b>
		94.7	33.4	<b>L2</b>
			95.2	<b>L3</b>

Sono state identificate 22 posizioni conservate nel dominio costante di tutti gli isotipi mentre nel dominio costante dell'isotipo L3 è stata osservata una inserzione all'estremità amminico terminale di uno o due residui aggiuntivi a quello degli isotipi L2 o L1, rispettivamente.

Le sequenze sono state analizzate con il programma NetPhos 2.0 ed è stato possibile individuare le ipotetiche posizioni di fosforilazione. Le serine che presentavano uno *score* superiore al 99 % sono state considerate siti di fosforilazione altamente probabili. Non è stato invece ottenuto uno *score* ugualmente elevato per alcun residuo di tirosina o treonina.

In particolare, una stringa di serine lungo dai 3 ai 10 residui è stato osservato sia per l'isotipo L1 che L2, dalla posizione 48 alla 57. Tutte le serine costituenti la stringa di serine dell'isotipo L1 potrebbero essere codificate dall'inserzione di un DNA microsatellite, che consiste in una ripetizione in tandem di triplette AGC. È stato notato infatti che questi codoni sono molto frequenti (40%) nelle ripetizioni di piccole sequenze (*short sequence repeats*, SSR) negli esoni di geni dei vertebrati (Li YC *et al.*, 2004). L'isotipo L3 mostra una stringa simile di cinque residui di serine dalla posizione 75 alla 79, tutte codificate dal codone AGC.

Non sono state osservate delle posizioni ipotetiche di N-glicosilazione ad eccezione del clone L320 che mostra un sito nella regione FR2 del dominio variabile.

In base all'allineamento, fatto con ClustalX 1.8, è stato costruito un albero filogenetico (Fig. 29) relativo al solo dominio CL che mostra l'esistenza di quattro cladi, ciascuno contenente le sequenze di un singolo isotipo.

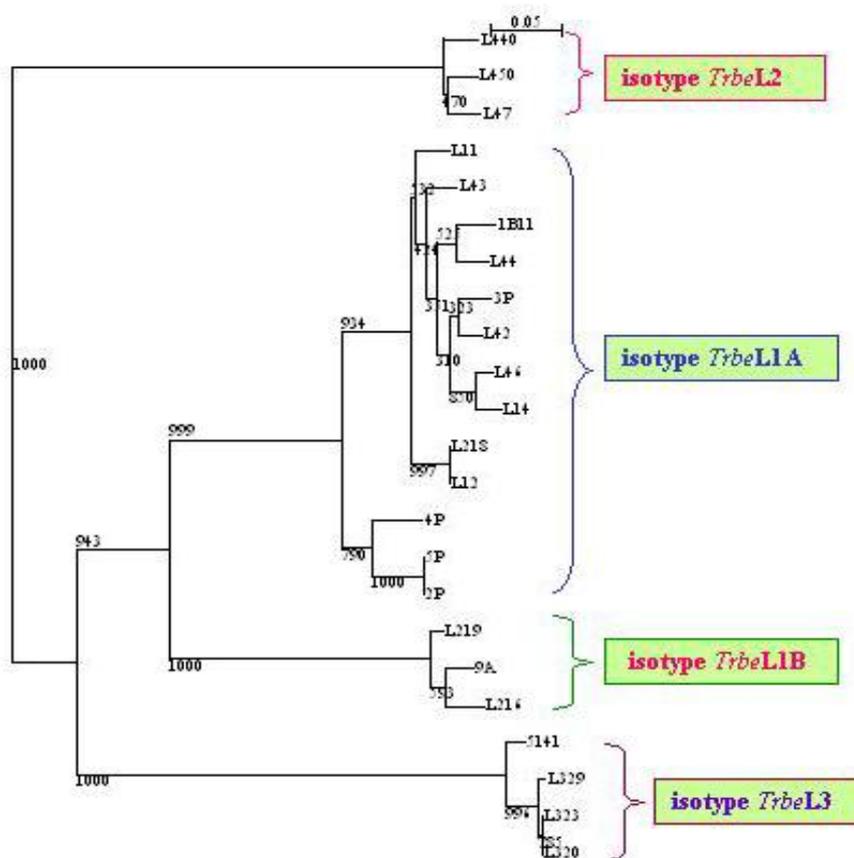


Figura 29. Albero filogenetico delle regioni costanti delle catene leggere della specie *T. bernacchii*.

I valori di *bootstrap* relativi alla divergenza di ciascun isotipo sono elevati e confermano un'elevata differenza tra le sequenze.

Per determinare il grado di variabilità amminoacidica per ciascuna posizione in tutte le sequenze ottenute, è stato utilizzato il metodo dell'entropia di Shannon (Fig. 30).

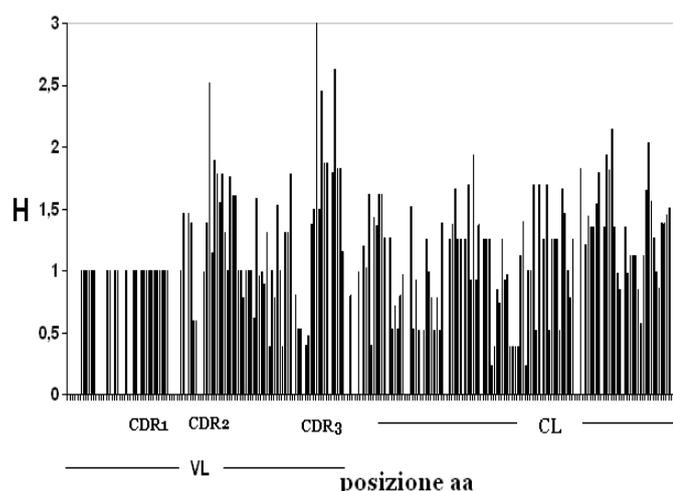


Figura 30. Istogramma dell'entropia di Shannon. In ascisse le posizioni sono distinte in base all'appartenenza al dominio variabile VL e costante CL. All'interno del dominio variabile sono indicate le CDR.

L'istogramma riportato in figura mostra sull'asse delle ordinate l'entropia (H), cioè il grado di variabilità amminoacidica, definito dall'algoritmo di Shannon, mentre sull'asse delle ascisse è riportata la posizione del residuo amminoacidico. Con questo metodo di analisi risulta che le regioni CDR2 e CDR3 sono le più variabili. L'istogramma presenta un picco con un valore di entropia pari a 2.5, corrispondente alla CDR2; valori alti di entropia si riscontrano anche in corrispondenza di due picchi relativi alle CDR3 che risultano essere pari a 3.0 e 2.62.

Appare evidente che la variabilità amminoacidica non è limitata alla regione variabile, ma anche alla regione costante delle catene L; tale variabilità è dovuta sia alle differenze tra gli isotipi che a quelle all'interno di ciascun isotipo.

### **3A.3. Confronto con gli isotipi delle altre specie**

Per chiarire le relazioni esistenti tra gli isotipi di catena L di *T. bernacchii* e quelli degli altri teleostei, è stato effettuato un allineamento con Clustal X 1.8 distinguendo così quattro gruppi di sequenze (Fig. 31)

Quattordici posizioni della regione CL sono in condivisione con le sequenze della regione CL di tutti i teleostei. In 7 posizioni ( 40, 45, 52, 76, 95, 96 e 102) le sequenze del gruppo 1A e 1B mostrano lo stesso residuo, assente nelle sequenze degli altri gruppi.



Un albero filogenetico relativo al dominio CL della specie in esame e a quello delle altre specie di teleostei, evidenzia come effettivamente ciascun isotipo di *T. bernacchii* ricada all'interno del clade dello stesso isotipo descritto nelle altre specie di teleostei (Fig. 32).

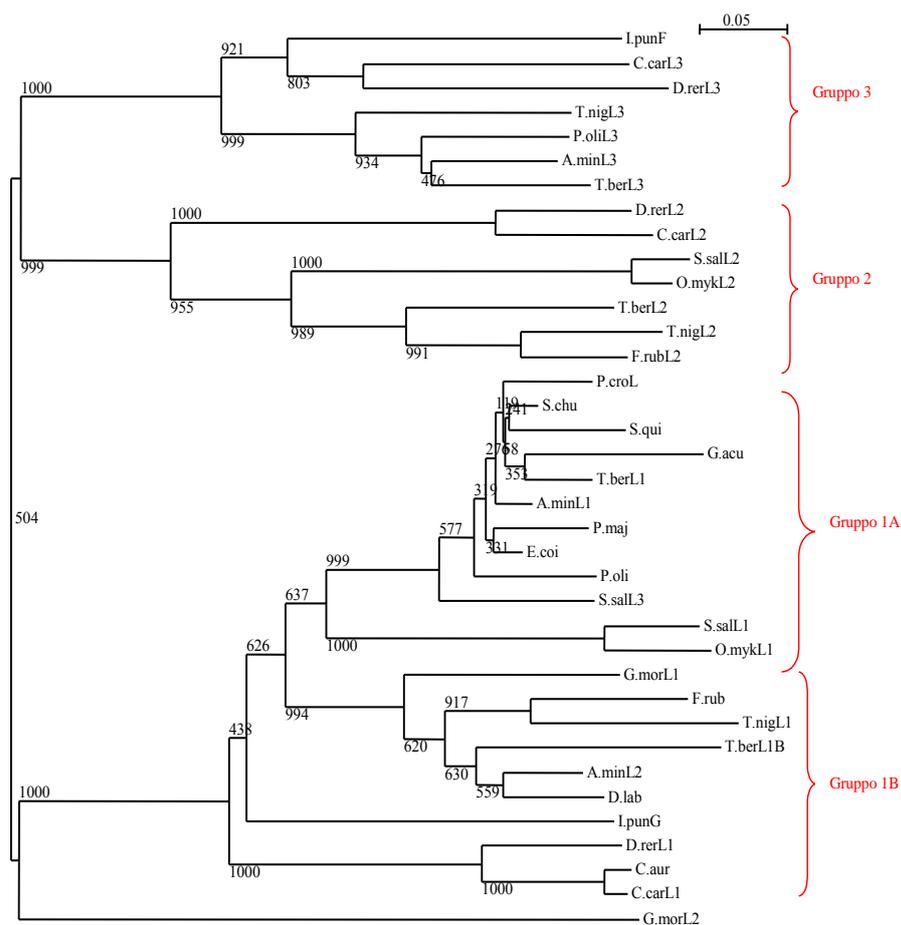


Figura 32. Albero filogenetico delle regioni costanti delle catene leggere della specie *T. bernacchii* e di altre specie di teleostei: *Tetraodon nigroviridis*, *Cyprinus carpio*, *Danio rerio*, *Paralichthys olivaceus*, *Anarhichas minor*, *Salmo salar*, *Oncorhynchus mykiss*, *Siniperca chuatsi*, *Seriola quinqueradiata*, *Gasterosteus aculeatus*, *Pagrus major*, *Epinephelus coioides*, *Gadus morhua*, *Dincentrarchus labrax*, *Ictalurus punctatus*, *Carassius auratus*, *Fugu rubripes*, *Pseudosciaena crocea*.

### 3A.4. Modelli molecolari relativi agli isotipi del dominio costante della catena L

Per ciascun isotipo L1A, L1B, L2 ed L3 di *T. bernacchii* sono stati costruiti modelli molecolari del dominio CL allo scopo di valutarne la conservazione strutturale, scegliendo una sequenza rappresentativa per ciascun isotipo (Fig. 33).

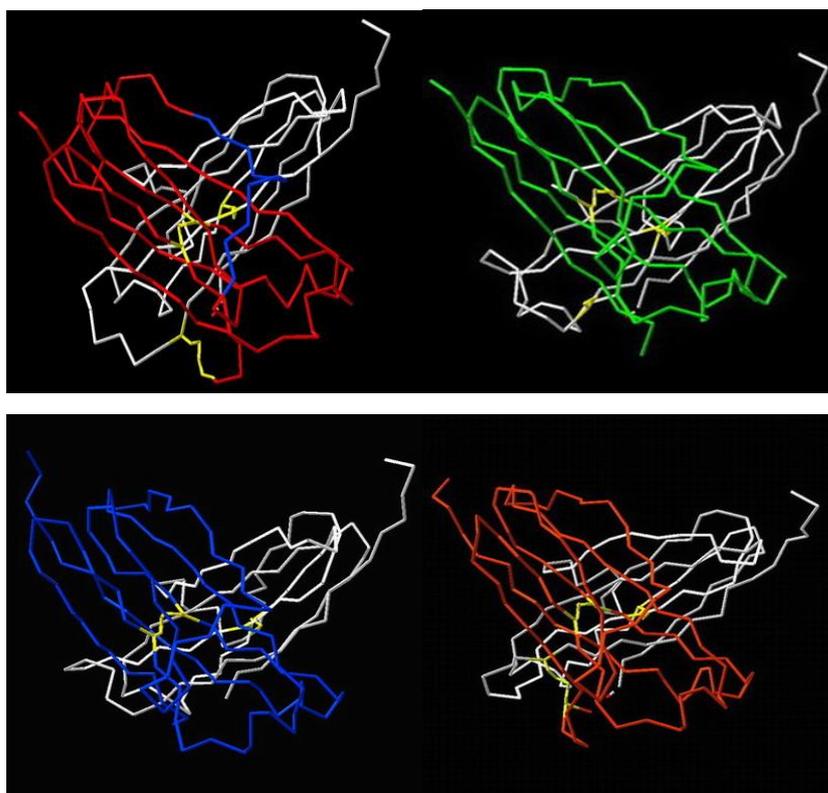


Figura 33. Modelli molecolari degli isotipi di catena leggera, L1A (in alto a sinistra), L1B (in alto a destra), L2 (in basso a sinistra) ed L3 (in basso a destra), con relativo orientamento del dominio CL (in rosso per L1A, verde per L1B, in blu per L2 e arancio per L3) rispetto al CH1 (bianco).

I modelli ottenuti sono stati analizzati e il tipo di appaiamento tra i domini CH1-CL risulta essere molto simile e compatto. È stata misurata la distanza fra gli atomi di zolfo delle cisteine coinvolte nel ponte disolfurico interdominio nei diversi isotipi e si è notato che tale valore oscilla tra 18,5 e 19,6 Å. L'orientamento reciproco dei domini

CH1 e CL è stato calcolato misurando gli angoli fra gli assi principali dei due domini nei 4 modelli; i valori trovati sono molto simili nei diversi isotipi (circa  $120^\circ$ ). Le dimensioni dei dimeri sono molto simili, essi sono tutti inseribili all'interno di una box ( $43 \text{ \AA} \times 42 \text{ \AA} \times 27 \text{ \AA}$ ) ma differiscono per la carica elettrostatica del sistema che è 0, -2, -1, -3 rispettivamente per gli isotipi L1A, L1B, L2, e L3.

Anche la distribuzione di carica appare diversa fra gli isotipi (Fig. 34).

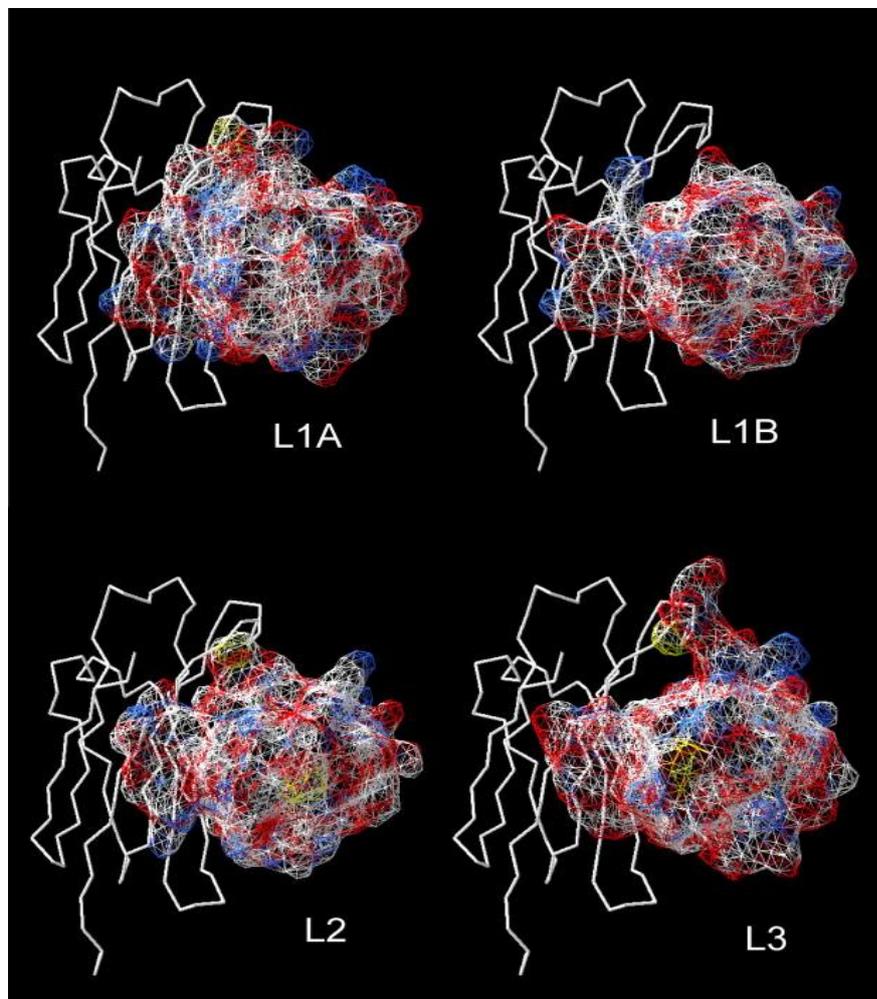


Figura 34. Distribuzione di carica superficiale dei modelli di isotipi di catena leggera. In rosso sono mostrate le cariche negative, in blu quelle positive mentre in bianco le cariche neutre. Per ogni dominio CH è mostrato solo lo scheletro carbossilico.

Il valore dell'RMSD per gli atomi di  $C_{\alpha}$  fra i vari modelli del dominio CL sovrapposti variano tra 1.32 e 1.65 Å (TabellaVI).

Tabella VI. RMSD espressa in Å degli atomi di  $C_{\alpha}$  fra i modelli dei diversi isotipi di catene leggere.

	L1B	L2	L3
L1A	1.32	1.54	1.53
L1B		1.40	1.62
L2			1.65

In contrasto con i bassi valori di RMSD ottenuti, è stato osservato che le posizioni degli scheletri carbossilici sono molto divergenti in numerosi *loop*: in particolare il *loop* che connette gli *strand* D ed E è molto esposto al solvente nel dominio CL di L1A. Infine, è stato osservato che gli ipotetici siti di glicosilazione sono tutti distribuiti in tre regioni precise (Fig. 35).

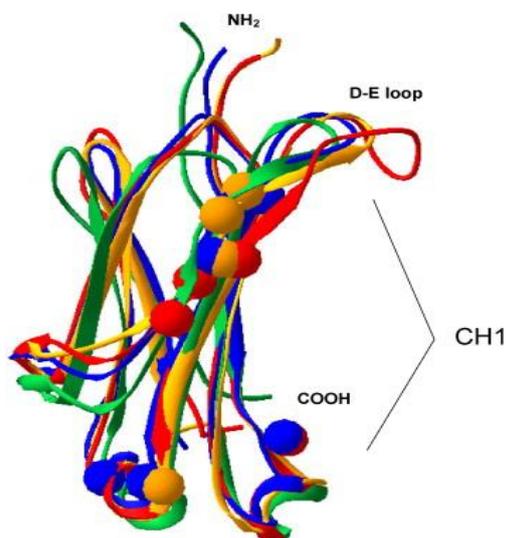


Figura 35. Sovrapposizione dei modelli molecolari dei domini costanti delle sequenze L218 (isotipo L1A, in rosso), L219 (isotipo L1B, in blu), L440 (isotipo L2, in verde), 5141 (isotipo L3, in arancio). I residui di serina, che presumibilmente sono fosforilati, sono evidenziati con delle sfere. Sono state indicate le estremità N- e C- terminale del *loop* che connette gli *strand* D ed E.

### **3B. Studio delle interazioni del dominio transmembrana delle Ig di *Chionodraco hamatus***

La regione transmembrana (IgTM) della catena pesante delle Ig di *C. hamatus* è stata studiata avvalendosi di metodi computazionali al fine di comprendere il tipo di interazioni coinvolte nell'associazione del dimero all'interno della membrana cellulare.

#### **3B.1. Costruzione dei modelli e studio di dinamica molecolare dell'elica transmembrana delle Ig**

Un modello molecolare della regione transmembrana (TM) delle Ig di *C. hamatus*, corrispondente alla sequenza NMGSTAITFILLFLITLLFSIGTTAFKVK era già stato costruito durante lo svolgimento della tesi sperimentale utilizzando come stampo l'elica H del centro di reazione fotosintetico di *R. sphaeroides* (identità di sequenza 33%). Il modello consiste in un'elica estesa, vicina all'idealità, in cui i residui amminoacidici Thr5, Phe9, Leu12, Thr16, Phe19 e Thr24 si trovano sulla stessa faccia dell'elica e sono esposti al solvente. Prima di studiare l'associazione è stata valutata la stabilità di una singola elica attraverso simulazioni di MD in POPC per 50 ns. durante tutta la simulazione, la struttura del modello viene conservata e rimane prossima all'idealità, come si evince dall'analisi della struttura secondaria, del raggio di girazione del raggio dell'elica e dal numero totale dei legami idrogeno in funzione del tempo. La deviazione quadratica media (RMSD, *root mean square deviation*) del modello risultante dopo la simulazione verso il modello di partenza è costante e di circa 2 Å. La regione centrale idrofobica

(residui 9-19) dell' $\alpha$ -elica è completamente immersa nel doppio strato lipidico mentre le estremità cariche (residui 1-4 e 27-29) sono ancorate all'interfaccia polare acqua-lipidi.

### **3B.2. Associazione elica-elica delle IgTM e loro stabilità**

Per poter studiare la formazione spontanea di un omodimero parallelo delle IgTM, sono state costruite e studiate mediante MD in POPC quattro configurazioni, che differiscono per la posizione relativa delle due eliche. I risultati ottenuti sono stati confrontati con i dati di simulazioni condotte in acqua ottenuti durante la mia tesi sperimentale.

Sono state condotte otto simulazioni nel doppio strato lipidico POPC (da 15 a 50 ns per simulazione, per un totale di oltre 150 ns) e tutte mostrano la formazione di un omodimero stabile. Questo risultato non sorprende dal momento che le aggregazioni spontanee elica-elica sono state già osservate in simulazioni dell'ordine di nanosecondi (Mottamal M *et al.*, 2006; Vereshaga YA *et al.*, 2007) mostrando così che le configurazioni di partenza non sono così lontane da quelle all'equilibrio. I dati ottenuti dalle simulazioni in POPC sono stati confrontati con quelli di dieci simulazioni in acqua condotte durante la tesi sperimentale, delle quali quattro non producono un appaiamento parallelo delle due eliche. In alcuni casi, in acqua, le strutture risultanti erano caratterizzate da un impacchettamento antiparallelo delle due eliche (in due casi), le due eliche rimanevano separate (in un caso) o le eliche interagivano in maniera irregolare (in un caso). Nella figura 36 sono mostrati sei modelli rappresentativi ottenuti dalle simulazioni in POPC (A e B) e

sei modelli ottenuti dalle simulazioni in acqua (C e D); nella sovrapposizione un'elica è mantenuta fissa nella stessa posizione per ciascuna struttura. I valori di RMSD dei residui amminoacidici presenti nella regione all'interfaccia delle due eliche (residui 12-19 delle due catene) fra i diversi modelli ottenuti dalle simulazioni in POPC cadono nell'intervallo tra 0,4 e 3.3 Å. Questi risultati indicano che le interfacce che stabilizzano l'omodimero ottenuto dalle simulazioni in POPC sono simili tra loro (Fig. 36A e B). Questa somiglianza suggerisce che i modelli ottenuti possono essere una rappresentazione plausibile dell'associazione delle IgTM nel BCR.

Gli omodimeri ottenuti in acqua, invece, presentano una grande variabilità strutturale (Fig. 36C e D). Le eliche interagiscono attraverso facce diverse; in alcuni casi si osserva una marcata distorsione della struttura dell'elica. L'RMSD dei residui 12-19 delle due catene di queste strutture copre un intervallo che varia tra 2.1 a 5.0 Å.

Le interazioni responsabili dell'associazione delle eliche in ambiente lipidico sono prevalentemente di tipo idrofobico: durante la simulazione la superficie totale accessibile al solvente dei residui idrofobici diminuisce rapidamente (da un valore medio di 1100 Å<sup>2</sup> a 950 Å<sup>2</sup>) mentre l'area accessibile al solvente dei residui idrofilici rimane praticamente immutata a circa 900 Å<sup>2</sup>. L'evoluzione nel tempo degli elementi di struttura secondaria, il raggio di girazione dei C<sub>α</sub> e il numero di legami idrogeno di ciascuna elica individualmente mostrano che i monomeri conservano la loro struttura durante l'associazione elica-elica, ad eccezione delle regioni terminali. Il passo per residuo delle eliche varia tra 1,32 e 1,61 ed in alcuni casi è prossimo al valore ideale di 1,5.

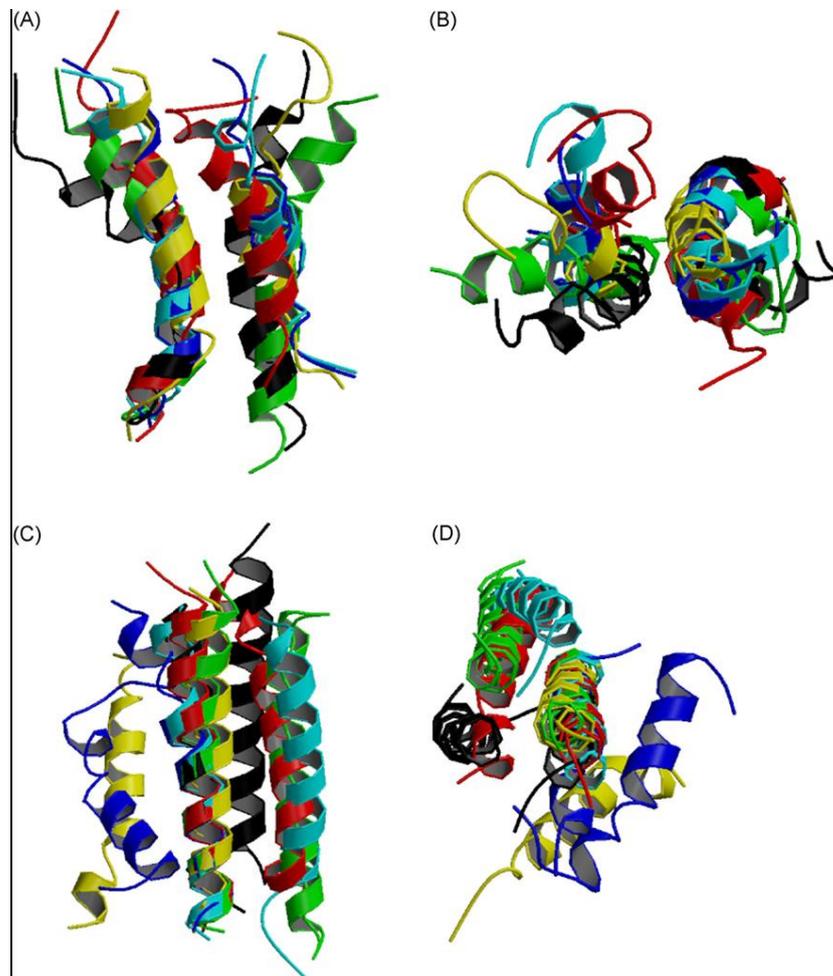


Figura 36. Confronto strutturale fra gli omodimeri IgTM ottenuti mediante simulazioni in ambienti diversi. Le sei strutture rappresentative ottenute dalla simulazione in POPC sono mostrate nei riquadri A e B mentre le sei strutture ottenute mediante simulazione in acqua sono mostrate nei riquadri C e D.

I residui coinvolti nell'interazione elica-elica si trovano all'interfaccia e sono stati identificati: la regione interfacciale non presenta legami idrogeno intermolecolare ma le due eliche interagiscono attraverso interazioni di van der Waals (Fig. 36A–C) fra gli stessi residui Phe9, Leu12, Phe13, Leu14, Leu17 e Phe19 delle due catene.

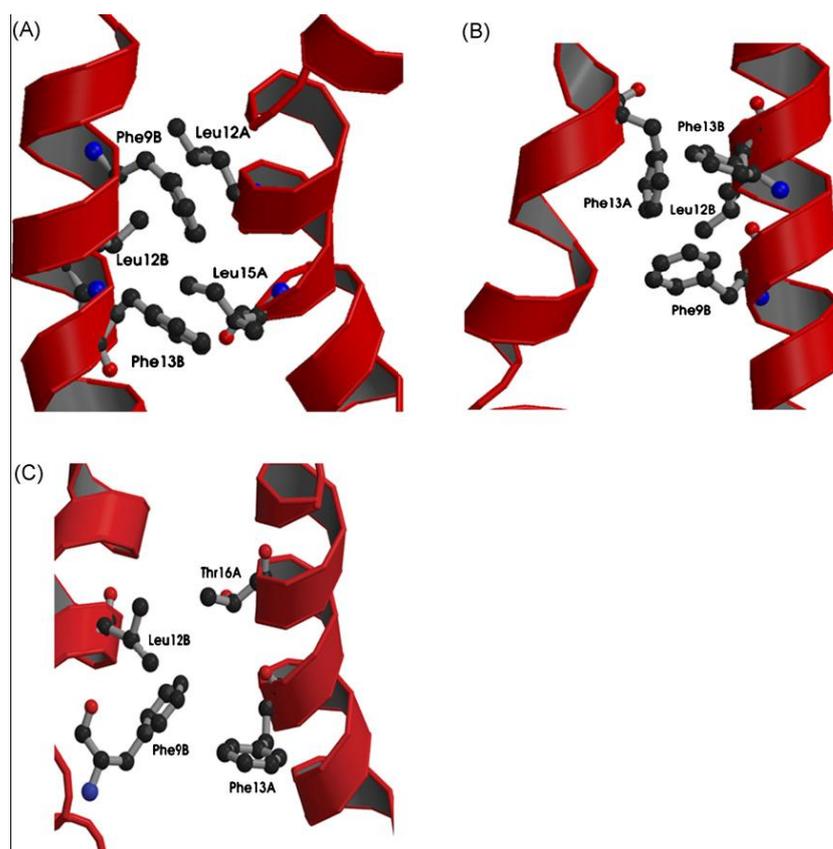


Figura 37. I residui amminoacidici determinanti nella struttura e nella stabilità degli omodimeri IgTM in modelli rappresentativi ottenuti mediante simulazione in POPC.

Sebbene le distanze intermolecolari medie fra questi residui chiave cambino durante le simulazioni, sono stati identificati i contatti che portano all'associazione delle IgTM. Le interazioni intermolecolari più conservate si formano tra le catene laterali dei residui Phe13 e Leu12 (Fig. 35A e B), inoltre è stata osservata un'occasionale interazione  $\pi$ - $\pi$  tra le due Phe19 dal momento che i due anelli aromatici sono paralleli ad una distanza di 4,3 Å (Burley SK and Petsko GA, 1986). Altre interazioni osservate spesso coinvolgono le catene laterali dei residui Leu12, Phe13, e Leu14 con Phe9 (Fig. 35C). Infine i residui Phe9-Phe9, Leu17-Phe19 e Phe13-Leu14 possono trovarsi a stretto contatto.

Nei modelli finali degli omodimeri le estremità N- e C- terminale sono esposte al solvente. All'estremità C-terminale Lys27 e Lys29 delle due catene interagiscono con gli atomi di ossigeno delle teste polari delle molecole di POPC sulla superficie del doppio strato lipidico, consentendo così l'ancoraggio del dimero alla membrana. Si è notato infatti durante la simulazione che questi residui sono liberi di muoversi ma sono costantemente sulla faccia esterna del doppio strato lipidico.

### **3B.3. Confronto della regione TM delle Ig di *Chionodraco hamatus* con quella di altre specie di vertebrati**

Per verificare se il tipo di interazione responsabile dell'associazione dei dimeri delle IgTM di *C. hamatus* fosse una caratteristica peculiare dei teleostei antartici, lo studio di tale regione è stato esteso a isotipi diversi di Ig appartenenti a tutte le classi di vertebrati

#### *Analisi delle sequenze amminoacidiche*

Sono state analizzate le sequenze amminoacidiche della forma di membrana delle IgH di diversi isotipi di numerose specie di vertebrati (Tabella S2 in appendice). Utilizzando metodi computazionali sono state predette la regione extracellulare prossimale alla membrana (EMPD), quella che attraversa la membrana (TM) e la regione citoplasmatica di ciascuna sequenza amminoacidica dedotta. La regione EMPD è estremamente variabile: la sua lunghezza varia nell'intervallo di 9-39 residui amminoacidici (il valore medio è 20,6). La composizione amminoacidica di ciascuna

regione mostra un alto contenuto di residui carichi negativamente (Glu + Asp 31,7%) nell'EMPD. La regione TM presenta un alto contenuto di Phe (14,9%) e di Ser+Thr (28,5%), un basso contenuto di residui con piccola catena laterale e l'assenza di cariche negative. La coda citoplasmatica, infine, mostra un alto contenuto di residui carichi positivamente (Lys 21,0%) (Tabella VII).

Tabella VII. Percentuale di composizione amminoacidica delle diverse regioni delle IgH, specifiche per la forma di membrane nelle 68 sequenze del dataset.

	<b>Total</b>	<b>EMPD</b>	<b>TMD</b>	<b>Cyt</b>
Ala	5.4	5.4	6.5	2.5
Arg	1.2	0.9	0.1	4.8
Asn	2.8	5.0	1.0	2.5
Asp	5.1	13.0	-	3.0
Cys	1.8	3.2	1.3	-
Gln	2.4	4.0	0.1	6.2
Glu	7.4	18.7	-	2.2
Gly	4.4	6.6	2.3	4.8
His	0.6	1.1	-	1.3
Ile	7.2	4.3	10.1	6.1
Leu	12.5	10.1	17.8	4.3
Lys	4.3	1.2	0.1	21.0
Met	2.0	2.1	1.5	2.9
Phe	8.2	2.8	14.9	3.0
Pro	2.0	3.0	0.4	4.6
Ser	8.2	6.3	10.6	6.1
Thr	10.3	3.8	17.9	5.4
Trp	2.1	1.0	2.9	2.4
Tyr	3.2	2.3	4.1	3.0
Val	8.4	6.0	8.3	13.9

Per determinare se la composizione delle IgTM sia una caratteristica comune delle eliche transmembrana, è stata misurata

la frequenza dei vari amminoacidi in 236 segmenti dell'elica transmembrana di proteine comprese nel PDBTM dataset (vedi Metodi) (Tabella VIII).

Tabella VIII. Confronto fra la composizione amminoacidica delle IgTM e quella di 236 catene di  $\alpha$ -eliche TM incluse nel PDBTM dataset.

	<b>IgTM</b>	<b>PDBTM</b>
Ala	6.5	11.4
Arg	0.1	3.1
Asn	1.0	2.2
Asp	–	1.9
Cys	1.3	1.1
Gln	0.1	2.5
Glu	–	3.3
Gly	2.3	7.1
His	–	1.4
Ile	10.1	9.6
Leu	17.8	14.7
Lys	0.1	2.9
Met	1.5	3.3
Phe	14.9	7.2
Pro	0.4	2.3
Ser	10.6	4.8
Thr	17.9	5.5
Trp	2.9	2.4
Tyr	4.1	3.4
Val	8.3	9.9

In un campione di 24.886 residui di eliche transmembrana è stato trovato il 10,3% di Ser+Thr, i residui carichi negativamente sono il 5,2% mentre quelli carichi positivamente sono il 6,0%, i residui con piccole catene laterali (Gly+Ala) sono il 18,5% infine il contributo di Phe è del 7,2%. Il confronto fra le sequenze dell'IgTM e la tendenza generale conferma che le IgTM posseggono una composizione di sequenza unica.

Una caratteristica importante delle sequenze IgTM è l'assenza di residui carichi positivamente che sono comuni nella maggioranza dei recettori immunologici e interagiscono, tramite ponti salini, con i dimeri atti al *signaling* (Call ME and Wucherpfennig KW, 2007). Questo dato suggerisce l'idea che le eliche IgTM si associno al dimero trasducente Ig $\alpha$ -Ig $\beta$  in un modo diverso. La conservazione dei residui nell'allineamento delle sequenze IgTM è stata analizzata con il metodo dell'entropia di Shannon (Fig. 38).

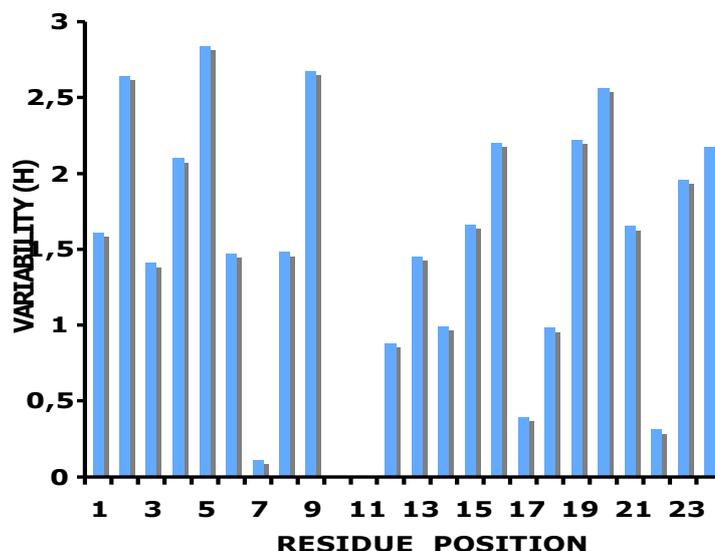


Figura 38. Grafico dell'entropia di Shannon (H) dell'allineamento di sequenze amminoacidiche del dataset di IgTM.

I valori più alti di entropia sono stati trovati in 12 posizioni delle regioni N- e C- terminale. In particolare sono stati identificati otto residui altamente conservati o il cui carattere è conservato tra le diverse specie (FXXLFLXS/TXXYSXXXT); tale motivo è visualizzato mediante il tool WebLogo utilizzato per l'analisi della variabilità delle posizioni (Fig. 39).



Figura 39. Posizioni conservate nell'allineamento di 68 sequenze di IgTM del dataset rappresentate mediante il tool WebLogo.

Per stabilire se tale motivo conservato ha un significato biologico per le molecole Ig, è stata valutata la sua frequenza nel dataset PDBTM e non è stato mai trovato.

### Analisi strutturale

Per costruire modelli per omologia di omodimeri IgTM, sono state selezionate 7 sequenze TM diverse comprendenti 29 residui amminoacidici, con valori di identità di sequenza tra il 37% e l'89% (Fig. 40).

							1	5	10	15	20	25																							
IgM <i>M.musculus</i>							N	L	W	T	A	S	T	F	I	V	L	F	L	L	S	L	F	Y	S	T	V	T	L	F	K	V	K		
IgM <i>I.punctatus</i>							N	M	A	N	T	A	L	T	F	V	F	L	F	L	I	T	L	F	Y	S	I	G	V	T	V	F	K	V	K
IgM <i>P.waltl</i>							N	I	W	S	T	T	S	T	F	I	I	L	F	L	L	S	L	S	Y	S	A	T	V	T	L	F	K	V	K
IgM <i>A.platyrrhynchos</i>							N	L	W	T	T	A	S	T	F	I	V	L	F	I	L	S	L	F	Y	S	A	T	V	T	L	I	K	V	K
IgG2A <i>M.musculus</i>							G	L	W	T	T	I	T	I	F	I	S	L	F	L	L	S	V	C	Y	S	A	S	V	T	L	F	K	V	K
IgNAR <i>G.cirratum</i>							N	S	S	L	S	I	M	T	F	V	I	L	F	I	L	S	M	L	Y	S	T	F	V	T	V	M	K	D	L
IgY <i>A. carolinensis</i>							H	L	W	M	T	I	F	I	F	I	L	F	L	L	S	V	C	Y	S	A	T	V	T	L	S	K	V	K	

	IgG2a <i>M.mus</i>	IgM <i>A.pla</i>	IgY <i>A.car</i>	IgM <i>P.wal</i>	IgM <i>I.pun</i>	IgNAR <i>G.cir</i>
IgM <i>M.mus</i>	68	89	62	79	62	41
IgG2a <i>M.mus</i>		65	75	65	44	37
IgM <i>A.pla</i>			62	75	55	41
IgY <i>A.car</i>				65	41	41
IgM <i>P.wal</i>					55	41
IgM <i>I.pun</i>						41

Figura 40. Allineamento e percentuale di identità di sequenza fra le regioni IgTM studiate in MD.

I modelli sono stati studiati mediante MD per un tempo totale di simulazione di circa 305 ns. Le simulazioni mostrano la formazione di omodimeri molto compatti (Fig. 41 e tabella IX).

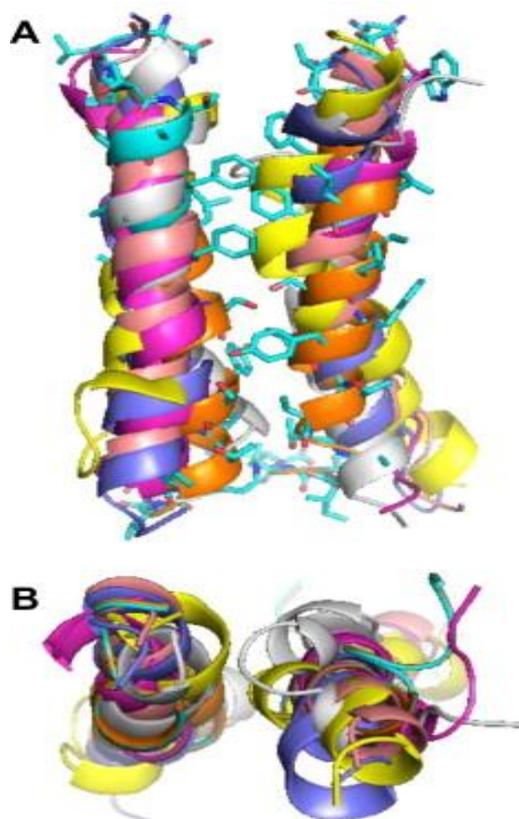


Figura 41. Confronto strutturale fra gli omodimeri IgTM ottenuti mediante simulazioni in POPC. (A) Visione frontale, (B) visione dall'alto.

Tabella IX. Parametri strutturali risultanti dalle simulazioni MD delle regioni TM di diversi isotipi di Ig.

	Lunghezza media dell'elica (Å)	Raggio medio dell'elica	Passo medio dell'elica (Å)	Raggio di girazione per residuo (Å)	Media del n° di legami idrogeno
IgM <i>M. musculus</i>	18.9 ± 0.4	2.61 ± 0.04	1.53 ± 0.03	14.2 ± 0.2	33.7
IgM <i>I. punctatus</i>	11.9 ± 0.4	2.57 ± 0.03	1.52 ± 0.03	13.0 ± 0.1	33.3
IgM <i>P. waltil</i>	22.0 ± 0.7	2.7 ± 0.1	1.37 ± 0.05	11.8 ± 0.4	33.9
IgM <i>A. platyrhynchos</i>	28.3 ± 0.5	2.64 ± 0.04	1.57 ± 0.03	13.9 ± 0.1	33.1
IgG2A <i>M. musculus</i>	25.7 ± 0.4	2.44 ± 0.04	1.55 ± 0.02	13.0 ± 0.1	35.6
IgNAR <i>G. cyrratum</i>	21.4 ± 0.3	2.43 ± 0.01	1.49 ± 0.02	13.3 ± 0.1	39.4

	<b>Lunghezza media dell'elica (Å)</b>	<b>Raggio medio dell'elica</b>	<b>Passo medio dell'elica (Å)</b>	<b>Raggio di girazione per residuo (Å)</b>	<b>Media del n° di legami idrogeno</b>
IgY <i>A. carolinensis</i>	31.9 ± 0.3	2.44 ± 0.03	1.52 ± 0.01	14.8 ± 0.1	38.5

I valori calcolati di RMSD per le strutture dei dimeri sono in un intervallo tra 1,5-2,7 Å, ad indicare che le strutture delle IgTM sono molto simili in specie diverse.

Sono stati studiati i principi fisici coinvolti nella formazione del dimero e sono stati identificati i contatti specifici che portano all'associazione delle due catene. Le catene laterali di un'elica sono complementari alle cavità dell'altra elica, permettendo un impacchettamento molto stretto (valori di  $S_C$  tra 0,61 e 0,86) stabilizzato prevalentemente da interazioni intermolecolari tra gli anelli aromatici di due residui di Phe9 e Phe13 delle due catene (Fig. 42).

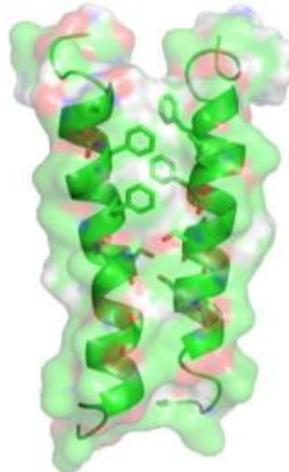


Figura 42. Le interazioni che determinano la formazione e la stabilità degli omodimeri IgTM in un modello rappresentativo ottenuto mediante simulazioni in POPC. Le interazioni specifiche sono prevalentemente mediate da una perfetta complementarità di superficie delle due eliche.

Questo è in accordo con il ben noto ruolo centrale dell' $\pi$ -*stacking* nei processi di auto-assemblaggio osservati in chimica e biochimica (McGaughey G *et al.*, 1998; Gazit E, 2002). Le interazioni osservate tra i residui di fenilalanina sono in contrasto con l'idea generale che catene laterali più piccole ottimizzino l'impacchettamento elica-elica (Jiang S and Vakser IA, 2004). Per stabilire il potenziale significato biologico della presenza di un motivo FXXXF nelle  $\alpha$ -eliche transmembrana, è stata ricercata la sua frequenza all'interno del Database di proteine transmembrana ed è stato notato che non è statisticamente rappresentativa. Questo risultato è in accordo con le analisi statistiche condotte da Senes e collaboratori (Senes A *et al.*, 2000).

Negli omodimeri IgTM è stata trovata inoltre la presenza di legami idrogeno tra le catene laterali di Ser/Thr16, Ser20 di una catena e i corrispondenti residui dell'altra catena. Durante la simulazione, le catene laterali dei residui di serina adottano prevalentemente la conformazione  $g^+$  ( $\chi_1 = 240-360^\circ$ ). Ciò

suggerisce che le interazioni  $\pi-\pi$  tra i residui di fenilalanina potrebbero essere la forza che guida l'associazione, mentre i residui di serina e treonina forniscono assistenza all'associazione formando legami idrogeno intercatena.

Nella regione citoplasmatica, all'estremità C-terminale dell'elica, due catene laterali di Lys27 e Lys29 sono coinvolte in un'interazione favorevole con gli atomi di ossigeno delle molecole di POPC sulla superficie del doppio strato lipidico.

Nel complesso i risultati delle simulazioni e delle analisi statistiche delle eliche transmembrana suggeriscono che l'associazione elica-elica delle IgTM dipendono dalle interazioni sequenza-specifiche che coinvolgono un motivo inusuale, FXXXFXXS/TXXXS. Questo motivo, mai identificato prima, sembra essere una sequenza che promuove l'associazione elica-elica in regioni transmembrana.

### **3C. Studio dello *splicing* atipico delle Ig di membrana di *Chionodraco hamatus***

#### **3C.1. Confronto della sequenza della catena pesante delle Ig dei teleostei**

Nel 2007 nel corso del mio lavoro di tesi di laurea sperimentale ho sequenziato la forma secretata e di membrana della catena pesante delle Ig di *C. hamatus* e dal loro confronto è risultato che la forma di membrana è mancante degli ultimi due domini costanti. Per verificare se il dato ottenuto fosse specifico per il gene dell'individuo utilizzato nell'esperimento, il trascritto codificante mIgM è stato sequenziato in campioni di milza di 8 individui diversi di *C. hamatus* ed i risultati ottenuti, a prescindere da limitate variazioni nucleotidiche attribuibili a polimorfismo allelico, indicavano la mancanza dei due domini. Per verificare se tale perdita fosse una specificità della specie o fosse comune anche in altre specie antartiche, è stato fatto un allineamento con altre 13 sequenze di cDNA di Ig di membrana di teleostei antartici e non antartici (Fig. 43).

Sui domini condivisi di tale allineamento è stato costruito un albero filogenetico che mostra come tutte le specie che non presentano l'esone CH3 nella forma di membrana appartengono allo stesso clade (Fig. 44).

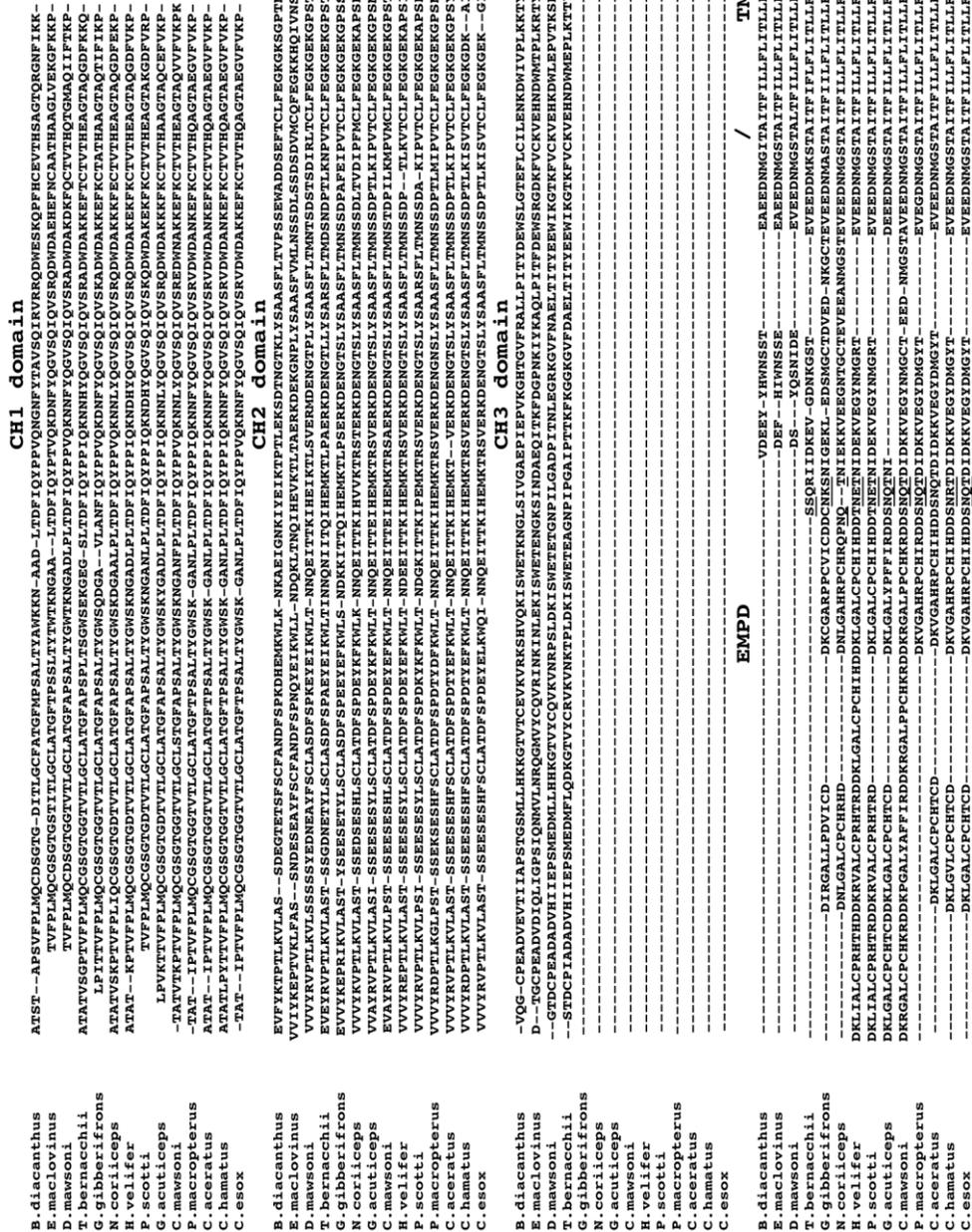


Figura 43. Allineamento delle sequenze amminoacidi che codificanti la forma di membrana della catena pesante delle Ig di 13 specie di Nototenioidi antartici e non antartici.

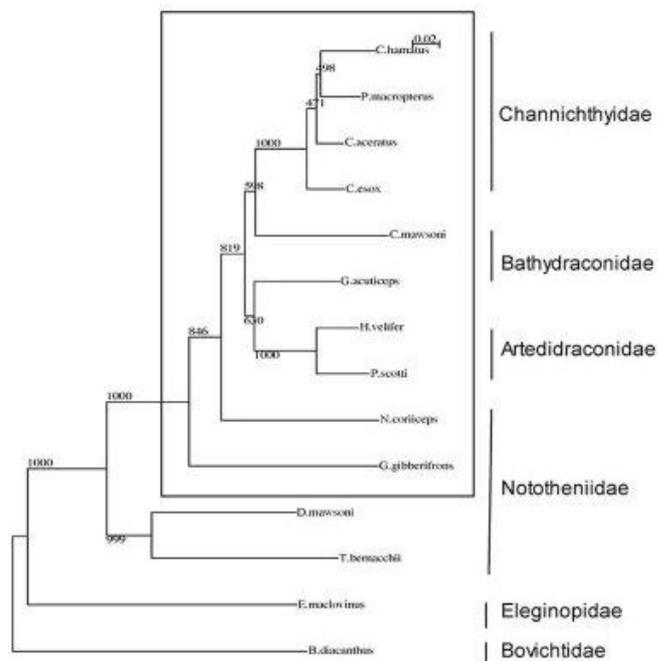


Figura 44. Albero filogenetico costruito con il metodo Neighbor-Joining mostra le distanze tra le sequenze di Ig nell'allineamento di figura 43. Nel riquadro sono comprese le specie che presentano una forma di membrana della catena pesante mancante dell'esone CH3

È stato interessante notare che all'assenza del terzo esone costante è associata la presenza di un braccio spaziatore extracellulare insolitamente lungo. Tale regione è composta da sequenze di 13 residui amminoacidici, ripetuti in numero variabile fra le specie da 1 a 4, la cui composizione è abbastanza conservata tra le specie. La composizione amminoacidica rivela un'abbondanza relativa di residui carichi negativamente (acido aspartico e acido glutammico sono il 23,7%). I residui di cisteina e prolina sono rispettivamente il 6,2 e 4,2 % del totale.

La sequenze dei 39 nucleotidi codificanti le ripetizioni di 13 residui amminoacidici mostrano un alto grado di conservazione (Fig. 45).

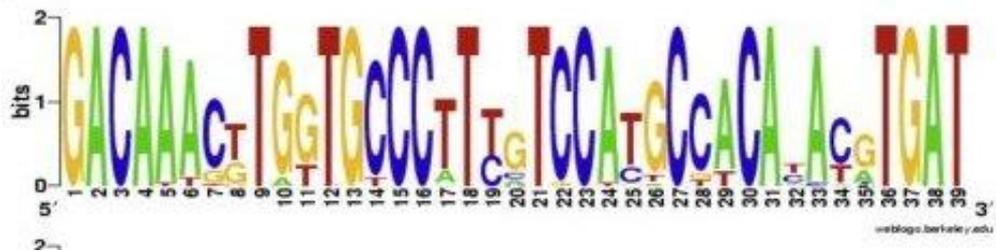
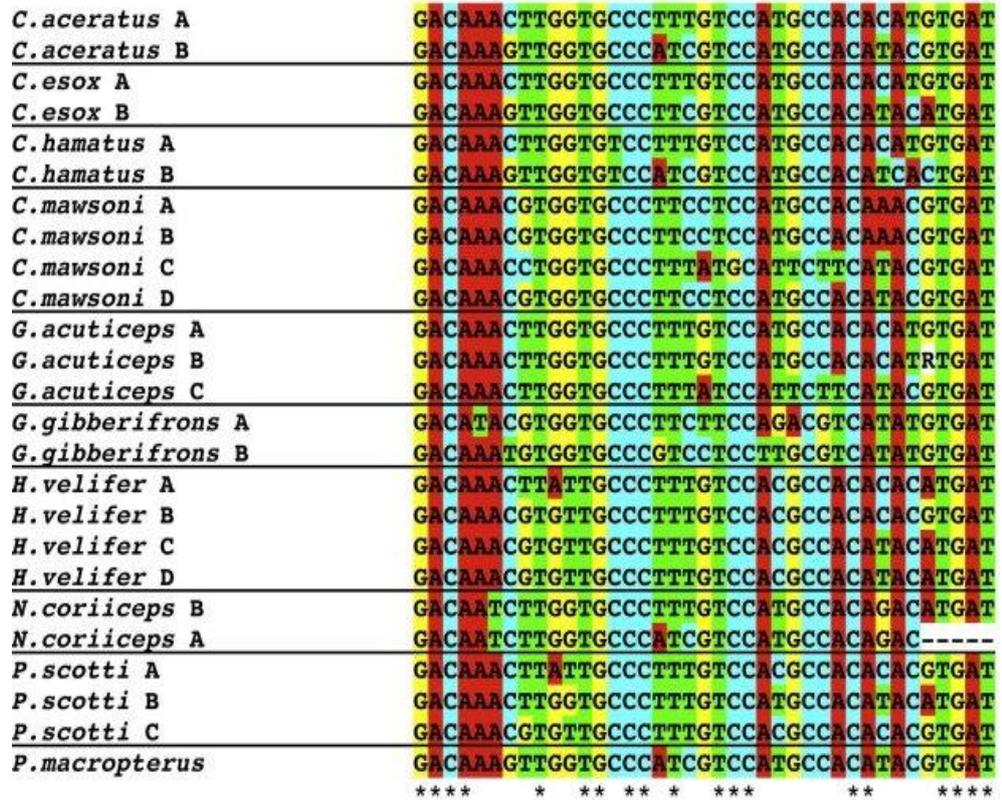


Figura 45. In alto: Allineamento delle sequenze di 39 nt. Le lettere (A-D) identificano la ripetizione da 39 nt in ciascuna specie dove A indica quella all'estremità C-terminale. Gli asterischi indicano le posizioni conservate. In basso: confronto della conservazione nella sequenze dell'allineamento, rappresentato dal *tool*/WebLogo

### **3C.2. Espressione di un peptide ricombinante comprendente le regioni CH2 e EMPD delle Ig di *Chionodraco hamatus***

La presenza di un braccio spaziatore extracellulare più lungo, osservato nelle Ig di *C. hamatus*, potrebbe avere un ruolo strutturale importante. Per comprendere la funzione di tale EMPD è stato espresso un frammento ricombinante della mIgM di *C. hamatus* comprendente il secondo esone costante, CH2, e la regione extracellulare prossimale alla membrana (EMPD). La regione di cDNA codificante il frammento proteico di interesse è stata amplificata mediante PCR utilizzando come stampo il clone S7, contenente l'intera sequenza della catena pesante delle mIg. I *primers* oligonucleotidici utilizzati sono stati *CH2Rec* (complementare ad una sequenza al 5' del dominio CH2) e *VerdeRec* (complementare alla sequenza al 3' dell'EMPD), in entrambe è stata sostituita una base presente nello stampo per poter inserire una tripletta di inizio ed un codone di stop. Dopo aver purificato il frammento dal gel d'agarosio, è stata condotta una reazione per convertirne le estremità in una forma *blunt* e fosforilata. Il frammento ottenuto è stato poi clonato all'interno del vettore di espressione *pETBlue-1* trasformando il ceppo di cellule *NovaBlue Singles Competent Cells* per poter verificare il costrutto. Il frammento, comprendente 497 nt, è stato sequenziato e tradotto (Fig. 46). È stato predetto il peso molecolare della proteina codificata che risulta essere pari a 17.966 Da. Per l'espressione inducibile con IPTG del gene, è stato necessario trasformare il plasmide ricombinante stabilizzato nelle cellule *NovaBlue* nel ceppo *Tuner(DE3)pLacI*. Sono state condotte colture cellulari e indotte con IPTG 0,1mM per una notte a +4°C. Le cellule recuperate dopo l'induzione sono state

separate dal terreno di coltura e separate in due frazioni: l'estratto cellulare e la frazione insolubile contenente i detriti cellulari e i corpi di inclusione. Per verificare l'effettiva espressione del peptide ricombinante, sono stati caricati su SDS-PAGE al 15% i seguenti campioni: coltura cellulare, cellule, estratto cellulare e frazione insolubile della coltura indotta con IPTG e della coltura non indotta utilizzata come controllo negativo. È stata ritrovata una banda di proteine, il cui peso era di oltre i 15 kDa, molto abbondante presente nella coltura, nelle cellule e nella frazione insolubile indotte con IPTG.

```

atggagggtgttcttcgtaatgccagtggtggtttatagagatccaactcttaaagtgtta
M E G V F V M P V V V Y R D P T L K V L
gcctccacttctctgaagaaagtgagttctcaacttctctgcttggccactgatttttca
A S T S S E E S E S H F S C L A T D F S
ctgacacatatgagttcaaatggctgacaaacaatcaagaaatcacaaccaaactccac
P D T Y E F K W L T N N Q E I T T K I H
gaaatgaaaactcgttctgtggaagaaaggacgagaatggaacttcaactgtacagcgca
E M K T R S V E R K D E N G T S L Y S A
gcaagtttctcactcagatgaattccagtgaccgcactttgaagatctcagtcacatgtctg
A S F L T M N S S D P T L K I S V T C I
tttgatgggaaagggataaagctatcgtgaatgacacgctcatcacaagacatgtcca
F D G K G D K A I V N D T L I T K T C P
atatgtcccgacaaaacttgggtgtcctttgtccatgccacacatgtgatgacaagttggt
I C P D K L G V L C P C H T C D D K V G
gccatcgtccatgccacatacatgatgactccaaccaaacggacattgacaagaagta
A H R P C H I H D D S N Q T D I D K K V
gaaggataagacatggg
E G -

```

Figura 46. La sequenza nucleotidica (in minuscolo) e la sequenza amminoacidica dedotta (in maiuscolo) dei domini CH2 (in verde) e EMPD (in giallo). In rosso è mostrata la tripletta di stop.

Per rivelare l'effettiva presenza del frammento ricombinante in tali frazioni, è stato ripetuto l' SDS-PAGE ed il gel è stato trasferito su filtro di nitrocellulosa per effettuare un *immunoblotting*. L'anticorpo primario usato è un anticorpo anti-catena pesante delle Ig di *C. hamatus*, prodotto nel laboratorio stesso. È stata osservata la presenza di una banda molto intensa alla stessa altezza nelle colture, nelle cellule e nella frazione insolubile della coltura indotta con IPTG (Fig. 45). Tale risultato, oltre a confermare l'avvenuta

espressione, permette anche di localizzare la presenza del peptide ricombinante all'interno della frazione insolubile.

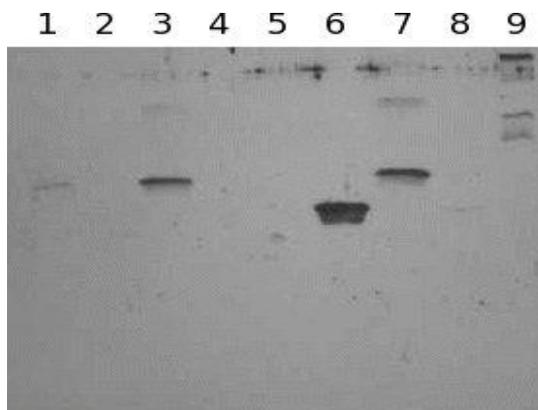


Figura 47. *Immunoblotting* delle frazioni cellulari delle colture indotte e non indotte con IPTG effettuato con un anticorpo primario anti-catena pesante delle Ig di *C. hamatus*. Sono stati caricati nell'ordine: 1) e 2) terreno di coltura della coltura indotta e non indotta; 3) e 4) cellule della coltura indotta e non indotta; 5) e 6) frazione solubile delle colture indotta e non indotta; 7) e 8) frazione insolubile della coltura indotta e non indotta; 9) Ig purificate da bile di *C. hamatus* utilizzate come controllo positivo.

### **3C.3. Analisi della regione genomica compresa tra gli esoni CH2 e TM in *Chionodraco hamatus***

#### *Determinazione della sequenza*

L'analisi genomica è stata condotta allo scopo di chiarire lo *splicing* atipico osservato nei teleostei antartici. A tale scopo è stato amplificato il frammento genomico comprendente gli esoni CH2 e TM in *C. hamatus* utilizzando *primers* appropriati. Per confrontarlo con quello di altre specie che non mostrano lo stesso tipo di *processing* è stato amplificato il frammento genomico relativo alla stessa regione del Nototenioideo non antartico, *B. diacanthus*. I prodotti amplificati per PCR mostrano altezze diverse (Fig. 48): quello ottenuto dal DNA di *B. diacanthus* è lungo circa 2.000 bp mentre quello ottenuto con il

DNA di *C. hamatus* è di circa 4.500 bp. Entrambe i prodotti sono stati clonati, sequenziati e analizzati (Numeri di accesso in banca dati rispettivamente EU884292 e EU925137).



Figura 48. Esperimenti di PCR usando come *primer* senso una sequenza oligonucleotidica riprodotte una sequenza nell'esone CH2 di *C. hamatus* e come antisense una sequenza nell'esone TM della stessa specie. I prodotti amplificati mediante PCR da DNA genomico sono stati caricati nel seguente ordine: *B. diacanthus* (linea 1) e *C. hamatus* (linea 2). La conferma che tali prodotti appartengano al locus delle IgH è stata ottenuta mediante ibridazione con una sonda marcata specifica per IgH di *C. hamatus*. La mobilità dei marcatori è indicata a sinistra.

Le sequenze relative alla regione genomica CH2-TM in *C. hamatus* e in *B. diacanthus* sono riportate in appendice rispettivamente come sequenza A1 e sequenza A2.

I risultati ottenuti sono mostrati schematicamente in Figura 49. Nel caso di *B. diacanthus*, il frammento da 2.022 nt presenta un'organizzazione genica simile a quella osservata in altre specie di teleostei (Fig. 49A). Per *C. hamatus* è stato ottenuto un frammento da 4.563nt (Fig. 49B) la cui sequenza presenta rispetto a quella di *B. diacanthus* un'inserzione. Da un'analisi più approfondita risulta che in *C. hamatus* il segnale di poliadenilazione specifico per la forma secreta (vedi sequenza A1 in appendice) segue al 3' una sequenza non codificante di 1.348nt comprendente il segnale di adenilazione e l'esone RB; l'esone RB di 39 nt; un introne di 371 nt fra gli esoni RB

ed RA; l'esone RA di 39 nt; un introne tra l'esone RA e l'esone TM1 di 1.066 nt ed infine una porzione dell'esone TM1.

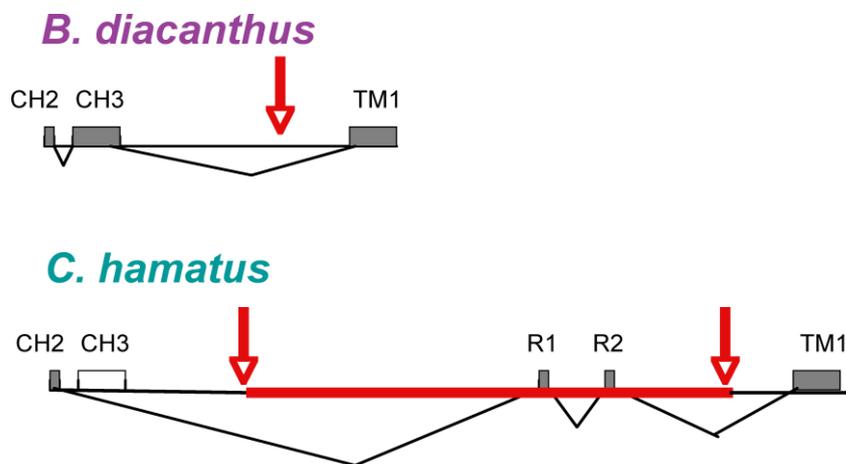


Figura 49. Confronto delle sequenze genomiche di *B. diacanthus* e *C. hamatus* comprese fra la fine dell'esone codificante l'esone CH2 e l'esone codificante la regione transmembrana. Gli esoni utilizzati nel trascritto maturo sono in grigio. La regione presente in *C. hamatus* ma non in *B. diacanthus*, e contenente i piccoli esoni addizionali R1 e R2, è in rosso. Le linee in basso uniscono gli esoni presenti nel trascritto maturo. Le frecce indicano la localizzazione dell'inserzione.

Una successiva analisi del frammento da 4.563 nt ha rivelato che l'esone RA e l'esone RB sono inclusi in due grandi sequenze ripetute, da 411 nt ciascuna, chiamate APS2 e APS3 la cui sequenza è molto simile alla sequenza antiparallela di una regione a monte nello stesso locus che inizia nell'introne CH2-CH3 e termina nell'introne CH3-CH4. Più a monte è presente una terza regione di 1.043 nt, chiamata APS1, identificata come antiparallela di una sequenza che comprende l'introne CH2-CH3 ed il 3' UTR (Regione S); all'interno della regione APS1 è presente una sequenza non codificante di 39 nt (RC) molto simile ad RA ed RB. Infine, sono state identificate due sequenze da 555 nt ciascuna, chiamate

rispettivamente m ed n, che presentano un'identità di sequenza pari al 99,6%. La figura 50 schematizza la situazione descritta.

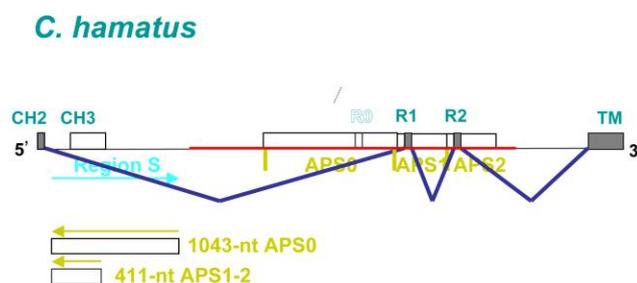


Figura 50. Schema della regione genomica di *C. hamatus* compresa tra la fine dell'esone codificante il dominio CH2 e l'esone codificante la regione transmembrana. Gli esoni utilizzati nel trascritto maturo sono in grigio. L'inserzione contenente R1 e R2, è in rosso. Le linee blu in basso uniscono gli esoni presenti nel trascritto maturo. Le regioni in giallo all'interno dell'inserzione (APS0, APS1 e APS2) sono antiparallele rispetto alla regione in azzurro; le frecce in basso indicano la loro sovrapposizione.

### 3C.4. Analisi comparativa delle sequenze antiparallele e delle relative controparti

Il confronto tra le sequenze nucleotidiche delle regioni APS0, APS1, APS2 e la rispettiva regione inversa complementare ha mostrato un elevato grado di identità con percentuali di identità comprese tra l'89,8 e il 96,9%.

Un'analisi filogenetica rivela che le regioni APS1 e APS2 sono molto più simili tra loro che ad APS0 (Fig. 51).

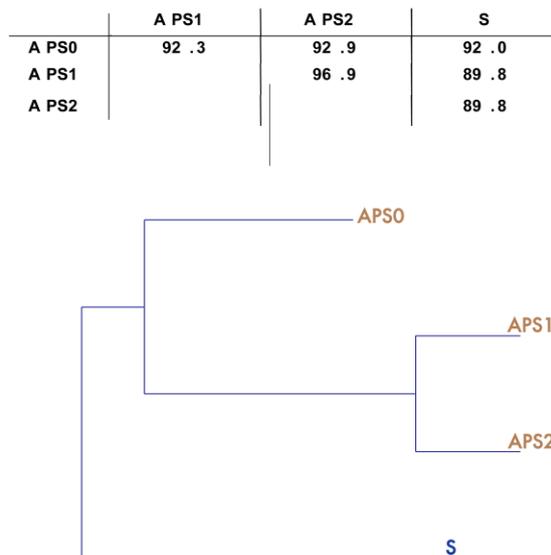


Figura 51. In alto è riportata la percentuale di identità nucleotidica fra le regioni APS0, APS1, APS2 e la rispettiva controparte inversa complementare nel dominio CH3 (la regione S). In basso è mostrato l'albero delle distanze relative che ne descrive le relazioni.

L'acquisizione di siti donatori ed accettori di *splicing* nelle sequenze inverse complementari comporta l'inclusione nel trascritto maturo di sequenze ripetute R1 ed R2 codificanti 13 residui amminoacidici. Sono stati analizzati i siti di *splicing*, il *branch site*, il tratto polipirimidinico e le ipotetiche regioni che interagiscono con i componenti dello spliceosoma U1, U5, e U2 presenti nelle regioni APS0, APS1, e APS2. Le sequenze canoniche di *splicing* sono presenti in tutte le regioni APS ma soltanto i due esoni R2 ed R1 sono stati ritrovati nel trascritto maturo, pertanto codificati nella proteina mentre l'esone R0, la sequenza omologa presente in APS1, è assente.

Basandosi sul confronto con la rispettiva controparte, sono stati identificate numerose sostituzioni nelle regioni R2, R1 ed R0 in *C. hamatus* (Fig. 51). La sequenza R0, che non viene inclusa nel trascritto maturo, mostra una sola sostituzione rispetto alla sua

controparte nella regione S mentre le sequenze R1 ed R2, che sono usate come esoni addizionali, mostrano rispettivamente 8 e 12 sostituzioni (Fig. 52). Diciannove delle venti sostituzioni in R1 ed R2 interessano la prima e/o la seconda posizione nel codone e risultano non sinonime. È interessante, inoltre, notare che 5 delle 20 sostituzioni introducono un codone codificante una cisteina, suggerendo una preferenza per questo residuo amminoacidico nella proteina sintetizzata.

S	GAC	AAA	CGT	GGT	GCC	CTT	TAT	CCA	TTC	TTC	ATA	CGT	GAT
R0	GAC	AAA	CTT	GGT	GCC	CTT	TAT	CCA	TTC	TTC	ATA	CGT	GAT
R1	GAC	AAA	CTT	GGT	GTC	CTT	TGT	CCA	TGC	CAC	ACA	TGT	GAT
R2	GAC	AAA	GTT	GGT	GTC	CAT	CGT	CCA	TGC	CAC	ATC	ACT	GAT

Figura 52. Allineamento di R0, R1, R2 con la sequenza inversa complementare della controparte nella regione S. Le 21 basi, che in R0, R1, R2 differiscono da quelle nella stessa posizione nella regione S, sono mostrate in grassetto. I codoni di R1 ed R2 sono ombreggiati alternatamente; in queste regioni 19/20 sostituzioni interessano la prima e/o la seconda posizione del codone. Le basi che introducono un codone codificante cisteina sono in corsivo.

Per verificare che la presenza di regioni invertite e ripetute all'interno del locus genico delle Ig non fosse una peculiarità della specie in esame ma una caratteristica comune alle specie antartiche, è stata sequenziata la regione genomica compresa tra gli esoni CH2-TM di un altro teleosteo antartico, *Cygnodraco mawsoni*. In tale specie è stata osservata la presenza di ben 4 regioni aventi un'alta omologia con la regione APS1, contenenti gli esoni da 39 nt codificanti per le ripetizioni da 13 residui amminoacidici dell'EMPD, ed una regione analoga ad APS0. La sequenza genomica ottenuta è mostrata in Appendice come sequenza A9. Anche in questo caso è presente una sequenza di 39 nt nella regione analoga ad APS0 che contiene gli stessi siti possibili di *splicing* delle sequenze da 39 nt nelle altre 4 regioni ma non è espressa.

Questi risultati suggeriscono che l'inserzione di sequenze ripetute e invertite nel locus delle Ig dei teleostei antartici è una caratteristica comune e una possibile conseguenza dell'adattamento al freddo.

### **3C.5. Studi sui frammenti di RNA complementari ad una porzione del locus delle Ig in *Chionodraco hamatus* e *Bovichtus diacanthus***

#### *Sintesi di frammenti di RNA*

Per comprendere se l'inserzione di circa 2.000 nucleotidi nel locus della catena pesante delle Ig di *C. hamatus* fosse coinvolto nello *splicing* atipico per la forma di membrana delle immunoglobuline, è stato sintetizzato un frammento di RNA riprodotto la sequenza genomica comprendente la regione CH2-TM. Il frammento genomico è stato clonato in un vettore plasmidico contenente un sito di legame per la RNA polimerasi T7 e la trascrizione è stata condotta usando come enzima una RNA polimerasi T7. Il frammento ottenuto comprendeva 4.564 nt, il peso predetto con un *tool* disponibile in rete era di 1.548 kDa, la composizione in GC è di 39,1%, la concentrazione per mole di molecola di RNA era di 0.147  $\mu$ M mentre la concentrazione calcolata per mole di nucleotide era di 0.671 mM.

Un analogo protocollo è stato usato per sintetizzare un frammento di RNA riprodotto parte del locus delle Ig della specie non antartica *B. diacanthus*, con l'obiettivo di comprendere le differenze a livello trascrizionale fra i due tipi di *processing*. Il

frammento comprendeva 2.073 nt, il suo peso molecolare predetto era di 597,8 kDa, la composizione in GC è di 35,6%, la concentrazione per mole di RNA era di 34.4  $\mu$ M mentre quella per mole di nucleotide era di 71.3 mM.

Le sequenze delle due specie sono mostrate rispettivamente in appendice come sequenze A3 e A4.

È stato inoltre sintetizzato un frammento di RNA riprodotto la regione genomica di *C. hamatus* comprendente l'esone CH2 e l'esone CH4, per utilizzarlo come elemento di riferimento nelle successive misure. La sua sequenza è mostrata in appendice in sequenza A5, dove sono evidenziati in grassetto le regioni complementari ai primers oligonucleotidici utilizzati. Tale frammento comprendeva 1.142 nt, il suo peso molecolare predetto era di 329,26 kDa, la composizione in GC è di 40,9%, la concentrazione per molecola era di 1,242  $\mu$ M mentre quella calcolata per mole di nucleotide era di 1,42 mM.

### **3C.5.1. Predizione di struttura secondaria**

Per determinare se vincoli strutturali possono essere coinvolti nella perdita dell'esone CH3 e del putativo esone RC nel trascritto maturo delle Ig di *C. hamatus* sono stati effettuati studi di predizione di struttura. La struttura secondaria delle molecole di RNA a diverse temperature è stata predetta utilizzando il *tool* computazionale *mfold* (Zuker M *et al.*, 2003). Le temperature di analisi sono 0°C e 25°C. A bassa temperatura, più prossima a quella fisiologica per le specie antartiche, la struttura a più bassa energia predetta per la molecola di RNA di *C. hamatus* presenta una regione a doppio filamento molto

estesa stabilizzata da un contributo di energia libera pari a -3.830 kcal/mole (Fig. 53). La struttura a doppia elica coinvolge la fine dell'esone CH2 e la fine dell'esone CH4 su un filamento e la regione APS0 sull'altro, per un totale di 1.980 nucleotidi coinvolti (vedi sequenza A6 in appendice). I nucleotidi impegnati nel doppio filamento sono  $154 \div 1.144$  e  $1.804 \div 2.796$  mentre il *loop* fra i due filamenti è composto da 658 nt. Il contributo di energia libera calcolato per la sola regione a doppio filamento è di -2.49 kcal/mole, più del 65% dell'energia totale. La sequenza del primo filamento include il sito accettore di *splicing* dell'esone CH3 e della sequenza R0, che in questo modo non sono disponibili ad un eventuale legame dei componenti del complesso dello spliceosoma. Questo impedimento strutturale potrebbe essere il motivo per il quale la forma di membrana delle immunoglobuline è mancante dell'esone CH3 e il possibile esone R0 non è espresso.

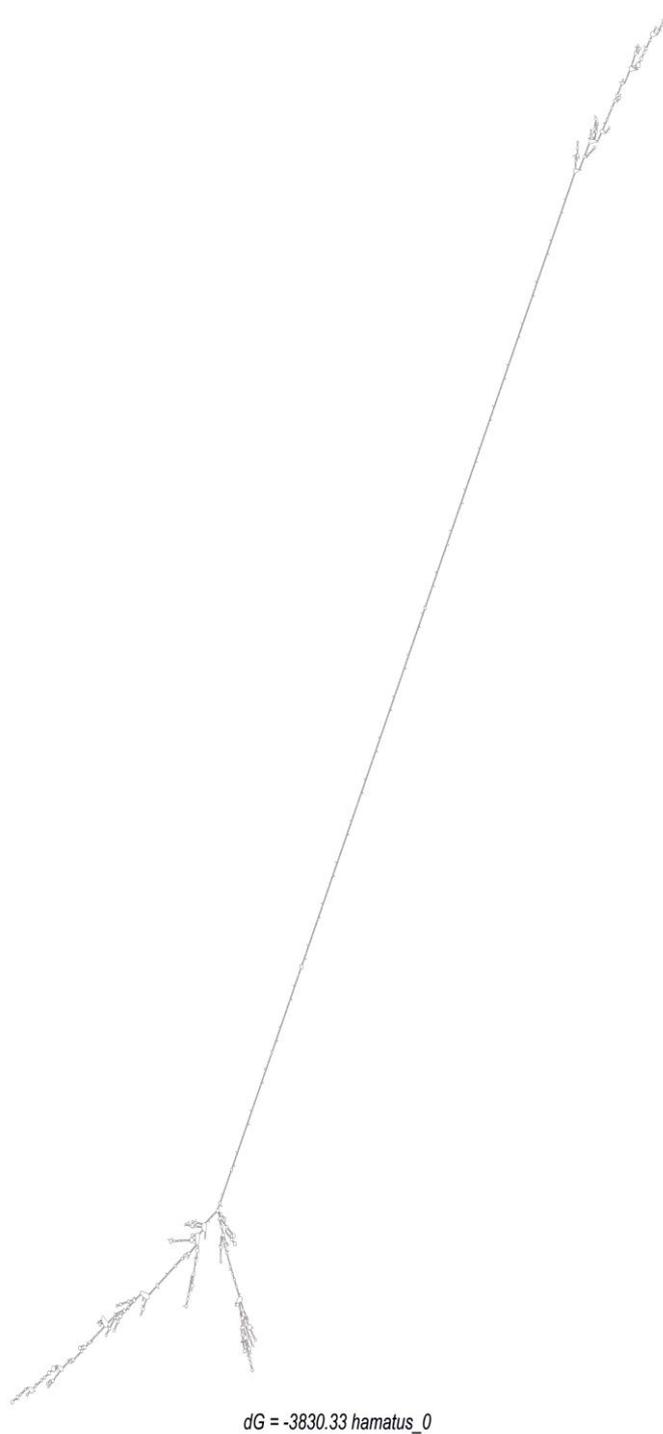


Figura 53. La struttura ad energia più bassa predetta per l'RNA di *C. hamatus* a 0 °C con il programma *mfold*. La presenza di una regione a doppio filamento molto estesa è evidente.

A temperature più alte, il risultato della predizione di struttura secondaria mostra la possibilità che il frammento di RNA di *C. hamatus* può presentarsi in due diverse conformazioni, legate alla capacità delle regioni APS1 e APS2 di formare una struttura a doppio filamento con la sua controparte nella regione CH3 (Fig. 54 e Fig. 55). La prima conformazione risultante dalla predizione a 25°C presenta una struttura a doppio filamento in cui la sequenza impegnata nel primo filamento coinvolge gli stessi nucleotidi del primo filamento della struttura a bassa temperatura, mentre il secondo filamento coinvolge una parte della regione APS0 e l'intera regione APS1 (vedi sequenza A7 in appendice). Il numero di nucleotidi coinvolti è lo stesso, 990 nt per filamento, e sono 154 ÷ 1144 su un filamento e 1.804 ÷ 2.554 e 2.965 ÷ 3.203 sull'altro. Il *loop* ha la stessa lunghezza di quello della struttura predetta a bassa temperatura ma in questa conformazione è presente un *bulge* al centro del secondo filamento composto da 409 nt (2.555 ÷ 2.964). Il contributo di energia libera calcolato per questa struttura è di -2.357 kcal/mol.

La seconda conformazione predetta ad alta temperatura mostra anch'essa una struttura a doppio filamento che coinvolge 1.880 nt, i nucleotidi 154 ÷ 1.144 su un filamento e 1.804 ÷ 2.554 e 3.361 ÷ 3.617 sull'altro (vedi sequenza A8 in appendice). In questo caso il secondo filamento coinvolge la prima parte della regione APS0 e l'intera regione APS2. La lunghezza del *loop* tra i due filamenti è la stessa (658 nt) mentre il *bulge* è più grande e comprende 805 nt. Il contributo di energia libera calcolato è di -2443 kcal/mole.

La presenza del *bulge* nelle due conformazioni a 25°C le rende meno stabili (la diminuzione di energia è di circa -1.473 kcal/mole) e pertanto termodinamicamente sfavorite.

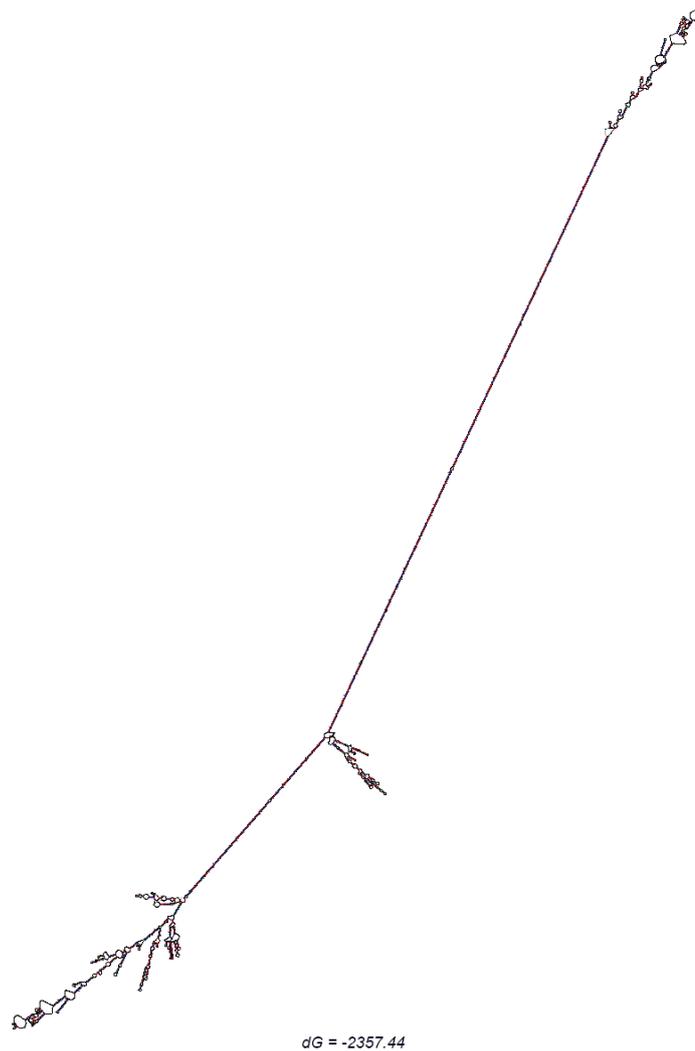
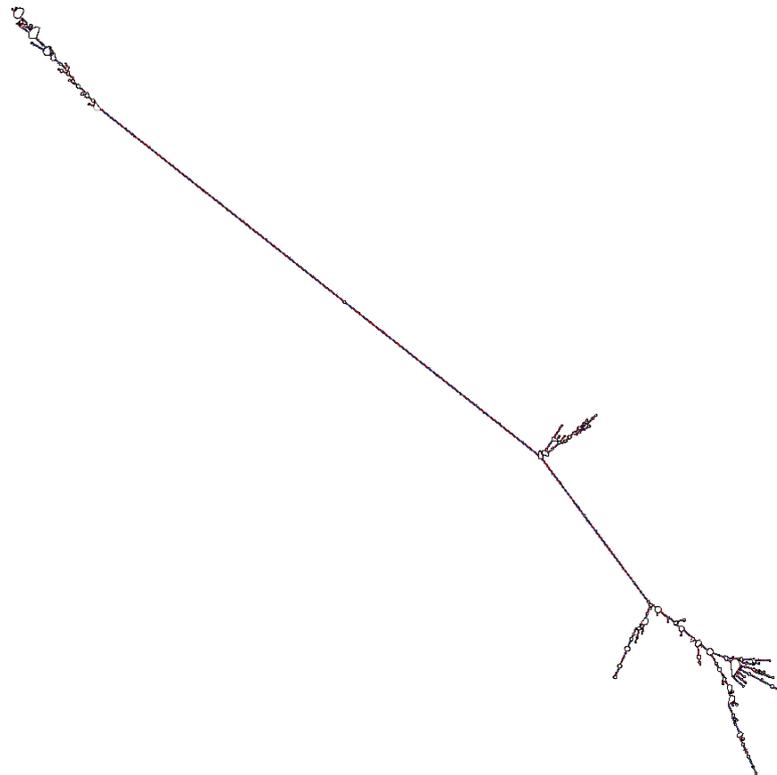


Figura 54. La prima delle due conformazione predette per il frammento di RNA di *C. hamatus* a 25 °C. La struttura a doppio filamento è presente anche a temperature alte ed è visibile un *bulge* al centro del secondo filamento.



$dG = -2443.15$

Figura 55. La seconda delle due conformazione predette per il frammento di RNA di *C. hamatus* a 25 °C. La struttura a doppio filamento è presente anche a temperature alte ed è visibile un *bulge* al centro del secondo filamento.

Per la specie non Antartica, *B. diacanthus*, le conformazioni predette ad alta e bassa temperatura sono meno strutturate: non sono evidenti strutture a doppio filamento molto estese né a bassa né ad alta temperatura a causa dell'assenza di regioni di sequenze inverse complementari. La struttura del frammento di RNA sembra essere globulare e il contributo di energia associato a questa conformazione è di soli -1.073 kcal/mole (Fig. 56).

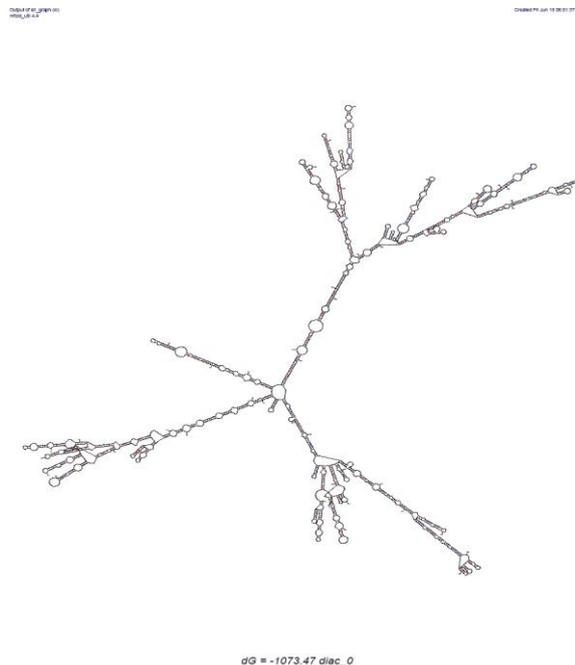
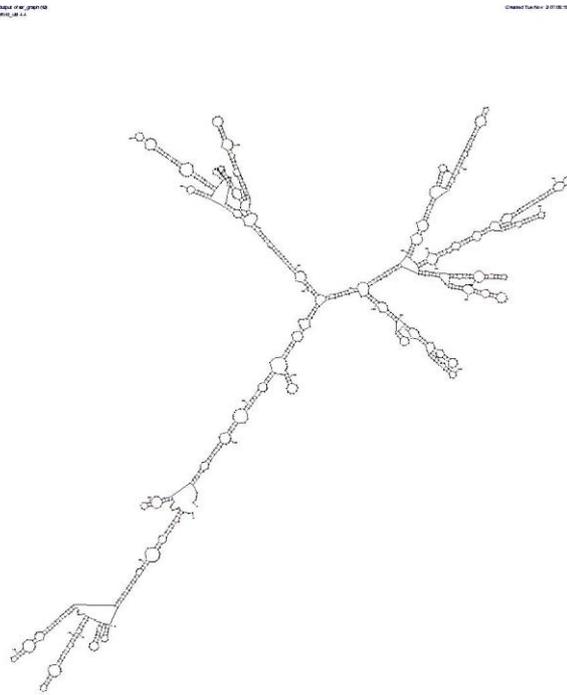


Figura 56. Predizione di struttura secondaria condotta sulla sequenza di RNA di *B. diacanthus* corrispondente alla regione genomica tra gli esoni CH2 e TM. È evidente l'assenza di strutture a doppio filamento esteso.

È stata inoltre predetta la struttura del frammento di RNA di *C. hamatus* comprendente gli esoni CH2-CH4: si suppone che tale regione non si strutturi a doppio filamento ed è preso come riferimento di una struttura a singolo filamento. La predizione mostra l'assenza del doppio filamento sia a basse che alte temperature ed il contributo di energia libera è molto più basso di quello delle

conformazioni di *C. hamatus* e *B. diacanthus* (-604,35 kcal/mole a 0°C)(Fig. 57).



*dG = -604.35 single strand 0*

Figura 57. Predizione di struttura del frammento di RNA di *C. hamatus* compresa tra gli esoni CH2-CH4 condotta a 0 °C. L'RNA è meno strutturato ed è assente una struttura a doppio filamento esteso.

### 3C.5.2. Dicroismo Circolare

Il Dicroismo Circolare (CD) è una tecnica spettroscopica comunemente usata per la caratterizzazione in soluzione di proteine ed acidi nucleici. È una tecnica in grado di fornire informazioni per una data struttura secondaria di peptidi, proteine ed acidi nucleici. È particolarmente utile per monitorare i cambiamenti strutturali in seguito a variazioni delle condizioni ambientali come temperatura, forza ionica e pH.

Gli spettri CD per le varie specie in esame sono stati registrati a diverse temperature (0, 10, 20, 30, 40, 50, 60, 70, 80, 90 e 100°C),

in un intervallo fra 220-320 nm, per poter seguire eventuali cambiamenti conformazionali causati dall'aumento della temperatura.

Le misure di CD sono state condotte prima sul frammento di RNA di *C. hamatus* riprodotte la regione genomica compresa tra gli esoni CH2-CH4. La sovrapposizione degli spettri è mostrata in figura 58. Lo spettro registrato a 0 °C presenta 3 picchi: uno molto intenso e positivo a 268 nm, uno positivo a 226 nm ed uno negativo a 236 nm.

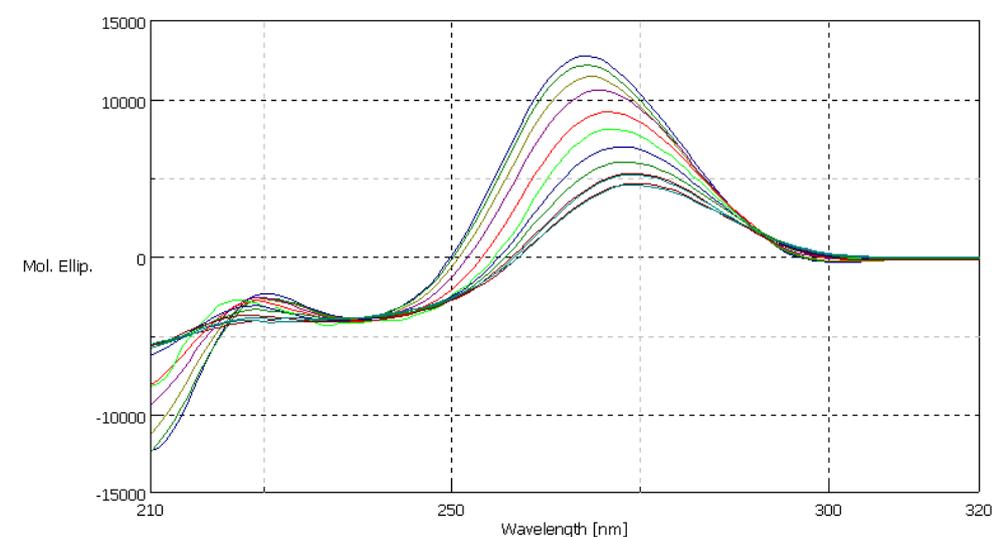


Figura 58. Sovrapposizione degli spettri CD del frammento di RNA di *C. hamatus* riprodotte la regione genomica compresa tra gli esoni CH2-CH4 a differenti temperature.

A temperature maggiori, il massimo dell'intensità del picco positivo si sposta in una regione a lunghezze d'onda maggiori (fino a 274 nm) e la sua intensità diminuisce notevolmente, mentre i picchi a 236 e 226 nm scompaiono progressivamente. Seguendo l'andamento dell'intensità a 268 nm in funzione della temperatura non è stato possibile identificare una temperatura di fusione del sistema (Fig. 59). A 108 °C è presente un unico picco a 274 nm che

può essere usato come riferimento per una struttura a singolo filamento.

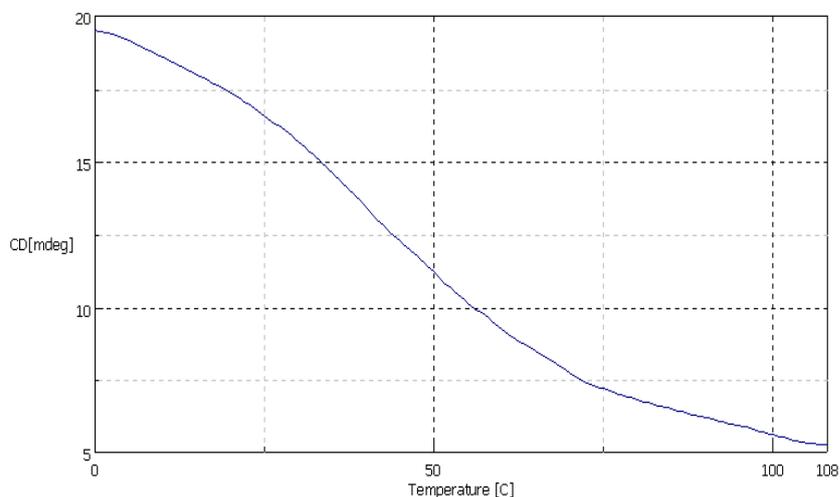


Figura 59. Curva termica che rappresenta l'andamento dell'intensità del segnale dicroico a 267 nm a in funzione dell'aumento della temperatura del frammento di RNA di *C. hamatus* riprodotto la regione genomica CH2-CH4.

Le stesse misure sono state condotte sul campione di RNA di *C. hamatus* corrispondente alla regione genomica di interesse. Lo spettro a più bassa temperatura mostra un picco molto intenso a 267 nm ed altri tre picchi meno intensi, due negativi a 210 nm e 238 nm ed uno positivo a 226 nm (Fig. 60).

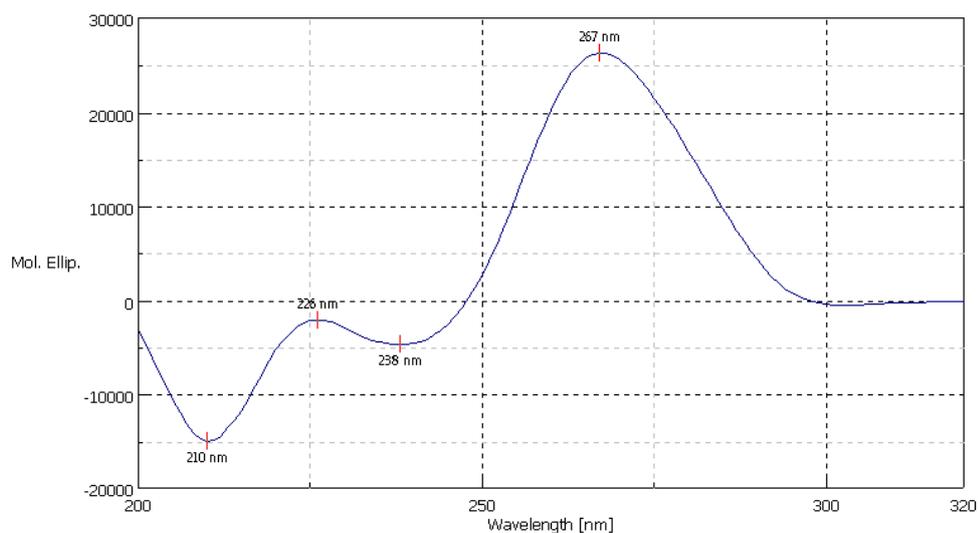


Figura 60. Spettro CD dell'RNA di *C. hamatus* a 0 °C. È evidente un picco positivo molto intenso a 267 nm, due picchi negativi rispettivamente a 210 e 238 nm ed un picco positivo meno intenso a 238 nm.

In seguito all'aumento della temperatura la lunghezza d'onda del massimo del picco più intenso si sposta verso valori maggiori (da 267 nm a 0°C fino a 275 nm a 108°C). Ciò potrebbe essere imputabile alla perdita del contributo di doppio filamento alla struttura della molecola (Fig. 61).

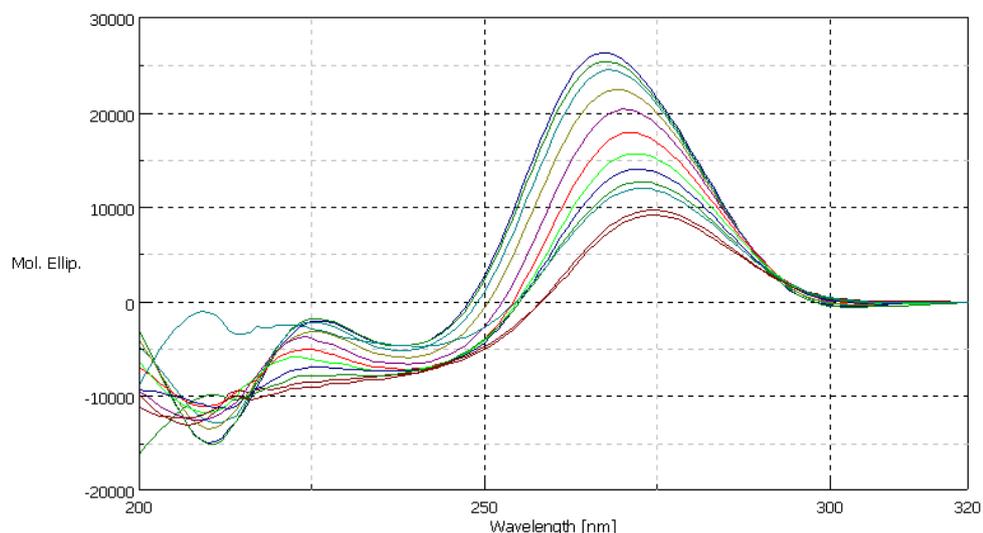


Figura 61. Sovrapposizione degli spettri CD del frammento di RNA di *C. hamatus* corrispondente alla regione genomica compresa tra gli esoni CH2-TM. Il massimo dell'intensità del picco positivo più intenso si sposta in regioni dello spettro a lunghezza d'onda maggiori, da 226 nm a 0 °C 275 nm a 108 °C. I due picchi a 238 e 226 nm scompaiono a temperature maggiori.

La denaturazione termica è stata seguita registrando l'intensità del segnale a 267 nm in funzione dell'aumento della temperatura in un intervallo fra 0 e 108 °C (Fig. 62). La curva non presenta un andamento sigmoidale, pertanto non è possibile identificare una temperatura di fusione, temperatura alla quale metà delle molecole di RNA sono denaturate. Ciò suggerisce che il processo di denaturazione potrebbe essere un processo non cooperativo e che la perdita di struttura secondaria interessera soltanto le estremità della molecola e non la struttura a doppio filamento. La sovrapposizione degli spettri a tutte le temperature analizzate è mostrato in figura 61.

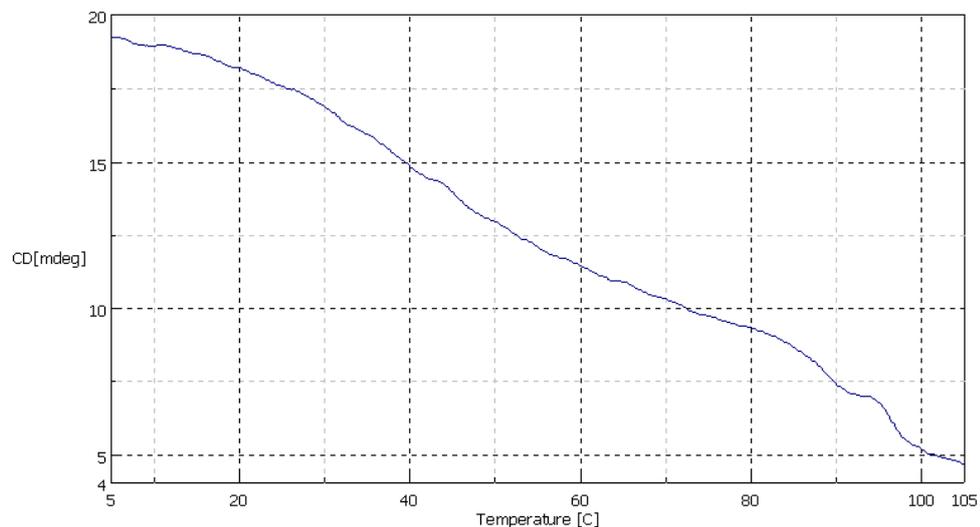


Figura 62. Denaturazione termica del frammento di RNA di *C. hamatus* corrispondente alla regione genomica compresa tra gli esoni CH2-TM. L'andamento della curva non è sigmoidale pertanto non è stato possibile identificare una temperatura di denaturazione.

Lo spettro CD registrato a 0 °C sul frammento di RNA di *B. diacanthus* riprodotto nella regione genomica di interesse mostra quattro picchi: un picco positivo molto intenso a 267 nm, un picco positivo meno intenso a 223 nm e due picchi negativi rispettivamente a 237 nm e 211 nm.

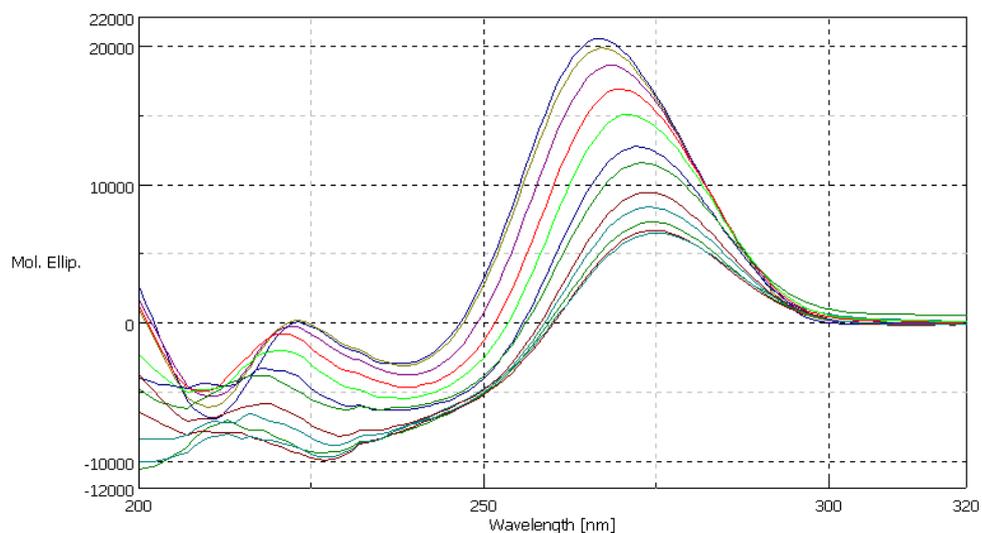


Figura 63. Sovrapposizione degli spettri CD registrati sul frammento di RNA di *B. diacanthus* corrispondente alla regione genomica compresa tra gli esoni CH2-TM a diverse temperature.

All'aumentare della temperatura la lunghezza d'onda del massimo del picco positivo più intenso si sposta verso valori maggiori e l'intensità del picco diminuisce. L'intensità del picco a 237 nm aumenta con la temperatura mentre l'intensità del picco a 211 nm diminuisce (Fig. 63).

Confrontando lo spettro a 108 °C dell'RNA di *C. hamatus* riprodotto la regione genomica CH2-TM con quello riprodotto la regione genomica CH2-CH4 si osserva che in entrambe i casi i picchi meno intensi scompaiono ad alte temperature (Fig. 63). È interessante notare che l'intensità del picco a 275 nm riprodotto la regione genomica CH2-TM è notevolmente maggiore di quello riprodotto la regione CH2-CH4, ad indicare un maggior grado di strutturazione di tale molecola anche a temperature così elevate.

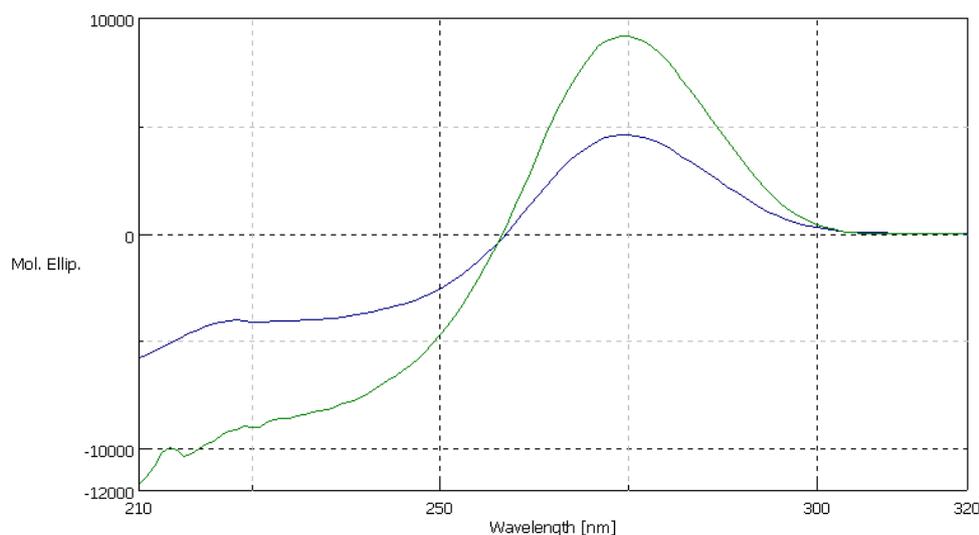


Figura 64. Sovrapposizione degli spettri CD registrati a 108 °C per l'RNA di *C. hamatus* corrispondente alla regione genomica CH2-TM (in verde) e CH2-CH4 (in blu).

### 3C.5.3. Microscopia a forza atomica

Allo scopo di verificare l'attendibilità delle predizioni di struttura secondaria sui frammenti di RNA sintetizzati sono stati condotti studi su tali molecole mediante microscopia a forza atomica (AFM), una tecnica sperimentale che permette di ottenere immagini di superfici con risoluzione di pochi nanometri.

Le analisi mediante AFM sul frammento di RNA di *C. hamatus* corrispondente alla regione genomica tra gli esoni CH2 e TM sono state effettuate in due condizioni diverse: in un primo caso la soluzione contenente l'RNA è stata riscaldata per 5 min a 90°C e poi bloccata in ghiaccio. Le immagini ottenute al microscopio relative a questo campione mostrano la presenza di molecole con una struttura filamentosa molto estesa che è assumibile ad una regione a doppio filamento di RNA alle cui estremità sono presenti due regioni globulari di diversa grandezza (Fig. 65). Sono state misurate 30 molecole e la lunghezza media della regione a doppio filamento è

risultata essere  $244 \pm 16$  nm. È stato mostrato in letteratura (Arnott JN *et al.*, 1973) che il passo per base nell'elica A (la classica struttura dell'RNA a doppio filamento) è di  $0,29 \pm 0,03$  nm, pertanto il numero di nucleotidi coinvolti nelle strutture osservate al microscopio risulta essere 841 per filamento. Tale risultato è compatibile con quello predetto dal *tool* computazionale *mfold* secondo il quale il numero di nucleotidi a doppio filamento coinvolti sono di 990 per filamento.

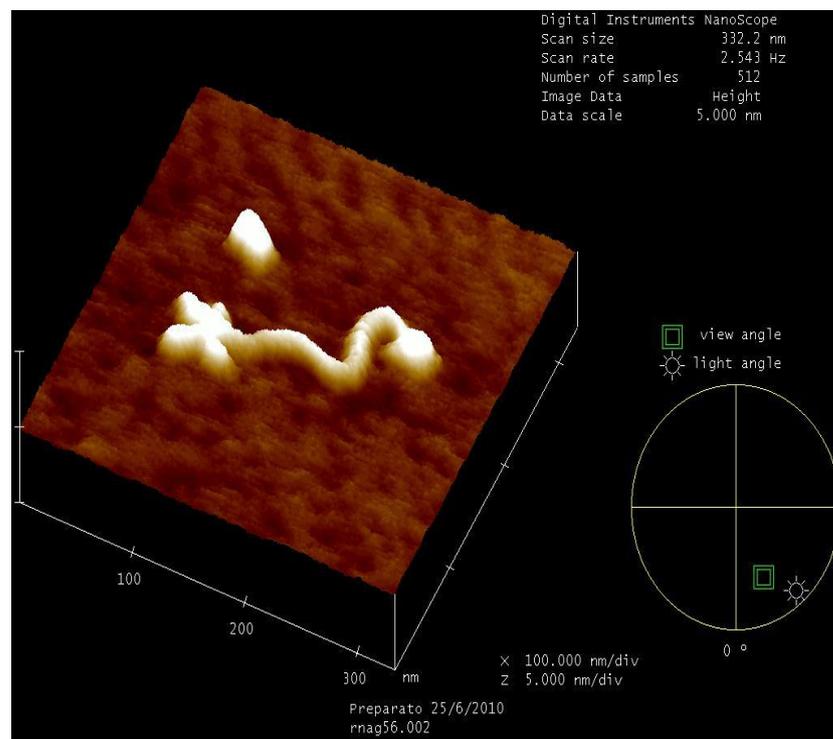


Figura 65. Immagine al microscopio a forza atomica di una molecola di RNA di *C. hamatus* corrispondente alla regione genomica compresa tra gli esoni CH2 e TM. La molecola è stata depositata sulla superficie di mica dopo un trattamento a  $95^{\circ}\text{C}$  per 5 min seguito da un'incubazione in ghiaccio. L'oggetto di forma globulare presente nell'immagine è una contaminazione proteica.

Il secondo tipo di trattamento previsto per il frammento di RNA di *C. hamatus* comportava un riscaldamento a  $95^{\circ}\text{C}$  per 5' e un lento raffreddamento fino al raggiungimento della temperatura ambiente. In tal caso erano attese strutture compatibili con le previsioni di

struttura condotte a 25 °C. Le immagini al microscopio mostrano la presenza di molecole aventi due regioni a doppio filamento interrotte da un *bulge*, alle cui estremità sono presenti regioni globulari (Fig. 66). Le misure condotte su 10 molecole di RNA mostrano una regione complessiva a doppio filamento di  $238 \pm 13$  nm per un totale di 821 nucleotidi impegnati nella struttura a doppio filamento.

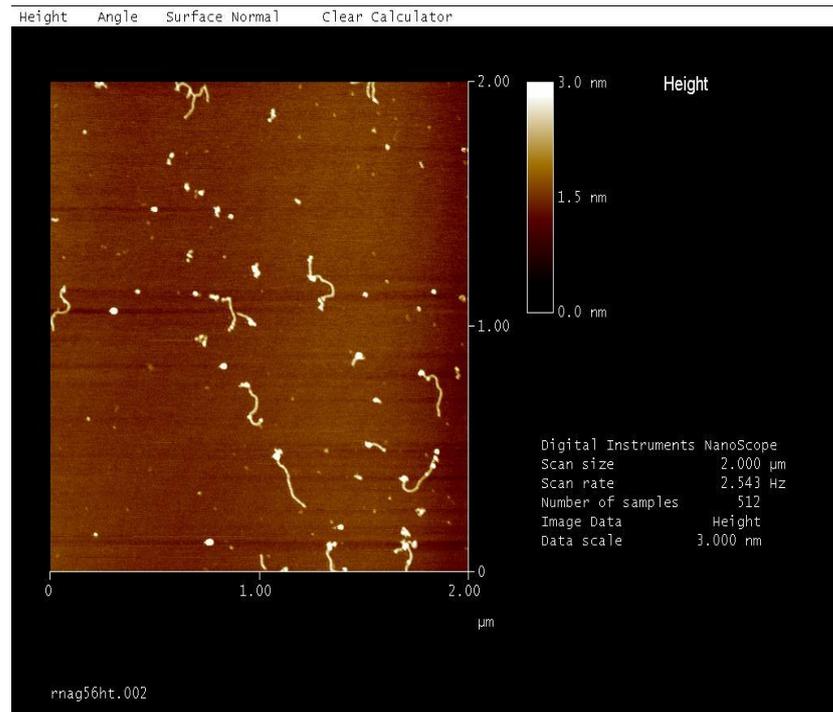


Figura 66. Immagine al microscopio a forza atomica di molecole di RNA di *C. hamatus* corrispondente alla regione genomica compresa tra gli esoni CH2 e TM. Prima della deposizione, il campione è stato riscaldato a 95 °C per 5' e successivamente raffreddato molto lentamente fino alla temperatura ambiente.

Con lo stesso approccio di microscopia è stato studiato il frammento di RNA di *B. diacanthus* riprodotto nella regione genomica compresa tra gli esoni CH2 e TM. In questo caso le immagini ottenute alla microscopia a forza atomica mostrano molecole quasi completamente globulari, in cui le regioni a doppio filamento sono difficili da distinguere (Fig. 67). Tale struttura è compatibile con la previsione condotta con il metodo computazionale

secondo il quale non sarebbero presenti regioni a doppio filamento esteso.

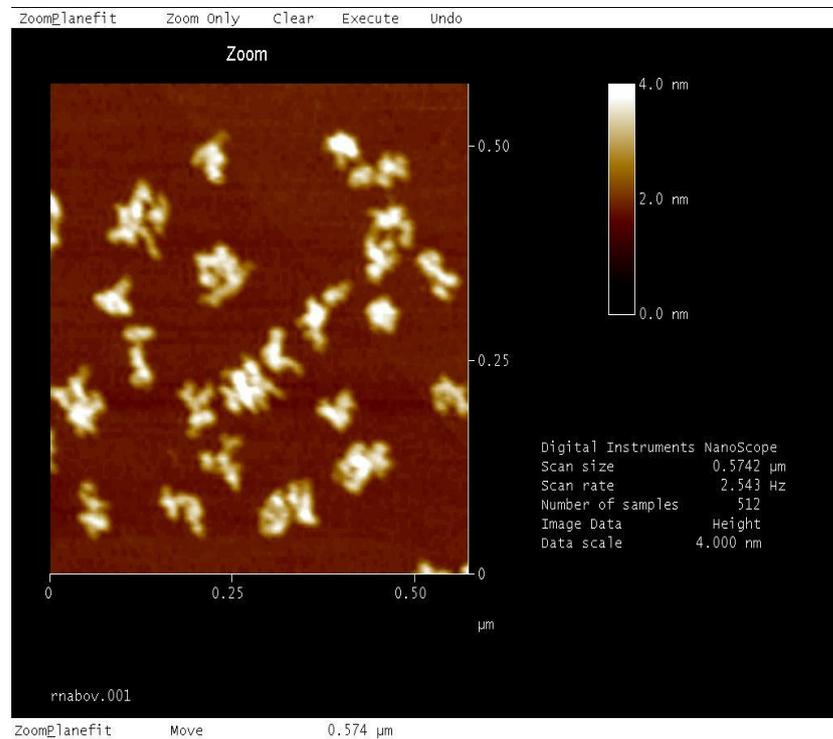


Figura 67. Immagine al microscopio a forza atomica di molecole di RNA di *B. diacanthus* corrispondente alla regione genomica compresa tra gli esoni CH2 e TM. Dall'immagine è possibile notare l'assenza di strutture con un doppio filamento esteso mentre le molecole mostrano tutte una struttura di tipo globulare.

## **4. CONCLUSIONE E DISCUSSIONE**

La molecola che è stata oggetto di studio del mio Dottorato di Ricerca in Scienze Chimiche è l'Immunoglobulina di teleostei antartici. È stata prestata attenzione ad entrambe le catene, leggera e pesante, delle Ig di più specie di Nototenioidi.

La scelta di studiare molecole appartenenti a specie antartiche diverse ha permesso di generalizzare i risultati e l'analisi di specie ancestrali attualmente viventi in mari non antartici hanno indicato che i risultati sono relazionabili all'evoluzione adattativa.

Il mio lavoro si è inserito all'interno di linee di ricerca già avviate ed ha contribuito a concludere progetti precedenti; tali lavori infatti sono stati oggetto di pubblicazioni su riviste scientifiche internazionali.

Il lavoro sulle catene leggere delle Ig di *T. bernacchii* ha permesso di identificare caratteristiche della struttura primaria che sembrano essere una conseguenza dell'adattamento al freddo di tale molecola: l'elevato numero di siti di fosforilazione e glicosilazione sembra essere una possibile soluzione al problema della bassa solubilità delle proteine a temperature prossime allo zero. L'analisi comparata di tutte le sequenze di catene leggere di teleostei ha permesso di identificare quando nell'evoluzione delle specie, si è verificata la distinzione dell'isotipo L1 in due sottoisotipi, uno dei quali contenente un'inserzione di un microsatellite.

L'analisi della regione TM delle Ig delle specie antartiche e di specie appartenenti a classi di vertebrati diverse ha permesso di conseguire risultati molto interessanti. Infatti, è stato possibile identificare in tale regione un motivo universale conservato in tutti gli isotipi di catene pesanti di tutte le classi di vertebrati; tale motivo è caratteristico per la molecola Ig ed è statisticamente assente nei tratti in  $\alpha$ -elica di tutte le altre proteine di membrana. Gli studi di MD

hanno suggerito la funzione di alcuni residui di tale motivo indicando che l'interazione tra due residui di fenilalanina sia la forza trainante dell'associazione delle catene pesanti delle Ig all'interno della membrana cellulare. Questo tipo di associazione non è stato descritto in altre proteine di membrana e potrebbe essere coinvolto nel complesso meccanismo di trasduzione che coinvolge il BCR. I risultati conseguiti potrebbero essere ampliati mediante ulteriori studi di MD della regione TM: a tale scopo potrebbe essere utile studiare l'associazione dell'Ig con l'eterodimero trasducente  $Ig\alpha$ - $Ig\beta$  in ambienti lipidici più simili a quello fisiologico, quali potrebbero essere doppi strati lipidici con composizione simile a quella del *lipid raft*, il microdominio di membrana in cui il BCR esplica la sua funzione.

I risultati più recenti riguardano lo *splicing* atipico del trascritto primario codificante la forma di membrana delle Ig. L'analisi genomica del locus codificante la catena pesante di *C. hamatus* ha svelato una struttura del tutto particolare. La presenza contemporanea di una regione nucleotidica e della sua antiparallela ripetuta un numero variabile di volte a seconda della specie, sembra essere una caratteristica unica dei teleostei antartici. Tale caratteristica è la causa del meccanismo di processamento del trascritto primario che abbiamo evidenziato. L'analisi chimico-fisica di un RNA sintetico corrispondente alla regione del trascritto primario interessata allo *splicing*, ci ha permesso di individuare nella struttura secondaria a doppia elica dell'RNA, la causa del meccanismo di processamento atipico. Tali studi potrebbero proseguire caratterizzando l'RNA nucleare estratto da cellule di rene cefalico di *C. hamatus* mediante ulteriori metodi chimico-fisici mirati ad evidenziarne la stabilità, come la calorimetria differenziale a scansione. Questo potrebbe costituire una nuova evidenza del ruolo giocato dalla temperatura nello stabilizzare la struttura dell'RNA e di

conseguenza modulare le funzioni cellulari. Infine la purificazione del peptide ricombinante potrebbe aprire la strada a caratterizzazioni strutturali che spiegherebbero le conseguenze dello *splicing* atipico a livello di struttura della proteina.

La mia ricerca ha dimostrato la validità di usare approcci differenziati allo studio di una molecola: sono stati usati metodi di biochimica e biologia molecolare; sono state condotte caratterizzazioni chimico-fisiche ed utilizzati approcci computazionali sia per l'analisi delle sequenze nucleotidiche ed amminoacidiche che per le elaborazioni strutturali. L'obiettivo era quello di identificare specificità, relazionabili all'adattamento evolutivo non solo a livello proteico ma anche a livello genomico e trascrizionale.

Questi risultati hanno confermato nozioni già consolidate che riguardano l'aumento di solubilità delle proteine di organismi adattati alle basse temperature, aumento della flessibilità tra i domini che le compongono; inoltre hanno anche evidenziato specificità della struttura del gene e dei meccanismi in cui è coinvolto.

## **5. BIBLIOGRAFIA**

Abelli L, Coscia MR, De Santis A, Zeni C, Oreste U. Evidence for hepato-biliary transport of immunoglobulin in the antarctic teleost fish *Trematomus bernacchii*. *Dev Comp Immunol*. 2005;29(5):431-42.

Arnold JN, Wormald RB, Rudd PM, Dwek RA. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol*. 2007;25:21-50.

Arnott S, Hukins DW, Dover SD, Fuller W, Hodgson AR. Structures of synthetic polynucleotides in the A-RNA and A'-RNA conformations: x-ray diffraction analyses of the molecular conformations of polyadenylic acid--polyuridylic acid and polyinosinic acid--polycytidylic acid. *J Mol Biol*. 1973;5;81(2):107-22.

Artéro S and Lefranc MP. The teleostei immunoglobulin light IGL1 and IGL2 V, J and C genes. *Exp Clin Immunogenet*. 2000;17(3):162-72.

Artéro S and Lefranc MP. The Teleostei immunoglobulin heavy IGH genes. *Exp Clin Immunogenet*. 2000;17(3):148-61.

Atkinson TP, Hall CG, Goldsmith J, Kirkham PM. Splice variant in TCRzeta links T cell receptor signaling to a G-protein-related signaling pathway. *Biochem Biophys Res Commun*. 2003;310:761-6.

Batista FD, Iber D, Neuberger MS. B cells acquire antigen from target cells after synapse formation. *Nature*. 2001;411:489-94.

Berendsen HJC, Postma JPM, van Gusteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external heat bath. *J Chem Phys*. 1984;81:3684-90.

Berendsen HJC, Postma JPM, van Gusteren WF, Hermans J. Interaction Models for Water in Relation to Protein Hydration. Reidel Dordrecht. 1981; 331-42.

Bestagno M, Vangelista L, Mandiola PA, Mukherjee S, Sepúlveda J, Burrone OR. Membrane immunoglobulins are stabilized by interchain disulfide bonds occurring within the extracellular membrane-proximal domain. *Biochemistry*. 2001;40:10686-92.

Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*. 1999;17;294(5):1351-62.

Blum JH, Stevens TL, De Franco AL. Role of the mu immunoglobulin heavy chain transmembrane and cytoplasmic domains in B cell antigen receptor expression and signal transduction. *J Biol Chem*. 1993;268:27236–45.

Brandt JP, Patapoff TW, Aragon SR. Construction, MD simulation, and hydrodynamics validation of an all-atom model of a monoclonal IgG antibody. *Biophysical Journal*. 2010;99:905-13.

Burley SK and Petsko GA. Amino-aromatic interactions in proteins. *FEBS Lett*. 1986;203:139–43.

Call ME and Wucherpfennig KW. Common themes in the assembly and architecture of activating immune receptors. *Nat Rev Immunol*. 2007;7(11):841-50.

Carrasco B, Garcia de la Torre J, Davis KG, Jones S, Athwal D, Walters C, Burton DR, Harding SE. Crystallohydrodynamics for solving the hydration problem for multi-domain proteins: open physiological conformations for human IgG. *Biophys Chem*. 2001;28;93(2-3):181-96.

Carrasco YR, Fleire SJ, Cameron T, Dustin ML, Batista FD. LFA-1/ICAM-1 interaction lowers the threshold of B cell activation by facilitating B cell adhesion and synapse formation. *Immunity*. 2004;20:589–99.

Chen L, DeVries AL, Cheng C-HCE. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc Natl Acad Sci USA*. 1997;94:3811–6.

Coscia MR, Borriello A, Oreste U. Plasma and bile antibodies of the teleost *Trematomus bernacchii*, specific for the nematode *Pseudoterranova decipiens*. *Dis Aquat Org*. 2000;41:37-42.

Coscia MR, Morea V, Tramontano A, Oreste U. Analysis of a cDNA sequence encoding the immunoglobulin heavy chain of the Antarctic teleost *Trematomus bernacchii*. *Fish Shellfish Immunol*. 2000;10:343-57.

Coscia MR and Oreste U. Humoral immune response of Antarctic fish to the nematode *Contracaecum osculatum*. *Fish & Shellfish Immunology*. 1998;8:295-302.

Coscia MR and Oreste U. Exploring Antarctic teleost immunoglobulin genes. *Marine Genomics*. 2009;2:19-25.

Coscia MR, Varriale S, De Santi C, Giacomelli S, Oreste U. Evolution of the Antarctic teleost heavy chain gene locus. *Mol Phylogenet Evol*. 2010;55:226-33.

Dal Porto JM, Gauld SB, Merrell KT, Mills D, Pugh-Bernard AE, Cambier J. B cell antigen receptor signaling 101. *Mol Immunol*. 2004;41:599–613.

Danilova N, Bussmann J, Jekosch K, Steiner LA. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat Immunol*. 2005;6:295-302.

Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys*. 1993;98:10089–92.

Eastman JT. The nature of diversity of Antarctic fishes. *Polar Biol.* 2005;28: 93-107.

Fleire SJ, Goldman JP, Carrasco YR, Weber M, Bray D, Batista FD. B cell ligand discrimination through a spreading and contraction response. *Science.* 2006;312:738–41.

Gazit E. A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J.* 2002;16:77–83.

Hansen JD, Landis ED, Phillips RB. Discovery of a unique Ig heavy chain isotype in rainbow trout : implications for a novel B-cell developmental pathway in teleost fish. *Proc Natl Acad Sci USA.* 2005;102:6919-24.

Harpaz Y and Chothia C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J Mol Biol.* 1994;238:528-39.

Hazel JR. Thermal adaptation in biological membranes: is homeoviscous adaptation the explanation? *Annu Rev Physiol.* 1995;57:19-42.

Hirono I, Nam BH, Enomoto J, Ughino K, Aoki T. Cloning and characterisation of a cDNA encoding Japanese flounder *Paralichthys olivaceus* IgD. *Fish Shellfish Immunology.* 2003;15:63-70.

Hofmann K and Stoffel W. TMbase - A database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler.* 1993;374,166.

Hong J, Lee A, Han H, Kim J. Structural characterization of immunoglobulin G using time-dependent disulfide bond reduction. *Analytical Biochemistry.* 2009;384(2):368-70.

Hordvik I, Voie AM, Glette J, Male R, Endresen C. Cloning and sequence analysis of two isotypic IgM heavy chain genes from Atlantic salmon, *Salmo salar* L. Eur J Immunol. 1992;22:2957-62.

Ishitani A, Geraghty DE. Alternative splicing of HLA-G transcripts yields proteins with primary structures resembling both class I and class II antigens. Proc Natl Acad Sci U S A. 1992;89:3947-51.

Kabsch W and Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577-637.

Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol. 2000;2: 499-520.

Jiang S and Vakser IA. Shorter side chains optimize helix–helix packing. Protein Sci. 2004;13:1426–9.

Jones TA, Bergdoll M, Kjeldgaard M. O: a macromolecule modeling environment. In: C. Bugg, S. Ealick (Eds.), Crystallogr Model Methods Mol Des Springer-Verlag Press. 1990;189–99.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst. 1993;26:283-91.

Lesk AM and Chothia C. Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. J Mol Biol. 1982;160:325-42.

Lesk AM and Chothia C. Elbow motion in immunoglobulins involves a molecular ball-and-socket joint. Nature 1988;335:188-90.

Li YC, Karol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function and evolution. Mol Biol Evol. 2004;21:991–

1007.

Lynch KW. Consequences of regulated pre-mRNA splicing in the immune system. *Nat Rev Immunol.* 2004;4:931-40.

Lu Y, Harding SE, Michaelsen TE, Longman E, Davis KG, Ortega A, Grossmann JG, Sandlie I, García de la Torre J. Solution conformation of wild-type and mutant IgG3 and IgG4 immunoglobulins using crystallohydrodynamics: possible implications for complement activation. *Biophys J.* 2007;93(11):3733-44.

Major JG Jr, Davis FM, Liou RS, Chang TW. Structural features of the extracellular portion of membrane-anchoring peptides on membrane-bound immunoglobulins. *Mol Immunol.* 1996;33:179-87.

McGaughey GB, Gagné M, Rappé AK. Pi-Stacking interactions. Alive and well in proteins. *J Bio. Chem.* 1998;273:15458–63.

McIntosh TJ, Vidal A, Simon SA. Sorting of lipids and transmembrane peptides between detergent-soluble bilayers and detergent-resistant rafts. *Biophys J.* 2003;85:1656–66.

Merlino A, Varriale S, Coscia MR, Mazzarella L, Oreste U. Structure and dimerization of the teleost transmembrane immunoglobulin region. *J Mol Graph Mod.* 2008;27:401-7.

Morea V, Leplae R, Tramontano A. Protein structure prediction and design. *Biotechnol Annu Rev.* 1998;4:177-214.

Mottamal M, Zhang J, Lazaridis T. Energetics of the native and non-native states of the glycophorin transmembrane helix dimer. *Proteins.* 2006;62:996–1009.

Nei M and Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986;3(5):418-26.

Nezlin R. Internal movements in immunoglobulin molecule. *Adv*

Immunol. 1990;48:1-40.

Oreste U and Coscia MR. Specific features of immunoglobulin *VH* genes of the Antarctic teleost *Trematomus bernacchii*. *Gene*. 2002;295:199-204.

Park JY and Jongstra-Bilen J. Interactions between membrane IgM and the cytoskeleton involve the cytoplasmic domain of the immunoglobulin receptor. *Eur J Immunol*. 1997;27(11):3001-9.

Peterson ML. Mechanisms controlling production of membrane and secreted immunoglobulin during B cell development. *Immunol Res*. 2007;37:33-46.

Pisano E, Coscia MR, Mazzei F, Ghigliotti L, Coutanceau JP, Ozouf-Costaz C, Oreste U. Cytogenetic mapping of immunoglobulin heavy chain genes in Antarctic fish. *Genetica*. 2007 130(1):9-17.

Pleiman CM, Chien NC, Cambier JC. Point mutations define a mIgM transmembrane region motif that determines intersubunit signal transduction in the antigen receptor. *J Immunol*. 1994;152:2837-44.

Pucci B, Coscia MR, Oreste U. Characterization of serum Immunoglobulin M of the Antarctic teleost *Trematomus bernacchii*. *Comp Bioch Physiol part B* 2003;135:349-57.

Reth M. Antigen receptors on B lymphocytes. *Annu Rev Immunol*. 1992;10:97-121.

Reth M. Oligomeric antigen receptors: a new view on signalling for the selection of lymphocytes. *Trends Immunol*. 2001;22:356-60.

Rycyzyn MA, Wilson MR, Warr GW, Clem LW, Miller NW. Membrane Immunoglobulin-associated molecules on Channel catfish B lymphocytes. *Devel Comp Immunol*. 1996;20:341-51.

Saitou N and Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;4(4):406-25.

Sandin S, Öfverstedt L, Wikström A, Wrange Ö, Skoglund U. Structure and flexibility of individual immunoglobulin G molecules in solution. *Structure.* 2004;12:409-15.

Saphire EO, Stanfield MD, Crispin MDM, Parren PWHI, Rudd PM, Dwek RA, Burton DR, Wilson IA. Contrasting IgG Structures reveal extreme asymmetry and Flexibility. *JMB.* 2002;319:9-18.

Senes A, Gerstein M, Engelman DM. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol.* 2000;296:921–36.

Sinha N and Smith-Gill SJ. Molecular dynamics simulation of a high-affinity antibody-protein complex. *Cell Biochemistry and Biophysics.* 2005;43:253-73.

Stanfield RL, Zemla A, Wilson IA, Rupp B. Antibody elbow angles are influenced by their light chain class. *J Mol Biol.* 2006;357:1566-74.

Stenvik J, Schrøder MB, Olsen K, Zapata A, Jørgensen TO. Expression of immunoglobulin heavy chain transcripts (VH-families, IgM, and IgD) in head kidney and spleen of the Atlantic cod (*Gadus morhua* L.). *Dev Comp Immunol.* 2001;25(4):291-302.

Stewart JJ, Lee CY, Ibrahim S, Watts P, Shlomchik M, Weigert M, Litwin S. A Shannon entropy analysis of immunoglobulin and T cell receptor. *Mol Immunol.* 1997;34(15):1067-82.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL\_X windows interface: flexible strategies for

multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997 15;25(24):4876-82.

Tolar P, Sohn HW, Liu W, Pierce SK. The molecular assembly and organization of signaling active B-cell receptor oligomers. *Immunol Rev.* 2009;232(1):34-41.

Tusnady GE and Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics.* 2001;17(9):849-50.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005;26:1701–18.

Vereshaga YA, Volynsky PE, Pustovalova JE, Nolde DE, Arseniev AS, Efremov RG. Specificity of helix packing in transmembrane dimer of the cell death factor BNIP3: a molecular modeling study. *Proteins.* 2007;69:309–25.

Warr GW. The immunoglobulin genes of fish. *Dev Comp Immunol.* 1995;19(1):1-12.

Warr GW, Miller NW, Clem LW, Wilson MR. Alternate splicing pathways of the immunoglobulin heavy chain transcript of a teleost fish, *Ictalurus punctatus*. *Immunogenetics.* 1992;35:253-6.

Wawrzynczak EJ, Cumber AJ, Parnell GD, Jones PT, Winter G. Blood clearance in the rat of a recombinant mouse monoclonal antibody lacking the N-linked oligosaccharide side chains of the CH2 domains. *Mol Immunol.* 1992;29:213-20.

Weber M, Treanor B, Depoil D, Shinohara H, Harwood NE, Hikida M, Kurosaki T, Batista FD. Phospholipase C-c2 and Vav cooperate within signaling microclusters to propagate B cell spreading in response to membrane-bound antigen. *J Exp Med.* 2008;205.

William AF and Barkeley AN. The immunoglobulin superfamily-

domains for cell surface recognition. *Annu Rev Immunol.* 1988;6:381-405.

Wilson MR, Marcuz A, van Ginkel F, Miller NW, Clem LW, Middleton D, Warr GW. The immunoglobulin M heavy chain constant region gene of the channel catfish, *Ictalurus punctatus*: an unusual mRNA splice pattern produces the membrane form of the molecule. *Nucleic Acids Res.* 1990;18:5227-33.

Wilson MR and Warr GW. Fish immunoglobulins and the genes that encode them. *Annu Rev Fish Dis.* 1992;2:201-21.

Ye J, Bromage ES, Kaattari SL. The strength of B cell interaction with antigen determines the degree of IgM polymerization. *J Immunol.* 2010;184:844-5.

Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acid Res.* 2003;31:3406-15.

## **6. APPENDICE**

TGTTTGAGGGGAAAGGGGAAAAAGCTATCGTGAATAAAACGCTCATCACA 50  
 AAGATATGTCCAATATGTCCCAGtcccggtaaatgactactttatcttta 100  
 aaaattgtatacaaaagagttcaattttcatgtttttgcagccatctctgt 150  
 ttaataattcttctactgtctcctcatttaaagatggcacagattgtcctg 200  
 tagcagatgCGGatgtacatatcatogaaccctcaatggaggacatgtta 250  
 ttacaccacaaaaggaacggatcttgtcaagtCACGGTTAACAACCcATC 300  
 tctggataagatttctgtgggagaccgtgactggaaatgcaatagatggg 350  
 ctgaccaattaccaaccttaaggaccgtaagggagtatttgttctgtaa 400  
 ctaccatcacgtatgaagaatggataaagggcaccacgtttgtctgcaa 450  
 ggtggaacataaagattggatggagccacttaaaaaatcctaccaaagga 500  
 ttaatggtaagaatctatggagcaattgttcaaacctgtttcaattcc 550  
 tcttctcatgCCatcaaacacgtgttatctttttgtgtattatgtagcag 600  
 ggattttcttctttagttcttctatctgtggcagaccctaattgttct 650  
 ctaatgcttgcagggaggaacgactcagcgtccatcagtgtttatgctacc 700  
 tccctagaacatactaaaaaagacacgggtgaccctgacttgcttagtga 750  
 aagacttctaccogaagtatgtgtttgtggcttggtctggtgatgatctg 800  
 gaatcaggcttacaatacaaccacaagccctatagaaaaccagggatc 850  
 ctattctgcttatggccagttactctcagccttgaacagtggaaacgaca 900  
 atgacactgtgtatagctgtgtgtttaccatgagtctatggtcAAAACA 950  
 ggaactaaagccattgtcagatccattgggcaagaacatttgaagaac 1000  
 caacctggTcaacctcaacatgaacatccccgatacgtgcaaggcccaca 1050  
 attagatgtgatctctgtctactgtgtcttcggctgtttgtgtttaatg 1100  
 tttgtgcttgttatatcacattgtgtttgtcgttttaatgcagattcaaaa 1150  
 ataaaaaaaaatctctttgcaactcagttgttgtatgtcgcagctgtgtaa 1200  
 tgtctttattattcaatattagtttgatctccatgtttatcatactttaa 1250  
 gtgtagagtcacagtgctcccctaataatgcaacacacattaggttaagccat 1300  
 ccgtgtgttaagcagtgattaaaaatagatcaaatttaacaacatcacctc 1350  
 acaaaatctaatagaagaaaatctatttgtttattgtttttgtatacata 1400  
 gtattaccattgtttgtgtgacttgtctttgtctacgtgtgtttctacta 1450  
 ctgggtcttgcagcccagacataagacaaatagttattcacggccaaattc 1500  
 cagctctatggTcaataatgaattgactgattcaagaagccaggtcatgg 1550  
 cacataaagggtctttcaaaagcaattcaagatcattgtaggatgactt 1600  
 taattatcataatcaattttagtgggaccttagcccagccttttactact 1650  
 tacaataatgcatacagttacattatttaaattttagttaaacttggat 1700  
 tgactgccattgatttaccgtagtttcacactgacctgaggttattaagc 1750  
 gtgggggaggggaagtgattatgatatggtaatcatctcatgagctatatt 1800  
 gcatctgcattaaaacgacaaaacaaatgtgatataacaagcacaacat 1850  
 taacaacaacacagccgaagacacagtgacacagaatcacatctaattgt 1900  
 gggccttgcaCGtatccgggatgttcatgttgaggttgaccaggttgggt 1950  
 ctttcaaatgttctgtgccaatggatctgacaatggcctttagtctctgt 2000  
 tttgaccatagactcatggtaaaacacacagctatacacagtgctcattgt 2050  
 cgttccactgttcaaggctgagagttaactggccataagcagaataggat 2100  
 ccctgggtttctatagggctgtgtgtgtgtatttgaagcctgattcctg 2150  
 atcatcaacaagccaagccacaaaacatacttcgggtagaagtctttta 2200  
 ctaagcaagtCagggTcaccatttcttttttagtatgttctaggggaggt 2250  
 agcataaaactgatggacgctgagtcgttccctcctgcaagcattagaga 2300  
 acaattagggctctggccacagatagacgaactaaagaagaaaaatccctg 2350  
 ctacataatacacaAAAAAataaacacgtgtttgatggcatgacaagagga 2400  
 attgaaaacatgttttgagcaattgctccatagattcttaccathtaattc 2450  
 tttggtagggcttTgaagtggctccatccaatctttatgtttccaccttg 2500  
 cagacaaaactTggtgcccctttatccattcttcatcagtgatggtaagttc 2550  
 agcattaaatactcccgtAaggtcctcaaggttggttaattgggtcagcac 2600  
 caagtattggatttccagctctcggctctcccacgaaatcttatccagagat 2650  
 agtttgTtaacctgacttgacaataataccgttccctgtgtgggttaataa 2700  
 catgccctccatgaggggtttgatgatatgtacatctgcacatctgtacag 2750  
 gacaatctgtgCCgCcttttttaaatgaggacacagtgagaatttatcct 2800  
 acgagtgTTTTgtattgttttacgtcatgtaggagaggtttttcagtctta 2850  
 coattaatcctttggtaggatcttTgaagtgggtgcatctaattctttatg 2900  
 ttccacctgCagACAAACTTGGTGTCTTTGTCCATGCCACACATGTGA 2950

```

TGgtaagttcagcatcaaatactcccttacggcccttaaggttggaatt 3000
gggtcagcaccaagtattggatttccagtcctcgggtctcccacaaaatctt 3050
atccagagataatttgtaaccttgacttgacaataatccctttcctttgtg 3100
gtgtaataacatgtcctccattgagggtttgatgatatgtacatccacat 3150
ctgctacaggacaatctgtgcccgccttttttaaatgaggacacagtgaag 3200
aattatcttactagtgttttgatgtttttacgtcatgtaggagagttt 3250
ttcagttcttaccattaatcctttggtaggatcttgtaagtgggtgcacatc 3300
aatctttatgttccaccttgacagACAAAGTTGGTGCCCATCGTCCATGCC 3350
ACATACATGATGgtaagttcagcatcaaatactcccttacggcccttaag 3400
gttggtaattgggtcagcaccaagtattggatttccagtcctcgggtctccc 3450
acaaaatcttatccagagatagtttgtaaccttgacttgacaataatacc 3500
gttcctgtgtggtgtaataacatgtcctccattgagggtttgatgatag 3550
tacaaccacatctgctacaggacaatctgtgcccgccttttttaaatgagg 3600
acacagtgagaattatcttactagtgttttgatgtttttacgtcatg 3650
taggagagtttttcag.tctgtggcttctgatctggtttcattttctcgt 3700
gttttacatttgtttcatgtcctgtctattttgatgtattcatgtcata 3750
ttttatacagttggtgatgttcagcactttgttcaactctgtttgtttt 3800
aaatgtgctatagaaataaaatggattggatactttaagtgtagagtcac 3850
agtgtctcctaatagaacacacattagggttaagccatccgtgtgtaag 3900
cagtgattaaaaatagatcaaatttaacaacatcacctcacaaaatcta 3950
gaagaaaatctatttgtttattgtttttgtatacatagattaccattg 4000
tttgtgtgacttgtctttgtctacgtgtgtttctactactgggtcttgca 4050
gcccgcataagacaaaatagttattcacggccaaattccagctctatgg 4100
caataatgaattgactgattcacaagccaggtcatggcacataaagggg 4150
tctttcaaaagcaattcaagatcattgtaggatgactttaattatcata 4200
tcaattttagtgaggacctttagcccagccttttacttacaataatgca 4250
tacagttacattatttaattttagtttaacttggattgactgccattg 4300
atctactgtagtttcaactgacctgaggttattaagcgtgggggagggga 4350
agtgattatgatatggtaatcatctcatgagctactgttttaagctggg 4400
tttgtcaggcatattgaatcatttctgtacatccagACTCCAACCAAAA 4450
CGGACATTGACAAGAAGGTAGAAGGATATGACATGGGATACACAGAGGTA 4500
GAAGAAGATAACATGGGAAGCACAGCCATCACCTTCATCCTTCTCTCTCT 4550
CATCACTCTGCTGT

```

Sequenza A1. La regione genomica compresa tra gli esoni CH2 e TM in *C. hamatus*. In giallo sono mostrate le sequenze APS0, APS1 e APS2, in blu è evidenziata la sequenza S, corrispondente all'inverso complementare della sequenza APS0. In maiuscolo è mostrata la sequenza codificante la forma di membrana delle Ig: la fine dell'esone CH2 e gli esoni R1, R2 e TM.

```

TGCTGTTTGAGGGGAAAGGTAAAAGCGGTCCAACAAACAAGTCATCATC 50
TGACTCACAAACCTTTGGATCCAACGtgtgagtaaattgttactttt 100
cattaaaaatggtgtgttgatgtaatttaaattctcatggtttttgcagc 150
catttatgttcaatcattttgaatgaattcctcactcaacagATGTCCAA 200
GGATGTCTGAAGCAGATGTAGAGGTAACATCATAGCACCCCAACGGG 250
GAGCATGCTCTTACACAAAAAAGGAACAGTAACGTGTGAGGTCAAGGTAC 300
GCAAATCACATGTTTCAAGATTTTCGTGGGAGACTAAGAATGGACTGTCA 350
ATAGTTGGTGCCGAGCCAATTGAACCCGTTAAAGGACATACAGGCGTATT 400
TCGCGCTCTACTTCCCATCACGTATGACGAATGGAGCCTGGGAACAGAGT 450
TCCTCTGCATTTGAAAATAAAGATTGGATTGTACCACTTAAGAAAACA 500
TACAAAAGGCCTATCGGtaagaaaactatttgtccgttcctgtgtttaat 550
tcctattcatgcatcaaacataatgtgagtgttaattctaaagtgtatt 600
ttgtagcataatcttctactgtgggactgtacttttaaagtgtggcagga 650
gccctaattgtcctctcatgcttgcaggaggacacactcagcgtccttcc 700
gtgtttatgttacctccactagaacaaactagaaaagaaatggtgaccct 750
gacttgctatgtgaaagacttccaccctcatgaggtttttgtggcttggc 800
ttgttgatgatgaagaagcagactcaacatacaagttccataaccacaaac 850
cctgtagaaagccatggatcctattctgcttatggccagttatctctcag 900
ccttgaccagtggaacacgctgacactgtgtacagctgtgtagttacc 950
acgagtcctgtgtaacacacacaaagctattgtcaggtccatcggccac 1000
agaacatttgaaaaatcaacctgggtcaacctcaacatccctgatactgt 1050
caagcccagtagatgttcttgtatgtcacgtgtcttctgccgtttgttg 1100
ttaatgtttgtgcttgtgatgacattgtgtgtcttttaatgcaga 1150
ttcaaatcaaaaataaaagccctttgcaactctgttgatgt 1200
atgtgcgcatgctgttcatgtcttattaaccaatattttgagctccatgt 1250
ttgccataagtgtagtcacagtgctatcctaataacacacaaatagggtg 1300
catctatcagtatgtttgagtttaagcacagtgatagaacaatatat 1350
caaatttcaatatcatattgtactcattagaataacacacaaaccagca 1400
gactatctttatgccacaattgaatgaaaataatcctatttgttcattgt 1450
tattttagacagtgcatcaatgtttgtgtggcttgtctttgtgtac 1500
ctgtgtttccactactggctcttgcacccccataaaggaagaaaaacat 1550
tccagctctaattgtctccagtcacaaattatgaatatattgacaagccaag 1600
tcatgcatgccacactaataaaggggtctttaaaaaagcaattcaggatc 1650
actgtaggaaaactttaaaatcaacacaatcaattttagaaggacattca 1700
aaactattagcttaggcctcatggtacactgataacaacgtaataagtaa 1750
cattgcatatgccagactgttcttttaattttagttagtctttttatac 1800
atggattgactgaccaattagatagactctgacctttaggttataaaat 1850
aagttacactactacaggaattgaatatgttgtctcatgagctactgctt 1900
aaagccgggtttgtctcaagtattaactgaatattaattgtacagTAGAT 1950
GAAGAATACTACCATTGGAACAGTTCCACGGAGGCAGAAGAAGATAACAT 2000
GGGAATCACAGCAATCACCTTCATCCTTCTCTCCTAATCACTCTGCTGT 2050
TCACCATTGGAACCACTGCTTTC

```

Sequenza A2. La sequenza genomica compresa tra gli esoni CH2 e TM in *B. diacanthus*. La sequenza codificante la forma di membrana delle Ig è riportata in maiuscolo: la fine dell'esone CH2 e gli esoni CH3 e TM.

UGUUUGAGGGGAAAGGGGAAAAAGCUAUCGUGAAUAAAACGCUCAUCACA 50  
 AAGAUAUGUCCAAUAUGUCCCCgucccgguaaaugacuauuuuauuuuu 100  
 aaaaauuguauacaagaguucauuuuucauguuuuuugcagccaucucugu 150  
 uuaaaauuuuucacugucucccauuuaaagauggcacagauuguccug 200  
 uagcagaugcggauuacaucaucgaaccucuauggaggcauguua 250  
 uuacaccacaaaaggaacgguaucuuugucaagucacgguaacaacccauc 300  
 ucuggauaagauuucgugggagaccgugacuggaaaugcauagauuggug 350  
 cugacccaauuaccaaaccuuaggaccguaaggaguuuuugucugaa 400  
 cuuaccaucacguugaagauggauaaagggcaccacguuugucugcaa 450  
 gguggaacauaaagauuggauggaccacuuaaaaaaucuaccaaagga 500  
 uuaaugguaagaaucauaggagcaauuguucaaaaccuguuuucaauucc 550  
 ucuuugcaugccaucaaaacaguguuauuuuuuguguauuauguagcag 600  
 ggauuuuucuuuuaguucuuucuaucuguggccagagcccauuuuuguc 650  
 cuuauugcuugcaggaggaacgacucagcgucccaucaguuuuuugcacc 700  
 uccccuagaacauacuaaaaaagacacggugacccugacuugcuuaguga 750  
 aagacuucuaaccgaaguuguuuguggcuuggcucguugaugaucug 800  
 gaucagggcuuacaauacaaccacacaagcccauagaaaaccagggaucc 850  
 cuauucugcuuauuggccaguuaacucucagccuugaacaguggaacgaca 900  
 augacacuguguaauagcuguguuuuuaccagagucuauggucaaaca 950  
 ggaaacuaagccauugucagauccauugggcaagaaacuuuugaagaaac 1000  
 caaccuggucaaaccuacaacauagaacaucaccgacugcagggccaca 1050  
 auuagauugauucugugucacugugucuuucggcuguuuuguuuuuauug 1100  
 uuugugcuuguuauaucacaauuguguuugucguuuuuuauugcagauucaa 1150  
 auaaaaaaaaucucuuugcaacucaguuuuuuuugucgcaugcuuuuu 1200  
 ugucuuuuuuuucauuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1250  
 uguuagagucacagugcucccuaauugcaacacacauuagguuaagccau 1300  
 ccguguguuuagcaguguuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1350  
 acaaaaucuaaagaagaaaaucuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1400  
 guuuuaccuu 1450  
 cugggucuuugcagcccgacauaagacaaauuuuuuuuuuuuuuuuuuuuu 1500  
 cagcucuauggucaaauuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1550  
 cacauuu 1600  
 uuu 1650  
 uacauuu 1700  
 ugacugccauuu 1750  
 guuu 1800  
 gcaucugcauu 1850  
 uuu 1900  
 gggccuugcagcuuuccgggaguuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1950  
 cuuu 2000  
 uuugaccuauuagcuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 2050  
 cguuuccacuguuuacagggcugaguuuuuuuuuuuuuuuuuuuuuuuuuuuu 2100  
 ccuuguu 2150  
 aucaucaaagccaaagccacaaacacauacuucggguagaagucuuuuuu 2200  
 cuuagcaagucagggucacuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 2250  
 agcauu 2300  
 acauu 2350  
 cuacauuu 2400  
 auuu 2450  
 uu 2500  
 cagacaaacuu 2550  
 agcauu 2600  
 caaguu 2650  
 aguu 2700  
 caugcccuuu 2750  
 gacaaucugugccgcuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 2800  
 acgaguguu 2850  
 ccuu 2900  
 uuucccuugcagACAAACUUGGUGUCCUUUGUCCAUGCCACACAUGUGA 2950  
 UGguuaguuuagcaucaaauacuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 3000



```

UGUCUGUUUGAGGGGAAAGGUAAAAGCGGUCCAACAAACAAGUCAUCAUC 50
UGUGACUCACAAACCUUUGGAUCCAACgugugaguaaaauuguuacuuuuu 100
cauuaaaaaugguguguugauguaauuuuuuuuucuaugguuuuuugcagc 150
cauuuauguucaaucauuuuugaauugaauuccucacucaacagAUGUCCAA 200
GGAUGUCCUGAAGCAGAUGUAGAGGUAAACUAUCAUAGCACCCUCAACGGG 250
GAGCAUGCUCUUACACAAAAAAGGAACAGUAACGUGUGAGGUCAAGGUAC 300
GCAAAUCACAUGUUCAGAAGAUUUCGUGGGAGACUAAGAAUGGACUGUCA 350
AUAGUUGGUGCCGAGCCAAUUGAACCCGUAAAAGGACAUACAGGCGUAUU 400
UCGCGCUCUACUUCUCAUCACGUUAGACGAAUGGAGCCUGGGAACAGAGU 450
UCCUCUGCAUUCUUGAAAAUAAAGAUUGGAUUGUACCACUUAAGAAAAACA 500
UACAAAAGGCCUAUCGguaagaaaacuauuuuguccguuccuguuuuuuuu 550
uccuauucaugccaucacaaucauuuugaguguuuuuuucaaaguguaau 600
uuguagcauaauuuuucuaugugggacuguaauuuuuuuuuuuuuuuuuuu 650
gcccauuuuguccucuaugcugcaggaggacacacucagcguccuucc 700
guguuuuuguuaccuccacuagaacaaacuagaaaagaaauggugacccu 750
gacuugcuaugugaaagacuuccacccucaugagguuuuuuugugcuggc 800
uuguugaugaugaaagcagacucaacauacaaguuccauaccacaaac 850
ccugugaaagccauggaucuauucugcuuauggccaguuauucucucag 900
ccuugaccaguggaaaaacgcugacacuguguaacagcuguguaauuacc 950
acgaguccguguuuacacaaccaaagcuauugucagguccaucggccac 1000
agaacauuuugaaaaaaaucaaccuggucaaccucaacaauccugauacgug 1050
caaaagcccagugauguucuguaugucacgugucuuucugccguuuugug 1100
uuuaauguuugugcuugugauaugacauugugugucuuuuuuuuuuuuuu 1150
uucaaaaucaaaaauaaaaaaagcccuuugcaacucuguuugaugu 1200
augucgcaugcuguucaugucuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1250
uugccauaaguguaugucacagugcuauccuaaugcaacacaaauuaggug 1300
caucuaucaguauguuugaguuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1350
cauuuuucaauuacauuuuguaucuaauagaauaacacaaacaaccagca 1400
gacuauuuuuuugccacaaugaaugaaaauaauccuauuuuuuuuuuuuu 1450
uauuuuguaagacagugcauuuacauuuuuuuuuuuuuuuuuuuuuuuuu 1500
cuguguuuccacuacuggcucugcaccuccuuuuuuuuuuuuuuuuuuuuuu 1550
uccagcucuaauugucuccagucuaauuuuugaauuuuuuuuuuuuuuuuu 1600
ucaugcaugccacacuaauuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1650
acuguaaggaaaacuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1700
aaacuauuuagcuuaggccuacuguaacacuguaaacaacguaauuuaaguu 1750
cauugcauauugccagacugucuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1800
auggauugacugaccauuuagauagacucugaccuuuuuuuuuuuuuuuuuu 1850
aaguuaacuaucuaacaggaauuugaauuugugucuaugagcuaucugcuu 1900
aaagccggguuuugucuaaguuuuuuuuuuuuuuuuuuuuuuuuuuuuuu 1950
GAAGAAUACUACCAUUGGAACAGUCCACGGAGGCAGAAGAAGAUAAACAU 2000
GGGAAUCACAGCAAUCACCUUCAUCCUUCUCUCCUAAUACUCUCUGCUGU 2050
UCACCAUUGGAACCACUGCUUUC

```

Sequenza A4. Il frammento di RNA riprodotto la regione genomica in *B. diacanthus* comprendente gli esoni CH2 e TM. Le sequenze evidenziate in verde sono quelle codificanti la forma di membrana delle IgM.















TGTTTGAGGGGAAAGGGGAAAAAGGTCCAAGTACCTTGAATAGAACGATC 50  
 ATCACAAGATATGTGAAGGACCCACTCCCgtaaatgactactttatct 100  
 ttcaaaattgtatacaagagttcaatthttcatgtttttgagccatctc 150  
 tgtttaataattcttcaactgtctccttttaagATGGCATAGGTTGTCAT 200  
 GAAGCAGATGCCGATGTACATATCATCGAACCCTCAATGGAGGACATGTT 250  
 ATTACACCACGAAGGAACAGTATTGTGTTTCAGTCAAGATTAACAAAGCAT 300  
 CTGTGTATAAGATTTCGTGGCAGACCAAGGCTGGAATCCAATACTGGT 350  
 GCTGAACCAATTACCGCCATAAGGAACATATGGAGGTATTTAATGCTGA 400  
 ACTTACCATCACGTATGAAGAATGGATAAAGGGTACCACGTTTGTCTGCA 450  
 AGGTGGAACAGAAAGATTGGATAGAGGCAGTTACAAAATCCTACCAAAGG 500  
 ATTAATGgtaagaatctatggagcaattggtcAAAACCTgttttcaattc 550  
 ctctttgtcatgccatcaaacataatgagtggttattgtttgtgtattat 600  
 gttagcagggttttcttctttagttcttctatctgtggccagagccctaa 650  
 ttgttctctctaattgcttgagGAGGAACGACTCAGCGTCCATCAGTGT 700  
 TATGCTACCTCCCCTAGAACATACTAAAAAGACACGGTGACCCTGACTT 750  
 GCTTAGTGAAAGACTTCTACCCGAAGTATGTGTTTGTGGCTTGGCTTGT 800  
 GATGATCTGGAATCAGCTTACACTACAACACCACAAGCCCTATAGAAAA 850  
 TGAGGGATCCTATTCTGCTTATGGCCAGTTAACTCTCAGCACTGAAAAGT 900  
 GGAACGACAATGACACTGTGTATAGCTGTGTTGTTTACCATGAGTCTATG 950  
 GTCAAAAACAGGAACTAAAGCCATTGTCAGATCCATTGGGCACAGAACATT 1000  
 TGAAGAACAACCTGGTCAACCTCAACATGAACATCCCCGATACGTGCA 1050  
 AGGCTACCTGCAATTAgatgtgattctgtgtcactgtgtcttcggctgtt 1100  
 tgttgtttaatgtgtgtgcttgttatatcacattgtgtttgtcgtttaa 1150  
 tgcagattcaaaaataaaaaaaagctctttgcaactcagttgtttgtatgt 1200  
 cgcattgtgtttatgtctttattttaaattagtttgatctccatgtt 1250  
 tatcatactttaagtgtagtgtcacagtgtctcctaatagaacacagacat 1300  
 taggttaagccatccgtgattaaaaatagatcaaatttaacaacatcacct 1350  
 caaaaaatctaataagaacatctattgtttattgtttttgtatacat 1400  
 agtattaccggttgtttgtgtgactgtctttgtgtatgtgtgtttctact 1450  
 accgggtcttgagcccgacataagacaaatagttattcacggccacact 1500  
 ccagctctatggtcgaataatgaattgactgattcacaagccaggtcatg 1550  
 acacataaaggggtctttcaaaagcaattcaagatcattgttaggatgact 1600  
 ttaattatcataatcaatthttagtyggaccttttagccagccttttacac 1650  
 ttacaataatgtatagttacattatthtttaattgtagttaaacttggat 1700  
 tgtctgacattgatttgtgtgtagttcacactgacctgaggttattaaga 1750  
 gtgagggaggggaagtgattatgatatggtaacctctcatgagctattht 1800  
 gatctgcatthaaacgacaaacacaatgtaataaacaagcagaaacat 1850  
 taacaacaaacagccgaagacacagtgacacagaatcacatctaattgc 1900  
 agggagccttgcaogtatcggggatgttcatgttgaggttgaccaggttg 1950  
 gttctttcaaatgtctgtgcccgaatggatctgacaatggcttttagttcc 2000  
 tgttttgaccgttagactcatggttaacaacacagctatacacagtgatc 2050  
 tgtcgttccactgttcagtgctgagagtttaactggccataagcagaatag 2100  
 gatccctgggttttctatagggcttgggtgtttagtgtaagcctgattc 2150  
 ctgatcatcaacaagccaagccacaaaacatacttcgggtagaagtctt 2200  
 tcaactaagcaagtcagggtcaccatttcttttttagtatgttctagggga 2250  
 gtagcataaaacactgatggacgctgagtcggtctcctcctgcaagcattag 2300  
 agaacaattagggctctggccacagatagaagaactaaagaagaaaatcc 2350  
 ctgctacataatacacaaacaataaacatgtgtttgatggcatgacaaga 2400  
 ggaattgaaaacatgttttgaacaattgctccatagattcttaccatthaa 2450  
 tcttttggtagggtctgttaagtggctccgtccaatctttatgttccacc 2500  
 ttgcagacaaacttgggtgcccctttatccattcttatacgtgatgtaag 2550  
 ttcagcatgaaatactcccttatggcccttaagggttgtaattgggtcgg 2600  
 caccaagtattggatttccagtcctcggctctcccacgaaatcttttcaga 2650  
 gatgttttgttaacctgacttgacaataataccgttccctttgtggtgtaa 2700  
 taacatgtcctccattgaggggttgatgatatgtacatccgcatctgcta 2750  
 caggacaatctgtgcccccttttttaaatgaggagacagtgagaatthaa 2800  
 cctactagggtttgtattgttttacctcatgtaggagaggtttttcagtc 2850  
 ttaccattaatcctttggtaggatcttgaagtgggtgcatctaatcttt 2900  
 atgttcaaccttgagACAAACGTGGTGCCCTTCCTCCATGCCACAACG 2950  
 TGATGgtaagttcagcatgaaatactcccttacggcccttaaggttgta 3000

attgggtcggcaccaagtattggatttccaatctcggctctcccacgaaat 3050  
 ctaatccagagatgttttgtaaccttgacttgacaataaccgctcctt 3100  
 tgtgggtgtaataacatgtcctccattgaggggttgatgatatgtacatcc 3150  
 gcatctgctacaggacaatctgtgcccccttttttaaatgaggacacag 3200  
 tgaataattatcttactagtgttttgattgtttttacgtcatgtaggag 3250  
 agtttttcagctctaccattaatcctttggtaggatcttgtaagtgggtgt 3300  
 catctaactctttatgttcaaccttgcaagACAAACCTGGTGCCCTTATGC 3350  
 ATTCTTCATACGTGATGgtaagttcagcatgaaatactcccttacggccc 3400  
 ttaaggttggttaattgggtcggcaccaagtatttgatttccaatctcgg 3450  
 ctcccacgaaatctttccagagatgttttgtaaccttgacttgacaat 3500  
 ataccgcttactttgtgggtgtaataacatgtcctccattgaggggttgatg 3550  
 atatgtacatccgcatctgctacaggacaatctgtgcccccttttttaa 3600  
 tgaggacacagtgaaataattatcttactagtgttttgattgtttttacg 3650  
 tcatgtaggagagtttttcagctctaccattaatcctttggtaggatctt 3700  
 gtaagtgggtgtcatctaactctttatgttcaaccttgcaagACAAACGTGGT 3750  
 GCCCTTCCTCCATGCCACAAACGTGATGgtaagttcagcatgaaatactc 3800  
 ccttacggcccttaaggttggttaattgggtcggcaccaagtattggattt 3850  
 ccagtctcgggtctcccacgaaatctttccagagatgttttgtaacctt 3900  
 gacttgacaataaccgctcctttgtgggtgtaataacatgtcctccattg 3950  
 agggtttgatgatgtacatccgcatctgctacaggacaatctgtgccc 4000  
 ccttttttaaatgaggagacagtgaaataattatcttactagtgttttgta 4050  
 ttgtttttacgtcatgtaggagagtttttcagctctaccattaatccttt 4100  
 ggtaggatcttgtaagtgggtgtcatctaactctttatgttcaaccttgca 4150  
 ACAAACGTGGTGCCCTTCCTCCATGCCACAAACGTGATGgtaagttcagc 4200  
 atgaaatactcccttacggcccttaaggttggttaattgggtcggcaccia 4250  
 gtattggatttccaatctcgggtctcccacgaaatctaataccagagatgtt 4300  
 ttgttaaccttgacttgacaataaccgctcctttgtgggtgtaataacat 4350  
 gtccctccattgaggggttgatgatgtacatccgcatctgctacaggac 4400  
 aatctgtgcccgccttttttaaatgaggacacagtgaaataattatattact 4450  
 agtgtttttgattgtttttacgtcatgtaggagagttttcagctcgtggc 4500  
 ttctgatctggtttcatttctcgtgttttacattgtttcatgtcctgtt 4550  
 tatttgatgtattcatgtctatattttatacagttgttgatgtacagca 4600  
 cttgttcaactcgtgtgttttttaaatgtgctatagaaataaaaaatgatt 4650  
 ggatactttaagtgtagagtcacagtgctctctaagcaacacacattag 4700  
 gtaagccatccgtgtgttaagcagtgattaaaatagatcaaatttaaca 4750  
 acatcacctcacaataatgaagaaaatctattgtttattgttttt 4800  
 tgtatacatagtataccggtgtttgtgtgactgtcctttgtgtatgtgt 4850  
 gtttctactactgggtcttgagcccgacataagacaaatagttattcat 4900  
 gggcacactccagctctatggctgaataatgaattgactgattcacaagc 4950  
 caggtcatgacacataaaggggtctttcaaaagcaattcaagatcattgt 5000  
 aggatgactttaattatcataatcaatttttagtgggaccttagccagc 5050  
 cttttacacttacaataatgtatagttacattatttaaatttgtagttaa 5100  
 acttggattgactgacattgatttctgttagtttcacactgacctgaggt 5150  
 tattaagagtgaggagggaagtgtgtgatatgtaacatctcatga 5200  
 gatactgttttaagctgggtttgtctcaggcattaactgaatcatttct 5250  
 gtacatccagACTCCAACCAACGGACATTGACAAGAAGGTAGAAGGATA 5300  
 TAACATGGGATGCACAGAAGAAGATAACATGGGAAGCACAGCGGTAGAAG 5350  
 AAGATAACATGGGAAGCACAGCCATCACCTTCATCCTTCTCTCCTCATC 5400  
 ACTCTGCTGTTCCGCAATTGGAACCACTGCTTTC

Sequenza A9. Regione genomica di *C. mawsoni* comprendente la fine dell'esone CH2 e l'esone TM. In vede gli esoni CH2, CH3, CH4, R1, R2, R3, R4 e TM. In giallo le regioni ripetute.

Sequenza	N° ACCESSO
Trematomus newnesi immunoglobulin mu heavy chain membrane form mRNA, partial cds. 553 bp linear mRNA	FJ685522.1
Trematomus pennellii membrane-bound immunoglobulin M heavy chain mRNA, partial cds. 234 bp linear mRNA	FJ652069.1
Gobionotothen gibberifrons immunoglobulin M heavy chain gene, exons CH2, CH3, CH4 and partial cds	FJ628417.1
Chionodraco rastropinosus immunoglobulin M heavy chain gene, exons CH2, CH3 and partial cds 258 bp linear DNA	FJ628416.1
Trematomus pennellii immunoglobulin M heavy chain gene, exons CH2, CH3 and partial cds. 268 bp linear DNA	FJ623171.1
Trematomus hansonii immunoglobulin M heavy chain gene, exons CH2, CH3 and partial cds 211 bp linear DNA	FJ623170.1
Trematomus newnesi immunoglobulin M heavy chain gene, exons CH2, CH3 and partial cds 297 bp linear DNA	FJ623169.1
Cygnodraco mawsoni immunoglobulin M heavy chain gene, exons CH2, CH3 and partial cds 660 bp linear DNA	FJ623168.1
Gymnodraco acuticeps immunoglobulin M heavy chain gene, exons CH2, CH3 and partial cds 279 bp linear DNA	FJ623167.1
Trematomus loennbergii membrane-bound immunoglobulin M heavy chain mRNA, partial cds 626 bp linear mRNA	FJ623166.1
Dissostichus mawsoni immunoglobulin M heavy chain secreted form mRNA, partial cds 1,461 bp linear mRNA	EU488726.1
Gymnodraco acuticeps immunoglobulin mu heavy chain secreted form mRNA, partial cds 1,494 bp linear mRNA	EU340099.1
Eleginops maclovinus immunoglobulin M heavy chain secreted form mRNA, partial cds 1,311 bp linear mRNA	EU221319.1
Cygnodraco mawsoni immunoglobulin M heavy chain secreted form mRNA, partial cds 1,510 bp linear mRNA	EU232636.1
Pagetopsis macropterus immunoglobulin mu heavy chain secreted form mRNA, partial cds 1,342 bp linear mRNA	EU195861.1
Trematomus bernacchii clone TS nonfunctional immunoglobulin light chain isotype 1 mRNA, partial sequence 587 bp linear mRNA	EF114791.1
Trematomus bernacchii clone 4932 immunoglobulin light chain isotype L2 (IgL) mRNA, partial cds 566 bp linear mRNA	DQ842622.1
Trematomus bernacchii immunoglobulin M heavy chain gene, exons, and partial cds, alternatively spliced 1,962 bp linear DNA	EU884293.1

Bovichtus diacanthus immunoglobulin M heavy chain secreted form mRNA, partial cds 1,491 bp linear mRNA	EU884292.1
Chionodraco hamatus immunoglobulin heavy chain gene, partial sequence 4,567 bp linear DNA	EU925137.1
Chionodraco hamatus immunoglobulin M heavy chain secreted form mRNA, partial cds 1,476 bp linear mRNA	EU192078.2
Chionodraco hamatus immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 318 bp linear mRNA	EU192079.1
Bovichtus diacanthus immunoglobulin M heavy chain membrane-bound form precursor, mRNA, partial cds 1,561 bp linear mRNA	EU240584.3
Trematomus bernacchii clone 180 immunoglobulin light chain isotype L1A mRNA, complete cds 743 bp linear mRNA	EF421439.1
Trematomus bernacchii clone 450 immunoglobulin light chain isotype 2 mRNA, partial cds 481 bp linear mRNA	EF421438.1
Trematomus bernacchii clone 440 immunoglobulin light chain isotype L2 mRNA, partial cds 487 bp linear mRNA	EF421437.1
Trematomus bernacchii clone 47 immunoglobulin light chain isotype L2 mRNA, partial cds 487 bp linear mRNA	EF421436.1
Trematomus bernacchii clone 110 immunoglobulin light chain isotype 1A mRNA, partial cds 384 bp linear mRNA	EF421435.1
Trematomus bernacchii clone L480 immunoglobulin light chain isotype 2 mRNA, partial cds 487 bp linear mRNA	EF114790.1
Trematomus bernacchii clone L470 immunoglobulin light chain isotype 2 mRNA, partial cds 487 bp linear mRNA	EF114789.1
Trematomus bernacchii clone L46 immunoglobulin light chain isotype 1 mRNA, partial cds 587 bp linear mRNA	EF114788.1
Trematomus bernacchii clone L44 immunoglobulin light chain isotype 1 mRNA, partial cds 583 bp linear mRNA	EF114787.1
Trematomus bernacchii clone L43 immunoglobulin light chain isotype 1 mRNA, partial cds 582 bp linear mRNA	EF114786.1
Trematomus bernacchii clone L420 immunoglobulin light chain isotype 2 mRNA, partial cds 487 bp linear mRNA	EF114785.1
Trematomus bernacchii clone L42 immunoglobulin light chain isotype L1A mRNA, partial cds 592 bp linear mRNA	EF114784.1
Trematomus bernacchii clone 3P immunoglobulin light chain isotype L1 (IgL) mRNA, partial cds 450 bp linear mRNA	DQ842635.1
Trematomus bernacchii clone 4P immunoglobulin light chain	DQ842634.1

isotype L1 (IgL) mRNA, partial cds 432 bp linear mRNA	
Trematomus bernacchii clone 2P immunoglobulin light chain isotype L1 (IgL) mRNA, partial cds 435 bp linear mRNA	DQ842633.1
Trematomus bernacchii clone 14 immunoglobulin light chain isotype L1 (IgL) mRNA, partial cds 407 bp linear mRNA	DQ842632.1
Trematomus bernacchii clone 218 immunoglobulin light chain isotype L1 (IgL) mRNA, partial cds 419 bp linear mRNA	DQ842631.1
Trematomus bernacchii clone L12 immunoglobulin light chain isotype L1 (IgL) mRNA, partial cds 419 bp linear mRNA	DQ842630.1
Trematomus bernacchii clone 5P immunoglobulin light chain isotype L1 (IgL) mRNA, partial cds 435 bp linear mRNA	DQ842629.1
Trematomus bernacchii clone L11 immunoglobulin light chain isotype L1 (IgL) mRNA, partial cds 398 bp linear mRNA	DQ842628.1
Trematomus bernacchii clone 1B11 immunoglobulin light chain isotype L1 (IgL) mRNA, partial cds 708 bp linear mRNA	DQ842627.1
Trematomus bernacchii clone 5141 immunoglobulin light chain isotype L3 (IgL) mRNA, complete cds 977 bp linear mRNA	DQ842626.1
Trematomus bernacchii clone 320 immunoglobulin light chain isotype L3 (IgL) mRNA, partial cds 402 bp linear mRNA	DQ842625.1
Trematomus bernacchii clone 323 immunoglobulin light chain isotype L3 (IgL) mRNA, partial cds 402 bp linear mRNA	DQ842624.1
Trematomus bernacchii clone 329 immunoglobulin light chain isotype L3 (IgL) mRNA, partial cds 405 bp linear mRNA	DQ842623.1
Trematomus bernacchii clone 216 immunoglobulin light chain isotype L2 (IgL) mRNA, partial cds 392 bp linear mRNA	DQ842621.1
Trematomus bernacchii clone 219 immunoglobulin light chain isotype L2 (IgL) mRNA, partial cds 392 bp linear mRNA	DQ842620.1
Dissostichus mawsoni immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 532 bp linear mRNA	EU240586.1
Gymnodraco acuticeps immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 374 bp linear mRNA	EU240585.1
Eleginops maclovinus immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 705 bp linear mRNA	EU221318.1
Cygnodraco mawsoni immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 381 bp linear mRNA	EU232635.1
Gobionotothen gibberifrons clone 1044 immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 348 bp linear mRNA	EU232634.1

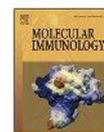
Gobionotothen gibberifrons clone G3 immunoglobulin M heavy chain secreted form mRNA, partial cds 1,371 bp linear mRNA	EU232633.1
Histiodraco velifer immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 390 bp linear mRNA	EU200778.1
Pogonophryne scotti immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 384 bp linear mRNA	EU192081.1
Pogonophryne scotti immunoglobulin M heavy chain secreted form mRNA, partial cds 1,331 bp linear mRNA	EU192080.1
Champscephalus esox immunoglobulin M heavy chain membrane-bound form mRNA, partial cds 308 bp linear mRNA	EU192077.1
Champscephalus esox immunoglobulin M heavy chain secreted form mRNA, partial cds 1,339 bp linear mRNA	EU192076.1
Histiodraco velifer immunoglobulin mu heavy chain secreted form mRNA, partial cds 1,359 bp linear mRNA	EU195863.1
Pagetopsis macropterus immunoglobulin mu heavy chain membrane-bound form mRNA, partial cds 270 bp linear mRNA	EU195862.1
Chaenocephalus aceratus clone Chm membrane bound immunoglobulin heavy chain mRNA, partial cds 312 bp linear mRNA	EU161632.1

Tabella S1. Elenco delle sequenze nucleotidiche da me ottenute durante il dottorato di ricerca e relativo numero di accesso in banca dati.

Chondrichthyes	<i>Gynglimostoma cirratum</i>	IgW	AY524278
		IgNAR	AY612428
	<i>Heterodontus francisci</i>	IgM9	X07785
	<i>Leucoraja erinacea</i>	IgM	M29678
Osteichthyes	<i>Acipenser baerii</i>	IgM	AJ245364
	<i>Amia calva</i>	IgM	X82070
	<i>Anarhichas minor</i>	IgM	AF138958
	<i>Chionodraco hamatus</i>	IgM	EU192079
	<i>Cyprinus carpio</i>	IgM	AB004106
	<i>Danio rerio</i>	IgM	AF281479
		IgZ/T	AY643750
	<i>Gadus morhua</i>	IgM	X58871
		IgD	AF155199
	<i>Hippoglossus hippoglossus</i>	IgD	AY077849
	<i>Ictalurus punctatus</i>	IgM	X52617
		IgD	AF363448
	<i>Oncorhynchus mykiss</i>	IgM1	X65263
	<i>Paralichthys olivaceus</i>	IgM	AB109029
		IgD	AB052658
	<i>Protopterus aethiopicus</i>	IgM	AF437735
	<i>Salmo salar</i>	IgM1	Y12391
		IgD1	AF278717
	<i>Takifugu rubripes</i>	IgT/Z	AB201355
	Amphibia	<i>Pleurodeles walii</i>	IgM
IgP			AM410552
<i>Xenopus laevis</i>		IgM	X90768
		IgY sp.	X90767
<i>Xenopus tropicalis</i>		IgX sp.	X90766
		IgD	DQ387453
		IgF	DQ350887
Reptilia	<i>Anolis carolinensis</i>	IgY	EF690360
	<i>Eublepharis macularius</i>	IgM	EU287911
		IgD	EU306881
		IgA	DQ523197
Aves	<i>Anas platyrhynchos</i>	IgM	AJ314754
		IgY	X78357
		IgA	AF324491
Mammalia	<i>Ailuropoda melanoleuca</i>	IgM	AY818393
	<i>Bos taurus</i>	IgM	U63637
	<i>Camelus dromedarius</i>	IgM	AB091834
		IgG2	AJ131945
	<i>Canis lupus</i>	IgM	EF563150
	<i>Equus caballus</i>	IgM	AY631941
	<i>Homo sapiens</i>	IgM	X14940
		IgD	K02881
		IgG1	X52847
		IgG2	AB006776
		IgG4	AL928742
		IgE	X63693
		IgA1	M60193
		IgM	AF046784
	<i>Macaca mulatta</i>	IgM	FJ972692
	<i>Mus musculus</i>	IgM	J00444
		IgD	J00450
		IgG1	J00454
		IgG2a	J00471
		IgG3	X00915
		IgE	X03624
		IgA	K00691
IgM		K01357	
<i>Oryctolagus cuniculus</i>	IgM	L04260	
<i>Ovis aries</i>	IgM	L04260	

		IgM	<b>AABR03049259</b>
		IgD	<b>AABR03049259</b>
		IgG2b	<b>AABR03048905</b>
		IgE	<b>K02901</b>
		IgA	<b>AABR03049801</b>
	<i>Sus scrofa</i>	IgM	<b>U50148</b>
	<i>Tursiops truncatus</i>	IgM	<b>AJ320193</b>

Tabella S2. Elenco delle sequenze amminoacidiche scelte per costruire il dataset di IgTM e relativo numero di accesso.



## Immunoglobulin light chain isotypes in the teleost *Trematomus bernacchii*

Maria Rosaria Coscia, Stefano Giacomelli, Concetta De Santi, Sonia Varriale, Umberto Oreste\*

Institute of Protein Biochemistry, CNR, via P. Castellino 111, 80131 Napoli, Italy

### ARTICLE INFO

**Article history:**  
Received 20 February 2008  
Received in revised form 10 March 2008  
Accepted 10 March 2008  
Available online 2 May 2008

**Keywords:**  
Immunoglobulin light chain isotypes  
Teleosts  
Antarctic teleosts  
Immunoglobulin light chain structure  
Immunoglobulin light chain evolution  
*Trematomus bernacchii*  
Real-time PCR  
Molecular models

### ABSTRACT

Three immunoglobulin light chain (IgL) isotypes *TrbeL1*, *TrbeL2*, and *TrbeL3* were identified in the Antarctic teleost *Trematomus bernacchii* by immunoscreening a cDNA expression library, and using RT-PCR, and 5' RACE. One of them was distinguished in two subisotypes *TrbeL1A* and *TrbeL1B*.

Real-time PCR experiments showed that the different isotypes were expressed in similar ratios in the various tissues analyzed. Interestingly, the expression level of *TrbeL1A* isotype was very high in all tissues.

Molecular models of the CH1–CL domain pairings were built and minimized for the different isotypes. Several differences were identified in the superimposable structures mainly in the loops. In addition, the isotype-specific residues determined a different distribution of the charges on the external CL domain surface.

Phylogenetic trees of 43 isotype representative sequences of CL domain from teleost species, built by different methods, indicated that all teleost light chain isotypes are distributed into three groups. Furthermore, the split of the group IgL1 into two subgroups, one of them carrying a micro-satellite DNA insertion, may have occurred in the Acanthopterygian ancestor.

© 2008 Elsevier Ltd. All rights reserved.

### 1. Introduction

The number of immunoglobulin light chain (IgL) isotypes present in several classes of vertebrates is generally accepted: three isotypes have been identified in amphibians ( $\rho$ ,  $\sigma$ ,  $\tau$ ), two in mammals ( $\kappa$ ,  $\lambda$ ), and only one in birds (Bengtén et al., 2000a; Pilström et al., 1998). In elasmobranchs, the most ancient jawed vertebrates possessing the immunoglobulin molecules, in addition to the three IgL isotypes previously defined (types I, II and III), a fourth has been recently identified (Criscitello and Flajnik, 2007) thus making it possible to include all vertebrate IgL isotypes in four ancestral clans.

In the past decade a large number of cDNA IgL sequences has been obtained from teleosts. The genomic organization of the IgL loci has been clarified and shown to consist in multiple clusters ( $V_L$ – $J_L$ – $C_L$ ) $_n$  (Ghaffari and Lobb, 1993; Daggfeldt et al., 1993; Timmusk et al., 2000). The combinatory mechanisms are thought to

be similar to those occurring in the IgH chain synthesis although the presence of many differently processed derivatives of sterile transcripts (non-rearranged or non-spliced) indicates that they are less controlled.  $V_L$  transcriptional orientation relative to  $J_L$  and  $C_L$  might be direct (re-arrangement by deletion) or opposite (re-arrangement by inversion). Presence of two  $J_L$  and two or three  $V_L$  segments in the same cluster has also been reported (Ghaffari and Lobb, 1993; Timmusk et al., 2000). Recently, the three L chain loci of *Danio rerio* have been found to be different to those of other teleosts: in two of them clusters carrying multiple (<6)  $V_L$  gene segments, in the same or opposite transcriptional orientation, have been identified (Hsu and Criscitello, 2006; Zimmerman et al., 2008). Finally, transcription regulatory elements (Octamer motifs, TATA box and enhancers) have been identified in several isotypes (Daggfeldt et al., 1993; Bengtén et al., 2000b; Timmusk et al., 2002).

Because of the extensive IgL diversity within the genome of each species and the very wide phylogenetic distance between teleost orders, criteria usually applied to classify IgL isotypes in homothermic species are difficult to apply to L chain sequences from teleost species. Since different numbers of IgL isotypes have been identified in different teleost species, it remains unclear at present whether all contain the same number of IgL isotypes. To date, IgL have been sequenced from 25 different species, as shown in Table 1.

The discrepancy in the number of IgL isotypes identified in each species may be explained by the incomplete screening of differently expressed isotype-specific mRNAs and by the different levels of diversity between the sequences identified in each species.

**Abbreviations:** CDR, Complementary Determining Region; CL, immunoglobulin light chain constant domain; C<sub>L</sub>, immunoglobulin light chain constant domain exon; IGH, immunoglobulin heavy chain; IGL, immunoglobulin light chain; J<sub>L</sub>, junctional immunoglobulin light chain gene segment; ML, Maximum-Likelihood; NJ, Neighbor-joining; SSC, saline tri-sodium citrate; FR, Framework Region; RACE, Rapid Amplification of cDNA Ends; RMSD, root mean square deviation; RT-PCR, reverse transcriptase-polymerase chain reaction; SSR, short sequence repeat; TBS, Tris-buffered saline; UAP, Universal Amplification Primer; UTR, Un-Translated Region;  $V_L$ , variable immunoglobulin light chain gene segment.

\* Corresponding author. Tel.: +39 0816132556; fax: +39 0816132629.  
E-mail address: u.oreste@ibp.cnr.it (U. Oreste).

0161-5890/\$ – see front matter © 2008 Elsevier Ltd. All rights reserved.  
doi:10.1016/j.molimm.2008.03.006

**Table 1**  
Immunoglobulin light chain isotypes sequenced to date in the Teleostei taxon species, as classified by NCBI taxonomy database; the accession number is reported only when any paper has been published

Superorder	Order	Species	Isotype	References
Euteleostei	Salmoniformes	<i>Oncorhynchus mykiss</i>	L1, L2, L3	Dagfeldt et al. (1993), Timmusk et al. (2000), Timmusk et al. (2002) and Partula et al. (1996)
		<i>Salmo salar</i>	L1, L2, L3	Solem and Jørgensen (2002)
		<i>Anarhichas minor</i>	L1, L2, L3	Espelid et al. (2001) and Espelid and Nygaard Grontvedt (2003)
		<i>Pagrus major</i>	L	Chen et al. (2004)
		<i>Epinephelus coioides</i>	L	[AY549459–AY549453]
		<i>Seriola quinqueradiata</i>	L	Okamoto et al. (2003)
	Perciformes	<i>Dicentrarchus labrax</i>	L1	dos Santos et al. (2001)
		<i>Sparus aurata</i>	L	[AY619988]
		<i>Siniperca chuatsi</i>	L	Zhang et al. (2003)
		<i>Pseudosciaena crocea</i>	L	[DQ862131]
		<i>Liza aurata</i>	L	[EF555085]
		<i>Sparus aurata</i>	L	[EF555066–EF555070]
	Atheriniformes	<i>Atherina boyeri</i>	L	[EF555073–EF555074]
		<i>Gadus morhua</i>	L1, L2	Dagfeldt et al. (1993) and Widholm et al. (1999)
	Gasterosteiformes	<i>Gasterosteus aculeatus</i>	L	[AY278356, BT026591]
		<i>Paralichthys olivaceus</i>	L	Kono and Sakay (2001)
	Pleuronectiformes	<i>Psetta maxima</i>	L	[DQ659628, DQ659634, DQ659636, DQ659639, DQ659651, DQ659655]
		<i>Tetraodon nigroviridis</i>	L1, L2, L3	[AJ57593–AJ57641; 804–807]
	Otocephala	Tetraodontiformes	<i>Takifugu rubripes</i>	L
<i>Cyprinus carpio</i>			L1A, L1B, L2, L3	Tomana et al. (1999, 2002) and Ishikawa et al. (2004)
Cypriniformes		<i>Danio rerio</i>	L1, L2, L3	Haire et al. (2000) and Zimmerman et al. (2008)
		<i>Ctenopharyngodon idella</i>	L	[CF751832]
		<i>Carassius auratus</i>	L	[AB201791]
Siluriformes		<i>Ictalurus punctatus</i>	F, G	Chaffari and Lobb (1993) and Lobb et al. (1984)
		<i>Ictalurus furcatus</i>	L2	[CK403931]

The functional significance of the diversity of IgL isotypes is still unknown, however recent data suggest the presence of different structural features of the antibody molecule depending on their light chain isotype (Stanfield et al., 2006). Variations in the light chain isotype expression ratio have been found in different tissue and developmental stages (Timmusk et al., 2000; Solem and Jørgensen, 2002). Moreover, individual differences in IgL expression have been reported in many species including teleosts (Espelid et al., 2001; Lobb et al., 1984), indicating that the environmental and/or genetic status might be responsible for them. Knowledge of the complete repertoire of IgL isotypes is, in turn, an essential requirement for a better understanding of the immunoglobulin light chain function.

In the present work we examined the light chain isotypes in *Trematomus bernacchii*, an Antarctic teleost of the Notothenioidei suborder (order Perciformes) and compared them to those of other teleosts in order to evaluate the evolutionary relationships between them.

## 2. Materials and methods

### 2.1. Species and sampling

*T. bernacchii* (Boulenger, 1902) belongs to the perciform suborder Notothenioidei, which is largely endemic to the Antarctic sea (*Trematomus* is a genus of the Notothenidae family). Fish specimens were collected in the seas of Victoria Land, during the XVIII expedition of the Italian National Program for Antarctic Research (2002–2003). Blood was collected by caudal venipuncture and serum was obtained by centrifugation (2000 × g for 10 min) and stored at –20 °C. The animals were killed by anesthetic (2-phenoxyethanol) overdose, afterwards spleen, headkidney, intestine, liver, skin, and muscle were collected, immediately frozen in liquid nitrogen, and stored at –80 °C.

### 2.2. Rabbit antisera

Two rabbits (RbA and RbB) were immunized with *T. bernacchii* IgL purified as previously described (Coscia et al., 2000). Each animal was injected intradermally with 100 µg of IgL in 400 µl of PBS emulsified with 500 µl of Complete Freund's adjuvant. The same amount of antigen, emulsified with Incomplete Freund's adjuvant, was injected intramuscularly, five times, every two weeks. Seven days after the last dose injection, the anesthetized specimens were bled and the sera were recovered by centrifugation at 2000 × g.

### 2.3. cDNA library immunoscreening

A *T. bernacchii* cDNA library was previously constructed (Coscia et al., 2000). Briefly, 5 µg of the affinity purified poly(A)<sup>+</sup> RNA from a single *T. bernacchii* spleen was used for cDNA synthesis. The *T. bernacchii* cDNA library was constructed using the ZAP Express™ cDNA Synthesis kit (Stratagene Cloning System, La Jolla, CA) and amplified. Nitrocellulose filters (Schleicher and Schuell, Germany) were first incubated with rabbit anti-*T. bernacchii* IgL antibodies, pre-adsorbed with *E. coli* extract (0.1 mg ml<sup>-1</sup>), and then, in sequence, with biotin-labeled goat anti-rabbit IgG (Kirkegaard and Perry Laboratories, MD) and peroxidase-labeled streptavidin (SP; Kirkegaard and Perry Laboratories). Antigen-antibody-antibody-SP complexes were detected using the chromogenic substrate 4-chloro-1-naphthol (HRP Color Development Reagent, BioRad Laboratories, Richmond, CA). After three consecutive immunoscreenings, several positive clones were identified and inserts cloned into the ZAP Express vector were excised out of the phage in accordance with the manufacturer's instructions (Stratagene).

### 2.4. RT-PCR

Total RNA was isolated from the spleen of a *T. bernacchii* specimen using the SV total RNA isolation system (Promega), and

**Table 2**  
Sequences of the oligonucleotides used as primers in RT-PCR experiments

Oligo	Sequence (5'–3')
FR2VLF	TG GTATAAACAGRYTCSAGG
L2f	TATGGTTCTGGTTTCTCATCTC
MCLr	AGGNTGCTCTCCARCTGTA
JLf	TG CAGCAGTTACTACTGTCTAGA
CL1Ar	GAACACTGGTCTCTCTCA
GSP2L1B	GAGAGGCTCAGCCCTTGAG
CL3r	CTCACACCGCTCTCTGC
CL2r	CACATTCAGACTTCTTGATGC

subjected to reverse transcription using M-MLV Reverse Transcriptase (Ambion, Austin, TX, USA). To accomplish PCR amplification of the double-stranded cDNA, the oligonucleotides listed in Table 2 were used as primers.

The forward primers, FR2VLF, MCLr and L2f were designed on conserved regions of teleost light chain gene sequences; the JLf was designed on the sequence of the *T. bernacchii* J<sub>L</sub> segments. The following isotype-specific reverse primers, complementary to the CL sequences of the sequenced IgL clones were used: CL1Ar, specific for CL1A isotype; GSP2L1B, specific for CL1B isotype; CL3r, specific for CL3 isotype; and CL2r, specific for the CL2 isotype. The amplification was performed as follows: 94 °C for 2 min, 30 cycles of 94 °C (1 min), 58 °C (or 63 °C for CL2 and CL3 isotypes) (1 min), and 72 °C (1 min) with a final extension at 72 °C for 10 min. The PCR products (about 500 bp) were then analyzed on 1% agarose gel, purified, and cloned into the pGEM-T Easy vector arms (Promega). Positive clones were sequenced on an ABI PRISM 3100 automated instrument at Primm (Naples, Italy).

## 2.5. 5' RACE

Anchored PCR was performed using a commercial kit (5' RACE System for Rapid Amplification of cDNA Ends, Invitrogen) and essentially following the manufacturer's instructions. Briefly, total spleen RNA was extracted; first-strand cDNA was synthesized using an antisense-specific primer (GSP1: 5'-GAACACTGGTCTCTCTCA-3') complementary to the 5' end of the *T. bernacchii* Ig CL isotype 1A, and Moloney murine leukemia virus reverse transcriptase (Superscript<sup>TM</sup> II, Invitrogen). A nested gene-specific 3' primer (GSP2L1A: 5'-GCCGAGTCTGGGAGCGG-3'), and the manufacturer's 5' Universal Amplification Primer (UAP) were used to accomplish PCR amplification of the double-stranded cDNA. Following amplification, 5' RACE products were cloned into pCRII-TOPO vector (Invitrogen).

## 2.6. Real-time PCR

The real-time PCR experiments were performed using the iCycler thermal cycling system (Bio-Rad). Five micrograms of total RNA extracted from each tissue were reverse-transcribed with oligo(dT) as primer and equivalent aliquots of the first strand cDNA reactions were then amplified. iQ SYBR Green Supermix (Bio-Rad) was used in a volume of 25 µl with 1 µl of first strand cDNA mixture and final concentration of 0.1 mM of each primer, following the manufacturer's instructions. Specific PCR primers were designed for the amplification of about 100–160 bp products from Ig light chain isotypes, *T. bernacchii* Ig heavy chain (AF094531) and β-actin (AJ532569) from a very close species *Chionodraco hamatus*, used as housekeeping gene. The primers used are listed in Table 3. The PCR cycle conditions were: 95 °C for 10 min, followed by 40 cycles of 95 °C for 45 s, 60 °C for 45 s and 72 °C for 45 s. Triplicate reactions were performed for each pair of isotype-specific primers and the template was replaced with water in all blank control reactions. In

**Table 3**  
Sequences of the oligonucleotides used as primers in Real-time PCR experiments

	Primer specificity	Sequence (5'–3')
1B11fw	L1A	TCTGCAGATCAGTGGAGGAAG
1B11rev		GGACTGGGAACACTGGTCTCT
219bfw	L1B	CAGCATCTGAGCTCTCTCG
219brev		GGTCCAGGCTTTGAGTGACA
440fw	L2	CAACCCGACAGACTTTTCAA
440rev		GGAGCCCTGAGTCACTTACA
323fw	L3	CTGACCGCTACTCTCTGA
323rev		CTGTAGCTCCGCTCTCTC
ChBACTfw	Beta-actin	CCCAGATCATCTCCGAGACC
ChBACTrev		CAGAGACGGACAGGAAG

all real-time PCR runs, melting curve analyses were performed and single specific melting peaks were observed indicating amplification specificity (data not shown). The specificity of the real-time PCR products was further confirmed by electrophoresis using a 2% (w/v) agarose gel. In all cases, the amplifications were specific and no amplification was observed in negative controls. Data were expressed as the mean value of triplicates; β-actin was used as the normalizer.

## 2.7. Computational methods

The percentage of nucleotide (nt) identity between *T. bernacchii* IgL sequence pairs was calculated with the LAJUN program, implementing the Huang and Miller algorithm (Huang and Miller, 1991). To translate the nucleotide sequences to deduced amino acid sequences the DNA Strider 1.1 program was used. The percentage of identity between *T. bernacchii* amino acid sequences was calculated using the SIM program. Sequence alignments were obtained with the ClustalX 1.8 program (Thompson et al., 1997) and the multiple sequence alignment was optimized manually. Site-specific diversity of the aligned amino acid sequences was studied using a program estimating Shannon entropy at each residue position (Stewart et al., 1997). Kinase-specific phosphorylation sites were predicted using the NetPhos 2.0 program (Blom et al., 1999).

Three-dimensional models of the CH1–CL pairings, where the CL domains were encoded by clones 1B11, 5141, 219, and L440, each representative of one of the IgL isotypes, were built using standard homology modeling techniques (Morea et al., 1998). Briefly, the protein sequence database at the NCBI (<http://www.ncbi.nlm.nih.gov/>) was searched with the BLAST program (Altschul et al., 1997) using the sequences of the target domains as queries. The proteins of known structure, providing the best hits to the domain queries, were downloaded from the Research Collaboratory for Structural Bioinformatic Protein Data Bank (Berman et al., 2000) and used as templates to model the main-chain of the conserved regions in the target-template alignment. All the steps required for structural analyses and model building were performed using the program Insight II (Dayringer et al., 1986) and its Biopolymer module (Accelrys Inc.). The main-chain atoms of loop regions were modeled using structural database searching techniques (Jones and Thirup, 1986). Side-chains of conserved amino acids between target and template were imported from the template. The conformation of the mutated residues was chosen from a rotamer library available with the program Insight II (Accelrys Inc.). The minimization procedure was performed using GROMACS 3.2 package (Van Der Spoel et al., 2005) and the GROMOS96 force field. The starting models were immersed in different boxes containing about 23,000 simple-point-charge water

molecules (Berendsen et al., 1986). The ionization state was set to mimic a neutral pH environment. The overall charge of each system was neutralized by adding the appropriate number of ions more than 7 Å away from the protein surface. In the first step of the equilibration process, the solvent was relaxed by energy minimization followed by a 20 ps Molecular Dynamics at 300 K. The overall system was then minimized without restraints. The electrostatic potential was calculated by Swiss-PdbViewer, with the Coulomb method, considering only the entire net charges and using the default dielectric constant.

Distance-based tree reconstruction was performed with the Neighbor-Joining (NJ) method as implemented in Mega 3.1 (Kumar et al., 2004). The JTT distance option was used, and positions containing indels were excluded in all comparisons.

Maximum-Likelihood (ML) analyses were as follows. The best-fitting substitution model was selected using ProtTest (Abascal et al., 2005). The resulting optimal model (WAG+I+G+F) (Whelan and Goldman, 2001), with proportion of invariant sites equal to 0.05, and four categories of rate with parameter alpha equal to 4.14 was implemented in Phylml 2.4.4 (Guindon and Gascuel, 2003).

Under both NJ and ML methods, a non-parametric bootstrap approach with 1000 replicates was used to test the robustness of the obtained tree.

### 2.8. Accession numbers

The *T. bernacchii* IgL sequences reported in this paper have been deposited in EMBL/GenBank database under accession numbers DQ842620–DQ842635, EF114784–EF114791, and EF421435–EF421439.

## 3. Results

### 3.1. Clones from a cDNA expression library

The *T. bernacchii* spleen expression library was screened with the rabbit antiserum RbA specific for the *T. bernacchii* IgL chain. Forty positive clones were isolated and sub-cloned. By a second immunoscreening, 19 clones were selected and subjected to further cloning. After the third screening, 8 clones, which showed the longest insert length, were amplified and sequenced. Three clones (1B11, 4932, 5141, GenBank # DQ842627, DQ842622, and DQ842626, respectively) were found to be very different to each other: amino acid identity of the overlapping regions was 54.7% between clones 1B11 and 4932, 33.0% between clones 1B11 and 5141, and 39.8% between clones 4932 and 5141, suggesting that these three clones are representatives of distinct isotypes. Because of their homology with IgL isotypes of other teleosts, resulting from BLAST search, 1B11 and 4932 clones defined the *TrbeL1* isotype, but distinguishing two subisotypes (*TrbeL1A*, *TrbeL1B*, respectively), while the clone 5141 was assigned to the *TrbeL3* isotype.

### 3.2. cDNA clones by PCR

To obtain a larger number of clones for each isotype, a PCR approach was chosen using pairs of oligonucleotides as primers, designed on *T. bernacchii* isotype-specific JL and CL sequences. The obtained PCR products were cloned: 13, 2 and 2 clones were obtained using the *TrbeL1A*, *TrbeL1B* and *TrbeL3* specific primers, respectively. In addition a transcript (GenBank # EF114791) was considered sterile because lacking part of the CL.

We also carried out 5' RACE using the GSP1 oligonucleotide primer, which was complementary to the 5' end of the *TrbeL1A* isotype sequence. The clone L180, obtained by cloning the 5' RACE products, was sequenced and found to comprise 58 nucleotides in

**Table 4**

Mean values of amino acid identity percentage within the CL region sequences belonging to the same or different isotype

	<i>TrbeL1A</i>	<i>TrbeL1B</i>	<i>TrbeL2</i>	<i>TrbeL3</i>
<i>TrbeL1A</i>	91.7	57.9	37.2	34.0
<i>TrbeL1B</i>		95.7	34.6	38.1
<i>TrbeL2</i>			94.7	33.4
<i>TrbeL3</i>				95.2

the 5' UTR. Its CL domain amino acid sequence shared 88.2–98.6% of its identity with the previously obtained *TrbeL1A* isotype clones.

### 3.3. Search for the *TrbeL2* isotype

Since no sequence had resulted from the immunoscreening that could be attributed to a *T. bernacchii* isotype ortholog to the IgL2 isotype described in several species, we followed a different strategy. Based on the alignment of IgL2 sequences from *Danio rerio*, *Cyprinus carpio*, *Salmo salar* and *Oncorhynchus mykiss*, we designed an oligonucleotides pair on conserved regions. The oligonucleotides L2f, designated at the 3' end of FR3 (5'-TATGGTTCTGGTTTCTCATCTC-3') and CL2r at the 3' end of the CL (5'-CACATTCAGACTTCTGATGC-3'), were used as primers in RT-PCR experiments. The PCR products were cloned and 8 clones were sequenced. Two out of 8 were found to be identical; the remaining ones shared between 95.6 and 99.8% of nucleotide identity. In contrast these sequences showed on average 51.6, 48.6 and 52.2% of nucleotide identity with the *TrbeL1A*, *TrbeL1B*, and *TrbeL3* sequences, respectively. When searching databases by BLAST, these sequences showed the highest scores of similarity with *O. mykiss* IgL isotype 2, and thus were defined as belonging to the *TrbeL2* isotype.

### 3.4. Analysis of cDNA clones

A total of 28 clones were sequenced and 18 were attributed to the *TrbeL1* (15 to the *TrbeL1A* and 3 to *TrbeL1B*), 6 to *TrbeL2*, 4 to *TrbeL3* isotype. The mean value of percentage of amino acid identity within the same isotype and between different isotype CL domains are reported in Table 4.

All clones, except clone 4932, encoded partial or entire VL domains. The variability plot for all sequenced clones was calculated using the Shannon entropy method (Stewart et al., 1997). Based on comparison with other teleost IgL sequences, the contribution of VL and JL boundaries were defined for each clone. The histogram in Fig. 1 shows the entropy value for each position.

Analysis of the plot showed evidence for highly variable positions in the CL as well as in the VL domains.

Deduced amino acid sequences of the CL domain encoded by all clones were aligned by ClustalX and the relative distance tree was built by the Neighbor-Joining method (Fig. 2).

Sixteen out of 116 positions were found to be conserved throughout the sequence. By multiple alignment, L3 constant domains resulted in two or one additional amino terminal residues with respect to L1 or L2 constant domains, respectively. The contribution of the L1 constant domain to the molecular weight, calculated on the deduced amino acid sequences was on average about 10 and 4% higher than those of the corresponding CL2 and CL3 domains, respectively. When the sequences were analyzed by the NetPhos 2.0 program, many serine residues were predicted to be potential kinase-specific phosphorylation sites, with score values up to 0.99. The KATLVCL stretch, which is well conserved in the C<sub>γ</sub> and T-cell receptor C<sub>β</sub> domains of many vertebrate species (André et al., 2000), is conserved only in the *TrbeL1A* isotype. In particu-

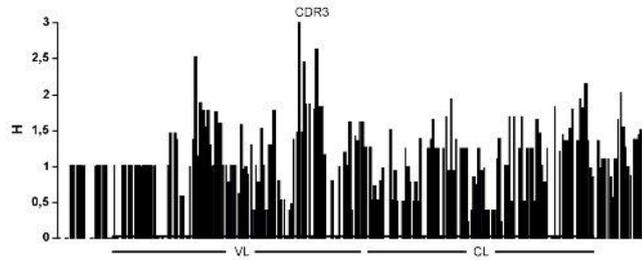


Fig. 1. Variability plot of *T. bernacchii* IgL sequences. The variability was calculated according to the Shannon entropy method. Entropy values are reported on the y-axis and the positions of the amino acid residues are on the x-axis.

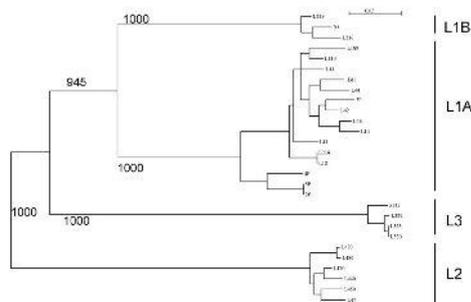
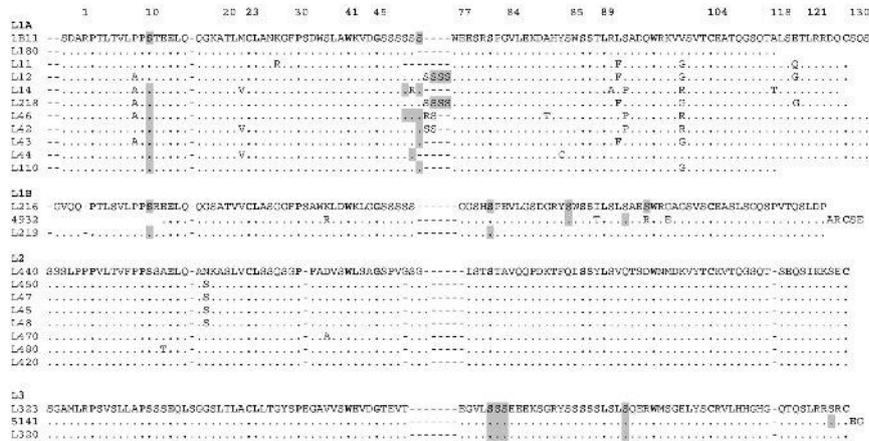


Fig. 2. At the top: amino acid sequences of the constant regions of *T. bernacchii* IgL cDNA clones aligned with ClustalX. Sequences are grouped into L1A, L1B, L2 and L3 isotypes. Identical residues are denoted by dots and gaps are indicated by bars. Potential phosphorylation sites showing the highest scores (>0.99 according to NetPhos 2.0 program) are shaded. Sixteen conserved residues are in bold. Amino acid residue position numbering is according to IMGT unique numbering for C-domain (Lefranc et al., 2005); the positions indicating the C-domain universal residues are in bold. On the bottom: distance tree, calculated by the Neighbor-joining method of *T. bernacchii* IgL sequences; the bootstrap incidences, relative to the major clades, are indicated on the branches.

```

1      10
JL1A
L11  YTFGGGOTRLDVG
L12  .....
L14  .....
L1B11
2P  L.....
3P  .....
4P  .....BP.
5P  L.....
L218 .....
L46  .....
L42  .....
L44  .....
JL1B
L216 .....K.IID
L219 .....K.IID
JL2
L450 WV...Q...K.IVT
L440 EV...Q...K.IVT
L47  WV...Q...K.IVT
L48  WV...Q...K.IVT
L49  WV...Q...K.IVT
JL3
5141 E.....K.IPH
L320 WP.....K.IPH
L323 WP.....K.IPH
L329 WP.....K.IPH

```

**Fig. 3.** ClustaX amino acid sequences alignment of the regions of *T. bernacchii* IgL cDNA clones presumably encoded by  $\lambda$  segments. Sequences are grouped into L1A, L1B, L2 and L3 isotypes. Residues identical to clone L11 are denoted by dots. Five conserved residues are in bold.

lar, a stretch of serines 3–10 residues long, was typically observed for both L1 and L2 isotype sequences encompassing positions 48 through 57. All serines present in the isotype *TrbL1A* stretch were encoded by the same codon (AGC) whereas in the *TrbL2* sequences, all showing five residues, were encoded by different codons.

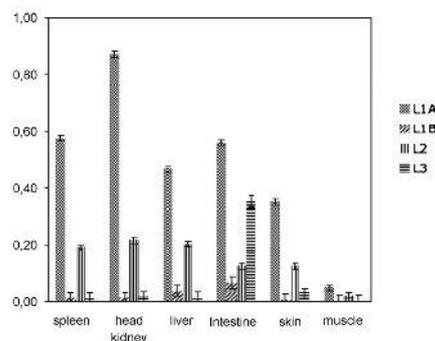
Isotype distribution was also assessed for the  $\lambda$ L segments, since JL1A, JL1B, JL2 and JL3 sequences were distinguished, as shown in Fig. 3. Of all  $\lambda$ L segments identified, those belonging to isotype L2 and L3 were most similar to each other.

### 3.5. Expression analysis of IgL isotypes in different tissues

The expression levels for all the IgL isotypes were analyzed in several *T. bernacchii* tissues by real-time PCR. Each experiment was performed in triplicate, and, to confirm amplification specificity, the PCR products from each primer pair were subjected to a melting curve analysis to verify that the product had a single melting peak, followed by polyacrylamide gel electrophoresis. The IgL isotype expression levels were normalized to  $\beta$ -actin levels and are shown in Fig. 4. The overall expression levels of the three isotypes were found to be very low in muscle, and very high in head/kidney. The highest expression level was found to be that of *TrbL1A* isotype followed by *TrbL2* isotype, whereas the lowest level was that of *TrbL3* isotype except in the intestine where it was significantly high.

### 3.6. Molecular models

To investigate the structural features of different isotypes, we built four molecular models of the CH1–CL pairing, each compris-



**Fig. 4.** Expression profiles of *TrbL1A*, *TrbL1B*, *TrbL2*, and *TrbL3* isotypes in different tissues obtained by real-time PCR experiments. Expression levels were expressed as a ratio relative to  $\beta$ -actin levels in each tissue. Data reported are the average value between three experiments.

**Table 6**

Root mean square deviation, expressed in Å, of the  $C_{\alpha}$  atoms between the different model structures

	L219 ( <i>TrbL1B</i> )	L440 ( <i>TrbL2</i> )	5141 ( <i>TrbL3</i> )
1B11 ( <i>TrbL1A</i> )	1.32	1.54	1.53
L219 ( <i>TrbL1B</i> )		1.40	1.62
L440 ( <i>TrbL2</i> )			1.65

ing a CL domain sequence representative of a different isotype. The appropriate templates, shown in Table 5, were selected, and molecular models were constructed by using standard homology modeling techniques. The obtained models were validated by minimization in neutralized and relaxed solvent using the GROMACS 3.2 package.

The resulting structures were characterized, and found to be similar in terms of CH1–CL domain pairing compactness, measured as distance between the sulfur atoms of the cysteines responsible for the interdomain disulfide bonds in the different models. In fact this value varied from 18.5 and 19.6 Å. The reciprocal orientation of the CH1 and CL domains was calculated by measuring the angles between the main axis of the two domains in the four models; the values were similar among the models (about  $120^{\circ}$ ).

The CL domains were analyzed in more detail in the models. They showed similar dimensions, being all insertable in the same box ( $43 \text{ \AA} \times 42 \text{ \AA} \times 27 \text{ \AA}$ ), but differed in the electrostatic charge of the system, which was 0, –2, –1, –3 for isotypes *TrbL1A*, *TrbL1B*, *TrbL2*, and *TrbL3*, respectively. The charge distribution also appeared to be different as seen in Fig. 5.

When superimposing the CL domains of the four models, the RMSD of the  $C_{\alpha}$  atoms varied from 1.32 and 1.65 Å (Table 6). Despite the low RMSD values, the backbones were found divergent in several loops (Fig. 6). In particular, the loop connecting the D and E strands was found extruded in the *TrbL1A* CL domain. Finally,

**Table 5**

Structures used as template to build the molecular models of CH1–CL domains pairing of sequences representative of each *Tb*IgL isotype

Target	Isotype	Template	Species	PDB code	Resolution (Å)	Sequence identity (%)	Insertion/deletion (%)	E-value
1B11	<i>Tb</i> IgL1A	Part of Fab 77	Mouse	1L7T	2.10	38	9	$9e-14$
L219	<i>Tb</i> IgL1B	Part of KOL antibody	Human	2IG2	3.00	39	.	$9e-8$
L440	<i>Tb</i> IgL2	Part of Fab HYB3	Human	1W72	2.15	44	1	$8e-14$
5141	<i>Tb</i> IgL3	Part of Fab SE1554	Mouse	1MFE	2.00	41	2	$2e-11$

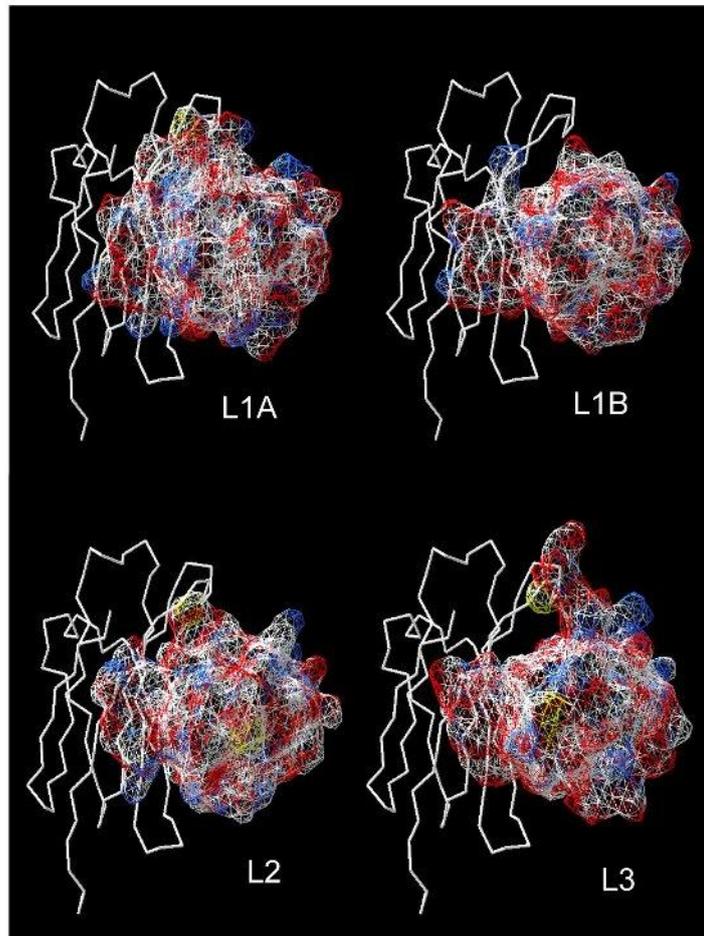


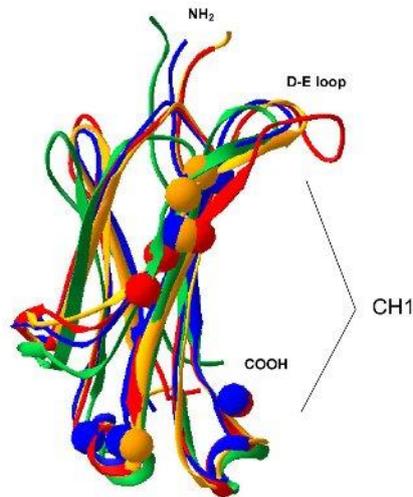
Fig. 5. Charge distribution on the CL surface of the CH1–CL pairing for the different isotypes; positive net charges are in blue, negative in red; the backbone of the CH1 domain is in white.

when analyzed, the putative phosphorylation sites were found to be distributed in three regions (Fig. 6).

### 3.7. Comparative analysis of teleost CL domains

To clarify the relationships between *T. bernacchii* isotypes and those of other teleost IgL, we constructed a phylogenetic tree using the Maximum-Likelihood (ML) method. All the CL sequences available for each species were aligned; the different isotypes were identified; a consensus sequence was determined for each isotype

and the sequence closest to the consensus was then used as isotype representative for the comparative analysis. Based on this criterion a multiple alignment of 43 CL sequences, each representative of a different isotype, from 25 teleost species, was obtained by ClustalX 1.8. To construct the tree, the positions where insertions/deletions were present were excluded. Results are shown in Fig. 7, part A. The CL sequences were distributed in three clades, each corresponding to a different isotype group (L1, L2, L3), confirming that three IgL isotypes are present in teleosts. The distinction of L1 in two subgroups is a specific feature of the Acanthopterygii. In fact, the



**Fig. 6.** Superimposition of the molecular models of the constant domain of L218 (isotype L1A, in red), L219 (isotype L1B, in blue), L440 (isotype L2, in green), 5141 (isotype L3, in orange) sequences. Presumptive phosphorylated serine residues are marked by balls. Amino and carboxi-terminal ends and the loop connecting the D and E  $\beta$ -strands are indicated. The heavy chain pairing side is on the right.

**Table 7**

Amino acid identity between the each isotype representative *T. bernacchii* constant domain sequence with CL sequences of other teleost, belonging to the same group

	Isotype group	Percent of amino acid identity	
		Mean value	Range
L811	CL1A	76.1	54.1–94.0
L216	CL1B	73.7	68.6–81.0
L440	CL2	58.0	49.1–73.9
L323	CL3	61.5	36.2–79.8

L1 sequences from orders such as Cypriniformes, Salmoniformes, Gadiformes and Siluriformes, belonging to other superorders, were found out of both subgroup clades (Fig. 7A).

Subsequently, to better investigate the Acanthopterygian subspecies divergence, we aligned isotype representative sequences from only the Acanthopterygian species by ClustalX (Fig. 8). Range and mean value of the amino acid identity between *T. bernacchii* and other teleost CL isotypes belonging to the same group are shown in the Table 7.

Fifteen out of the 16 conserved positions in *T. bernacchii* sequences were shared by all the teleost CL sequences analyzed. On the contrary, *T. bernacchii* sequences belonging to groups CL1A, CL1B, CL2 and CL3 differed from those of all the other species by 5, 10, 4, 8 positions, respectively. It is to be noted that the amino acid differences seen, did not generally imply a different character of the side chain.

Using the same alignment as mentioned above, and excluding the positions where gaps occurred, a distance tree was constructed by the Neighbor-joining method. The result shown in Fig. 7, part B, confirmed that obtained using the Maximum-Likelihood method.

#### 4. Discussion

In the present work we have identified light chain isotypes from the perciform teleost *T. bernacchii* a model species we have chosen in the framework of studies on Antarctic fish immunity. By sequencing 28 different cDNA clones we analyzed the isotype distribution, and evaluated specific sequential and structural features for each isotype.

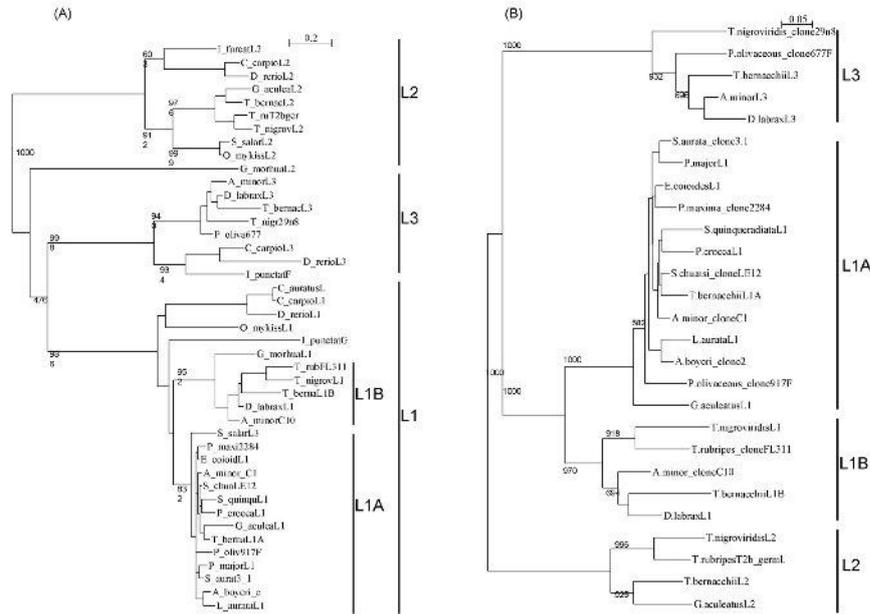
The VL–JL–CL cluster organization, a common trait of the IgL chains of all the so far investigated teleost species, is shared by *T. bernacchii* as we found recombination between isotype-specific JL and CL gene segments.

The variability degree of the IgL chain amino acid sequences encoded by different clusters is certainly a measure of the complexity of the overall immunoglobulin structure. In a previous paper we reported that *T. bernacchii* IgL chains were found to be highly heterogeneous since their isoelectric point range was very wide, probably as a result of the gene cluster organization (Pucci et al., 2003). To assess the variability of the 28 deduced amino acid sequences, a Shannon entropy plot was calculated. The incompleteness of many IgL sequences analyzed in the present work did not enable a correct estimation of the contribution of the amino terminal portion of the VL domain to the variability of the entire L chain. High values were observed in the constant domain exceeding 2 (2.14) at the beginning of strand F.

Our experiments provide unequivocal evidence for the presence of three distinct isotypes: the cDNA sequences clearly fell into three groups *TrbeL1*, *TrbeL2* and *TrbeL3* (Fig. 2B), their constant domains sharing high amino acid identity (91.7–95.2%) within the same isotype, but quite low identity between different isotypes (33.4–37.2%). *TrbeL1* constant domain sequences can be divided into two subgroups, *TrbeL1A* and *TrbeL1B*, depending on their relative sequence homology and on the presence/absence of a putative micro-satellite DNA insertion. The distinction of the two L1 subspecies, further supported by significantly high bootstrap values, might suggest that this event, probably derived by a gene locus duplication, dated about 90 MYA, that is the time evaluated for the Acanthopterygian divergence.

To verify whether specific isotypes were differentially expressed in different tissues we performed real-time PCR using primer pairs specific for each isotype and cDNA from spleen, headkidney, liver, intestine, skin, and muscle; the  $\beta$ -actin gene transcript was used as calibrator. Similar expression levels of the three isotypes in the tissues analyzed do not allow us to hypothesize any isotype-specific functionality. We found that the transcripts belonging to the *TrbeL1A* isotype were very abundantly expressed in all the tissues analyzed. This is in agreement with the high occurrence of sequences belonging to this isotype in our data set (15/28); it could be due to a higher number of *TrbeL1A* clusters and/or to a more efficient transcriptional activity. The higher diversity of the nucleotide sequence belonging to this isotype suggests the occurrence of a higher number of clusters. The predominance of this isotype in other Acanthopterygian species also suggests an evolutionary success due to specific structural features.

The *T. bernacchii* L1A is characterized by a serine stretch, starting at position 45.1 (IMGT numbering) (Lefranc et al., 2005) which might be encoded by insertion of a micro-satellite DNA, consisting in tandem repeats of the AGC triplet. It is to be noted that this codon is the most frequently occurring (40%) in the short sequence repeats (SSR) of vertebrate gene exons (Li et al., 2004). The variable length of this stretch in different clusters is in agreement with the dynamic behavior of these genomic elements. In fact it has been demonstrated that in SSRs, mutational events consisting in elongation/shortening are ten-fold more frequent than point mutational events (Borstnik and Pumpernik, 2002). The molecular



**Fig. 7.** (Part A) Phylogenetic tree of CL sequences representative of different isotypes of teleost species, obtained by the Maximum-Likelihood analysis; the bootstrap incidences are indicated on the branches. The sequences shown are from the following species under the accession numbers reported in parenthesis: *Anathichthys minor*: clone c10 (AF137397), clone c1 (AF137398) and L3 (Espelid and Nygaard Grontvedt, 2003, p. 318); *Athenina boyeri* clone 2 (EF555074); *Carassius auratus* (AB201791); *Cyprinus carpio*: L1 clone Cpb4 (AB015905), L2 clone Cpj3 (AB091120), L3 Cpd1 (AB035730); *Danio rerio*: L1 clone LC1.8 (AF246185), L2 clone NLC3 (AF246162), L3 clone LC1.7 (AF246193); *Dicentrarchus labrax*: L1 (AJ400216), L3 (DQ361010); *Epinephelus coioides* L1 (AY549453); *Gadus morhua*: L1 (X68516), L2 (AJ293808); *Gasterosteus aculeatus*: L1 (AY278356), L2 (BT026591); *Ictalurus punctatus*: F (U25705), G (L25533); *Ictalurus furcatus* L2 (CK403931); *Lea aurata* L1 (EF555085); *Oncorhynchus mykiss*: L1 (X68522), L2 (U69987); *Pagrus major* L1 (AY190706); *Paralichthys olivaceus*: L3 clone 677F (AL091026), L1 clone 917F nt 445–441 (AL091142); *Pseudosciaena crocea* L1 (DQ862131); *Psetta maxima* clone 2284 (DQ659655); *Salmo salar*: L2 (AF297518), L3 clone III-4 (AF406956); *Seriola quinqueradiata* L1 (AB062619); *Siniperca chuatsi* clone LE12 (AF454468); *Sparus aurata* clone 3.1 (EF555068); *Takifugu rubripes*: L1 clone FL311 (AB126061), L2 T2b germline (DQ471454); *Tetodon nigroviridis*: L1 (CR721909), L2 (AJ575637), L3 (AJ575806); *Tramatonus bernacchii*: L1A (DQ842627), L1B (DQ842620), L2 (EF114785), L3 (DQ842626). (Part B) Distance-based tree, performed with the Neighbor-Joining method of CL sequences of only Acanthopterygian species. They are the same as in Part A.

models we built indicated that, as a consequence of this insertion, the next loop, connecting the strands D and E, resulted lengthened (Fig. 6). The micro-satellite insertional event might have occurred before Acanthopterygian radiation; in fact this tandem repeats sequence is shared by different Acanthopterygian orders (Perciformes, Atheriniformes, Gasterosteiformes and Pleuronectiformes) and absent in the Paracanthopterygian (Gadiformes) as well as Proacanthopterygian (Salmoniformes) teleost orders. Among all the sequences determined, the longest SSR consists of 13 AGC triplets, and it is present in the IgL sequence of clone 30 from *Sparus aurata* (GenBank EF555067).

The highest serine content, despite the insertion of many serine codons in the *TrbeL1A* isotype, was found in the *TrbeL3* constant domain, up to 19.2% on average. This is the highest value seen so far in any other species. For instance, in human  $\kappa$  and  $\lambda$  isotypes, the serine content is 15.0 and 13.2%, respectively. A specific feature of several serine residues of both *TrbeL1* bisotypes and *TrbeL3*, is the high propensity to be sites for kinase phosphorylation, as predicted by the NetPhos 2.0 tool considering the high stringency score >0.99. No threonine or tyrosine residues were predicted as kinase targets. In our molecular models, all the identified phosphorylation sites

are exposed to the solvent, which may be important in increasing the solubility of the antibody molecules at the low temperatures that the *T. bernacchii* species has adapted to.

With respect to differences in the amino acid sequence among different isotypes, we found 16 out of 116 positions conserved. Most of them were located in the strands. Six out of 16 positions are universal among IgL sequences: they include cys23 and cys104 involved in the intradomain disulfide bond, cys129 responsible for the binding to the H chain, trp41, leu89 and leu121. Several positions have been found to be isotype specific: 26 positions for the *TrbeL1A* isotype, 25 for the *TrbeL1B*, 50 for the *TrbeL2* and 44 for the *TrbeL3*; the majority of them are located in loops.

The length of the amino acid stretch connecting VL and CL domains was found to be different among isotypes, the *TrbeL3* and *TrbeL2* two or one residues longer than in *TrbeL1A* and *TrbeL1B*, respectively. This region has been demonstrated to be crucial in the rearrangement of antibody domains upon antigen binding (Sottriffer et al., 2000). It has been suggested that the different lengths of mammalian  $\kappa$  and  $\lambda$  IgL in this region are responsible for the different Fab mobilities of IgG (Stanfield et al., 2006).

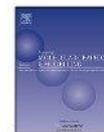


- Criscitello, M.F., Flajnik, M.F., 2007. Four primordial immunoglobulin light chain isotypes, including  $\lambda$  and  $\kappa$ , identified in the most primitive living jawed vertebrates. *Eur. J. Immunol.* 37, 2683–2694.
- Daggefeldt, A., Bengtén, E., Pilström, L.A., 1993. Cluster type organization of the loci of the immunoglobulin light chain in Atlantic cod (*Gadus morhua* L.) and rainbow trout (*Oncorhynchus mykiss* Walbaum) indicated by nucleotide sequences of cDNAs and hybridization analysis. *Immunogenetics* 38, 199–209.
- Dayringer, H.E., Tramontano, A., Sprang, S.R., Fletterick, R.J., 1986. Interactive program for visualization and modeling of protein, nucleic acid and small molecules. *J. Mol. Graph.* 4, 82–87.
- dos Santos, N.M.S., Hermsen, T., Rombout, J.H.W.M., Pilström, L., Stet, R.J.M., 2001. Ig light chain variability in DNP<sub>254</sub>-KLH immunised sea bass (*Dicentrarchus labrax* L.): evidence for intra-molecular induced suppression. *Dev. Comp. Immunol.* 25, 387–401.
- Espeli, S., Halse, M., Solem, S.T., Jørgensen, T.Ø., 2001. Immunoglobulin genes and antibody response in the spotted wolf fish (*Anarhichas minor* Olafsen). *Fish Shellfish Immunol.* 11, 399–413.
- Espeli, S., Nygaard Grontvedt, R., 2003. Immunoglobulin VH families and light chain isotypes in the spotted wolffish (*Anarhichas minor* Olafsen). *Fish Shellfish Immunol.* 15, 311–323.
- Ghaffari, S.H., Lobb, C.J., 1993. Structure and genomic organization of immunoglobulin light chain in the channel catfish: An unusual genomic organizational pattern of segmental genes. *J. Immunol.* 151, 6900–6912.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Haire, R.N., Ras, J.P., Litman, R.T., Litman, G.W., 2000. Characterization of three isotypes of immunoglobulin light chains and T-cell antigen receptor  $\alpha$  in zebrafish. *Immunogenetics* 51, 915–923.
- Hsu, E., Criscitello, M.F., 2006. Diverse immunoglobulin light chain organizations in fish retain potential to revise B cell receptor specificities. *J. Immunol.* 177, 2452–2462.
- Huang, X.Q., Miller, W., 1991. A Time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* 12, 337–357.
- Ishikawa, J., Imai, E., Morimoto, T., Nakao, M., Yano, T., Tomana, M., 2004. Characterization of a fourth immunoglobulin light chain isotype in the common carp. *Fish Shellfish Immunol.* 16, 369–379.
- Jones, T.A., Thirup, S., 1986. Using knowsubstructures in protein model building and crystallography. *EMBO J.* 5, 819–822.
- Kono, T., Sakay, M., 2001. The analysis of expressed genes in the kidney of Japanese flounder, *Paralichthys olivaceus*, injected with the immunostimulant peptidoglycan. *Fish Shellfish Immunol.* 4, 357–366.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics and sequence alignment. *Briefing Bioinform.* 5, 150–163.
- Lefranc, M.P., Pommie, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., Jean, C., Ruiz, M., Da Piedade, I., Rouard, M., Foulquier, E., Thouvenin, V., Lefranc, G., 2005. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.* 29, 185–203.
- Li, Y.C., Karol, A.B., Fahima, T., Nevo, E., 2004. Microsatellites within genes: structure, function and evolution. *Mol. Biol. Evol.* 21, 991–1007.
- Lobb, C.J., Olson, M.O., Clem, W., 1984. Immunoglobulin light chain classes in a teleost fish. *J. Immunol.* 132, 1917–1923.
- Morea, V., Leplae, R., Tramontano, A., 1998. Protein structure prediction and design. *Biotechnol. Annu. Rev.* 4, 177–214.
- Okamoto, K., Ikemura, H., Savan, R., Sakay, M., 2003. Cloning, sequence and variability analysis of expressed immunoglobulin light chain genes from yellowtail *Seriola quinqueradiata*. *Fish Shellfish Immunol.* 14, 55–70.
- Partula, S., Schwager, J., Timmusk, S., Pilström, L., Charlemaigne, J., 1996. A second immunoglobulin light chain isotype in the rainbow trout. *Immunogenetics* 45, 44–51.
- Pilström, L., Lundqvist, M.L., Wermenstam, N.E., 1998. The immunoglobulin light chain in poikilothermic vertebrates. *Immunol. Rev.* 166, 123–132.
- Pucci, B., Coscia, M.R., Oreste, U., 2003. Characterization of serum immunoglobulin M of the Antarctic teleost *Trematomus bernacchii*. *Comp. Biotech. Physiol. Part B* 135, 349–357.
- Saha, N.R., Suetake, H., Suzuki, Y., 2004. Characterization and expression of the immunoglobulin light chain in the fugu: evidence of a solitary type. *Immunogenetics* 56, 47–55.
- Solem, S.T., Jørgensen, T.Ø., 2002. Characterization of immunoglobulin light chain cDNAs of the Atlantic salmon, *Salmo salar* L.: evidence for three IgL isotypes. *Dev. Comp. Immunol.* 26, 635–647.
- Sottriffer, C.A., Rode, B.M., Varga, J.M., Liedl, K.R., 2000. Elbow flexibility and ligand-induced domain rearrangements in antibody Fab N6.8: large effects of a small hapten. *Biophys. J.* 79, 614–628.
- Stanfield, R.L., Zemla, A., Wilson, I.A., 2006. Antibody elbow angles are influenced by their light chain class. *J. Mol. Biol.* 357, 1566–1574.
- Stewart, J.J., Lee, C.Y., Ibrahim, S., Watts, P., Shlomchik, M., Weigert, M., Litwin, S., 1997. A Shannon entropy analysis of immunoglobulin and T cell receptor. *Mol. Immunol.* 34, 1067–1082.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Timmusk, S., Partula, S., Pilström, L., 2000. Different genomic organization and expression of immunoglobulin light-chain isotypes in the rainbow trout. *Immunogenetics* 51, 905–914.
- Timmusk, S., Strömberg, S., Pilström, L., 2002. Light chain promoter regions in rainbow trout *Oncorhynchus mykiss*: function of a classical and atypical Ig promoter. *Dev. Comp. Immunol.* 26, 785–796.
- Tomana, M., Nakao, M., Morimoto, T., Fujiki, K., Yano, T., 1999. Isolation of cDNA encoding immunoglobulin light chain from common carp (*Cyprinus carpio* L.). *Fish Shellfish Immunol.* 9, 71–80.
- Tomana, M., Ishikawa, J., Imai, E., Morimoto, T., Nakao, M., Yano, T., 2002. Characterization of immunoglobulin light chain isotypes in the common carp. *Immunogenetics* 54, 120–129.
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J., 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718.
- Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a Maximum-Likelihood approach. *Mol. Biol. Evol.* 18, 691–699.
- Widholm, H., Lundbäck, A.S., Daggefeldt, A., Magnadottir, B., Warr, G.W., Pilström, L., 1999. Light chain variable region diversity in Atlantic cod (*Gadus morhua* L.). *Dev. Comp. Immunol.* 23, 231–240.
- Zhang, Y.A., Nie, P., Luo, H.Y., Wang, Y.P., Sun, Y.H., Zhu, Z.Y., 2003. Characterization of cDNA encoding immunoglobulin light chain of the mandarin fish (*Siniperca chuatsi*). *Vet. Immunol. Immunopathol.* 95, 81–90.
- Zimmerman, A.M., Yeo, G., Howe, K., Maddox, B.J., Steiner, L.A., 2008. Immunoglobulin light chain (IgL) genes in zebrafish: genomic configurations and inversional rearrangements between (V<sub>L</sub>-J<sub>L</sub>-C<sub>L</sub>) gene clusters. *Dev. Comp. Immunol.* 32, 421–434.



Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: [www.elsevier.com/locate/JMGM](http://www.elsevier.com/locate/JMGM)

## Structure and dimerization of the teleost transmembrane immunoglobulin region

Antonello Merlino<sup>a,\*</sup>, Sonia Varriale<sup>a,b</sup>, Maria Rosaria Coscia<sup>b</sup>, Lelio Mazzarella<sup>a,c</sup>, Umberto Oreste<sup>b</sup>

<sup>a</sup> Department of Chemistry, University of Naples 'Federico II', Via Cinthia, I-80126, Naples, Italy

<sup>b</sup> Institute of Protein Biochemistry, CNR, Via Pietro Castellino 111, I-80131, Naples, Italy

<sup>c</sup> Institute of Biostructures and Biomages, CNR, Via Mezzocannone 16, I-80134, Naples, Italy

### ARTICLE INFO

#### Article history:

Received 15 May 2008

Received in revised form 8 July 2008

Accepted 10 July 2008

Available online 22 July 2008

#### Keywords:

Membrane protein

Transmembrane domain

Molecular dynamics

Explicit lipid bilayer

Protein–protein interactions

Helix–helix packing

### ABSTRACT

The immune system cells express activating receptors, which consist of a dimeric ligand-binding molecule associated with a signal transducing dimer. The communication between the receptor partners depends primarily on the interactions between their membrane-embedded segments. In the B cell receptor (BCR) the sequence traversing the lipid bilayer of the immunoglobulin (IgTM) is highly conserved among species. We have investigated the association of the IgTM regions of the BCR of the Antarctic teleost *Chionodraco hamatus*. The nucleotide sequence of the entire immunoglobulin chain has been determined and the length, polarity, and structure of the IgTM region have been thoroughly analyzed. Structural models of the IgTM homodimer were also obtained by performing several MD simulations in a lipid bilayer using, as a starting model, two copies of the IgTM helix placed at various relative orientations and distances. Despite a certain degree of conformational heterogeneity, the predicted models of the IgTM homodimer display similar packing interfaces, characterized by a high degree of surface complementarity. The residues presumably responsible for the interaction and, consequently for the receptor stability have been identified in this manner.

© 2008 Elsevier Inc. All rights reserved.

### 1. Introduction

Membrane proteins play an essential role in several cellular functions, including cell division, intra- and inter-cellular signaling, macromolecule and energy transport. Although membrane proteins can be exceedingly complex, their transmembrane (TM) regions are often relatively simple. Two different basic architectures can be attributed to the TM regions: the  $\alpha$ -helix-bundle and the  $\beta$ -barrel [1]. Biogenesis of helix-bundle and  $\beta$ -barrel proteins follows different steps. In particular, ribosomes making helix-bundle proteins bind, during translation, to a protein complex known as translocon that assures the translocation of the nascent chain in the endoplasmic reticulum where it folds [2]. The protein then moves to the lipid bilayer as monomer or dimer depending on its hydrophobicity and tendency to aggregate.

The function of membrane proteins often depends on the interactions between their transmembrane regions [3]. The types of helix–helix interactions determining the structure of the TM region are sequence specific. The helix–helix assembly can be

driven by van der Waals contacts [4,5], but electrostatic interactions [6] as well as hydrogen bonds [7] are sometimes observed. Specific TM helices packing motifs have been identified: the GxxxG and AxxxA motifs, present in the middle region of associated TM helices [8,9] and the SxxSSxxT or SxxxSSxxT motifs [10]. Examples of membrane proteins whose structure and assembly depend on helix–helix packing in the lipid bilayer include the homodimeric Rhodopsin [11,12], the glycoporphin A (GpA) [13], the cell death factor BNIP3 [14].

Due to the difficulty in crystallizing membrane proteins for X-ray diffraction studies, multidisciplinary approaches that combine experimental and computational techniques are frequently used to gain insights into the structures of TM domains and to analyze the energetically favorable and tightly packed transmembrane  $\alpha$ -helices [15]. Computational methods have been used to identify the interactions between TM helices when biochemical and biophysical data are not available. In this respect, molecular dynamics simulations have been demonstrated to be a very useful approach to provide a better understanding of TM helices association [16,17]. Dodecane has most often been used for the simulations to create a hydrophobic environment [18]. Recently, a more appropriate environment for many membrane proteins has been produced by using lipid molecules organized in a pre-

\* Corresponding author. Tel.: +39 081674276; fax: +39 081674090.  
E-mail address: [antonello.merlino@unina.it](mailto:antonello.merlino@unina.it) (A. Merlino).

equilibrated bilayer system: dodecyl phosphocholine (DPC), dimyristoyl phosphatidylcholine (DMPC), palmitoyl oleoyl phosphatidylcholine (POPC), dipalmitoyl phosphatidylcholine (DPPC), palmitoyl linoleoyl phosphatidylcholine (PLPC) [16,17].

Also membrane protein complexes such as immune receptors have been recently investigated [19–21]. They share a distinctive assembly consisting of a ligand-binding molecule non-covalently associated with a signal transducing dimer, that carries cytoplasmic phosphorylation motifs [19]. In particular, within the T cell receptor  $\alpha\beta$  heterodimer, the  $\zeta\zeta$  dimer interactions have been well characterized [20]. At present, there are no data available about the B cell antigen receptor complex (BCR), the molecule responsible for the first step of immune response cascade. BCR is composed of a membrane-bound immunoglobulin molecule (Ig), consisting of two covalently linked heavy chains (IgH) and two light chains (IgL), associated with the transducing molecular partner formed by the disulphide bridged heterodimer Ig $\alpha$ /Ig $\beta$  [22]. The two covalently linked heavy chains wherein the extracellular environment traverse the lipid bilayer and three amino acid residues lean out of the cell membrane in the cytoplasm. The region of Ig embedded in the lipid bilayer (IgTM) is highly conserved among species and contains a sequential motif common to all immune receptors, known as the CART (conserved antigen receptor transmembrane) motif [23]. Although it is well known that the IgTM domain is crucial either for the BCR assembly or for the signal transduction, little is known about the structure of this region.

We have investigated the association of the IgTM regions in the BCR of *Chiono draco hamatus*, a fish living in the Antarctic sea, which represents the model species we have chosen for studying immunity. The primary sequence of the Ig was obtained. Using different computational methods, the length and the polarity of the  $\alpha$ -helical IgTM region within the cell membrane were predicted, and a molecular model was built. The stability of the model was evaluated by molecular dynamics (MD) simulations performed both in water and in lipid bilayers. Models for the association of two IgTM helices were also investigated. Moreover, the results of the MD simulations were discussed on the bases of experimental data collected on the IgTM region of mammals.

## 2. Experimental methods

### 2.1. Cloning and sequencing of the *C. hamatus* IgH chain

Total RNA was isolated from the head kidney of a *C. hamatus* specimen, and subjected to reverse transcription using an oligo (dT) primer linked to an appropriate adapter (oligoT-AP: 5'-CGGAGATCTCCAATGTGATGGGAAATCTTTTTTTTTTTTTTTT-3'). Double-stranded cDNA was amplified using a 5' primer (5'-CTATAGGACAGGACTGC-3') complementary to a sequence in the leader region of a related species (*Trematopus bernacchii*) and a 3' primer (5'-CGGAGATCTCCAATGTGATGGGAAATTC-3') complementary to the adapter. The amplification was performed as follows: 94 °C for 2 min, 30 cycles of 94 °C (1 min), 61 °C (1 min), and 72 °C (1 min) with a final extension at 72 °C for 10 min. The PCR products were then analyzed on 1% agarose gel: two bands were observed, one corresponding to the secreted form (1500 bp) and a second one, of the expected size of 1000 bp, corresponding to the membrane-bound form. Both were cloned into pCRII-TOPO vector (Invitrogen). Sequences of positive clones were determined with an ABI PRISM 3100 automated sequencer at PRIMM (Naples, Italy).

The accession number for the sequence encoding the secreted form is EU192078 and that for the sequence encoding the membrane form is EU192079.

### 2.2. Homology modeling

Sequence similarity searches were performed using the data from Swiss-Prot [24] and Protein Identification Resource (PIR) databanks [25]. The transmembrane region was predicted using TMpred ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)) [26] and HMMTM (<http://biophysics.bio.uoa.gr/HMM-TM/index.jsp>) [27] tools. Secondary structure predictions of the TM region were performed using the GOR program [28]. Models of *C. hamatus* IgTM region were built with the program 3D-PSSM (<http://www.bmm.icnet.uk/~3dpssm>) [29], using the structure of the helix H of the photosynthetic reaction center of *Rhodobacter sphaeroides* as a template (sequence identity 33%). The structure was validated with PROCHECK [30], and a summary table of the output is reported in Table 1.

MD simulations in water and in the POPC bilayer have been performed to evaluate the stability of the model. POPC is a good mimic of the in vivo environment. It was chosen because it includes both a saturated (C16) and an unsaturated (C18) fatty acid, thus better resembling the cell membrane composition and thickness. The simulations in water have been performed simply to compare the different approaches.

Furthermore, in order to study the spontaneous formation of a parallel IgTM homodimer, four different initial configurations, in which the relative positions of the two helices were slightly varied, were built using the program O [31] and submitted to MD simulations both in water and in POPC environments. The first dimer was generated with the helix axes perfectly aligned. This initial construction was used to generate three other configurations by rotating one helix around its own axis of a few degrees, and by translating the same helix of a few Å. In all cases the distances between the centers of mass of the two helices were in the range 14–16 Å. The minimum distance between atoms belonging to the two helices was 5.2 Å.

### 2.3. Molecular dynamics simulations

All the computer simulations reported in this study were performed using GROMACS 3.2 package [32] and the GROMOS96 force field. The starting models were immersed in different boxes containing 3500–16,000 simple-point-charge water molecules. The ionization state was set to mimic a neutral pH environment. The overall charge of the system was neutralized by adding the appropriate number of ions 7 Å farther from the protein surface. In the first step of the equilibration process, the solvent was relaxed by energy minimization followed by a 20 ps MD at 300 K. The overall system was then minimized without restraints, before the productive run. All bond lengths were constrained by lincs [33]. Newton's equations of motion were integrated with a time step of 2 fs and atomic coordinates were saved for analyses every 0.5 ps. A dielectric constant of 1 was used. A cut-off of 14 Å was used for the

**Table 1**  
Procheck summary

Ramachandran plot	
Residues in core region	100.0%
Residues in generously allowed regions	0.0%
Residues in disallowed regions	0.0%
Bad contacts	0
G-factors	
Dihedrals	0.18
Covalent	0.16
Overall	0.19
M/c bond lengths within limits	99.3%
M/c bond angles within limits	94.9%

treatment of both electrostatic and Lennard-Jones interactions. All systems were simulated in the isobaric–isothermal (NPT) ensemble at 300 K, using periodic boundary conditions in the three coordinate directions. The pressure was controlled using the Berendsen algorithm at 1 bar with a coupling constant  $\tau_p = 1$  ps at 300 K. The temperature was kept constant using the Berendsen thermostat with a coupling constant  $\tau_p = 1$  ps [34].

The process of model insertion in the lipid bilayer and the MD runs in lipid environments were performed as suggested by other authors [35,36]. Specifically, the starting models were placed in a box containing a mixture of POPC lipids and water molecules. Before model insertion, the box contained 128 lipids and 3755 water molecules. After the protein insertion in the centre of the box, water molecules with oxygen atoms closer to 0.40 nm from a non-hydrogen atom of the protein, and lipid molecules with at least one atom closer to 0.4 nm from a non-hydrogen atom of the model were removed. Removal of these atoms introduces small voids between the model and water or lipid molecules that are easily removed during the first steps of the equilibration process. During this step, a few simulations of 10 ps were performed, with the atomic coordinates of the protein restrained to the initial positions. This resulted in final systems containing 110 lipids and about 3730 water molecules. For the simulations in POPC, the Part Mesh Ewald (PME) method [37] was used for the treatment of electrostatic interactions for atoms at a distance greater than 9 Å. The force field and the coordinates for POPC were downloaded from the Tieleman group web-site (<http://mouse.bio.ucalgary.ca>).

All the trajectories were estimated to assess the quality of the simulations using GROMACS routines and in-house programs. We evaluated the surface complementarity by calculating the statistics  $S_c$  value [38], and by quantifying the geometrical packing of protein interfaces on representative models using the CCP4 suite [39]. The global structural stability of the simulations was confirmed by analyzing several structural parameters (i.e. secondary structure, radius of gyration and total number of hydrogen bond) as a function of time. The stability of the simulations was also assessed by measuring the convergence in the essential space. Essential degrees of freedom were extracted from the trajectories according to the essential dynamics method [40]. The convergence in the essential space (first 10 eigenvectors) was checked by calculating the root mean squared inner products (RMSIP) between two halves of the trajectory (the RMSIP values are >0.59) as previously reported [41,42] and by measuring mean squared fluctuation versus time block analysis. The RMSIP is defined as:

$$\sqrt{\frac{1}{10} \sum_{i=1}^{10} \sum_{j=1}^{10} (\eta_i^t \eta_j^t)^2}$$

where  $\eta_i^t$  and  $\eta_j^t$  are the *i*th and *j*th eigenvectors, obtained from diagonalization of the covariance matrix, from the first and second half of the trajectory, respectively (for further details see references [41,42]).

### 3. Results

#### 3.1. Analysis of the membrane-bound domain

The transcripts of both the secreted and membrane-bound IgM heavy chain were cloned and sequenced. By sequence analysis the membrane form was shown to possess a different carboxyl-terminal region, which is responsible for the interaction with the cell membrane. This region showed the longest size (79-amino acid residues) among all species (ranging between 43 residues in mammals and 49 residues in puffer fish). Based on comparison with other vertebrates, the identity percentage for the overlapping regions, ranged between 87.5 (spotted wolffish) and 53.3% (man).

The prediction of the extra cellular, membrane-crossing and cytoplasmic regions obtained by both tools mentioned in the methods, were similar, and are reported in Fig. 1, where the *C. hamatus* membrane-bound region is aligned together with those from other vertebrates available in data banks. The membrane-crossing region as well as the cytoplasmic tail are highly conserved, whereas the extracellular spacer is highly divergent either between teleosts and mammals or within teleosts. The membrane-crossing region, predicted to have an  $\alpha$ -helical conformation, includes the majority of the conserved positions (in bold in Fig. 1).

#### 3.2. Homology modeling and molecular dynamics of a single IgM helix

A molecular model of the *C. hamatus* IgTM region, corresponding to the sequence NMGSTAITFILLFLITLFSIGTIAFKVK, was built using the helix H of the photosynthetic reaction centre of *R. sphaeroides* as a template (sequence identity 33%). The modeled IgTM structure consists of an extended, nearly ideal,  $\alpha$ -helix. As shown in Fig. 2, the conserved residues (Thr5, Phe9, Leu12, Thr16, Phe19 and Thr24, according to the numbering scheme used in this paper) are all located on one face of the helix and are solvent exposed.

As a preliminary investigation of the IgTM helix–helix association, the stability of the IgTM structure in POPC was monitored by performing a 50-ns-duration simulation. The IgTM molecule remained folded and close to the ideality during the simulation timescale, as suggested by the analysis of secondary structure, radius of gyration, helix radius and total number of hydrogen bonds as a function of time. The  $C_\alpha$  root mean square

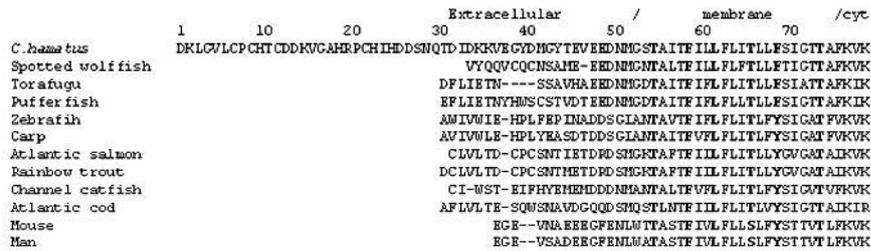


Fig. 1. Sequence alignment of IgTM region of different species. Conserved residues are shown in bold.



Fig. 2. Ribbon diagram of IgTM  $\alpha$ -helix. Conserved residues are shown in ball-and-stick.

deviation (RMSD) of the structure versus the starting model is constant at about 2 Å. The hydrophobic patch (residue 9–19) of the  $\alpha$ -helix is completely immersed in the membrane and the charged terminal portions (residues 1–4 and 27–29) are anchored to the water-membrane polar interface.

In order to characterize the structure of the IgTM helix in the different environments, two simulations in water were also performed. The results of these simulations revealed that a local unfolding of the N- and C-termini of the  $\alpha$ -helix occurred in water.

### 3.3. IgTM helix–helix association, structure and stability

In order to study the spontaneous formation of a parallel IgTM homodimer, four configurations, which differ for the relative positions of the two helices, were generated and used in MD studies performed in water and in POPC.

Eight simulations were performed in POPC bilayer (from 15 to 50 ns per simulation, for a total of more than 150 ns), and all showed the formation of a stable homodimer. This result is not surprising since spontaneous helix–helix aggregations were already observed in nanosecond time simulations, providing that the starting configurations are not too far away [13,14]. On the other hand, 4 out of 10 simulations performed in water, for a total of about 120 ns, did not produce a parallel pairing of the two helices. In such cases, the resulting structures were characterized by an antiparallel packing of the two helices (two cases), separated folded  $\alpha$ -helices (one case) or helices interacting in an irregular fashion (one case).

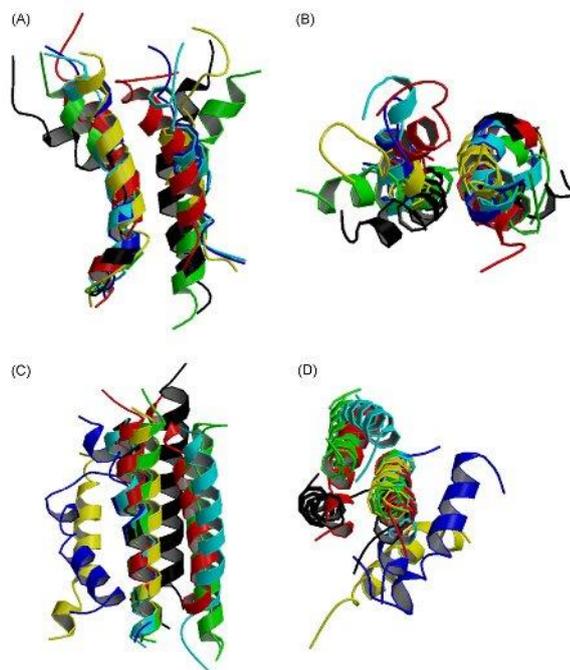
Fig. 3 depicts views of six representative models obtained by the POPC simulations (panels A and B) and of the six models obtained by MD simulations in water (panels C and D). In the figure, one helix is fixed in the same position for each structure so that the differences in the orientation between the helices can be easily grasped. The root mean square deviations of the residues involved in the interfacial region (residues 12–19 of the two chains) among the different models obtained from the POPC simulations fall in the range from 0.4 to 3.3 Å. These data indicate that the interfaces which stabilize the homodimers obtained by the simulations in POPC are similar to each other (Fig. 3A and B). This similarity suggests that the models obtained can be a plausible representation for the IgTM assembly in the BCR.

On the contrary, the homodimers obtained in water presented a larger structural variability (Fig. 3C and D). The helices interacted through different faces. In some cases, a deep distortion of the helical structure occurred. The root mean square deviation of residues 12–19 of the two chains among these structures covered the range from 2.1 to 5.0 Å.

The MD simulations in the lipid environment were investigated to single out the relevant interactions that lead to IgTM homodimer association and structural stability. As expected, the pairing of the two helices is driven by hydrophobic interactions. In fact, the total solvent accessible surface area of the hydrophobic residues rapidly changes its value from an average of 1100 Å<sup>2</sup> to an average of about 950 Å<sup>2</sup>, whereas the solvent exposed area of hydrophilic residues remains practically unchanged at about 900 Å<sup>2</sup>. The time evolution of the secondary structure elements, the gyration radius of the C <sub>$\alpha$</sub>  atoms and the number of hydrogen bonds of each separated helix indicate that the monomers retain their structure during the helix–helix association, except for the expected unwinding at the terminal tails. The rise per residue of the helices ranges from 1.32 to 1.61 and in most cases is close to the ideal value of 1.5.

The total accessible surface area, the number of hydrogen bonds (37–40) and the surface complementarity *sc* (0.63–0.69) of the homodimeric IgTM models show that they have a compact structure. The two helices are formed by an average of 22 residues, two residues more than those required to span the 30 Å of the hydrophobic membrane. The presence of longer stretches of residues in some cases is associated to a little helix distortion (helix kink). The helix kink results in an increased radius of gyration and in a decreased number of hydrogen bonds per helix. Interfacial residues involved in the helix–helix recognition have been identified by visual inspection of the MD trajectories and by the plot of the distance between side chain atoms of residues belonging to the two chains as function of time. The intermolecular interface comprises 6–8 residues. The interfacial region lacks intermolecular hydrogen bonds. The two helices interact essentially via van der Waals interactions (Fig. 4A–C) between the same interfacial residues Phe9, Leu12, Phe13, Leu14, Leu17 and Phe19 of the two chains. Although the mean intermolecular distances between these key residues change across the simulations, specific contacts that drive the IgTM recognition can be identified. The most conserved intermolecular interaction forms between side chains of residues Phe13 and Leu12 (Fig. 4A and B). Furthermore, a  $\pi$ – $\pi$  interaction between the two Phe19 is intermittently observed, since the two aromatic rings are parallel at a distance of 4.3 Å [43]. Other interactions, often observed, involve the side chains of residues Leu12, Phe13, and Leu14 with Phe9 (Fig. 4C). Finally, residues Phe9–Phe9, Leu17–Phe19 and Phe13–Leu14 can be in close contact.

In the final homodimeric models, the N- and C-termini are solvent exposed. At the C-terminal tails, Lys27 and Lys29 of the two chains make favorable interactions with oxygen atoms of the POPC



**Fig. 3.** Structural comparison between the IgTM homodimers obtained from simulations in different environments. Six representative structures obtained from POPC simulations are shown in panel A and B, whereas six structures obtained from water simulations are shown in panel C and D.

molecules on the surface of the lipid bilayer, thus anchoring the dimer to the membrane. In fact we noted that during the simulations, these residues were allowed to move but constantly facing the lipid headgroups. At the amino-terminus an Asn residue seems to interact with the external phosphate groups.

#### 4. Discussion

The difficulty of obtaining experimental structures of membrane protein molecules has stimulated the application of theoretical procedures as a tool to provide basic clues regarding the proteins assembly in membrane. In particular, MD simulations can provide useful insights into the factors governing helix–helix association in lipid bilayers. Although incomplete sampling, inaccurate force fields, and approximate treatment of long-range interactions intrinsically limit these studies, intriguing results on the mechanisms of TM helix recognition have been recently reported [13,14,44–47].

In this paper, we used a computational approach that combines homology modeling and molecular dynamics simulations to predict the structure of the transmembrane region of the immunoglobulin and analyzed the interactions stabilizing its association in the membrane. The results are relevant to the problem of receptor stability, signal transduction cascade inside the cell, and could also be used to design therapeutic agents for

potential pharmaceutical applications. We have focused on the transmembrane region of *C. hamatus* Immunoglobulin, whose sequence is reported here for the first time. The results indicate that the immunoglobulin transmembrane region adopts an  $\alpha$ -helical structure, which is significantly more stable in the lipid bilayer than in solution. The MD simulations also suggest that individual IgTM helices are stable, as observed in the case of glycophorin A [18].

Moreover, several simulations independently carried out on the IgTM homodimers reveal that the spontaneous formation of the transmembrane Ig helix–helix dimer of IgTM is more likely to form inside the lipid bilayer, rather than in water, where a number of antiparallel/non-physiological structures have been also obtained. These data suggest that in the case of BCR, the lipid environment does exert a major effect on the TM helix packing.

Given the high sequence similarity between the IgTM region of *C. hamatus* and that of other species, our results could be extended to the folding and assembly of different species BCR. The present data suggest that the parallel helix–helix structure of IgTM in BCR is stabilized basically by the van der Waals interactions of the two Phe19 side chains and by the insertion of Leu12 side chain of one chain between Phe9 and Phe13 of the partner chain at the helix–helix interface. The tight packing of these side chains results in a small (about  $150 \text{ \AA}^2$ ), but highly complementary ( $sc = 0.65$ ) contact surface area, that lacks inter-helical hydrogen bonds. These data

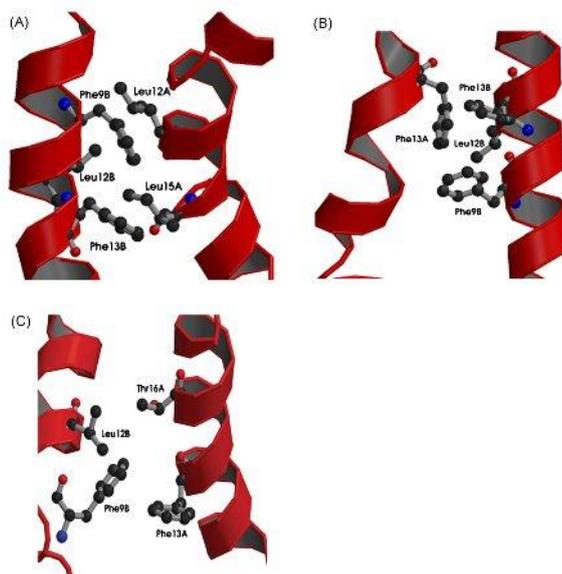


Fig. 4. Residues determining the structure and stability of the IgTM homodimers in representative models obtained from POIC simulations.

are in agreement with mutagenesis studies that revealed that the substitution of polar residues at the N-terminus (residues 4–8 in our numbering scheme) of human IgM with alanine or valine does not affect the BCR association and signaling [48–50], whereas the double mutation of residues 19 and 20 (Tyr-Ser in mouse, Phe-Ser in *C. hamatus*) to Val, in the transmembrane domain of mouse membrane Ig, weakens the association of Ig with Ig $\alpha$ /Ig $\beta$  and destabilizes the BCR structure [51]. It should be recalled that both the single mutations Y19F and S20A of mouse IgM do not hamper the signal transduction [49,52]. These observations strongly validate the final model obtained for the dimer, which highlights the inter-helix direct interactions between the aromatic side chains of residue 19 as a stabilizing motif for the dimer. Interestingly, these data are also in line with a recent suggestion that the aromatic nature of Trp, Tyr and Phe can be a critical feature in self-association of TM helices [20,53]. Vice versa, the simulations also suggest that the three conserved threonines (residues 5, 16 and 24 in our numbering scheme) provide little assistance in the association of the helices. This result agrees with mutagenesis data which show that the replacement of Thr5 or Thr16 with a hydrophobic residue is well tolerated [52,54], and that the point mutation of T24V does not alter the BCR association, although it decreases the sensitivity of signal transduction [52,55]. This also agrees with the finding that, although serine and threonine are found in TM helices more frequently than other polar residues, they prefer to form intra helical hydrogen bonds to main-chain carboxyl oxygen rather than inter-helical contacts [55,56]. Altogether these results strengthen the hypothesis that the MD homodimeric structure may indeed depict the IgTM association.

Interestingly, in some models one of the helices immersed in the bilayer is kinked in the middle. It should be noted that this

result has been obtained without imposing helix kinks. This finding could be explained by the 'hydrophobic match' effect [57] that may affect the protein conformation, inducing a conformational change in the helices. In the case of the IgTM, an important role in driving the helix kink could be also played by the interactions that the lysine residues at the C-terminal regions form with the oxygen atoms of the lipid bilayer. Notably, C-terminal residues have also been considered important for BCR structure/function by other authors. In fact, experimental data have shown that the deletion of the KVK region (residues 27–29 of our model) produces a loss of mouse BCR function [52,58], and that the substitution of the lysines 27 and 29 with negatively charged residues affects the signal transduction, whereas the replacement of the lysines with positively charged residues does not [49,50].

#### Acknowledgment

The Italian National Program financially supports this study for Antarctic Research (PNRA), Project 2005.1.2.

#### References

- [1] A. Elofsson, C. von Heijne, Membrane protein structure: prediction vs reality, *Annu. Rev. Biochem.* 76 (2007) 125–140.
- [2] S.H. White, C. von Heijne, Transmembrane helices before, during and after insertion, *Curr. Opin. Struct. Biol.* 15 (2005) 378–386.
- [3] B.J. Bormann, D.M. Engelman, Intramembrane helix-helix association in oligomerization and transmembrane signalling, *Annu. Rev. Biophys. Biomol. Struct.* 21 (1992) 223–242.
- [4] K.R. MacKenzie, J.H. Prestegard, D.M. Engelman, A transmembrane helix dimer: structure and implications, *Science* 276 (1997) 131–133.
- [5] M.A. Lemmon, J.M. Flanagan, J.F. Hunt, B.D. Adair, B.J. Bormann, C.E. Dempsey, D.M. Engelman, Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices, *J. Biol. Chem.* 267 (1992) 7683–7689.

- [16] H. Grakowski, J.D. Lear, W.F. DeGrado, Polar side chains drive the association of model transmembrane peptides, *Proc. Natl. Acad. Sci. USA* 98 (2001) 880–885.
- [17] F.X. Zhou, M.J. Cocco, W.P. Russ, A.T. Brunger, D.M. Engelman, Interhelical hydrogen bonding drives strong interactions in membrane proteins, *Nat. Struct. Biol.* 7 (2000) 154–160.
- [18] G. Kleiger, R. Grothe, P. Mallick, D. Eisenberg, GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles, *Biochemistry* 41 (2002) 5990–5997.
- [19] D. Schneider, D.M. Engelman, Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions, *J. Mol. Biol.* 343 (2004) 799–804.
- [20] J.P. Dawson, J.S. Weinger, D.M. Engelman, Motifs of serine and threonine can drive association of transmembrane helices, *J. Mol. Biol.* 316 (2002) 799–805.
- [21] T. Okada, I. Le Trong, B.A. Fox, C.A. Behnke, R.E. Stenkamp, K. Palczewski, X-Ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles, *J. Struct. Biol.* 130 (2000) 73–80.
- [22] A.B. Patel, E. Crocker, P.J. Reeves, E.V. Getmanova, M. Eilers, H.G. Khorana, S.O. Smith, Changes in interhelical hydrogen bonding upon rhodopsin activation, *J. Mol. Biol.* 347 (2005) 803–812.
- [23] M. Mottamal, J. Zhang, T. Lazaridis, Energetics of the native and non-native states of the glyophorin transmembrane helix dimer, *Proteins* 62 (2006) 996–1009.
- [24] Y.A. Vreshaga, P.E. Volynsky, J.E. Pustovalova, D.E. Nolde, A.S. Arseniev, R.G. Efremov, Specificity of helix packing in transmembrane dimer of the cell death factor Bnip3: a molecular modeling study, *Proteins* 69 (2007) 309–325.
- [25] J.A. Leeds, D. Boyd, D.R. Huber, G.K. Sonoda, H.T. Luu, D.M. Engelman, J. Beckwith, Genetic selection for and molecular dynamic modeling of a protein transmembrane domain multimerization motif from a random *Escherichia coli* genomic library, *J. Mol. Biol.* 313 (2001) 181–195.
- [26] J.M. Cuthbertson, P.J. Bond, M.S. Sansom, Transmembrane helix-helix interactions: comparative simulations of the glyophorin A dimer, *Biochemistry* 45 (2006) 14298–14310.
- [27] T. Rog, K. Murzyn, M. Karttunen, M. Pasenkiewicz-Gierula, Nonpolar interactions between trans-membrane helical EGF peptide and phosphatidylcholines, sphingomyelins and cholesterol, *Molecular dynamics simulation studies*, *J. Pept. Sci.* 14 (2008) 374–382.
- [28] J. Henin, A. Pohorille, C. Chipot, Insights into the recognition and association of transmembrane alpha-helices. The free energy of alpha-helix dimerization in glyophorin A, *J. Am. Chem. Soc.* 127 (2005) 8478–8484.
- [29] M.E. Call, K.W. Wucherpfennig, Common themes in the assembly and architecture of activating immune receptors, *Nat. Rev. Immunol.* 7 (2007) 841–850.
- [30] M.E. Call, J.R. Schnell, C.Xu, R.A. Lutz, J.J. Chou, K.W. Wucherpfennig, The structure of the  $\zeta_2$  transmembrane dimer reveals features essential for its assembly with the T cell receptor, *Cell* 127 (2006) 355–368.
- [31] J. Feng, M.E. Call, K.W. Wucherpfennig, The assembly of diverse immune receptors is focused on a polar membrane-embedded interaction site, *PLoS Biol.* 4 (2006) e142.
- [32] J. Hombach, T. Tsubata, L. Leclercq, H. Stappert, M. Reth, Molecular components of the B-cell antigen receptor complex of the IgM class, *Nature* 343 (1990) 760–762.
- [33] K.S. Campbell, B.T. Backstrom, G. Tiefenthaler, E. Palmer, CART: a conserved antigen receptor transmembrane motif, *Semin. Immunol.* 6 (1994) 393–410.
- [34] A. Bairoch, R. Apweiler, The SWISS-PROT protein sequence data bank and its supplement TrEMBL, *Nucleic Acids Res.* 25 (1997) 31–36.
- [35] K.E. Sidman, D.C. George, W.C. Barker, L.T. Hunt, The protein identification resource (PIR), *Nucleic Acids Res.* 16 (1988) 1869–1871.
- [36] K. Hofmann, W. Stoffel, TMbase – a database of membrane spanning proteins segments, *Biol. Chem.* 374 (1993) 166.
- [37] P.G. Bages, T.D. Liakopoulos, S.J. Hamodrakas, Algorithms for incorporating prior topological information in HMMs: application to transmembrane proteins, *BMC Bioinform.* 7 (2006) 189.
- [38] A. Kloczkowski, K.L. Ting, R.L. Jernigan, J. Garnier, Combining the GOR V algorithm with evolutionary information for protein secondary structure prediction from amino acid sequence, *Proteins* 49 (2002) 154–166.
- [39] L.A. Kelley, R.M. MacCallum, M.J.E. Sternberg, Enhanced genome annotation using structural profiles in the program 3D-PSM, *J. Mol. Biol.* 299 (2000) 499–520.
- [40] R.A. Laskowski, M.W. MacArthur, M.D. Moss, J.M. Thornton, PROCHECK: A program to check the stereochemical quality of protein structure, *J. Appl. Crystallogr.* 26 (1993) 283–291.
- [41] T.A. Jones, M. Bergdoll, M. Kjelgaard, O. a macromolecule modeling environment. In: C. Bugz, S. Ealick (Eds.), *Crystallogr. Model Methods Mol. Des.* [Pap. Symp.], Springer-Verlag Press, 1990, pp. 189–199.
- [42] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, *J. Comput. Chem.* 26 (2005) 1701–1718.
- [43] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction Models for Water in Relation to Protein Hydration, *Reidel, Dordrecht*, 1981, pp. 331–342.
- [44] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, J.R. Haak, Molecular dynamics with coupling to an external heat bath, *J. Chem. Phys.* 81 (1984) 3684–3690.
- [45] C. Kandt, W.L. Ash, D.P. Tieleman, Setting up and running molecular dynamics simulations of membrane proteins, *Methods* 41 (2007) 475–488.
- [46] L. Vaccaro, K.J. Cross, J. Kleinjung, S.K. Straus, D.J. Thomas, S.A. Wharton, J.J. Skehel, F. Fraternali, Plasticity of influenza haemagglutinin fusion peptides and their interaction with lipid bilayers, *Biophys. J.* 88 (2005) 25–36.
- [47] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [48] M.C. Lawrence, P.M. Colman, Shape complementarity at protein/protein interfaces, *J. Mol. Biol.* 234 (1993) 946–950.
- [49] Collaborative Computational Project, Number 4, The CCP4 suite: programs for protein crystallography, *Acta Crystallogr. Sect. D* 50 (1994) 760–763.
- [50] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Essential dynamics of proteins, *Proteins* 17 (1993) 412–425.
- [51] A. Merlino, L. Vitagliano, M.A. Ceruso, I. Mazzarella, Subtle functional collective motions in pancreatic-like ribonucleases: from ribonuclease A to angiogenin, *Proteins* 53 (2003) 101–110.
- [52] A. Merlino, L. Vitagliano, M.A. Ceruso, I. Mazzarella, Dynamic properties of the N-terminal swapped dimer of ribonuclease A, *Biophys. J.* 86 (2004) 2383–2391.
- [53] S.K. Burley, G.A. Petsko, Amino-aromatic interactions in proteins, *FEBS Lett.* 203 (1986) 139–143.
- [54] E. Sparr, W.L. Ash, P.V. Nazarov, D.T. Rijkers, M.A. Hemminga, D.P. Tieleman, J.A. Killian, Self-association of transmembrane alpha-helices in model membranes: importance of helix orientation and role of hydrophobic mismatch, *J. Biol. Chem.* 280 (2005) 39324–39331.
- [55] J. Gumbart, Y. Wang, A. Aksimentiev, E. Tajkhorshid, K. Schulten, Molecular dynamics simulations of proteins in lipid bilayers, *Curr. Opin. Struct. Biol.* 15 (2005) 423–431.
- [56] J. Henin, B. Maigret, M. Tarek, C. Escrieut, D. Fourny, C. Chipot, Probing a model of a GPCR/ligand complex in an explicit membrane environment: the human cholecystokinin-1 receptor, *Biophys. J.* 90 (2006) 1232–1240.
- [57] O. Samra, Soumana, N. Garnier, M. Genest, Molecular dynamics simulation approach for the prediction of transmembrane helix-helix heterodimers assembly, *Eur. Biophys. J.* 36 (2007) 1071–1082.
- [58] B.J. Cherayil, K. MacDonald, G.L. Waneck, S. Pillai, Surface transport and internalization of the membrane IgM H chain in the absence of the Mb-1 and B29 proteins, *J. Immunol.* 151 (1993) 11–19.
- [59] A.C. Shaw, R.N. Mitchell, Y.K. Weaver, J. Campos-Torres, A.K. Abbas, P. Leder, Mutations of immunoglobulin transmembrane and cytoplasmic domains: effects on intracellular signaling and antigen presentation, *Cell* 63 (1990) 381–392.
- [60] S.A. Grupp, K. Campbell, R.N. Mitchell, J.C. Cambier, A.K. Abbas, Signaling-defective mutants of the B lymphocyte antigen receptor fail to associate with Ig-alpha and Ig-beta/gamma, *J. Biol. Chem.* 268 (1993) 25776–25779.
- [61] T.L. Stevens, J.H. Blum, S.P. Foy, L. Matsuuchi, A.L. DeFranco, A mutation of the mu transmembrane that disrupts endoplasmic reticulum retention. Effects on association with accessory proteins and signal transduction, *J. Immunol.* 152 (1994) 4397–4406.
- [62] J.H. Blum, T.L. Stevens, A.L. DeFranco, Role of the  $\mu$  immunoglobulin heavy chain transmembrane and cytoplasmic domains in B cell antigen receptor expression and signal transduction, *J. Biol. Chem.* 268 (1993) 27236–27245.
- [63] N. Sal-Man, D. Gerber, I. Bloch, Y. Shai, Specificity in transmembrane helix-helix interactions mediated by aromatic residues, *J. Biol. Chem.* 282 (2007) 19753–19761.
- [64] C.M. Pleiman, N.C. Chien, J.C. Cambier, Point mutations define a mlgM transmembrane region motif that determines intersubunit signal transduction in the antigen receptor, *J. Immunol.* 152 (1994) 2837–2844.
- [65] J.A. Ballesteros, X. Deupi, M. Oliveira, E.E. Haaksma, L. Pardo, Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation, *Biophys. J.* 79 (2000) 2754–2760.
- [66] T.M. Gray, B.M. Matthews, Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins, *J. Mol. Biol.* 175 (1984) 75–81.
- [67] T.M. Weiss, P.C. van der Wel, J.A. Killian, R.E. 2nd Koeppe, H.W. Huang, Hydrophobic mismatch between helices and lipid bilayers, *Biophys. J.* 84 (2003) 379–385.
- [68] R.N. Mitchell, A.C. Shaw, Y.K. Weaver, P. Leder, A.K. Abbas, Cytoplasmic tail deletion converts membrane immunoglobulin to a phosphatidylinositol-linked form lacking signaling and efficient antigen internalization functions, *J. Biol. Chem.* 266 (1991) 8856–8860.



## Evolution of the Antarctic teleost immunoglobulin heavy chain gene

Maria Rosaria Coscia<sup>a</sup>, Sonia Varriale, Concetta De Santi, Stefano Giacomelli, Umberto Oreste

<sup>a</sup>Institute of Protein Biochemistry, CNR, Via P. Castellino 111, 80131 Naples, Italy

### ARTICLE INFO

**Article history:**  
Received 7 July 2009  
Revised 28 September 2009  
Accepted 29 September 2009  
Available online 2 October 2009

**Keywords:**  
Notothenioidei  
Alternative mRNA splicing  
Immunoglobulin gene structure  
Cold adaptation  
Antarctic teleost evolution

### ABSTRACT

Notothenioid teleosts underwent major modifications of their genome to adapt to the cooling of the Antarctic environment. In order to identify specific features of the Antarctic teleost immunoglobulin, transcripts encoding the constant region of the IgM heavy chain from 13 Antarctic and non-Antarctic notothenioid species were sequenced. The primary mRNA splicing for the membrane form was found to be atypical in the majority of Antarctic species, because it led to exclusion of two entire constant exons, and to inclusion of 39-nucleotide exons encoding an unusually long Extracellular Membrane-Proximal Domain (EMPD). Genomic DNA analysis revealed that each 39-nucleotide exon fell within a long sequence that was the reverse complement of an upstream region. Deduced amino acid sequence analysis lead to the identification of cysteine encoding codons in the 39-nucleotide exons, but not in the respective sequence counterpart, suggesting that these residues might play an important role in the folding of the EMPD.

© 2009 Elsevier Inc. All rights reserved.

### 1. Introduction

Studies on the teleosts which inhabit the cold seawater of the Antarctic continental shelf ( $-1.87^{\circ}\text{C}$ ) have revealed specific features, induced by the evolutionary adaptation in their morphology, ecology and biology. A recent study strongly suggests that evolution under constant cold conditions has resulted in dramatic genomic expansions of specific protein gene families (Chen et al., 2008). The amplitude of gene modifications is well documented by the synthesis of novel genes such as those encoding the antifreeze glycoproteins or by the loss of the hemoglobin gene locus in several species (Cheng and Detrich, 2007).

Comprised of eight families and 120 species, the majority of Antarctic teleosts belong to the suborder Notothenioidei (Eastman, 1993). It is a common opinion that the phylogenetically basal families of the notothenioid suborder are those of Bovichtidae, Pseudaphritidae, and Eleginopidae. Their emergence predates the severe decrease in the water temperature that occurred about 24 MYA (Near, 2004). Bovichtidae, Pseudaphritidae, and Eleginopidae currently survive in non-Antarctic areas at higher temperatures.

We previously sequenced both light (Coscia et al., 2008) and heavy (Coscia et al., 2000a; Oreste and Coscia, 2002) chain of Immunoglobulin M from one species, *Trematomus bernacchii*, studied the amplitude of the antibody repertoire (Coscia and Oreste, 2002), the antibody specificity toward parasitic antigens (Coscia and Oreste, 1998; Coscia et al., 2000b), and the mechanism of the hepatobiliary transport (Abelli et al., 2005). This study is aimed

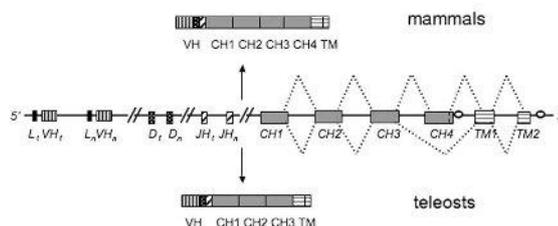
at investigating the Immunoglobulin M heavy chain (IgM) throughout notothenioid radiation.

Due to the plasticity of its gene locus, immunoglobulin is an appropriate protein candidate for the characterization of evolutionary modifications. Unpredicted domain assemblies have been recently identified in immunoglobulin molecules from different species and new heavy chain isotypes have been identified in lower vertebrates (Rumfelt et al., 2001; Danilova et al., 2005; Hansen et al., 2005; Savan et al., 2005a,b; Zhao et al., 2006). In addition, the IgD isotype was found to be different in the number and composition of constant domains among teleost species (Bentg en et al., 2002; Hordvik, 2002; Saha et al., 2004; Danilova et al., 2005).

Alternative pre-mRNA splicing mechanisms are a common source of diversity and also represent a mechanism underlying the regulation of immune gene expression (Lynch, 2004). In mammals IgM,  $\mu$  heavy chain primary transcripts are alternatively spliced into mature transcripts, encoding either the membrane receptor or the secretory antibody, according to the balance efficiencies of the membrane or the secretory polyA site, located at the 3'-end of TM and CH4 exon, respectively (Peterson, 2007). In teleosts, contrary to mammals, it has been demonstrated that the CH4 exon sequence does not contain the cryptic donor splice site, and that the membrane form is synthesized by using the donor site at the 3'-end of the CH3 exon (Wilson and Warr, 1992) (Fig. 1). As a result, the synthesized protein lacks the entire CH4 domain.

The IgH locus has been analyzed in many teleost species, and membrane IgM transcripts have been sequenced (Artero and Lefranc, 2000). Two exons, TM1 and TM2, encode the 45–48 amino acid residues specific for the membrane form: hydropathy plot indicates that 17–22 amino acid residues located at the the

\* Corresponding author. Fax: +39 08161 32629.  
E-mail address: [mr.coscia@ibp.cnr.it](mailto:mr.coscia@ibp.cnr.it) (M.R. Coscia).



**Fig. 1.** Schematic representation of the splicing mechanism generating the membrane-bound IgH. Coding exons are represented as shaded boxes, introns as lines. An open circle denotes the polyA site. The somatic recombination of VH, D, and JH gene segments generates the variable (VH) domain. The mRNA splicing pathway in mammals and teleosts, is indicated by dashed lines, and the resulting proteins are shown. A dashed vertical line depicts the cryptic donor site in the CH4 exon, that has been lost in teleosts.

N-terminus can be attributed to the Extracellular Membrane-Proximal Domain (EMPD<sup>1</sup>); the following 21 residues cross the cell membrane, whereas the last three C-terminal residues comprise the cytoplasmic tail, which mediates interaction with the actin filaments of the cytoskeleton (Park and Jongstra-Bilen, 1997). Among all vertebrate species, EMPD does not exhibit high sequence conservation and varies in length. Moreover most of the available sequences share some common features such as the membrane proximal acid rich region and cysteine and proline residues typically found in spacer regions of immunoglobulins.

The species analyzed in the present work belong to the families Bovichtidae (*Bovichtus diacanthus*), Eginopidae (*Eleginops maclovinus*), Nototheniidae (*Dissostichus mawsoni*, *Trematomus bernacchii*, and *Gobionotothen gibberifrons*), Bathydraconidae (*Cygnodraco mawsoni* and *Gymnodraco acuticeps*), Artedidraconidae (*Histiadraco velifer* and *Pogonophryne scotti*), Channichthyidae (*Chionocephalus aceratus*, *Champocephalus esox*, *Chionodraco hamatus*, and *Pagetopsis macropterus*), *Notothenia coriiceps* (Nototheniidae) secretory and membrane-bound Igμ, and *Chionocephalus aceratus* (Channichthyidae) secretory Igμ sequences, previously determined by Ota et al. (2003) are also included in the analysis.

## 2. Materials and methods

### 2.1. Biological samples

Fish specimens of the species *T. bernacchii*, belonging to the family Nototheniidae, and those belonging to the Bathydraconidae (*G. acuticeps*), Artedidraconidae (*H. velifer*) and Channichthyidae families (*C. hamatus*, and *P. macropterus*) were collected in the sea of Victoria Land during the Italian Antarctic Expeditions (2002–2007). A *P. scotti* specimen was sampled during the PNRA 2005 cruise in the Ross Sea; *G. gibberifrons*, *D. mawsoni*, and *C. aceratus* specimens were caught in the sea near Palmer Station (Peninsula), and *B. diacanthus*, *C. esox*, and *E. maclovinus* specimens were collected at Tristán de Cuña, Falkland, or South Georgia islands, during the ICEFISH expedition (International Collaborative Expedition to collect and study Fish Indigenous to Sub-Antarctic Habitats: 2004 cruise supported by NSF).

<sup>1</sup> Abbreviations used: APS1-3, anti-parallel sequence 1-3; CH1-4, heavy chain constant domain 1-4; DG, Digoxigenin; EMPD, extracellular membrane-proximal domain; IgD, class D Immunoglobulin; IgH, Immunoglobulin heavy chain; IgM, class M Immunoglobulin; Igμ, μ isotype Immunoglobulin heavy chain; MASP, Mannose-binding protein-associated Serine proteases; MYA, million years ago; NJ, neighbor-joining; POPC, dipalmitoyl phosphatidylcholine; RT-PCR, Reverse Transcriptase polymerase chain reaction; polyA, polyadenylation; RA-C, repeat A-C; RACE, rapid amplification cDNA ends; SSC, saline-Sodium Citrate; TAE, Tris-Acetate-EDTA; TM, transmembrane; UTR, untranslated region.

Fish specimens were sampled by use of gill nets, traps or hand lines and kept in aquaria with running, aerated seawater before being sacrificed. After collection, tissue samples (spleen and testis) were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until used.

### 2.2. Sequence determination of secretory and membrane Igμ constant region transcripts of 13 notothenioid species

Total RNA was isolated from head kidney or spleen using the SV total RNA isolation system (Promega), and subjected to reverse transcription using M-MLV Reverse Transcriptase (Ambion, Austin, TX, USA) according to the manufacturer's instructions. Double-stranded cDNA PCR amplification was carried out by using different combinations of the following sense primers: leader, 5'-CTATAGGACAGGACTGC-3'; FR4, 5'-GAAAAGGAACAATGGTGACCGTC-3'; iCH1Th, 5'-ACTGTGTTCTCTGATGCAA-3'; and anti-sense primers: anti-CH4endTh, 5'-CGGGGATGTCATGTTGAG-3'; TManti-3, 5'-GAAAGCAGTGGTCCA-3'. In most cases the primers used had been designed based on the *T. bernacchii* sequence. The amplification was performed as follows:  $92^{\circ}\text{C}$  for 2 min, 30 cycles of  $92^{\circ}\text{C}$  (1 min), annealing at  $62^{\circ}\text{C}$ , or  $58^{\circ}\text{C}$ , or  $52^{\circ}\text{C}$  (1 min), and  $72^{\circ}\text{C}$  (1 min) with a final extension at  $72^{\circ}\text{C}$  for 10 min.

In a few cases, partial sequences were obtained with the 3'RACE cDNA Amplification Kit (Invitrogen); species-specific primers were then designed based on the partial sequences obtained, whenever possible, to amplify the full-length sequence by RT-PCR. Products were then cloned into pSC-A by using the StrataClone PCR cloning kit (Stratagene) and positive clones were identified by blue/white selection. All sequences were determined on an ABI PRISM 3100 automated instrument at Primm (Naples, Italy).

### 2.3. Analysis of genomic regions

Genomic DNA was isolated from testis of *C. hamatus*, *T. bernacchii*, and *C. mawsoni* specimens according to Sambrook et al. (2001). PCR amplification on genomic DNA was conducted by using 200 ng of DNA and the Expand Long Template PCR system (Roche), following the manufacturer's instructions. Sense primers used were as follows: CH2 (5'-TGT CTG TTT GAG GGG AAA GG-3') or CH4 specific for *T. bernacchii* (5'-CTG ATC TGC TTA GTG AAA-3'), CH4h specific for *C. hamatus* (5'-AACTAAAGCCATTGTCAGATCCA-3'). Anti-sense primers used were as follows: TManti-3 specific for *T. bernacchii* (5'-GAA AGC AGT GGT TCC A-3') or TManti-h specific for *C. hamatus* (5'-AACAGCAGAGTGATGAGGAAGAG-3'). The amplification was performed as follows: initial denaturation at  $92^{\circ}\text{C}$  for 2 min, 10 cycles of denaturation at  $92^{\circ}\text{C}$  for 10 s, annealing at  $53^{\circ}\text{C}$  (or  $60^{\circ}\text{C}$ , or  $63^{\circ}\text{C}$ ) for 40 s, elongation at  $68^{\circ}\text{C}$  for 4 min; 20 cycles of denaturation at  $92^{\circ}\text{C}$  for 15 s, annealing at

53 °C (or 60 °C, or 63 °C) for 30 s, elongation at 68 °C for 4 min and 20 s increase for each successive elongation cycle, final elongation at 68 °C for 7 min. The genomic DNA PCR products, were separated on a 0.8% agarose gel in TAE buffer (40 mM Tris–HCl, pH 8.0, 1 mM EDTA). The gel was then incubated in denaturation buffer (0.5 M NaOH containing 1.5 M NaCl) for 30 min with gentle agitation. After neutralization in 0.5 M Tris–HCl, pH 7.5, containing 1.5 M NaCl, DNA was transferred onto a Hybond-N membrane (Amersham, GE Healthcare) by capillary action by incubation for 16 h using a 20× SSC solution as transfer buffer. Membranes were baked at 80 °C for 2 h to fix the nucleic acids. DIG-labeled cDNA probe specific for *C. hamatus* CH2-CH4 domains was obtained by PCR using the DIG High Prime DNA labeling kit (Boehringer Mannheim). The membrane was incubated in DIG Easy Hyb solution for 2 h at 42 °C with gentle agitation. 20 ng/ml of labeled probe was denatured by boiling for 5 min and rapidly cooling on ice water. The probe was then incubated with the membrane at 42 °C overnight with gentle agitation. Post-hybridization washes were 2 × 5 min in 2× SSC, 0.1% SDS at room temperature; 2 × 15 min in 0.1× SSC, 0.1% SDS at 68 °C under agitation. Immunological detection was performed with the Detection Starter Kit II (Boehringer Mannheim). The membrane was air dried and exposed to a Fuji Super RX film with an intensifying screen at room temperature for 20 min. *T. bernacchii* and *C. hamatus* positive bands were cloned into pSC-A by using the StrataClone blunt PCR cloning kit (Stratagene) and positive clones were identified by blue/white selection. All sequences were determined on an ABI PRISM 3100 automated instrument at Primm (Naples, Italy) using the vector primers and successive internal primers.

#### 2.4. Computational methods

The sequence alignments were performed by the ClustalX 1.83 program (Thompson et al., 1997) and analyzed by the WebLogo tool (<http://weblogo.berkeley.edu/>). Variation in the protein alignment was measured by the Shannon Entropy analysis (Stewart et al., 1997) (<http://hcv.lanl.gov/content/sequence/ENTROPY/entropy.html>).

Putative *N*-glycosylation sites were identified by looking for the consensus motif, referred to as the sequon NXS/T, where *X* is an amino acid residue different from proline.

Distance-based tree reconstruction was performed with the Neighbor-Joining (NJ) method (Saitou and Nei, 1987) with the alignment generated by ClustalX, using the BLOSUM weight matrix. Positions with gaps were excluded from the alignment. The observed distances were corrected for multiple substitutions using Motoo Kimura's formula. 1000 bootstrap replicates were made.

The protein sequence similarity and the nucleotide sequence identity were determined by SIM, and LALIGN tools respectively, both available at <http://www.expasy.ch/>.

The synonymous and non-synonymous substitution rate was calculated in the nucleotide alignment of cDNA sequences, as described in Nei and Gojobori (1986) for all pairwise comparisons of sequences in an alignment. When comparing two sequences, the number of synonymous and non-synonymous codon changes was counted, as well as the number of potential synonymous and non-synonymous changes. Ambiguous codons or codons with insertions were excluded from the tally of compared codons. The tool used was the SNAP program (<http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html>).

The computation of the accessibility plot profile was obtained by the ProtParam tool (<http://www.expasy.ch/>).

The Extracellular Membrane-Proximal Domain (EMPD) and cell membrane crossing residues were identified in the I $\mu$  sequences by comparing results obtained with two predictive tools: TMPRED

(<http://www.ch.embnnet.org>), based on the statistical analysis of a database of naturally occurring transmembrane proteins, and TMHMM (<http://www.cbs.dtu.dk/>) using a hidden Markov model. Overall, the results obtained by the two methods were in agreement with each other.

The RNA folding was predicted using the Zuker's algorithm (Zuker, 2003) (<http://mfold.bioinfo.rpi.edu/>) at working temperature of 273 or 310 K. To calculate the RNA duplex free energy, palindromic sequences separated by a 5-adenine loop were constructed.

The model of the region comprising the C-terminal six residues of *C. hamatus* CH2 and the two 13-aa repeats of the EMPD was built using the crystal structure of the CUB1-EGF-CUB2 region of MASP2 (Mannose-binding protein-Associated Serine Proteases 2, pdb ID: 1NT0), refined at a resolution of 2.7 Å, as template. The template was searched with the PHYRE program (<http://www.sbg.bio.ic.ac.uk/phyre>). The structure was validated with the WHAT IF package (<http://swift.cmbi.kun.nl/watthif/>). The starting model was immersed in a box containing 3888 simple-point-charge water molecules. The overall charge of the system was neutralized by adding the appropriate number of ions more than 7 Å away from the protein surface. The energy minimization procedure was performed using the GROMACS 3.2 package and the GROMOS96 force field by a 20 ps Molecular Dynamics at 300 K. The dimeric structure of the minimized model was with the assumption that Cys 5 and Cys 13 are the only residues capable of forming the interchain disulfide bridges necessary to link the two heavy chains of the immunoglobulin molecule. The minimization procedure of the dimer model was performed as described above for the monomer.

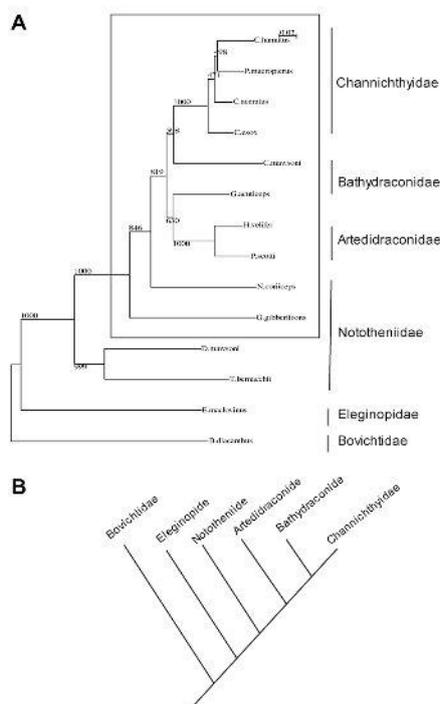
### 3. Results

#### 3.1. Analysis of the secretory I $\mu$ transcripts in the notothenioid species

Using a PCR strategy, mature transcripts encoding the constant domains of secretory  $\mu$  chain were sequenced from 12 different species, belonging to the suborder Notothenioidei. Alignment of the deduced amino acid sequences corresponding to each constant domain is shown in Supplementary Fig. 1A. The domain distribution of *N*-glycosylation sites was analyzed. Compared to non-notothenioids, notothenioid species possess more sites in the CH2 domain and generally lack those in CH3. Different I $\mu$  domains showed a different degree of diversity between the species as indicated by the Shannon Entropy plot (Supplementary Fig. 2), with CH4 being the most conserved domain, as reported for all teleost species (Wilson and Warr, 1992).

Based on the alignment of the CH4 domain amino acid sequences, a distance-based tree was constructed by the Neighbor-Joining method (Fig. 2A). The topology was found to be consistent with the evolutionary radiation of notothenioid families (Fig. 2B) (Near and Cheng, 2008).

The amino acid sequences of *B. diacanthus* and *C. hamatus*, belonging to the most distant families Bovichtidae (non-Antarctic) and Channichthyidae (Antarctic), respectively, displayed 61.4% similarity. It is to be noted that when the *B. diacanthus* sequence was compared to sequences from several species belonging to suborders other than Notothenioidei, amino acid similarity values were found to be higher: 64.7% with *Siniperca chuatsi* (Percoidae), 63.4% with *Epinephelus coioides* (Percoidae), 62.2% with *Anarhichas minor* (Zoarcoidei). These identity values indicate that a high degree of diversification occurred during the speciation of the phylogenetically basal non-Antarctic and the recent Antarctic species (Table 1).



**Fig. 2.** (A) Neighbor-joining tree showing the distances between 14 notothenioid IgH sequences. The tree was constructed by using the amino acid sequences of the CH4 domain (see Supplementary Fig. 1A). The clade boxed comprises the species that showed membrane-bound IgH sequences lacking the CH3 exon, as described in paragraph 2 in the Section 3. (B) Schematic drawing of the phylogenetic tree of the families belonging to the suborder Notothenioidei.

### 3.2. Membrane-bound IgH lacks the CH3 domain in most notothenioid species

Mature transcripts encoding constant domains of the membrane-bound IgH chain were also amplified and sequenced from 13 notothenioid species. The deduced amino acid sequences encoded by TM exons are showed in the Supplementary Fig. 1B. The mRNA splicing occurs in some species in the same manner as generally observed in other teleost fish, whereas in the majority of species it follows a different pattern, leading to the exclusion of

both CH3 and CH4 exons. In parallel with the CH3–CH4 exclusion, an elongation of the EMPD, consisting of 13-aa tandem repeats, occurred. With the exception of *C. exos*, the species following this atypical pattern live in the Antarctic environment and all belong to the clade boxed in the tree shown in Fig. 2A.

An extra *N*-glycosylation site was observed in all sequences. In the case of *T. bernacchii* and *D. mawsoni*, this is generated by the CH3 domain C-terminal Asn and the first Ser encoded by the TM1 exon. The EMPD is somewhat unusual, with an average length of 54.9-aa residues, whereas in other teleosts the mean length is 22.2 aa (21–24 aa). Furthermore, the amino acid composition reveals a relative abundance of negatively charged residues (aspartic and glutamic acid, 23.7%). Cysteine and proline residues account for 6.2 and 4.2% of the total, respectively.

To assess whether the splicing type is a species-specific feature, we sequenced membrane IgH cDNA from different tissues of several *C. hamatus* and *T. bernacchii* individuals. The mRNA splicing pattern was found to be specific for each species (data not shown).

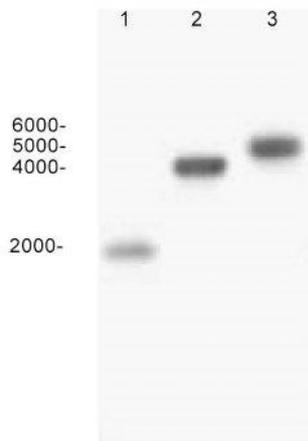
### 3.3. Analysis of the genomic region encompassing the CH2 and TM exons in *C. hamatus*

Genomic analysis was conducted in an attempt to elucidate the different splicing mechanisms found in different species. To this end, genomic fragments encompassing the CH2 and TM exons in some species showing the classical (*T. bernacchii*) or the atypical (*C. hamatus*, *C. mawsoni*) splicing mechanism were amplified by PCR, using the appropriate oligonucleotides. The PCR products were found to differ in length (Fig. 3); that obtained with *T. bernacchii* DNA was about 2000-bp long; the products obtained with the DNA of *C. hamatus*, and *C. mawsoni* were about 4500 and 5400-bp long, respectively. The *T. bernacchii* and *C. hamatus* PCR products were cloned, sequenced and analyzed (Accession numbers EU884293 and EU925137, respectively).

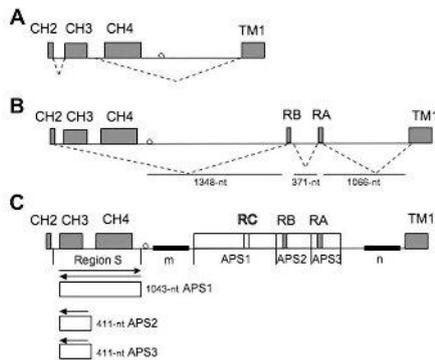
The results obtained are schematically represented in Fig. 4. In the case of *T. bernacchii*, the fragment of 2022-nt shows a gene organization similar to that seen in all the other teleost species (Fig. 4A). By contrast, in the case of *C. hamatus* (Fig. 4B), a 4567-nt fragment was obtained. Its sequence was similar to that of *T. bernacchii* in the upstream portion of the secretory polyA signal. The downstream sequence analysis revealed a 1348-nt non-coding sequence encompassing the secretory polyA signal and the RB exon; the RB 39-nt exon; a 371-nt intron between the RB and RA exons; the RA exon; a 1066-nt intron between the RA exon and the TM1 exon; and lastly the TM1 exon portion including the sequence complementary to the oligonucleotide used as PCR primer. A further analysis of this 4567-nt fragment revealed that the RA and RB exons were included in two larger repetitive sequences, each 411-nt long, denoted as APS2 and APS3, that were highly similar in sequence to the anti-parallel of an upstream region beginning in the CH2–CH3 intron, and ending in the intron between CH3 and CH4. Further upstream, a third region, 1043-nt long, denoted as APS1 was identified as the anti-parallel of the sequence located between the CH2–CH3 intron and the 3'UTR (Region 5); within APS1 a non-coding 39-nt sequence (RC) very similar to RA

**Table 1**  
Percentage similarity between secretory IgH sequences of notothenioid species belonging to different families.

	<i>B. diacanthus</i>	<i>E. maclovinus</i>	Nototheniidae	Artedidraconidae	Bathydraconidae	Channichthyidae
<i>B. diacanthus</i>		59.0	59.5	60.5	62.0	61.0
<i>E. maclovinus</i>			61.2	60.5	62.0	69.4
Nototheniidae				79.1	75.5	80.2
Artedidraconidae					86.5	83.2
Bathydraconidae						84.1
Channichthyidae						



**Fig. 3.** PCR experiments, using as forward primer an oligonucleotide sequence reproducing a sequence in the *T. bernacchii* CH2 exon and as reverse primer a sequence in the TM exon of the same species. Samples corresponding to PCR products of the template genomic DNA were loaded as follows: *T. bernacchii* (lane 1), *C. hamatus* (lane 2), and *C. mawsoni* (lane 3). The PCR products were confirmed as being part of the IgH locus by hybridizing the blotted bands to a labeled *C. hamatus* Igμ probe, as described in the Section 2. Marker mobilities are indicated on the left.



**Fig. 4.** Schematic representation of intron-exon structure of the CH2 to TM1 gene region in *T. bernacchii* (A) and *C. hamatus* (B). Open circle denotes the secretory polyA site and dashed lines indicate the splicing pattern observed in the two species. Shaded boxes indicate coding exons and a thin line, to scale, indicates introns. In C, duplicated regions are labeled "m" and "n" and are represented by thick horizontal lines; the anti-parallel sequences (APS1, APS2, and APS3) are represented by open rectangles; the 39-nt exons RA and RB are designated as narrow filled rectangles; the 39-nt homolog non-coding RC region in APS1 is designed as narrow open rectangle. In C, the APS1 to APS3 boxes are also reported below the gene structure to show the exon-intron regions they overlap.

and RB, was observed. Finally two 555-nt sequences denoted as m and n, sharing 99.6% identity were defined (Fig. 4C). The complete nucleotide sequence of the 4567-nt fragment is reported in the Supplementary Fig. 3.

### 3.4. Comparative analysis of the anti-parallel sequences and their counterpart

The percentage of nucleotide identity between the 411-nt homologous regions of APS1, APS2, APS3 and their respective reverse complement counterpart together with the relative distances tree is reported in Fig. 5.

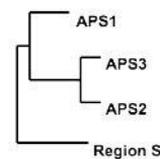
Upon analysis of the nucleotide substitutions in APS1, APS2 and APS3 and their counterpart in the Region S, higher mutability in coding than in non-coding regions was observed; in fact while 14.1% of positions are changed in coding regions, only 2.6% of non-coding positions are mutated. These results indicate that coding sequences underwent a different evolutionary pressure compared to non-coding segments.

The gaining of donor and acceptor splice sites in the reverse complement sequence results in a mature transcript comprising the sequences encoding the 13-aa repeats RB and RA present therein. The splice sites, the branch site, the polypyrimidine tract, and the putative regions which interact with the spliceosomal components U1, U5, and U2 in the APS1, APS2, and APS3 were analyzed (Supplementary Fig. 4). Although the canonical splice sequences are present in all cases, RB and RA are found in the mature transcript, whereas RC, the homologous sequence present in APS1, is absent. In order to determine whether structural constraints are the underlying cause of the missing 39-nt RC, folding of the primary transcript region encompassing the sequenced region between CH2 and TM exons was predicted by using the mfold computational tool. The results showed a very compact structure (Supplementary Fig. 5) stabilized by a free energy of  $-3835.10$  kcal/mole at the Antarctic physiological temperature of 273 K. The presence of a long duplex comprising the entire CH2 and CH4 exons bound to the APS1 sequence, accounts for a free energy contribution of  $-2497.68$  kcal/mole (65% of the total). As a consequence, since the CH3 exon 5' splice site and the RC 39-nt region 3' splice site are included in this duplex, they are not available to bind the spliceosomal components. Although APS2 and APS3 are intrinsically capable of forming a duplex with their counterpart sequence in the CH3 region, the resulting structures are less stable ( $-811.88$  kcal/mole) and thus thermodynamically unfavorable.

### 3.5. The 13-aa repeats showed a positive selection

Based on comparison with their counterpart, several substitutions were identified in *C. hamatus* RA, RB, and RC 39-nt regions (Fig. 6). The RC sequence, which was not included in the mature transcript, showed only one substitution with respect to its counterpart, while the sequences RB and RA, which were used as addi-

	APS2	APS3	Region S
APS1	92.3%	92.9%	92.0%
APS2		96.9%	89.8%
APS3			89.8%



**Fig. 5.** Above, percentage nucleotide identity between the shared sequences of the reverse-complement regions APS1, APS2, APS3 and their counterpart in the CH3 domain (Region S) is reported. Below, the distance-based tree describing their relationships is shown.

```

S      gacaaaacttggtgccctttatccattcttcatacgtgat
RC     gacaaaacttggtgccctttatccattcttcatacgtgat
RB     GACAAACTTGGTGTCCCTTGTCCATGCCACACATGTGAT
RA     GACAAAGTTGGTGTCCATCGTCCATGCCACATCACTGAT

```

**Fig. 6.** Alignment of RA, RB, RC with the reverse complement counterpart (in the Region 5). The 21 bases in RA, RB, RC, which differ from those in the same position in Region 5, are in bold. Codons of RA and RB are alternatively shaded; in these regions, 19/20 substitutions occur in the first and/or second codon positions and thus are non-synonymous. The bases introducing a cysteine codon are in bold, italic.

tional exons, showed 8 and 12 substitutions, respectively (Fig. 6); 19 of 20 substitutions in RB and RA, occurring in the first and/or second codon positions, were found to be non-synonymous. It is also noteworthy that 5 of 20 substitutions introduced a cysteine codon, indicating a preference for this amino acid in the synthesized protein.

The analysis of the 39-nt sequences was extended to all species showing atypical splicing and insertion of 39-nt sequences. Each repeat shares, on average, 89.7% nucleotide identity with the respective counterpart in the CH3 exon. The ratio of synonymous to non-synonymous base substitutions was calculated by the SNAP program and found to be  $dN/dS = 1.82$ . The alignment of the 39-nt sequences from all species analyzed, indicates that the extent of sequence conservation is smaller than the relative counterpart sequence alignment (Supplementary Fig. 6). This observation leads to the suggestion that a high level of diversity developed in the notothenioid genomes during or after these insertion events.

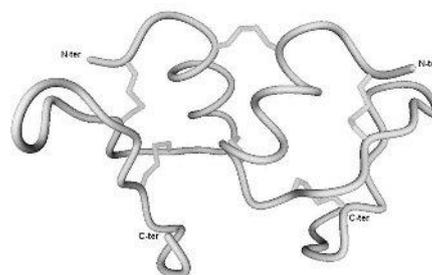
### 3.6. Structural analysis of the 13-aa repeats arrangement

The consensus sequence of all 13-aa repeats was found to be the following: DKLGALCPCHIRD. Displaying a clearly symmetrical hydrophobic character, charged hydrophilic residues are at the ends whereas the hydrophobic residues form the core. Thus, the region resulting from the assembly of several repeats is composed of alternate hydrophilic and hydrophobic parts, as demonstrated by the hydrophathy plot calculated by using the Proparam tool and the Kyte-Doolittle indexes (data not shown).

In a recent paper, we reported the modeling of the helix of *C. hamatus* Ig $\mu$  crossing the membrane and investigated its dimerization in a POPC lipid bilayer model system by Molecular Dynamics (Merlino et al., 2008). This work was carried out in an attempt to understand the structural significance of the presence and character of the 13-aa repeats.

For this purpose a molecular model of the region comprising the C-terminal six residues of *C. hamatus* CH2 and the two 13-aa repeats of the EMPD was built. The model was constructed by homology modeling using a region of the crystal structure of MASP2 as template; though showing a low sequence identity (24%), the region was chosen due to the presence of several features shared with our sequence, namely the presence of cysteine residues in the same position, the location between two well defined domains and the role in homodimerization promotion. The model built was validated by the WHAT IF package and minimized for 20 ps (Supplementary Mov. 1).

A model of the homodimer of the two heavy chains was built by imposing two interchain disulfide bridges between the two Cys5 and the two Cys13 residues, being the only cysteine residues available to covalently link the two heavy chains of the membrane-bound *C. hamatus* Ig. The resulting homodimer model was also minimized for 20 ps (Supplementary Mov. 2) resulting in an atypical domain, unique to the Ig structure, whose symmetrical structure showed peripheral hydrophilic regions and a compact hydrophobic core (Fig. 7).



**Fig. 7.** Molecular model of the dimer of *C. hamatus* Ig $\mu$  region including the C-terminal six amino acid residues of the CH2 domain and the two 13-aa tandem repeats.

## 4. Discussion

The results presented here suggest an important gene modification event that appeared in the most recent common ancestor of *Disostichus-Trematomus* genera and remaining notothenioids, during the radiation of Nototheniidae (Fig. 2), initiated about 24 MYA (Near, 2004). It consisted in a large insertion in the Ig $\mu$  gene locus in parallel with a significant modification of the mRNA splicing mechanism resulting in a membrane antigen receptor of only two constant Ig domains. This event occurred during the major cooling period thus suggesting an adaptive requirement. However, two points do not match well with the hypothesis that the event was driven by cold adaptation. *C. esox*, a species living in sub-antarctic regions, shows the same features as those observed in the Antarctic species. This could be explained by the fact that this species shared an original Antarctic habitat with all the Channichthyidae. The other inconsistent result was the splicing mechanism found in two species living in the Antarctic seas, *T. bernacchii* and *D. mawsoni*, which differed from the pathway followed by all the other Antarctic species analyzed in the present paper. The observations relating to *T. bernacchii* and *D. mawsoni* suggest that the new splicing mechanism was not strictly necessary for species survival, but was maintained later on, probably due to its advantageous nature.

To date, this shortened form resembles only a few cases that have been previously described: (i) in mouse, membrane IgD heavy chain from a plasmacytoma tumor was shown to possess only two constant domains (Tucker et al., 1980); (ii) in sharks, IgNAR mRNA is alternatively spliced to produce two distinct membrane-bound forms, one of them displaying two constant domains (Rumfelt et al., 2004); (iii) in the chondrosteian *Acipenser baeri*, three different splice patterns producing the membrane Ig $\mu$  form have been characterized, one of them utilizing the CH1 exon 3' splice site (Lundqvist et al., 2009).

The contribution that a shorter antigen receptor makes to functional improvement is difficult to ascertain. The antigen occupancy of the antibody site communicates with the transducing machinery that resides in the cytoplasm. This signal could be more efficiently transmitted through a shorter receptor chain, since the conformational variations would be less than those usually required; consequently, the amount of free energy involved in the signal transduction could be smaller.

Furthermore, the question remains as to whether the addition of new cysteine residues is necessary for the assembly of the two heavy chains in the absence of the CH3 domain. The molecular model built here supports just such a hypothesis. In the absence

of experimental validation, this remains a hypothetical possibility which is nonetheless consistent with a computational validation, by 20 ps minimization and WHAT IF analysis. We suggest that there are two parallel disulfide bridges linking the two heavy chains. However, the possibility of the presence of multiple isoforms, differing in disulfide connectivity, as demonstrated for human IgG2 (Wypych et al., 2008), cannot be excluded. The presence of many outward facing charged residues would imply some electrostatic interactions with the heterodimer CD79.

Additional exons encoding domain-connecting regions, such as the hinge region of the human IgG3, are known to be derived from a homologous nucleotide sequence which is present in the same gene (Huck et al., 1986). Indeed, to our knowledge, the insertion of a long reverse complement sequence in a gene structure, such as that reported in the present work, is unique. The high level of homology shared by the APS elements and their counterparts is particularly surprising, as is the greater propensity of the short exons to vary.

The atypical splicing mechanism may result from the difficulty the splicing complex has in interacting with the CH3 exon 5' splice site, probably because it resides in a strong duplex structure formed by the APS1 and its counterpart (Supplementary Fig. 5). The loss in the mature transcript of the RC 39-nt sequence included in APS1, as confirmed by sequencing transcripts from eight individuals, might have a similar explanation: despite possessing the same splice sites as those that are present in the other two expressed 39-nt sequences, both of these sites are, however, inaccessible due to the fact that they reside in the same strong duplex as the CH3 splice sites, and thus are not used. It is worth noticing also that the non-expressed RC 39-nt sequence of APS1 does not present the mutations found in the expressed 39-nt sequences. The results of thermodynamic calculations indicate that the pre-mRNA folding is temperature-dependent. In fact, it was observed that at higher temperature (310 K) the pre-mRNA structure is less stable ( $\Delta G = -2357.54$  kcal/mol) and the duplex formation occurs between different complementary regions (Region 5 interacts partially with APS1 and partially with APS2). This leads to the conclusion that the splicing mechanism is dependent on pre-mRNA folding, which only at the physiological subzero temperatures, assumes a structure accounting for the atypical splicing.

The evolutionary history of the insertion is likely to have involved a stepwise progression which initiated with a gene duplication event. Duplications of genes are frequent in notothenioids: in fact we have previously demonstrated an IgH gene locus duplication in several Antarctic species, the duplicated genes being both expressed and resident on different chromosomes (Pisano et al., 2007). Whether copies of APS sequences have been inserted in the duplicated gene once, or on a number of occasions remains a mystery. Sequencing the genomic DNA of many species would allow the evolutionary history of the insertion to be traced back in time.

Further areas of exploration would be to determine whether the genome of each species possesses the same number of APS elements and whether a longer APS element that does not contribute to the mature transcript is always present. Preliminary data indicate that genomic PCR products obtained from different species show different sizes, consistent with the different number of repeats found in the cDNA sequences. In fact, as shown in Fig. 3, the *C. mawsoni* PCR genomic fragment was about 900-bp longer than the corresponding *C. hamatus* fragment. Two more APS elements in *C. mawsoni* might account for the size difference, in agreement with the two additional exons found in the cDNA sequence.

Our next objective is to address these questions, which would certainly be useful to the understanding of both phylogenetic relationships, and, more generally, the adaptive molecular mechanisms which underlie these relationships.

The results reported here provide new data relating to the structure of the Antarctic teleost immunoglobulin, whilst describing novel mechanisms that occurred throughout evolution.

#### 4.1. Data deposition

The sequences reported in this paper have been deposited in GenBank database (Accession Nos: *B. diacanthus* EU884292, EU240584; *C. aceratus* EU161632; *C. esox* EU192076–7; *C. hamatus* EU192078–9, EU925137; *C. mawsoni* EU232635–6; *D. mawsoni* EU488726, EU240586; *E. maclovinus* EU221318–9; *G. acuticeps* EU240585, EU340099; *H. vetifer* EU195863, EU200778; *P. macropterus* EU195861–2; *P. scotti* EU192080–1; *T. bernacchii* AF094531, EU040247, EU884293).

#### Acknowledgments

We are grateful to our colleagues Dr. D. de Pascale and E. Cocca for having provided us with tissue samples from some species. We thank Dr. Martin Flajnik for helpful discussions and critical reading of the manuscript. This study was carried out in the context of the international multidisciplinary program EBA (Evolution and Biodiversity in the Antarctic: <http://www.eba.aq/>). PNRA (Antarctic Research National Program: <http://www.pnra.it/>) supported this work with the project 2005.2.1.oreste.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.09.033.

#### References

- Abelli, L., Coscia, M.R., De Santis, A., Zeni, C., Oreste, U., 2005. Evidence for hepatobiliary transport of immunoglobulin in the Antarctic teleost fish *Trematomus bernacchii*. *Dev. Comp. Immunol.* 29, 431–442.
- Artero, S., Lefranc, M.P., 2000. The teleostei immunoglobulin heavy IGH genes. *Exp. Clin. Immunogenet.* 17, 148–161.
- Bengtén, E., Quiniou, S.M., Stuge, T.B., Katagiri, T., Miller, N.W., Clem, L.W., Warr, G.W., Wilson, M., 2002. The IgH locus of the Channel catfish, *Ictalurus punctatus*, contains multiple constant region gene sequences: different genes encode heavy chains of membrane and secreted IgD. *J. Immunol.* 169, 2488–2497.
- Cheng, C.H., Detrich, H.W., 2007. Molecular ecophysiology of Antarctic notothenioid fishes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 2215–2232.
- Chen, Z., Cheng, C.H., Zhang, J., Cao, L., Chen, L., Zhou, L., Jin, Y., Ye, H., Deng, C., Dai, Z., Xu, Q., Hu, P., Sun, S., Shen, Y., Chen, L., 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. USA* 105, 12944–12949.
- Coscia, M.R., Oreste, U., 1998. Humoral immune response of Antarctic fish to the nematode *Contracaecum osculatum*. *Fish Shellfish Immunol.* 8, 295–302.
- Coscia, M.R., Borriello, A., Oreste, U., 2000a. Plasma and bile antibodies of the teleost *Trematomus bernacchii*, specific for the nematode *Pseudoterranova decipiens*. *Dis. Aquat. Org.* 41, 37–42.
- Coscia, M.R., Morea, V., Tramontano, A., Oreste, U., 2000b. Analysis of a cDNA sequence encoding the immunoglobulin heavy chain of the Antarctic teleost *Trematomus bernacchii*: evidence of unusual features. *Fish Shellfish Immunol.* 10, 343–357.
- Coscia, M.R., Oreste, U., 2002. Limited diversity of the immunoglobulin heavy chain variable domain of the emerald rockcod. *Fish Shellfish Immunol.* 14, 71–92.
- Coscia, M.R., Giacomelli, S., De Santi, C., Variale, S., Oreste, U., 2008. Immunoglobulin light chain isotypes in the teleost *Trematomus bernacchii*. *Mol. Immunol.* 45, 3096–3106.
- Danilova, N., Bussmann, J., Jekosch, K., Steiner, L.A., 2005. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat. Immunol.* 6, 295–302.
- Eastman, J.T., 1993. In: *Antarctic Fish Biology: Evolution in a Unique Environment*. Academic Press, San Diego, USA.
- Hansen, J.D., Landis, E.D., Phillips, R.B., 2005. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: implications for a distinctive B cell developmental pathway in teleost fish. *Proc. Natl. Acad. Sci. USA* 102, 6919–6924.
- Hordvik, L., 2002. Identification of a novel immunoglobulin delta transcript and comparative analysis of the genes encoding IgD in Atlantic salmon and Atlantic halibut. *Mol. Immunol.* 39, 85–91.
- Huck, S., Fort, P., Crawford, D.H., Lefranc, M.P., Lefranc, G., 1986. Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene.

- comparison with the other human C gamma gene. *Nucleic Acids Res.* 14, 1779–1789.
- Lundqvist, M., Stromberg, S., Bouchenot, C., Pålström, L., Boudinot, P., 2009. Diverse splicing pathways of the membrane *IgHM* pre-mRNA in a Chondrosteian, the Siberian sturgeon. *Dev. Comp. Immunol.* 33, 507–515.
- Lynch, K.W., 2004. Consequences of regulated pre-mRNA splicing in the immune system. *Nat. Rev. Immunol.* 4, 931–940.
- Merlino, A., Varriale, S., Coscia, M.R., Mazzarella, L., Oreste, U., 2008. Structure and dimerization of the teleost transmembrane immunoglobulin region. *J. Mol. Graph. Model.* 27, 401–407.
- Near, T.J., 2004. Estimating divergence times of notothenioid fishes using a fossil-calibrated molecular clock. *Antarctic Sc.* 16, 37–44.
- Near, T.J., Cheng, C.H., 2008. Phylogenetics of notothenioid fishes (Teleostei: Acanthomorpha): inferences from mitochondrial and nuclear gene sequences. *Mol. Phylog. Evol.* 47, 832–840.
- Nei, M., Gjobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.
- Oreste, U., Coscia, M.R., 2002. Specific features of immunoglobulin VH genes of the Antarctic teleost *Trematomus bernacchii*. *Gene* 295, 199–204.
- Ota, T., Nguyen, T.A., Huang, E., Detrich, H.W., Amemiya, C.T., 2003. Positive Darwinian selection operating on the immunoglobulin heavy chain of Antarctic fishes. *J. Exp. Zool.* 295B, 45–58.
- Park, J.Y., Jongstra-Bilen, J., 1997. Interactions between membrane IgM and the cytoskeleton involve the cytoplasmic domain of the immunoglobulin receptor. *Eur. J. Immunol.* 27, 3001–3009.
- Peterson, M.L., 2007. Mechanisms controlling production of membrane and secreted immunoglobulin during B cell development. *Immunol. Res.* 37, 33–46.
- Pisano, E., Coscia, M.R., Mazzei, F., Ghigliotti, L., Coutanceau, J.P., Ozouf-Costaz, C., Oreste, U., 2007. Cytogenetic mapping of immunoglobulin heavy chain genes in Antarctic fish. *Genetica* 130, 9–17.
- Rumfelt, L.L., Avila, D., Diaz, M., Bartl, S., McKinney, E.C., Flajnik, M.F., 2001. A shark antibody heavy chain encoded by a nonsomatically rearranged VDJ is preferentially expressed in early development and is convergent with mammalian IgG. *Proc. Natl. Acad. Sci. USA* 98, 1775–1780.
- Rumfelt, L.L., Diaz, M., Lohr, R.L., Mochon, E., Flajnik, M.F., 2004. Unprecedented multiplicity of Ig transmembrane and secretory mRNA forms in the cartilaginous fish. *J. Immunol.* 173, 1129–1139.
- Saha, M.R., Suetake, H., Kikuchi, K., Suzuki, Y., 2004. Fugu immunoglobulin D: a highly unusual gene with unpredicted duplication in its constant region. *Immunogenetics* 56, 438–447.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Sambrook, J., Fritsch, E.F., Maniatis, T., 2001. *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press.
- Savan, R., Aman, A., Nakao, M., Wanatuki, H., Sakai, M., 2005a. Discovery of a novel immunoglobulin heavy chain gene chimera from common carp. *Cyprinus carpio* L. *Immunogenetics* 57, 458–463.
- Savan, R., Aman, A., Sato, K., Yamaguchi, R., Sakai, M., 2005b. Discovery of a new class of immunoglobulin heavy chain from fugu. *Eur. J. Immunol.* 35, 3320–3331.
- Stewart, J.J., Lee, C.Y., Ibrahim, S., Watts, P., Shlomchik, M., Weigert, M., Litwin, S., 1997. Entropy analysis of immunoglobulin and T cell receptor. *Mol. Immunol.* 34, 1067–1082.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Tucker, P.W., Liu, C.P., Mushinski, J.F., Blattner, F.R., 1980. Mouse immunoglobulin D: messenger RNA and genomic DNA sequences. *Science* 209, 1353–1360.
- Wilson, M.R., Warr, G.W., 1992. Fish immunoglobulins and the genes that encode them. *Ann. Rev. Fish Dis.* 2, 201–221.
- Wypych, J., Li, M., Guo, A., et al., 2008. Human IgG2 antibodies display disulfide-mediated structural isoforms. *J. Biol. Chem.* 283, 16194–16205.
- Zhao, Y., Pan-Hammarström, Q., Yu, S., Wertz, N., Zhang, X., Li, N., Butler, J.E., Hammarström, L., 2006. Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*. *Proc. Natl. Acad. Sci. USA* 103, 12087–12092.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.



Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



## An evolutionary conserved motif is responsible for Immunoglobulin heavy chain packing in the B cell membrane

Sonia Varriale<sup>a,b,1</sup>, Antonello Merlino<sup>b,c,1,\*</sup>, Maria Rosaria Coscia<sup>a</sup>, Lelio Mazzarella<sup>b,c</sup>, Umberto Oreste<sup>a</sup>

<sup>a</sup> Institute of Protein Biochemistry, CNR, via P. Castellino 111, 80131 Napoli, Italy

<sup>b</sup> Department of Chemistry University of Naples "Federico II", via Cintia, 80126 Napoli, Italy

<sup>c</sup> Institute of Biostructures and Bioimages, CNR, via Mezzocannone 16, 80100 Napoli, Italy

### ARTICLE INFO

Article history:  
Received 3 June 2010  
Revised 27 September 2010  
Accepted 28 September 2010  
Available online xxx

Keywords:  
Immunoglobulin heavy chain  
Transmembrane helix  
POPC lipid bilayer  
Immunoglobulin evolution  
Molecular dynamics  
Immunoglobulin isotypes  
Purifying selection

### ABSTRACT

All species of vertebrates synthesize immunoglobulin molecules, which differ in a number of aspects but also share a few common features responsible for their function, such as the presence of a transmembrane domain in the membrane bound form of the immunoglobulin heavy chain (IgTMD) that ensures communication with the signal transducing Ig $\alpha$ -Ig $\beta$  peptides. We have analyzed the gene sequence encoding the IgTMD of different heavy chain isotypes of very distant species, from shark to mammals. The IgTMD sequences show a high degree of sequence identity and their encoding nucleotide sequences were shown to be subject to purifying selection at most sites. We have built molecular models of seven IgTMDs from different vertebrate species and have investigated the formation of homodimer in a palmitoyl oleoyl phosphatidylcholine (POPC) lipid bilayer by molecular dynamics simulations. We found that the conserved FXXXFXS/TXXXS motif, never observed to date in protein transmembrane chains, is responsible for the two heavy chains association through two pairs of Phe–Phe hydrophobic interactions and two pairs of Ser/Thr–Ser/Ser hydrogen bonds. This interaction pattern, which stabilizes the dimer conformation in the lipid bilayer, was unique, being different from any other pattern identified in transmembrane helices to date.

© 2010 Elsevier Inc. All rights reserved.

### 1. Introduction

The evolutionary success of the Immunoglobulin (Ig) molecule is due not only to its ability to discriminate different epitopes, but also to its structural ability to transduce the message derived from the antibody site occupancy to the cell interior. The Ig molecule has diversified its amino acid sequence throughout vertebrate evolution generating different heavy chain isotypes, each with specific functionality. Isotypes differ in the number of constant domains, degree of polymerization and glycosylation, linking of additional peptide components and ability to recognize different

Fc receptors. The isotypes present in each vertebrate class are summarized in Table 1. Despite the enormous heterogeneity found among the IgH isotypes, a number of universal elements are identifiable: the presence of a V domain at the amino terminus; the presence of at least two C1 type domains of the IgSF domains; the dimeric structure IgH–IgH. In addition, the immunoglobulin membrane-bound form of all isotypes communicates the antigen binding event to the B cell interior via the same Ig $\alpha$ -Ig $\beta$  dimer.

The transmembrane domain of the membrane-bound form of the heavy chain (IgTMD) is jointly responsible for the functionality of the IgH dimer and for the interaction with Ig $\alpha$ -Ig $\beta$ . Campbell defined the Conserved Antigen Receptor Transmembrane (CART) motif, as the sequence shared by all antigen receptors (sequon), and suggested a common outward orientation of its amino acid side chains in the transmembrane helix–helix dimer (Campbell et al., 1994). The conservation of this motif throughout the vertebrate evolution suggests that it should have a functional role (Rumfelt et al., 2004). However, at present, the mechanism responsible for IgTMD helices packing remains unsolved.

Interaction of the transmembrane helices is a key topic that has been extensively investigated in many membrane proteins. Specific TM helices packing motifs have been identified: the GXXXG and AXXXA motifs, located in the central region of associated TM helices (Kleiger et al., 2002; Schneider and Engelman, 2004) and

Abbreviations: BCR, B cell receptor; CART, conserved antigen receptor transmembrane; CCP4, collaborative computational project number 4; DSSP, dictionary of protein secondary structure; EMPD, extra-cellular membrane proximal domain; FEL, fixed effects likelihood; Ig, immunoglobulin, IgH, immunoglobulin heavy chain; IgL, immunoglobulin light chain; IGSF, immunoglobulin superfamily, IgTMD, immunoglobulin transmembrane domain; LG, Le and Gascuel; LINC, linear constraint solver; MD, Molecular dynamics; ML, maximum-likelihood; NJ, neighbor-joining; PDB, protein data bank; POPC, palmitoyl oleoyl phosphatidylcholine; REL, random effects likelihood; RMSD, root mean square deviation; SLAC, single-likelihood ancestor counting; VLR, variable lymphocyte receptor.

\* Corresponding author at: Department of Chemistry, University of Naples "Federico II", via Cintia 80126 Napoli, Italy. Fax: +39 081674090.

E-mail address: antonello.merlino@unina.it (A. Merlino).

<sup>1</sup> These authors equally contribute to this work.

1055-7903/\$ - see front matter © 2010 Elsevier Inc. All rights reserved.  
doi:10.1016/j.ympev.2010.09.022

Please cite this article in press as: Varriale, S., et al. An evolutionary conserved motif is responsible for Immunoglobulin heavy chain packing in the B cell membrane. Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.09.022

**Table 1**  
Immunoglobulin isotypes distribution in the different vertebrate classes.

	IgM	IgW/IgD	IgV/IgC/IgE	IgA/IgX	IgNAR	IgT/IgZ
Chondrichthyes	●	●			●	
Osteichthyes	●	●	●			●
Amphibia	●	●	●	●		
Reptilia	●	●	●	●		
Aves	●	●	●	●		
Mammalia	●	●	●	●		

the SXSSSXT or SXXXSXXT motifs (Dawson et al., 2002). Different clusters of transmembrane packing motifs have been identified by Walters (Walters and DeGrado, 2006). The helix–helix assembly can also be driven by van der Waals contacts (MacKenzie et al., 1997; Lemmon et al., 1992), though electrostatic interactions (Gratkowski et al., 2001) as well as hydrogen bonds (Zhou et al., 2000) are occasionally observed. Harrington and Ben-Tal (Harrington and Ben-Tal, 2009) showed that closely packed small residues, hydrogen bonds, salt bridges, aromatic interactions and closely packed hydrophobic residues precisely determine the packing of transmembrane helices in 15 diverse well-studied membrane proteins. Examples of membrane proteins whose structure and assembly depend on helix–helix packing in the lipid bilayer include homodimeric rhodopsin (Okada et al., 2000; Patel et al., 2005), glycoporphin A (GpA) (Mottamal et al., 2006), and the cell death factor Bnip3 (Vereshaga et al., 2007).

Molecular dynamics simulations have been shown to be a powerful approach to understanding the association of TM helices (Cut-hbertson et al., 2006; Henin et al., 2005). Our aim is to undertake a thorough investigation of IgTMD conservation across the evolution at both nucleotide and amino acid levels and to correlate the conserved residues to their structural functions. To this end, we used molecular dynamics (MD) simulations of IgTMD dimers from different species to more accurately understand the correlation between residue conservation and packing mode. This approach allowed the identification of a conserved FXXXXFXS/TXXXS motif that is required for transmembrane Ig domain dimerization. We suggest that this motif may be considered as an universal feature that could help to trace the ancestor of the Ig molecule in its evolutionary history.

## 2. Methods

### 2.1. IgTMD dataset

The dataset used in the present study was constructed by searching for both cDNA and genomic DNA sequences encoding membrane-bound IgH in databases. In the case of cDNA sequences, the region specific for the membrane-bound form was identified by comparison with the corresponding sequence encoding the secreted IgH form. The search criterion set was to collect sequences from all the isotypes of all the vertebrate classes; the IgM isotype was represented from more species being the clearest evolutionary reference point. The 68 sequences selected in this study are listed in Table S1.

### 2.2. PDBTM dataset

A comprehensive nonredundant dataset (PDBTM dataset) of membrane protein chains possessing experimentally determined structures was generated using the protein Data Bank of transmembrane proteins (PDBTM <http://pdbtm.enzim.hu/>) as of March 10, 2010. Criteria of selection were: annotation of at least one transmembrane helix; less than 25% sequence identity to each other, exclusion of proteins whose structure has been inferred

using the Homology Modelling method. Helical fragments in protein structures were identified by DSSP, the Dictionary of Protein Secondary Structure (Kabsch and Sander, 1983). To guarantee at least two turns in all helices, only those with a number of residues greater than seven were selected for analysis. The final dataset contained 236 protein chains and 24,886 amino acid residues.

### 2.3. Computational tools

To predict the “out”, “in” and membrane-spanning (transmembrane) regions TMpred ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)) (Hofmann and Stoffel, 1993) and HMMTM (<http://biophysics.biol.uoa.gr/HMM-TM/index.jsp>) (Bagos et al., 2006) programs were used. The amino acid and nucleotide composition was calculated using the ProtParam tool at <http://expasy.org/tools/protparam.html> (Gasteiger et al., 2005). Nucleotide sequence identity was calculated by CLUSTALW 2 (Thompson et al., 1997) at <http://www.ebi.ac.uk/Tools/clustalw2/index.html> (Larkin et al., 2007). Multiple alignments of amino acid residues were obtained with CUSTALX 1.8. Positional diversity was calculated using the Shannon algorithm (Stewart et al., 1997) at [http://hcv.lanl.gov/content/sequence/ENTROPY/entropy\\_one.html](http://hcv.lanl.gov/content/sequence/ENTROPY/entropy_one.html). Web-Logo was used to visualize residue conservation (Crooks et al., 2004) and performed at <http://weblogo.berkeley.edu/>. The consensus nucleotide sequence was determined at <http://coot.embl.de/Alignment/consensus.html> (Schneider and Stephens, 1990).

### 2.4. Measuring codon evolutionary selection

Phylogenetic trees were inferred using the Maximum Likelihood (ML) method (Felsenstein, 2004). For amino acid sequences, the LG model of protein evolution (Le and Gascuel, 2008) was used. For nucleotide sequences, ML analysis was performed by application of the GTR+I+G model. ML analyses were performed with the Phylml 3.0 program (Guindon and Gascuel, 2003).

Tests for positive/purifying selection were carried out by comparing the number of non-synonymous substitutions per non-synonymous site (dN) with the number of synonymous substitutions per synonymous site (dS) in a ML framework. Three likelihood methods, incorporating variation in the rate of synonymous substitution, single-likelihood ancestor counting (SLAC) and random effects likelihood (REL) measuring purifying selection, and fixed effects likelihood (FEL) measuring positive selection (Kosakovsky Pond and Frost, 2005), were implemented on the DATAMONKEY (<http://www.datamonkey.org/>) web server.

### 2.5. MD simulations

Model building of the IgTM regions was manually performed, using the structure of *C. hamatus* IgTMD helix, previously reported (Merlino et al., 2008), as a template. The models were validated by PROCHECK (Laskowski et al., 1993). The homodimers were then obtained by MD simulations in palmitoyl oleoyl phosphatidylcholine (POPC) bilayers using two copies of the IgTMD, with the helix axes aligned, as a starting structure. The starting orientation of the two helices were driven by the successful MD simulations that reproduced the spontaneous formation of parallel IgTM homodimers of *C. hamatus* (Merlino et al., 2008). The process of model insertion in the lipid bilayer and the molecular dynamics (MD) runs in lipid environments were performed as suggested by us and other authors (Vaccaro et al., 2005; Merlino et al., 2008). During this process, a few simulations of 10 ps were performed, with the atomic coordinates of the protein restrained to the initial positions. This resulted in final systems containing 110 lipids and about 3730 water molecules. A summary of the MD simulations performed is reported in Table S2. MD simulations have been

Please cite this article in press as: Varriale, S., et al. An evolutionary conserved motif is responsible for immunoglobulin heavy chain packing in the B cell membrane. Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.09.022

performed with the GROMACS 3.2 package using the procedure described elsewhere (Merlino et al., 2005,2006,2008). Briefly, the temperature was set to 300 K by coupling the system to an external bath (Berendsen et al., 1984). Bond lengths were constrained by the Linear Constraint Solver (LINCS) algorithm (Hess et al., 1997). Electrostatic interactions were calculated with the particle-mesh Ewald algorithm with a cutoff of 0.9 nm (Darden et al., 1993). Lennard-Jones interactions were calculated with a 0.9/1.2 nm twin-range cutoff. A dielectric constant of 1 and a time step of 2 fs were used. The force field and the coordinates for POPC were downloaded from the Tieleman group website (<http://moose.bio.ucalgary.ca/>).

All the trajectories were estimated to assess the quality of the simulations using GROMACS routines and in-house programs. Surface complementarity of protein–protein interfaces was evaluated by calculating the statistical  $S_c$  value (Lawrence & Colman, 1993), using the Collaborative Computational Project Number 4 (CCP4) suite (1994). The statistic  $S_c$  depends both on the relative shape of the surfaces with respect to each other and on the extent to which the interaction brings individual elements of the opposing surfaces into proximity.

### 3. Results

#### 3.1. Amino acid sequences analysis

We analyzed 68 membrane bound IgH amino acid sequences of different isotypes of 33 different vertebrate species listed in Table S1. Using computational tools, we predicted the extracellular membrane proximal (EMPD), the crossing membrane (TMD) and the cytoplasmic (Cyt) regions in each deduced amino acid sequence (Fig. S1).

The EMPD was found to be extremely variable; its length varied in the range of 9–39 aa residues (mean value 20.6). The amino acid composition of each region is reported in Table S3; it is worthy to note (i) a high content in the EMPD of negatively charged residues (Glu + Asp 31.7%); (ii) the high content of Phe (14.9%) and Thr + Ser (28.5%), the absence of negative charges and low content of small side chain residues (Gly and Ala) in the IgTMD; (iii) the high content of positively charged residues (Lys 21.0%) in the cytoplasmic tail.

To determine whether the composition of the IgTMD reflects the trend observed in helical transmembrane regions, we measured the occurrence of the amino acid side chains in 236 helical segments of transmembrane proteins available from the PDBTM dataset (see Methods). In a sample of 24,886 transmembrane helical residues we found 10.3% Ser + Thr, 5.2% negatively charged residues, 6.0% positively charged residues, 18.5% small side chain residues Gly + Ala and 7.2% Phe. The comparison between the IgTMD sequences and the general trend confirms that the IgTMD has a unique sequence composition (Table 2).

An important feature of the IgTMD sequences is the absence of the positively charged residue that is common to the majority of immune receptors and interacts, by way of a salt bridge, with the signaling dimer (Call and Wucherpfennig, 2007). This finding suggests that IgTM helices assemble to Ig $\alpha$ –Ig $\beta$  transducing dimer in a different fashion.

We analyzed the residue conservation in the CLUSTALX alignment of the IgTMD sequences by using the Shannon entropy method. A plot of entropy at each position is shown in Fig. S2. The highest entropic values were found for the 12 positions of the N- and C-terminal regions when compared to the same number of positions in the middle part (total entropic values 24.4 and 12.8, respectively). In particular, we visualized eight conserved residues or conserved residue's character (FXXLFLXS/TXXYSXXXT) by the WebLogo tool we used for position variability analysis (Fig. 1).

**Table 2**

Comparison of the amino acid composition of IgTMD and that of 236  $\alpha$ -helical TM chains included in the PDBTM dataset.

	IgTMD	PDBTM
Ala	6.5	11.4
Arg	0.1	3.1
Asn	1.0	2.2
Asp	–	1.9
Cys	1.3	1.1
Gln	0.1	2.5
Glu	–	3.3
Gly	2.3	7.1
His	–	1.4
Ile	10.1	9.6
Leu	17.8	14.7
Lys	0.1	2.9
Met	1.5	3.3
Phe	14.9	7.2
Pro	0.4	2.3
Ser	10.6	4.8
Thr	17.9	5.5
Trp	2.9	2.4
Tyr	4.1	3.4
Val	8.3	9.9

To evaluate whether this motif has a biological significance for the Ig molecules, we examined its occurrence in the reference PDBTM dataset and never found it.

#### 3.2. Structural analysis

To develop a model for the IgTMD homodimers, we selected seven different 29 residue sequences including the TMDs, with amino acid identity scores ranging between 37% and 89% (Fig. 2) and built homology models of their transmembrane helices. These models were then submitted to MD studies for a total of about 305 ns of simulation. Our simulations show that all the models form well packed homodimers (Fig. 3 and Table 3). The RMSD values calculated for the resulting structures of the dimers are in the range 1.5–2.7 Å, thus indicating the IgTMDs have very similar structures in different Igs.

We investigated the physical principles underlying IgTMD dimer formation and the molecular nature of the IgTMD homodimer assembly. Specific contacts that drive the IgTMDs recognition have been identified (Fig. 4). Side chains from one helix fit into holes on the other helix resulting in dense packing ( $S_c$  values between 0.61 and 0.86), which is mainly stabilized by intermolecular interactions occurring between aromatic rings of two Phe9 and two Phe13 (Fig. 4), respectively (for the numbering scheme see Fig. 2). This is in line with the well-known central role of  $\pi$ -stacking interactions in self-assembly processes in the fields of chemistry and biochemistry (McGaughey et al., 1998; Gazit, 2002). The observed interaction between Phe residues is in contrast to the general idea that shorter side chains optimize helix–helix packing (Jiang and Vakser, 2004). To evaluate the potential biological significance of the presence of the FXXXF motif in transmembrane  $\alpha$ -helices, we examined its potential overrepresentation in the Database of protein transmembrane helices that we have built and we found that this motif is not overrepresented. This result is in line with the statistical analysis performed by Senes and coworkers (Senes et al., 2000).

In the IgTMD homodimer, we also found the presence of hydrogen bonds between the side chains of Ser/Thr16, Ser20 and the corresponding residues of the partner chain. During the simulations, the side chains of Ser residues preferentially adopt the  $g^+$  conformation ( $\chi_1 = 240^\circ$ – $360^\circ$ ).

Please cite this article in press as: Varriale, S., et al. An evolutionary conserved motif is responsible for immunoglobulin heavy chain packing in the B cell membrane. *Mol. Phylogenet. Evol.* (2010), doi:10.1016/j.ympev.2010.09.022

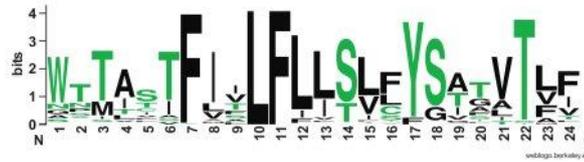


Fig. 1. Sequence conservation, represented by WebLogo tool, in the alignment of the 68 IgTMD sequences of the dataset.

	1	5	10	15	20	25
IgM <i>M.musculus</i>	NLWTTASTFIVLFLLSLFYSTTTLFKVK					
IgM <i>I.punctatus</i>	NMANTALTFVLFELITLFPYSIGVTVFKVK					
IgM <i>P.waltl</i>	NIWSTTSTFIIFLLSLSYSATVTLFKVK					
IgM <i>A.platyrrhynchos</i>	NLWTTASTFIVLFLLSLFYSATVTLIKVK					
IgG2a <i>M.musculus</i>	GLWTTITIFILSFLLSVCYSASVTLFKVK					
IgNAR <i>G.cirratum</i>	NSSLSIMTFVILFLLSMLYSTFVTVMKDL					
IgY <i>A.carolinensis</i>	HLWMTIFIFFILFLLSVCYSATVTLISKVK					

	IgG2a <i>M.mus</i>	IgM <i>A.pla</i>	IgY <i>A.car</i>	IgM <i>P.wal</i>	IgM <i>I.pun</i>	IgNAR <i>G.cir</i>
IgM <i>M.mus</i>	68	89	62	79	62	41
IgG2a <i>M.mus</i>		65	75	65	44	37
IgM <i>A.pla</i>			62	75	55	41
IgY <i>A.car</i>				65	41	41
IgM <i>P.wal</i>					55	41
IgM <i>I.pun</i>						41

Fig. 2. Alignment and percent of similarity among the IgTMD sequences used in MD simulations.

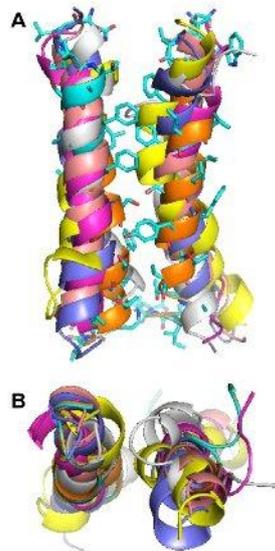


Fig. 3. Structural comparison among the IgTMD homodimers obtained by simulations in POPC. (A) frontal view (B) lateral view.

These findings suggest that  $\pi$ - $\pi$  interactions between Phe residues should drive IgTM helix association whereas Ser and Thr residues provide assistance in the association, by forming interhelical hydrogen bonds.

Regarding the cytoplasmatic region, at the C-terminal tail of the helix, Lys27 and 29 of the two chains make favourable interactions with the oxygen atoms of the POPC molecules on the surface of the lipid bilayer.

Altogether, the results of the simulations and the statistical analysis of the transmembrane helices suggest that the IgTM helix-helix association depends on sequence-specific interactions that involve the unusual FXXXFXXS/TXXXX motif. This motif, never identified so far, has been shown in the present study to act as a sequence promoting transmembrane helix-helix association.

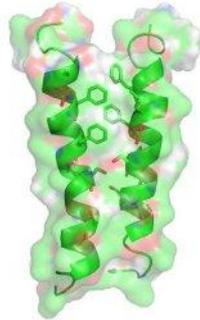
### 3.3. Phylogenetic analysis

To investigate the conservation of the codons encoding the motif, we analyzed the 72-nt region encoding the IgTMD in the nucleotide sequences of the data set listed in Fig. S3, and whose total nucleotide composition is reported in Table S4. The high occurrence of pyrimidine is worth noting (67.8%). In particular, T content is 31.7% in the 72-nt region, whereas it is 42.1% in its 33-nt core. This result agrees with the hypothesis that regions marked with high U content drive the primary transcript preferentially to the ribosomes responsible for the transduction of polypeptides to be exported via the endoplasmic reticulum to the cell surface (Prilusky and Eitan, 2009). In the 72-nt region, 13 positions are conserved in 97.3% of the sequences. Ten out of 13 conserved bases are T.

Phylogenetic trees were constructed using nucleotide or amino acid sequences and NJ or ML methods. Fig. 5 shows the ML tree

**Table 3**  
Structural parameters resulting from the MD simulations.

	Mean helix length (Å)	Mean helix radius (Å)	Mean helix rise (Å)	Radius of gyration per residue (Å)	Mean number of hydrogen bonds
IgM <i>M. musculus</i>	18.9 ± 0.4	2.61 ± 0.04	1.53 ± 0.03	14.2 ± 0.2	33.7
IgM <i>I. punctatus</i>	11.9 ± 0.4	2.57 ± 0.03	1.52 ± 0.03	13.0 ± 0.1	33.3
IgM <i>P. waltii</i>	22.0 ± 0.7	2.7 ± 0.1	1.37 ± 0.05	11.8 ± 0.4	33.9
IgM <i>A. platyrhynchos</i>	28.3 ± 0.5	2.64 ± 0.04	1.57 ± 0.03	13.9 ± 0.1	33.1
IgG2a <i>M. musculus</i>	25.7 ± 0.4	2.44 ± 0.04	1.55 ± 0.02	13.0 ± 0.1	35.6
IgNAR <i>G. cyrratum</i>	21.4 ± 0.3	2.43 ± 0.01	1.49 ± 0.02	13.3 ± 0.1	39.4
IgY <i>A. carolinensis</i>	31.9 ± 0.3	2.44 ± 0.03	1.52 ± 0.01	14.8 ± 0.1	38.5



**Fig. 4.** Interactions that determine the formation and the stability of the IgTMD homodimers in a representative model obtained by POPC simulations. Specific interactions are predominantly mediated by a perfect match between the surface of the two helices.

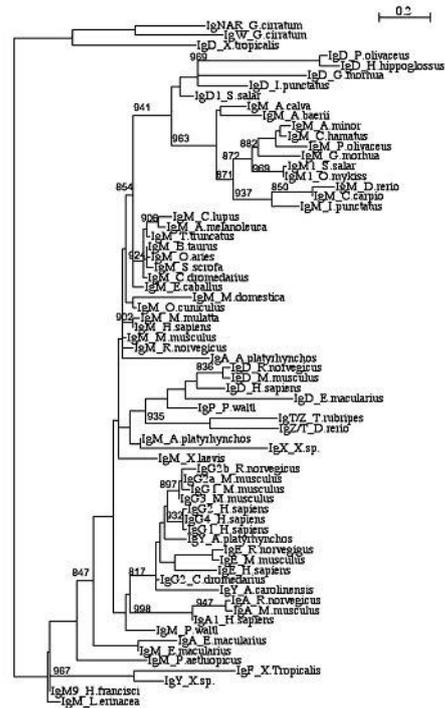
obtained using the nucleotide sequences. The analysis was performed by applying the GTR + I + G model.

Tests for positive/purifying selection were performed on ML or NJ trees, using three different methods (SLAC, FEL and REL): 12 codons underwent purifying selection according to at least one method. Codon positions with the highest probability of being under purifying selection were identified by setting the threshold  $p$  to 0.001. The codons found to be negatively selected by the most severe conditions (SLAC  $p$ -values = 0.000) were 3, 7, 10, 11, 14, 17, 18, 19 and 22. Table 4 shows the values obtained by the three methods based on the ML tree on protein coding sequences. The values obtained on the NJ and ML trees on nucleotides sequences gave similar results (Table S5, Table S6). The negatively selected codons comprise those encoding the amino acid residues identified to be functional in the conserved FXXXFXS/TXXXS motif.

#### 4. Discussion

Our study led to identification of an amino acid motif (FXXXFXS/TXXXS) of the transmembrane domain as universal feature of immunoglobulin molecules. The evolutionary conservation of the codons encoding the motif suggests that it plays a crucial role in immunoglobulin function. Based on MD simulations performed in POPC bilayer we found that this motif determines the stability of the packed transmembrane helices in a manner that is different from that used by other immune receptor components, either homodimeric or heterodimeric, and all known transmembrane proteins.

These results raised a number of questions as to what the origin of this novelty introduced 450 MYA with the adaptive immune response? In the non-vertebrate world, is there any membrane



**Fig. 5.** Phylogenetic Tree constructed by ML method using aligned nucleotide sequences encoding the IgTMD region.

molecule carrying this feature in its transmembrane region? Through preliminary search, the 72-nt consensus sequence was found to share 49.3% identity with a *Eptatretus stoutii* VLRB sequence (AY965571), and 8 of the 13 conserved nucleotide positions of the 72-nt sequence encoding the IgTMD were identified in the VLRB sequence. VLR are variable immune receptors of Agnathan lymphocyte cells (Pancer et al., 2005) phylogenetically unrelated to immunoglobulins, thus our observation requires more detailed investigation.

The obtained tree, shown in Fig. 5, represents a good draw of Ig isotype distances among different species and is consistent with the current hypothesis on isotype differentiation during evolution.

Please cite this article in press as: Varriale, S., et al. An evolutionary conserved motif is responsible for Immunoglobulin heavy chain packing in the B cell membrane. *Mol. Phylogenet. Evol.* (2010), doi:10.1016/j.jympev.2010.09.022

**Table 4**  
Negative selection (at least one method) on 12 codons based on ML tree constructed using protein coding sequences.

Codon number	SLAC dN-dS	SLAC p-value	FEL dN-dS	FEL p-value	REL dN-dS	REL Bayes factor
1	-2.796	0.005	-0.053	0.000	-0.750	0.000
3	-2.818	0.000	-0.067	0.000	-0.915	0.000
6	-1.335	0.012	-0.043	0.000	-0.857	0.000
7	-2.025	0.000	-0.056	0.000	-0.597	0.000
10	-1.511	0.000	-0.029	0.000	-0.904	0.000
11	-1.270	0.000	-0.030	0.000	-0.975	0.000
14	-2.687	0.000	-0.085	0.000	-0.892	0.000
17	-2.629	0.000	-0.108	0.000	-1.040	0.000
18	-2.478	0.000	-0.113	0.000	-0.905	0.000
19	-1.338	0.020	-0.042	0.000	-0.788	0.000
21	-1.175	0.015	-0.029	0.000	-0.834	0.000
22	-2.188	0.000	-0.032	0.000	-0.850	0.000

We extensively studied the assembly of the IgTMD by MD and our findings suggest that the helices pack in a sequence specific manner. Specific interactions between the helices are mainly mediated by a perfect fit between the surfaces of the two helices.  $\pi$ - $\pi$  interactions play a crucial role in the assembly of the two transmembrane helices. After the two helices have come into close contact, hydrogen bonds between Ser/Thr residues can further stabilize the helix dimer. The presence of intermolecular hydrogen bonds in the tail of the helices could be important for transferring signals across the membrane. It can be speculated that the presence of hydrogen bonds could serve as a basis for structural changes. Breaking and reforming hydrogen bonds may allow conformational changes, potentially allowing more flexibility than in the case of close packing between the helices.

Other residues of the IgTM region, besides those involved in the dimerization are conserved throughout the evolution. These include Leu12, Tyr/Phe19, Thr24 and Lys27. Different roles could be attributed to these residues to account for their conservation. Leu12, by its side chain, assists the Phe13–Phe13 hydrophobic interaction in the dimer. The residues at position 19 and 24 have their side chains positioned outwards at the dimer interface and may be good candidates for interact with the Ig $\alpha$ /Ig $\beta$  heterodimer; their high fluctuation scores (data not shown) bear out this hypothesis. Lys27 is always present, except in the mammalian IgA, and clearly is involved in an electrostatic interaction with the polar heads of the phospholipids; by consequence the backbone is bent at this position allowing the last positive residue (Lys/Arg29) to move below the lipid layer encountering different negative phosphate charges.

The pattern of interactions we identified refers to the IgTMD alone and does not consider the influence of the extra-cellular portion of the molecule; we do not exclude the possibility that the structures we described may vary according to the ligand (antigen) binding. This hypothesis, though attractive as it may go towards explaining BCR activation, remains very difficult to verify. We have planned a more extensive structural study, including determination of the Ig $\alpha$ -Ig $\beta$  transmembrane domain structure as well as the IgTMD dimer/Ig $\alpha$ -Ig $\beta$  structural relationships to verify the hypothesis that alternative conformations may occur according to resting (in the absence of ligand) or signaling (in the presence of ligand) receptor status.

We described the IgTMD structure within a POPC environment; this lipid bilayer mimics the lipid bilayer of the cell membrane due to the presence of both saturated and unsaturated acyclic chain. At present we are also able to perform MD simulations in a raft-like lipid environment, which resembles the context in which the BCR is found to act for signal transduction; we have also planned to compare the IgTMD dimer structure using MD in different lipid bilayers resembling raft and non-raft (POPC) cell membrane to verify whether the structure of the dimer depends on the specific lipid

environment. However, we are aware that computational studies need to be confirmed by experiments. For this purpose, we have planned to verify the functional role of the residues of the motif by site-directed mutagenesis in mouse Igs or by introducing the motif in other mouse transmembrane proteins.

Taken together, the results presented here highlight a key motif that has been demonstrated to possess a number of interesting structural implications in the IgTMD dimer assembly and, in turn, in the function of BCR.

#### Acknowledgments

We would like to thank Dr Luca Bargelloni (University of Padua, Italy) for helping us to carry out the phylogenetic analyses.

#### Appendix A. Supplementary data

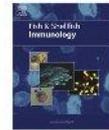
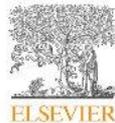
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.09.022.

#### References

- Bagos, P.G., Liakopoulos, T.D., Hamodrakas, S.J., 2006. Algorithms for incorporating prior topological information in HMMs: application to transmembrane proteins. *BMC Bioinf.* 7, 189.
- Berendsen, H.J.C., Postma, J.P.M., van Gasteren, W.F., Di Nola, A., Haak, J.R., 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690.
- Call, M.E., Wucherpfennig, K.W., 2007. Common themes in the assembly and architecture of activating immune receptors. *Nature* 7, 841–850.
- Campbell, K.S., Backstrom, B.T., Tiefenthaler, G., Palmer, E., 1994. CART: a conserved antigen receptor transmembrane motif. *Semin. Immunol.* 6, 393–410.
- Collaborative Computational Project Number 4, 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D* 50, 760–763.
- Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190.
- Cuthbertson, J.M., Bond, P.J., Sansom, M.S., 2006. Transmembrane helix–helix interactions: comparative simulations of the glycoprotein a dimer. *Biochemistry* 45, 14298–14310.
- Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092.
- Dawson, J.P., Weinger, J.S., Engelman, D.M., 2002. Motifs of serine and threonine can drive association of transmembrane helices. *J. Mol. Biol.* 316, 799–805.
- Felsenstein, J., 2004. *Inferring Phylogenies*. Sinauer, Sunderland, MA.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: Walker, John M. (Ed.), *The Proteomics Protocols Handbook*. Humana Press, pp. 571–607.
- Gazit, E., 2002. A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J.* 16, 77–83.
- Gratkowski, H., Lear, J.D., DeGrado, W.F., 2001. Polar side chains drive the association of model transmembrane peptides. *Proc. Natl Acad. Sci. USA* 98, 880–885.
- Guindon, S., Gascuel, O.A., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Harrington, S.E., Ben-Tal, N., 2009. Structural determinants of transmembrane helical proteins. *Structure* 17, 1092–1103.

Please cite this article in press as: Varriale, S., et al. An evolutionarily conserved motif is responsible for immunoglobulin heavy chain packing in the B cell membrane. *Mol. Phylogenet. Evol.* (2010), doi:10.1016/j.ympev.2010.09.022

- Henin, J., Pohorille, A., Chipot, C., 2005. Insights into the recognition and association of transmembrane alpha-helices. The free energy of alpha-helix dimerization in glycophorin A. *J. Am. Chem. Soc.* 127, 8478–8484.
- Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINC: A linear constraint solver for molecular simulations. *J. Comp. Chem.* 18, 1463–1472.
- Hofmann, K., Stoffel, W., 1993. TMbase-A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* 374, 166.
- Jiang, S., Vakser, I.A., 2004. Shorter side chains optimize helix–helix packing. *Protein Sci.* 13, 1426–1429.
- Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.
- Kleiger, G., Grothe, R., Mallick, P., Eisenberg, D., 2002. CXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry* 41, 5990–5997.
- Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. ClustalW and ClustalX version 2. *Bioinformatics* 23 (21), 2947–2948.
- Laskowski, R.A., MacArthur, M.W., Moss, M.D., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291.
- Lawrence, M.C., Colman, P.M., 1993. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* 234, 946–950.
- Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320.
- Lemmon, M.A., Hanagan, J.M., Hunt, J.F., Adair, B.D., Bornmann, B.J., Dempsey, C.E., Engelman, D.M., 1992. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J. Biol. Chem.* 267, 7683–7689.
- MacKenzie, K.R., Prestegard, J.H., Engelman, D.M., 1997. A transmembrane helix dimer: structure and implications. *Science* 276, 131–133.
- McGaughey, G.B., Gagné, M., Rappé, A.K., 1998. Pi-Stacking interactions. Alive and well in proteins. *J. Biol. Chem.* 273, 15458–15463.
- Merlino, A., Mazzarella, L., Di Fiore, A., Caramante, A., Notomista, E., Di Donato, A., Sica, F., 2005. The importance of dynamic effects on the enzyme activity: X-ray structure and molecular dynamics of onconase. *J. Biol. Chem.* 280, 17953–17960.
- Merlino, A., Esposito, L., Viragliano, L., 2006. Polyglutamine repeats and beta-helix structure: molecular dynamics study. *Proteins* 63, 918–927.
- Merlino, A., Varriale, S., Coscia, M.R., Mazzarella, L., Oreste, U., 2008. Structure and dimerization of the teletost transmembrane immunoglobulin region. *J. Mol. Graph. Model.* 27, 401–407.
- Mottamal, M., Zhang, J., Lazaridis, T., 2006. Energetics of the native and non-native states of the glycophorin transmembrane helix dimer. *Proteins* 62, 996–1009.
- Okada, T., Le Trong, I., Fox, B.A., Behnke, C.A., Stenkamp, R.E., Palczewski, K., 2000. X-Ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. *J. Struct. Biol.* 130, 73–80.
- Pancer, Z., Saha, N.R., Kasamatsu, J., Suzuki, T., Amemiya, C.T., Kasahara, M., Cooper, M.D., 2005. Variable lymphocyte receptors in hagfish. *Proc. Natl. Acad. Sci. USA* 102, 9224–9229.
- Patel, A.B., Crocker, E., Reeves, P.J., Getmanova, E.V., Eilers, M., Khorana, H.G., Smith, S.O., 2005. Changes in interhelical hydrogen bonding upon rhodopsin activation. *J. Mol. Biol.* 347, 803–812.
- Prilusky, J., Eitan, B., 2009. Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs. *Proc. Natl. Acad. Sci. USA* 106, 6662–6666.
- Rumfelt, L.L., Diaz, M., Lohr, R.L., Mochon, E., Flajnik, M.F., 2004. Unprecedented multiplicity of Ig transmembrane and secretory mRNA forms in the cartilaginous fish. *J. Immunol.* 173, 1129–1139.
- Schneider, D., Engelman, D.M., 2004. Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *J. Mol. Biol.* 343, 799–804.
- Schneider, T.D., Stephens, R.M., 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100.
- Senes, A., Gerstein, M., Engelman, D.M., 2000. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J. Mol. Biol.* 296, 921–936.
- Stewart, J.J., Lee, C.Y., Ibrahim, S., Watts, P., Shlomchik, M., Weigert, M., Litwin, S., 1997. Entropy analysis of immunoglobulin and T cell receptor. *Mol. Immunol.* 34, 1067–1082.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL\_X windows interface. Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Vaccaro, L., Cross, K.J., Kleinjung, J., Straus, S.K., Thomas, D.J., Wharton, S.A., Skehel, J.J., Fraternali, F., 2005. Plasticity of influenza haemagglutinin fusion peptides and their interaction with lipid bilayers. *Biophys. J.* 88, 25–36.
- Vereshaga, Y.A., Volynsky, P.E., Pustovalova, J.E., Noid, D.E., Arseniev, A.S., Efremov, R.G., 2007. Specificity of helix packing in transmembrane dimer of the cell death factor BNIP3: a molecular modeling study. *Proteins* 69, 309–325.
- Walters, R.F., DeGrado, W.F., 2006. Helix-packing motifs in membrane proteins. *Proc. Natl. Acad. Sci. USA* 103, 13658–13663.
- Zhou, F.X., Cocco, M.J., Russ, W.P., Brunger, A.T., Engelman, D.M., 2000. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat. Struct. Biol.* 7, 154–160.



## Antarctic teleost immunoglobulins: More extreme, more interesting

Maria Rosaria Coscia\*, Sonia Varriale, Stefano Giacomelli, Umberto Oreste

*Institute of Protein Biochemistry, CNR, Via P. Castellino 111, 80131 Naples, Italy*

### ARTICLE INFO

#### Article history:

Received 19 July 2010  
Received in revised form  
12 October 2010  
Accepted 22 October 2010  
Available online xxx

#### Keywords:

Immunoglobulin heavy chain  
Immunoglobulin light chain  
Gene duplication  
Parasite-specific antibody response  
IgM membrane-bound form  
Extracellular membrane proximal domain  
Alternative mRNA splicing  
Antarctic teleost  
VH gene family

### ABSTRACT

We have investigated the immunoglobulin molecule and the genes encoding it in teleosts living in the Antarctic seas at the constant temperature of  $-1.86^{\circ}\text{C}$ . The majority of Antarctic teleosts belong to the suborder Notothenioidei (Perciformes), which includes only a few non-Antarctic species. Twenty-one Antarctic and two non-Antarctic Notothenioidei species were included in our studies. We sequenced immunoglobulin light chains in two species and  $\mu$  heavy chains, partially or totally, in twenty species. In the case of heavy chain, genomic DNA and the cDNA encoding the secreted and the membrane form were analyzed. From one species, *Trematomus bernacchii*, a spleen cDNA library was constructed to evaluate the diversity of VH gene segments. *T. bernacchii* IgM, purified from the serum and bile, was characterized. Homology Modelling and Molecular Dynamics were used to determine the molecular structure of *T. bernacchii* and *Chionodraco hamatus* immunoglobulin domains. This paper sums up the previous results and broadens them with the addition of unpublished data.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

Antarctic fish belong to the perciform Notothenioidei suborder and are the largely endemic, dominant fish taxa in the cold waters surrounding Antarctica. The majority of notothenioids live at seawater temperatures between  $-1.86^{\circ}\text{C}$  and  $0^{\circ}\text{C}$ , but some sub-polar species inhabit waters as warm as  $10^{\circ}\text{C}$  in south of New Zealand and South America.

Over the past 55 million years, the temperature of the Southern Ocean has undergone a progressive reduction from about  $15^{\circ}\text{C}$  to the present-day  $-1.86^{\circ}\text{C}$ . During the boundary between Oligocene and Miocene, about 22–25 million years ago, the Southern Ocean was cut off as a consequence of the opening of the Drake passage, which separates Antarctica from South America, and the formation of the Antarctic Circumpolar Current, the so called Polar Front, a kind of thermal hydrographic barrier [1,2]. In consequence, many temperate fish species became extinct following the destruction of inshore habitat and changes in trophic relationships caused by climatic changes that led to freezing water temperatures [3]. Thus, fish diversity was limited and new niches became available. In

particular, one group of teleost fish, the Notothenioidei, succeeded in adapting to the environmental conditions, namely to low temperatures, and were diversified *in situ*. With little niche competition, Notothenioidei spread widely becoming the predominant Antarctic fish group today [4]. That is why, at present, species of the notothenioid suborder comprise 46% of all the fish species in the Southern Ocean. However in many areas of the continental platform, including the highest latitudes, notothenioids constitute 77% of the species and 90–95% of fish abundance and biomass [3].

Eight notothenioid families, including five that are predominantly Antarctic, encompass a total of 44 genera and 129 species, 101 Antarctic and 28 non-Antarctic. Although species diversity is low compared to other continental shelf habitats, the nature of the adaptive diversification among notothenioids is noteworthy. Adaptive changes in notothenioids required for cold survival have been reported in a number of studies. One key evolutionary adaptation was the *de novo* acquisition of antifreeze glycoproteins [5,6], other examples include adaptive modification of enzyme protein structures [7], cold-efficient microtubule assembly [8], cold-adapted protein translocation [9], and high mitochondrial densities [10]. Evolution under constant cold conditions was also accompanied by striking gene loss, most notably hemoprotein loss in the icefish family Channichthyidae [11] and independent losses of cardiac myoglobin in a subset of icefish species [12].

\* Corresponding author. Tel.: +39 081 6132556; fax: +39 081 6132629.  
E-mail addresses: [mr.coscia@ibp.cnr.it](mailto:mr.coscia@ibp.cnr.it) (M.R. Coscia), [s.varriale@ibp.cnr.it](mailto:s.varriale@ibp.cnr.it) (S. Varriale), [s.giacomelli@ibp.cnr.it](mailto:s.giacomelli@ibp.cnr.it) (S. Giacomelli), [uo.oreste@ibp.cnr.it](mailto:uo.oreste@ibp.cnr.it) (U. Oreste).

111 Chen et al. [13] recently reported a large-scale transcriptomic  
112 up-regulation of 177 gene families in *Dissostichus mawsoni*,  
113 belonging to the Nototheniidae family, compared to their orthologs  
114 in temperate and tropical fish. At the genomic level, 118 gene  
115 families, in several notothenioid species, were found to have  
116 undergone significant gene duplication. These findings suggest that  
117 the genomes of notothenioids are evolutionarily dynamic, thus  
118 contributing to the overall success of the group.

119 Due to the plasticity of its gene locus, Immunoglobulin is an  
120 ideal candidate to contribute to the characterization of the evolu-  
121 tionary modifications undergone by the notothenioid genome.

122 The Antarctic species analyzed in the present work belong to  
123 the families Bovichtiidae (*Bovichtus diacanthus*), Eginopidae  
124 (*Eleginops maclovinus*), Nototheniidae (*D. mawsoni*, *Trematomus*  
125 *bernacchii*, *Trematomus newnesi*, *Trematomus borchgrevinkii*, *Trem-*  
126 *atomus eulepidotus*, *Trematomus hansonii*, *Trematomus lepidorhinus*,  
127 *Trematomus loennbergii*, *Trematomus nicolai*, *Trematomus scotti* and  
128 *Gobionotothen gibberifrons*), Bathydraconidae (*Cygnodraco mawsoni*  
129 and *Gymnodraco acuticeps*), Artedraconidae (*Histiadraco velifer*  
130 and *Pogonophryne scotti*), Channichthyidae (*Chaenocephalus acer-*  
131 *atus*, *Champscephalus esox*, *Chionodraco hamatus*, *Chionodraco*  
132 *rastrorhinus*, *Pseudochaenichthys georgianus* and *Pagetopsis mac-*  
133 *ropterus*), *Notothenia coriiceps* (Nototheniidae) secretory  
134 and membrane-bound Igu and *C. aceratus* (Channichthyidae) secretory  
135 Igu sequences, previously determined by Ota et al. [14] are also  
136 included in the analysis.

## 137 2. Secreted IgM

138 Although teleost IgM share some structural and functional  
139 similarities with mammalian IgM, an interesting difference lies in  
140 the fact that the IgM of a number of teleost fish associate into larger  
141 polymers in order to produce heterogeneous mixtures by simply  
142 varying the number of disulfide bonds linking monomers or half-  
143 mer subunits. This heterogeneity in the basic structure is referred to  
144 as redox forms and has some significant implications for the anti-  
145 body affinity maturation [15].

146 A biochemical analysis of *T. bernacchii* antibody heterogeneity  
147 was performed. For this purpose, IgM was purified by sequential  
148 steps of Ammonium Sulphate precipitation, thiophilic adsorption  
149 chromatography and size-exclusion FPLC on a column pre-  
150 calibrated with marker proteins [16]. The mean concentration value  
151 of *T. bernacchii* IgM determined in the serum by ELISA was 2.7 mg/  
152 ml corresponding to 9.6% of the total serum proteins [17] which  
153 was comparable to the values reported for other fish species  
154 [18–20]. Purified IgM was analyzed by SDS-Polyacrylamide gel  
155 electrophoresis, isoelectrophoresis and 2-D electrophoresis.  
156 Relative molecular mass of the polymeric form was 830 kDa; that of  
157 separated IgH and IgL chains was respectively 78 and 25 kDa. The  
158 relative molecular mass of the IgH chain was found to be slightly  
159 higher than that of non-polar species, which is about 70 kDa [21].  
160 The disparity for the IgH chain relative molecular mass observed  
161 between polar and non-polar fish might be attributed to different  
162 carbohydrate content. The carbohydrate content of the entire  
163 *T. bernacchii* IgH chain was in fact estimated to be 12.8%. The  
164 isoelectric point of the whole molecule ranged from 4.4 to 6.5, that  
165 of separated IgH chains was between 4.0 and 6.0. Differences  
166 observed between the theoretical and experimental pI values can  
167 be attributed to the contribution of the carbohydrate moieties not  
168 considered by the computational tools used. Moreover, pI values for  
169 the IgH chain falling into a similar range were revealed by 2-D  
170 electrophoresis [17].

171 Purified *T. bernacchii* IgM was digested with trypsin and the  
172 cleavage site of trypsin was found at the end of the CH1 domain, as  
173 confirmed by the N-terminal amino acid sequence of one of the

174 resultant peptides and was a conserved trypsin target site shared  
175 by the majority of different teleost species heavy chains [22]. The  
176 protective effect of the carbohydrate moiety from enzyme attacks  
177 could explain the presence of residual undigested H chain after 20 h  
178 of hydrolysis at high enzyme/substrate ratio.

179 The presence of carbohydrates was analyzed in the H and L  
180 chains labeled with digoxigenin. Glycosyl residues were detected  
181 only in the H chain. Purified Igs were hydrolyzed by N-glycosidase F  
182 at different conditions and at least four different hydrolytic sites  
183 were revealed by limited deglycosylation, although it is not possible  
184 to determine which asparagine residue is involved in binding the  
185 glycan chains. Based on previously collected data, it was high-  
186 lighted that the *T. bernacchii* H chain amino acid sequence con-  
187 tained seven putative N-glycosylation sites, which is the highest  
188 number among all the vertebrate Ig investigated to date  
189 (Supplementary Fig. 1). This degree of glycosylation may account  
190 for higher solubility of the Ig molecule.

191 *T. bernacchii* IgM was purified and quantified also from the bile;  
192 its occurrence in the liver tissue as well as in the bile canaliculi  
193 and pre-ductules strongly indicated a hepato-biliary transport of this  
194 molecule to protect the intestinal epithelium [23].

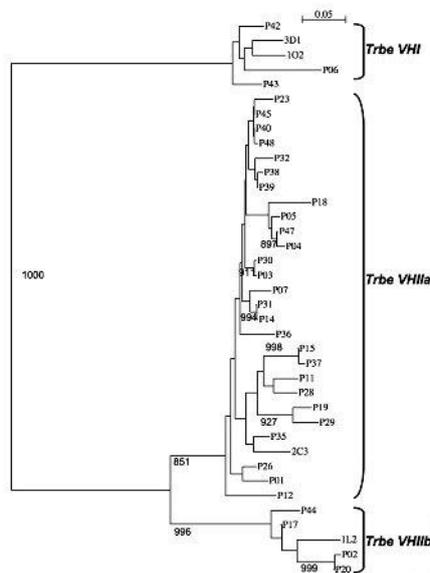
195 By immunohistochemical analysis B cells were shown mainly to  
196 seed the head kidney, thymus and spleen, while fewer were present  
197 in the gut-associated lymphoid tissue. Concentration of Ig positive  
198 cells in the thymus is a striking finding, clearly exceeding the  
199 concentration found in all fish species studied so far. Differential  
200 distribution of membrane- and cytoplasmic-Ig positive cells in the  
201 different vascular districts is particularly relevant (Abelli, in  
202 preparation).

## 203 3. Diversity of the heavy chain variable region

204 Besides the biochemical analysis of the antibody molecule, the  
205 cDNA and genomic sequences encoding the IgM heavy chain in  
206 *T. bernacchii* were characterized. A genomic library was constructed  
207 from *T. bernacchii* blood cells in the Lambda DASH II vector and  
208 screened using a cDNA clone encoding the *T. bernacchii* VH domain  
209 as probe. One of four hybridizing recombinant clones was digested  
210 with *NotI* and *HindIII* and subcloned. The clone NH6, containing VH  
211 gene segments was partially sequenced. By nucleotide sequence  
212 analysis three VH gene segments with flanking recombination  
213 signal sequences (RSS) were identified. One VH segment was clearly  
214 a pseudogene as a stop signal occurred in the coding region. The  
215 RSS were similar to those found in other teleost species [24].

216 To define the basis for the generation of antibody diversity in  
217 *T. bernacchii*, a spleen library was constructed by 5'RACE and forty-  
218 five cDNA clones, containing complete or partial sequences of  
219 rearranged VH(D)/JH segments, were analyzed [25]. All the VH  
220 sequences analyzed shared, on average, 80% of nucleotide identity.  
221 Based on the percentage of nucleotide identity, VH sequences were  
222 grouped into two families: most sequences (89%), fell into a family  
223 referred to as *Trbe VH II*; the remaining sequences belonged to the  
224 *Trbe VH I* family. In particular, the *Trbe VH II* family could be divided  
225 into two subfamilies: *Trbe VH IIa*, comprising 30 different VH gene  
226 segments sharing 81–99% nucleotide identity, and *Trbe VH IIb*  
227 consisting of only 5 gene segments with a nucleotide identity  
228 between 82 and 99% [25]. A Neighbor-Joining (NJ) tree showing the  
229 relative distances between the VH families was constructed on the  
230 basis of multiple alignment of the *T. bernacchii* cDNA VH sequences  
231 generated by the ClustalX program (Fig. 1). In contrast to other  
232 species possessing VH gene segments distributed in many families,  
233 up to 13 [27], only two VH gene families were identified in *T. ber-*  
234 *nacchii*, and if antibody diversity depends on the number of distinct  
235 families, this can lead to the conclusion that the expressed antibody  
236 repertoire in this Antarctic species is limited.

241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305

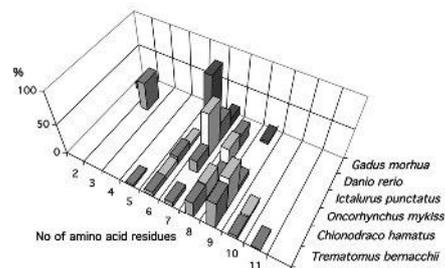


**Fig. 1.** Neighbor-joining tree of the 38 unique and complete *T. bemaichii* VH cDNA clones. The reconstruction was performed on the nucleotide sequences aligned by ClustalX. The observed distances were corrected for multiple substitutions using Motoo Kimura's formula [26]. 1000 bootstrap replicates were made. The reliable branch nodes, supported by bootstrap values higher than 80% are shown.

In most VH sequences the AUGAUG sequence was found as translation starting motif. A statistical analysis on a large set of vertebrate transcripts revealed that this motif was preferred compared to AUGXAUG or AUGXXAUG indicating that the double AUG contributes to translation more efficiently.

At present, four distinct putative D segments, with lengths ranging from 4 to 19 nucleotides, have been defined; D1, the most utilized segment, exhibited all reading frames encoding two to five glycines. These residues are important for the flexibility of CDR3. Eight different putative JH gene segments, showing a percentage of nucleotide identity ranging from 67 to 98%, were identified. The proportions of each JH segment were very different, however, JH1 (24.4%), JH2 (22.2%), and JH6 (26.6%) were the three major segments found to be equally associated to different VH segments of both families [25].

Although antibody diversity in *T. bemaichii* may be limited by gene family number, significant differences found in the CDRs suggested that the high length heterogeneity of CDR2 may contribute to a wider diversification of the antibody repertoire. The Shannon analysis revealed that while FRs were highly conserved, CDR3 showed the highest entropy values, and CDR2 contributed to variability in a significant manner showing a mean entropy value higher than that of CDR1. CDR2 size, determined in accordance with the IMGT demarcations [28], varied from 5 to 11 residues in *T. bemaichii*, and from 6 to 10 in *C. hamatus*, with mean lengths of 8.4 and 8.6 residues, respectively, resulting in the longest CDR2 regions if compared to other teleost species, as shown in Fig. 2. Large numbers of insertions and deletions, especially of triplets to



**Fig. 2.** CDR2 length analysis. The percentage of differently long CDR2s from *T. bemaichii*, *C. hamatus* and other Acanthopterygian species VH sequences are plotted versus their lengths. *I. punctatus* CDR2 sequences are from the IMGT database. *T. bemaichii* CDR2 sequences are from [25]; *D. rerio* CDR2 sequences are from [29]; *O. mykiss* CDR2 sequences are from [30]; *G. morhua* CDR2 sequences are from [31]. *C. hamatus* CDR2 sequences have been deposited under the GenBank accession numbers HQ384351–HQ384364.

maintain reading frame, found in the Antarctic species analyzed, are usually very rarely observed in other teleosts. Tandem repeats (direct or inverted) up to 8 bp and palindromic sequences were found essentially in CDR1 and CDR2. CDR3 length varied between 7 and 13 residues with a mean length of 11.6 residues, resulting in smaller size heterogeneity compared to other species. The presence of putative N- and P-region nucleotides in the VH/D and DJH junctions suggests that the junctional mechanisms, contributing to the CDR3 diversity, also occur in *T. bemaichii*.

The occurrence of the RGYW motif (Purine-Guanine-Pyrimidine-Adenine or Thymine), identified as the preferential site of mutations [32], was examined [33]. The RGYW motif was 3.5% out of a total of 11,141 quartets analyzed, and higher than that calculated for the entire constant region (1.8%). When CDRs and FRs were analyzed separately, the frequency was 0.4% in FR1 and 11.4% in CDR2.

To compare the occurrence of the RGYW motif in *T. bemaichii* and *Oncorhynchus mykiss* respectively, a set of VH gene segments, representing the *O. mykiss* gene complexity, for a total amount of 7800 bp, was selected from the databanks. The motif frequencies were calculated for each FR or CDR region. The highest frequency of RGYW motif was found in the CDR2 region of *T. bemaichii* (11.4%) and in the CDR1 region of *O. mykiss* (8.7%), suggesting that the mutational machinery targets different regions in a biased manner in these species [33].

Serine codons were found preferentially within this hypermutation hot spot motif. The serine codons AGY (where Y is C/T) and TCN (where N is A/T/C/G) in *T. bemaichii* VH gene segments were 45.6% and 54.4%, respectively. Notably, the usage of AGY codon in *T. bemaichii* is higher in CDRs and lower in FRs compared to *Salmo salar* [33].

It was also found evidence for Somatic Hypermutation comparing clones carrying the same V/D/J recombination, but differing in few bases in CDR1 or CDR2 [33].

The presence of "hybrid" IgH clones, which express different regions of members of different VH families in the same sequence, was observed, indicating the occurrence of the receptor revision mechanism (Supplementary Fig. 2). These data confirm the recently reported results for other teleost fish [34].

#### 4. Diversity of the heavy chain constant region

A great deal of our research has focused on VH/D/JH somatic recombination and the resulting VH repertoire, part of the work has

306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370

371 been dedicated to examining the constant domain exons which are  
372 known to be relatively invariant. In some fish species such as  
373 salmonids [35], ictalurids [36], and cyprinids [37], the molecular  
374 structure of the IgH gene locus has suggested the possibility that  
375 duplication events could have shaped their present organization. To  
376 date, the complexity of the IgH genes architecture has been  
377 investigated mainly by molecular biology approaches.

378 Duplications of genes are frequent in notothenioids: to investi-  
379 gate IgH gene locus multiplicity in notothenioids, the chromosomal  
380 location of the IgH locus has been analyzed in several species by  
381 Fluorescence In Situ Hybridization (FISH) [38]. Two IgH probes  
382 (specific for VH and CH regions, respectively) were prepared from  
383 the species *T. bernacchii* and mapped onto the chromosomes of ten  
384 species belonging to the same genus *Trematomus* and in the out-  
385 group species *N. coriiceps*, through one-color and two-color FISH.  
386 Since the IgH locus was found to be on one chromosome pair in the  
387 majority of the species examined, including the outgroup *N. cori-  
388 iceps*, this "single locus" pattern is more likely to be the more  
389 ancient condition for the IgH gene locus position. In four species,  
390 *T. scotti*, *T. newnesi*, *Trematomus pennellii*, and *T. lepidorhinus*, an  
391 additional IgH locus was detected. For a more detailed  
392 analysis of IgH locus in *T. newnesi*, Southern blot and RT-PCR  
393 experiments have been carried out. Both analyses confirmed the  
394 results obtained by FISH. In particular, RT-PCR experiments  
395 revealed the presence of three allelic variants in *T. newnesi*. When  
396 considering the structural aspects, it is worth noting that, regard-  
397 less of the single or double chromosomal location of IgH locus, the  
398 hybridization signals were equally strong. This suggests a similar  
399 gene segment multiplicity.

400 Furthermore, allelic polymorphism in the IgH constant domain  
401 coding sequences from 10 *T. bernacchii* individuals was investigated  
402 [24]: in the overall 1495-nt long sequence, 36 positions showed  
403 polymorphisms, 20 of which, accounting for 56% of the total, were  
404 concentrated in a 63-nt region located between the CH2 and CH3  
405 exons, formerly the hinge region (Fig. 3). Most substitutions were  
406 non-synonymous (dn/ds = 2.5) suggesting that selection acts on  
407 the polymorphic sites in order to favour amino acid replacements

436 and maintain high allelic polymorphism. A polymorphic position  
437 A/C changing Lysine to Asparagine was observed at the end of the  
438 CH3 domain and, interestingly, at the same position in the two  
439 *Cyprinus carpio* IgH sequences, A and B (GenBank accession  
440 numbers AB004105 and AB004108), indicating that this allotype  
441 marker is very old, in relation to the teleost evolution. Poly-  
442 morphism was also observed at some potential N-glycosylation  
443 sites. By sequencing genomic DNA some polymorphic positions in  
444 the splicing sites were found and that, consequently, gave rise to  
445 phenotypes showing different amino acid residue numbers.

446 Allelic polymorphism was also investigated in *C. hamatus* and  
447 found two alleles in 10 individuals differing in several positions  
448 within the CH2–CH3 hinge (unpublished data). The two alleles  
449 were differently occurring: in fact only a heterozygous individual  
450 was carrying the rare allele; the rarity of this allele could be partly  
451 explained by the absence of a cysteine residue in the hinge that may  
452 presumably affect the structural stability.

### 5. Comparative analysis of the secreted IgH constant domains from Antarctic and non-Antarctic species

454 At present, cDNA sequences encoding the constant domains of  
455 the secreted immunoglobulin heavy chain have been completely  
456 determined from 34 different teleost species. Our group contrib-  
457 uted by determining the IgH complete sequences of 15 notothenioid  
458 species. Species from which the IgH sequence has been  
459 determined are listed in Supplementary Table 1; the species belong  
460 to 9 orders. The most represented order is the Perciformes  
461 including 21 species belonging to 4 suborders (15 Notothenioids,  
462 4 Percoids, and one from Channoids and Zoarcoids, respectively).  
463 Thirteen out of 15 notothenioid species are Antarctic, *B. diacanthus*  
464 and *E. maclovinus* being the only two non-Antarctic.

465 Sequences coding for each constant domain have been aligned  
466 by ClustalX (Supplementary Fig. 1). The fourth domain has been  
467 found to be the most conserved whereas the second and the third  
468 have been found to be the most variable. All the Antarctic species  
469 showed the same residue in 8/591 positions, whereas the other

```

408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500

```

Fig. 3. Complete cDNA sequence encoding the IgH constant region from 10 *T. bernacchii* individuals. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. An asterisk below the sequence indicates that the amino acid residue is missing in some individuals. Polymorphic positions are reported as R corresponding to A or G; Y to C or T; K to G or T; M to A or C; S to C or G; W to A or T; V to not T. Numbers indicate every possible amino acid found at these positions: 1 = N/H/V; 2 = D/E/K/T; 3 = N/K/D/T/E; 4 = G/E/Q; 5 = E/K/I/V; 6 = S/F/L/P; 7 = S/C/I; 8 = A/E/G. The most polymorphic region is shaded.

Please cite this article in press as: Coscia MR, et al., Antarctic teleost immunoglobulins: More extreme, more interesting, Fish & Shellfish Immunology (2010), doi:10.1016/j.fsi.2010.10.018

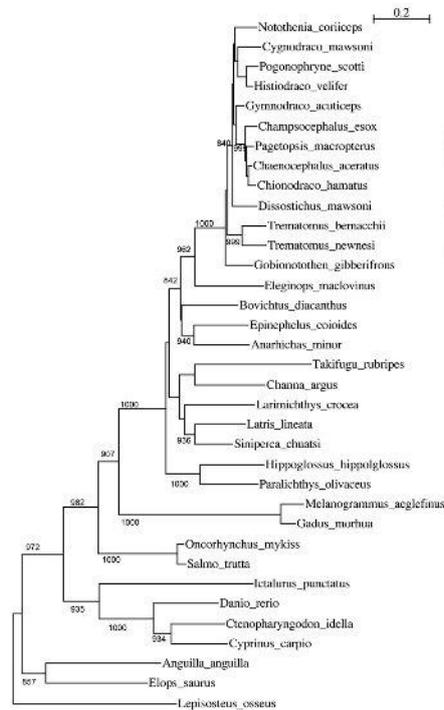
501 species, including the non-Antarctic notothenioids, showed  
502 different residues in the same positions. These positions were  
503 located both in loops and  $\beta$ -sheets of CH1 (1 position), CH3 (3  
504 positions) and CH4 (4 positions).

505 The overall distances among the I $\mu$  constant domain sequences  
506 are represented by the NJ tree shown in the Fig. 4.

507 The occurrence of N-glycosylation sites in the I $\mu$ H constant  
508 domains was analyzed by distinguishing the overall putative sites  
509 from those with high prediction score, determined using the  
510 NetNGly tool with a threshold of 65% (Supplementary Table 2).  
511 Comparison of the domain distribution in Antarctic and non-  
512 Antarctic species revealed that Antarctic notothenioids present  
513 heavily glycosylated CH2. In particular, four putative sites were in  
514 the *T. bema* CH2 domain.

515 The cysteine residues, presumed responsible for the H–H  
516 disulfide bridges, were located specifically in the CH2–CH3  
517 boundary of the Antarctic species sequences, whereas different  
518 localizations have been frequently found in other species.

519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565



560 **Fig. 4.** Distance tree of the I $\mu$  constant domain sequences from several teleost species.  
561 A Neighbor-joining tree was reconstructed by the alignment generated by ClustalX,  
562 using the BLOSUM weight matrix. Positions with gaps were excluded from the  
563 alignment. The observed distances were corrected for multiple substitutions using  
564 Motoo Kimura's formula [26]. 1000 bootstrap replicates were made. The reliable  
565 branch nodes, supported by bootstrap values higher than 80% are shown. The holo-  
566 stean *Lepisosteus osseus* sequence has been used as outgroup.

566 The most relevant differences between the Antarctic and non-  
567 Antarctic I $\mu$ H sequences rather than in overall amino acid compo-  
568 sition or post-translation modifications, reside in the length of the  
569 regions connecting domains, which ensure higher flexibility to the  
570 cold-adapted molecules. In fact one insertion of two residues was  
571 found at the VH-CH1 boundary affecting the so called "elbow  
572 motion" of the antigen combining region, that is the movement of  
573 the VL-VH dimer with respect to the CL-CH1 dimer, and a second  
574 insertion, localized at the CH2–CH3 boundary [39]. The latter is  
575 a proline and cysteine rich region and is hypothesized to act as  
576 a "hinge", that is a spacer peptide separating CH2 from CH3 (Fig. 5).  
577 This segment is considered to be the main mediator of flexibility of  
578 the immunoglobulin molecule; it is presumptively highly exposed  
579 to the solvent, and, thus, the presence of carbohydrate chains in this  
580 region may play a protective role against proteolytic attacks. The  
581 most distinct differences between the Antarctic and non-Antarctic  
582 fish species analyzed are the amino acid composition and length of  
583 the hinge region, which ranges between 14 and 22 residues (mean  
584 18.7) in Antarctic species versus 5–14 residues (mean 10.1) in non-  
585 Antarctic species.

## 6. Comparative analysis of the membrane-bound I $\mu$ constant domains from Antarctic and non-Antarctic species

586 One more interesting insertion regarded the membrane-bound  
587 form which was found to be due to an important gene modification  
588 event that appeared during the radiation of Nototheniidae, initiated  
589 about 24 million years ago [40]. It consisted in a large insertion in  
590 the I $\mu$  gene locus in parallel with a significant modification of the  
591 mRNA splicing mechanism resulting in a membrane antigen  
592 receptor of only two constant I $\mu$  domains [41]. This event occurred  
593 during the major cooling period, and, thus, may have an adaptive  
594 key significance.

595 Membrane I $\mu$  transcripts were sequenced in 13 notothenioid  
596 species, the majority of them showed atypical membrane I $\mu$   
597 coding sequences, since lacking the entire CH3 and CH4 constant  
598 exons, and possessing an unusually long Extracellular Membrane  
599 Proximal Domain (EMPD), with an average length of 49.2-aa resi-  
600 dues (mean length 20.3 aa in other teleosts) (Supplementary  
601 Table 3). The elongation of the EMPD was generated by the inser-  
602 tion of 13-aa tandem repeats, that were encoded by 39-nt addi-  
603 tional exons, present in various numbers (up to 4) according to the  
604 species analyzed. Analysis of the EMPD amino acid composition  
605 showed quite abundant negatively charged residues (aspartic and  
606 glutamic acid, 23.7%), and cysteine and proline residues being 6.2  
607 and 4.2% of the total, respectively [41].

608 The mRNA splicing generating the membrane form occurred in  
609 non-Antarctic notothenioids in the same way as generally observed  
610 in other teleost fish. However, in the majority of the species  
611 analyzed in the present work, it followed the atypical pattern. We  
612 were thus prompted to elucidate the different splicing mechanisms  
613 observed. Genomic DNA analyses were performed in two Antarctic  
614 species, *C. hamatus* and *C. mawsoni*, and in one non-Antarctic  
615 notothenioid, *B. diacanthus* [41]. While the non-Antarctic species  
616 sequence did not show any peculiar feature, those from the  
617 Antarctic species revealed that each 39-nt additional exon fell  
618 within a 411-nt long sequence that was the reverse complement  
619 sequence of an upstream region beginning in the CH2–CH3 intron,  
620 and ending in the region located between the CH2–CH3 intron and  
621 the 3'UTR. The *C. hamatus* 39-nt sequences, referred to as RB and  
622 RA, showed 8 and 12 substitutions, respectively, with respect to the  
623 antiparallel sequence; a third 39-nt sequence, named RC sequence,  
624 which was located upstream the RB sequence but was not included  
625 in the mature transcript, showed only one substitution compared to  
626 its reverse complement counterpart, indicating that coding regions  
627  
628  
629  
630

6

MR. Coscia et al. / Fish &amp; Shellfish Immunology xxx (2010) 1–9

	β-strand	CH2	hinge	CH3	
631		GGGGGGGGG		AAAAAAAAAAAAA	BBBBBBBB
632	<i>Trematomus newnesi</i>	CLFEGKGGKSTASRLNITKCDRQPTFECKVTIGCIADADVHIIIEPSMEDMLLHHEGTVTC	22		
633	<i>Trematomus bernacchii</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLQDQGTVC	20		
634	<i>Trematomus loannbergii</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLQDQGTVC	20		
635	<i>Trematomus pennelli</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLQDQGTVC	22		
636	<i>Histiadraco valifer</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	14		
637	<i>Pogonophryne scotti</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	17		
638	<i>Cynodraco mawsoni</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	17		
639	<i>Chionodraco hamatus</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	15		
640	<i>Chaenoccephalus aceratus</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	16		
641	<i>Champsoccephalus exos</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	16		
642	<i>Pagetopsis macropterus</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	16		
643	<i>Gymnodraco acuticeps</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	16		
644	<i>Notothenia coriiceps</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	16		
645	<i>Dissostichus mawsoni</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	14		
646	<i>Gobionotothen gibberifrons</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	17		
647	<i>Eleginops maclovinus</i>	COFEGKRRGIVNSTHGRQPIINPTER-----DTCCEADADVHIIIEPSMEDMLLHHEGTVLC	12		
648	<i>Borichthys discanthus</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	14		
649	<i>Anarhichas minor</i>	COFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	14		
650	<i>Epinephelus coioides</i>	CEFTKGGKGGPTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	12		
651	<i>Siniperca chuatsi</i>	CEFTKGGKGGPTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	10		
652	<i>Larimichthys crocea</i>	CLFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	11		
653	<i>Channa argus</i>	COFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	11		
654	<i>Paralichthys olivaceus</i>	COFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	12		
655	<i>Hippoglossus hippoglossus</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	14		
656	<i>Takiyapu rubripes</i>	COFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	11		
657	<i>Gadus morhua</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	6		
658	<i>Melanogrammus aeglefinus</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	7		
659	<i>Salmo trutta</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	8		
660	<i>Oncorhynchus mykiss</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	8		
661	<i>Anguilla anguilla</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	9		
662	<i>Cyprinus carpio</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	7		
663	<i>Carassius auratus</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	8		
664	<i>Ctenopharyngodon idella</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	10		
665	<i>Danio rerio</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	8		
666	<i>Ictalurus punctatus</i>	CFEFGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	7		
667	<i>Elops saurus</i>	COFEGKGGKSTVNRSLNITKCDKREGGL--IGSTDCPIADADVHIIIEPSMEDMLLHHEGTVLC	5		

Fig. 5. The hinge region of Antarctic and non-Antarctic teleosts. The region comprised between the last conserved cysteine residue of the CH2 domain and the first one of the CH3 domain was aligned from the IgH amino acid sequences of 17 notothenioid and 20 non-notothenioid species. The names of the Antarctic notothenioid species are in bold; those of non-Antarctic notothenioids are underlined. The presumptive location of the hinge region separating the two constant domains and the β-strands are indicated at the top; the cysteine residues available for the disulfide bonds in the IgH dimer are in bold; the proline residues in the hinge are shaded in grey. The hinge length is indicated on the right.

underwent strong selective pressure. Although the canonical splice sequences were present in all cases, RB and RA were found in the mature transcript, whereas RC was absent. In order to verify if there were any structural constraints that caused the missing of 39-nt RC, folding of the primary transcript region encompassing the sequenced region between CH2 and TM exons was predicted by using computational tools. The results showed the formation of a long duplex, due to the presence of complementary reverse sequences, stabilized by free energy of  $-3835.10$  kcal/mol at the Antarctic physiological temperature of  $0^{\circ}\text{C}$ . Since the CH3 exon 5' splice site and the RC 39-nt region 3' splice site are included in this duplex, they are not available to bind the spliceosomal complex [41]. Nineteen out of 20 substitutions in RB and RA were found to be non-synonymous, and 5 substitutions introduced a cysteine codon. The presence of cysteine encoding codons in the 39-nt additional exons, but not in the respective sequence counterpart, suggested that these residues may play an important role in the folding of the EMPD. The consensus sequence DKLGALCPCHIRD of the 13-aa repeats of all the Antarctic species sequenced showed a clearly symmetrical hydrophobic character. Therefore, the assembly of several repeats gives rise to a region consisting of alternate hydrophilic and hydrophobic parts [41].

In an attempt to understand the structural significance of the presence and character of the 13-aa repeats, first a molecular model of a *C. hamatus* region including CH2 and the two 13-aa repeats of the EMPD was built. Then, a model of the homodimer of the two heavy chains was obtained by imposing two interchain disulfide bridges between the two Cys5 and the two Cys13 residues, being the only cysteine residues available to covalently link the two heavy chains of the *C. hamatus* membrane-bound Ig. The resulting homodimer model showed an atypical domain, unique to the Ig structure, with hydrophilic regions facing outwards and a compact hydrophobic

core [41]. Molecular Dynamics experiments on this model showed that the equilibrium was reached with a  $60^{\circ}$  rotation of the constant domains with respect to the EMPD region (Fig. 6). This result indicates a higher flexibility of the antigen recognizing portion of the molecule than that of the portion bound to the cell membrane.



Fig. 6. Homology Modelling of the *C. hamatus* dimer encompassing the CH2 domain and the EMPD domain. The two constant domains are in dark grey; the two EMPD domains are in light grey. The axis of the constant domains dimer forms an angle of about  $60^{\circ}$  with the axis of the EMPD dimer.

Please cite this article in press as: Coscia MR, et al., Antarctic teleost immunoglobulins: More extreme, more interesting, Fish & Shellfish Immunology (2010), doi:10.1016/j.fsi.2010.10.018

## 7. The region crossing the cell membrane

Another crucial region of the Immunoglobulin molecule is the heavy chain domain crossing the B cell membrane (IgTM) within the transducing complex comprising the peptides Ig $\alpha$ –Ig $\beta$ . The communication between the receptor partners depends primarily on the interactions between their membrane-embedded segments. We have investigated the association of the two TM regions of *C. hamatus* [42], and *T. bernacchii* species. Model building was performed using the structure of the helix H of the photosynthetic reaction center of *Rhodospirillum rubrum* as a template. Models, consisting of extended, nearly ideal,  $\alpha$ -helices, were submitted to Molecular Dynamics simulations in water and in the palmitoyl oleoyl phosphatidylcholine (POPC) bilayer. Furthermore, in order to study the spontaneous formation of a parallel IgTM homodimer, different initial configurations were built and submitted to MD simulations both in water and in POPC environments. The obtained structures of the IgTM homodimers display similar packing interfaces, characterized by a high degree of surface complementarity through two pairs of Phe–Phe hydrophobic interactions and two pairs of Ser/Thr–Ser/Thr hydrogen bonds. This interaction pattern, which stabilizes the dimer conformation in the lipid bilayer, was unique, being different from any other identified in transmembrane helices to date. Fig. 7 depicts a view of a representative model obtained by a simulation of *C. hamatus* IgTM dimer in POPC.

## 8. Immunoglobulin light chain isotypes

Different teleost species express different numbers of immunoglobulin light chains (IgL) isotypes, usually more than one. The presence of more than two IgL isotypes in teleosts was initially

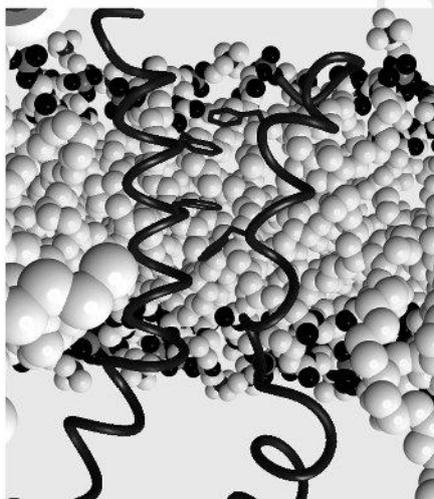


Fig. 7. Molecular model of *C. hamatus* IgH regions crossing the cell membrane. The structure was obtained by MD simulations after insertion of the IgTM dimer model into a POPC lipid bilayer. The protein helices and the phosphorus atoms are in dark grey, the lipid molecules in light grey. The Phenylalanine residues participating in the pairing of the two chains are shown.

suggested by a serological approach investigating the presence of structurally and antigenically different light chains in serum of *O. mykiss* [43]. This hypothesis was later supported by the isolation of cDNAs representing a third IgL isotype in *Dissostichus rerio* [44].

Our data provide clear evidence for the presence of three distinct isotypes in *T. bernacchii*: *TrbeL1*, *TrbeL2*, and *TrbeL3* were identified by immunoscreening a cDNA expression library, and using RT-PCR, and 5'RACE [45]. The constant domains shared high amino acid identity (91.7–95.2%) within the same isotype, but quite low identity between different isotypes (33.4–37.2%). *TrbeL1* constant domain sequences can be divided into two subgroups, *TrbeL1A* and *TrbeL1B*, depending on their relative nucleotide identity. The tissue distribution analysis performed by Real-time PCR experiments showed that the overall expression levels of the three isotypes were found to be very low in muscle, and very high in head kidney. Interestingly, the *TrbeL1A* transcripts were abundantly expressed in all the tissues analyzed, indicating that this isotype could be expressed both in primary and secondary lymphoid tissues. Indeed, the *TrbeL3* transcripts showed the lowest level of expression except in the intestine where it was significantly high, suggesting that this isotype might be specifically involved in mucosal immunity [45].

The organization of the IgL locus in teleosts has been reported to be of multicluster type, in which each gene cluster contains *VL*, *JL*, and *CL* segments [46]. As in *T. bernacchii* isotype-specific *JL* gene segments were found to rearrange with *CL* gene segments specific for the same isotype, it may be suggested that the gene segments of all three isotypes are organized in multiple clusters in *T. bernacchii* as well [45].

The higher diversity of the nucleotide sequences belonging to the *TrbeL1A* isotype suggests the occurrence of a higher number of clusters. The predominance of this isotype in other Acanthopterygian species also suggests an evolutionary success due to specific structural features. The *T. bernacchii* L1A is characterized by a serine stretch, starting at position 45.1 (IMGT numbering) [28] which might be encoded by insertion of a micro-satellite DNA, consisting in tandem repeats of the AGC triplet. It is to be noted that this codon is the most frequently occurring (40%) in the short sequence repeats (SSR) of vertebrate gene exons. Moreover, the molecular models we built indicated that, as a consequence of this insertion, the next loop, connecting the strands D and E, were lengthened. The micro-satellite insertional event might have occurred before Acanthopterygian radiation: in fact this tandem repeats sequence is shared by different Acanthopterygian orders (Atheriniformes, Gasterosteiformes and Pleuronectiformes) and absent in the Paracanthopterygian (Gadiformes) as well as Protacanthopterygian (Salmoniformes) teleost orders. The highest serine content, despite the insertion of many serine codons in the *TrbeL1A* isotype, was found in the *TrbeL3* constant domain, up to 19.2% on average. This represents the highest value found so far in any other species [45]. As predicted by the NetPhos 2.0 tool considering the high stringency score >0.99, several serine residues of both *TrbeL1* subisotypes and *TrbeL3* show high propensity to act as preferential sites for kinase phosphorylation. No threonine or tyrosine residues were predicted as kinase targets. In the molecular models we built, all the identified phosphorylation sites are exposed to the solvent, which may be important in increasing the solubility of the antibody molecules at the low temperatures which the *T. bernacchii* species has adapted to.

The length of the amino acid stretch connecting VL and CL domains was found to be different among isotypes, the *TrbeL3* and *TrbeL2* two or one residues longer than in *TrbeL1A* and *TrbeL1B*, respectively. This region has been demonstrated to be crucial in the arrangement of antibody domains upon antigen binding [47].

cDNA sequences encoding IgL chain (GenBank accession numbers: DQ842615.1–DQ842619.1) were determined from another

891 Antarctic species, *G. acuticeps*, belonging to the same suborder as  
892 *T. bemaichii*, and were found to be highly homologous to those of  
893 *T. bemaichii*, more than any other species (unpublished data).

### 9. Antibody response to nematode parasites

896 Parasitism is a very frequent event in the marine ecosystem and  
897 nematodes (Nematoda) belong to the most common parasites in  
898 demersal as well as pelagic fish throughout the world.

899 The most common nematode larvae found in Antarctic regions  
900 belong to the *Anisakis*, *Contracaecum*, and *Pseudoterranova* genus.  
901 The most infected fish family is the Channichthyidae.

902 To investigate the presence of antibodies to *Contracaecum*  
903 *osculatum* antigens in the plasma of several Antarctic fish, the  
904 immunoglobulin fraction was first purified from a plasma pool of  
905 *C. hamatus* specimens [16]. Immunoglobulins from *C. hamatus*  
906 plasma recognized *C. osculatum* somatic antigens either blotted  
907 onto nitrocellulose or immobilized on a HiTrap NHS-activated  
908 column. The relative molecular mass of the major *C. osculatum*  
909 antigen was 91 kDa. The presence of nematode-specific antibodies  
910 in Antarctic fish plasma was demonstrated by double immunodif-  
911 fusion using individual plasma from specimens of *C. hamatus*,  
912 *N. coriiceps*, *Notothenia gibberifrons*, *C. aceratus*, *P. georgianus*,  
913 *T. bemaichii* against *C. osculatum* proteins [16].

914 Moreover, the occurrence of antibodies against protein antigens  
915 of the other most common nematode parasite *Pseudoterranova*  
916 *decipiens* in the plasma and bile of *T. bemaichii* was investigated  
917 [48]. Three different *P. decipiens* protein solutions were prepared:  
918 excreted/secreted proteins from live larvae (ESP); surface-associ-  
919 ated proteins obtained by mild extraction of larval bodies (SAP);  
920 and cuticular soluble proteins recovered by extraction under strong  
921 reducing conditions (CSP). Using different immunoassays, these  
922 three preparations were tested for their ability to bind fish anti-  
923 body. As determined by ELISA, the specific antibody binding  
924 activity was higher in SAP than in CSP. As determined by dot-blot  
925 immunoassay, the specific antigen binding activity versus SAP was  
926 higher for bile than for plasma antibodies. A different number of  
927 antigenic components of SAP and ESP were identified by immu-  
928 noblotting performed with plasma or bile antibodies. These results  
929 suggest that *T. bemaichii* parasitism by nematodes involves plasma  
930 and bile anti-parasite antibodies. Furthermore, bile antibodies were  
931 found to be more reactive and more heterogeneous than plasma  
932 antibodies. It is noteworthy that the high titers of bile antibodies  
933 detected, and the fact that they recognized an antigenic repertoire  
934 larger than that recognized by plasma antibodies could be attrib-  
935 uted to the greater topic immune response at intestinal level.

### 10. Conclusions

936 To summarize our results indicate that the immunoglobulins  
937 from Antarctic teleosts present several specific features; the  
938 diversity of VH gene segments is lower than that of non-Antarctic  
939 species; several intrinsic features of the hypermutation mechanism  
940 described in mammals were identified in *T. bemaichii* VH  
941 sequences. These include high occurrence of mutations in hot spot  
942 sequences, a bias for transitions over transversions, high concen-  
943 tration of nucleotide changes to CDRs, and an unusual prevalence  
944 of insertions and deletions. Taken together these properties provide  
945 clues for somatic diversification. At cDNA sequence level, the  
946 regions connecting different immunoglobulin domains have been  
947 found to be longer than those of non-Antarctic species to allow for  
948 a higher flexibility; moreover, high number of preferential sites of  
949 N-glycosylation contributes to increase the solubility of the anti-  
950 body molecule at low temperatures; at genomic level, the presence  
951 of significant insertions of reverse complementary sequences that

952 can drive the pre-mRNA splicing mechanisms generating the heavy  
953 chain membrane-bound form was highlighted; at chromosomal  
954 level, an IgH gene locus duplication was demonstrated in several  
955 Antarctic species by Fluorescence In Situ Hybridization, the dupli-  
956 cated genes being both expressed and resident on different chro-  
957 mosomes; at biological level, the host-nematode parasite  
958 relationship offered a very interesting model as, despite an anti-  
959 body response to parasites was demonstrated, a very high level of  
960 parasitism was well documented in several Antarctic fish species.

### Appendix. Supplementary material

Supplementary data related to this article can be found online at  
doi:10.1016/j.fsi.2010.10.018.

### References

- [1] Livermore R, Nankivell A, Eagles G, Morris P. Paleogene opening of Drake passage. *Earth Planet Sci Lett* 2005;236:459–70.
- [2] Shevenell AE, Kennett JP, Lea DW. Middle Miocene southern ocean cooling and antarctic cryosphere expansion. *Science* 2004;305:1766–70.
- [3] Eastman JT. The nature of the diversity of Antarctic fishes. *Polar Biol* 2005;28:93–107.
- [4] Eastman JT, McCune AR. Fishes on the Antarctic continental shelf: evolution of a marine species flock? *J Fish Biol* 2000;57:84–102.
- [5] DeVries AL. Glycoproteins as biological antifreeze agents in Antarctic fishes. *Science* 1971;172:1152–5.
- [6] Chen L, DeVries AL, Cheng C-HC. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc Natl Acad Sci USA* 1997;94:3811–6.
- [7] Fields PA, Somero GN. Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc Natl Acad Sci USA* 1998;95:11476–81.
- [8] Detrich 3rd HW, Parker SK, Williams Jr RC, Nogales E, Downing KH. Cold adaptation of microtubule assembly and dynamics. Structural interpretation of primary sequence changes present in the alpha- and beta-tubulins of Antarctic fishes. *J Biol Chem* 2000;275:37038–47.
- [9] Romisch K, Collie N, Soto N, Logue J, Lindsay M, Scheper W, et al. Protein translocation across the endoplasmic reticulum membrane in cold-adapted organisms. *J Cell Sci* 2003;116:2875–83.
- [10] O'Brien KM, Sidell BD. The interplay among cardiac ultrastructure, metabolism and the expression of oxygen-binding proteins in Antarctic fishes. *J Exp Biol* 2000;203:1287–97.
- [11] Sidell BD, O'Brien KM. When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefishes. *J Exp Biol* 2006;209:1791–802.
- [12] Sidell BD, Vayda ME, Small DJ, Moylan TJ, Londraville RL, Yuan MI, et al. Variable expression of myoglobin among the hemoglobinless Antarctic icefishes. *Proc Natl Acad Sci USA* 1997;94:3420–4.
- [13] Chen Z, Cheng CH, Zhang J, Gao L, Chen L, Zhou L, et al. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc Natl Acad Sci USA* 2008;105:12944–9.
- [14] Ota T, Nguyen TA, Huang E, Detrich HW, Amemiya CT. Positive Darwinian selection operating on the immunoglobulin heavy chain of Antarctic fishes. *J Exp Zool* 2003;295B:45–58.
- [15] Ye J, Bromage ES, Kaattari SL. The strength of B cell interaction with antigen determines the degree of IgM polymerization. *J Immunol* 2010;184:844–50.
- [16] Coscia MR, Oreste U. Presence of antibodies specific for proteins of *Contracaecum osculatum* (Rudolphi, 1908) in plasma of several Antarctic teleosts. *Fish Shellfish Immunol* 1998;8:295–302.
- [17] Pucci B, Coscia MR, Oreste U. Characterization of serum Immunoglobulin M of the Antarctic teleost *Trematomus bemaichii*. *Comp Biochem Physiol Part B* 2003;135:349–57.
- [18] Magnadóttir B, Guðmundsdóttir BK. A comparison of total and specific immunoglobulin levels in healthy Atlantic salmon (*Salmo salar* L.) and in salmon naturally infected with *Aeromonas salmonicida* subsp. *Achromogenes*. *Vet Immunol Immunopathol* 1992;32:179–89.
- [19] Magnadóttir B, Jónsdóttir H, Helgason S, Björnsson B, Jørgensen TØ, Pilstrom L. Humoral immune parameters in Atlantic cod (*Gadus morhua* L.). The effects of environmental temperature. *Comp Biochem Physiol* 1999;122:173–80.
- [20] Uchida D, Hirose H, Chang PK, Aranishi F, Hirayabu E, Mano N, et al. Characterization of Japanese eel immunoglobulin M and its level in serum. *Comp Biochem Physiol* 2000;127:525–32.
- [21] Wilson MR, Warr GW. Fish immunoglobulins and the genes that encode them. *Annu Rev Fish Dis* 1992;2:201–21.
- [22] Magnadóttir B, Kristjánssdóttir H, Guðmundsdóttir S. Characterisation of monoclonal antibodies to separate epitopes on salmon IgM heavy chain. *Fish Shellfish Immunol* 1996;6:185–98.

Please cite this article in press as: Coscia MR, et al., Antarctic teleost immunoglobulins: More extreme, more interesting. *Fish & Shellfish Immunology* (2010), doi:10.1016/j.fsi.2010.10.018

- 1021 [23] Abelli L, Coscia MR, De Santis A, Zeni C, Oreste U. Evidence for hepato-biliary  
1022 transport of immunoglobulin in the Antarctic teleost fish *Trematomus ber-*  
1023 *nacchii*. *Dev Comp Immunol* 2005;29:431–42.
- 1024 [24] Coscia MR, Oreste U. Exploring Antarctic teleost immunoglobulin genes. *Mar*  
1025 *Genomics* 2009;2:19–25.
- 1026 [25] Coscia MR, Oreste U. Limited diversity of the immunoglobulin heavy chain  
1027 variable domain of the emerald rockcod *Trematomus bernacchii*. *Fish Shellfish*  
1028 *Immunol* 2003;14:71–92.
- 1029 [26] Kimura M. A simple method for estimating evolutionary rates of base  
1030 substitutions through comparative studies of nucleotide sequences. *J Mol Evol*  
1031 1980;16:111–20.
- 1032 [27] Brown GD, Kaattari IM, Kaattari SL. Two new Ig VH gene families in *Oncorhynchus mykiss*. *Immunogenetics* 2006;58:933–6.
- 1033 [28] Lefranc MP, Pommié C, Kaas Q, Duprat E, Bosc N, Guiraudou D, et al. IMGT  
1034 unique numbering for immunoglobulin and T cell receptor constant domains  
1035 and Ig superfamily C-like domains. *Dev Comp Immunol* 2005;29:185–203.
- 1036 [29] Danilova N, Hohman VS, Kim EH, Steiner LA. Immunoglobulin variable-region  
1037 diversity in the zebrafish. *Immunogenetics* 2000;52:81–91.
- 1038 [30] Roman T, Andersson E, Bengtén E, Hansen J, Kaattari S, Pilström L, et al. Unified  
1039 nomenclature of Ig VH genes in rainbow trout (*Oncorhynchus mykiss*): defini-  
1040 tion of eleven VH families. *Immunogenetics* 1996;43:325–6.
- 1041 [31] Stenwik J, Lundbäck A, Jørgensen TØ, Pilström L. Variable region diversity of  
1042 the Atlantic cod (*Gadus morhua* L.) immunoglobulin heavy chain. *Immunogenetics* 2000;51:670–80.
- 1043 [32] Foster SJ, Dörner T, Lipsky PE. Somatic hypermutation of V<sub>H</sub> rearrange-  
1044 ments: targeting of RGYW motifs on both DNA strands and preferential  
1045 selection of mutated codons within RGYW motifs. *Eur J Immunol* 1999;29:  
1046 4011–21.
- 1047 [33] Oreste U, Coscia MR. Specific features of immunoglobulin VH genes of the  
1048 Antarctic teleost *Trematomus bernacchii*. *Gene* 2002;295:199–204.
- 1049 [34] Lange MD, Waldbieser GC, Lobb CJ. Patterns of receptor revision in the  
immunoglobulin heavy chains of a teleost fish. *J Immunol* 2009;182:  
5605–22.
- [35] Hansen JD, Landis ED, Phillips RB. Discovery of a unique Ig heavy-chain iso-  
type (IgT) in rainbow trout: implications for a distinctive B cell developmental  
pathway in teleost fish. *Proc Natl Acad Sci USA* 2005;102:6919–24.
- [36] Bengtén E, Quiniou S, Hikima J, Waldbieser G, Warr GW, Miller NW, et al.  
Structure of the catfish IGH locus: analysis of the region including the single  
functional IGHM gene. *Immunogenetics* 2006;58:831–44.
- [37] Nakao M, Morimoto T, Tomana M, Fujiki K, Yano T. Isolation of cDNA encoding  
the constant region of the immunoglobulin heavy-chain from common carp  
(*Cyprinus carpio* L.). *Fish Shellfish Immunol* 1998;8:425–34.
- [38] Pisano E, Coscia MR, Mazzei F, Ghigliotti L, Coutanceau JP, Ozouf-Costaz C,  
et al. Cytogenetic mapping of immunoglobulin heavy chain genes in Antarctic  
fish. *Genetica* 2007;130:9–17.
- [39] Coscia MR, Morea V, Tramontano A, Oreste U. Analysis of a cDNA sequence  
encoding the immunoglobulin heavy chain of the Antarctic teleost *Trem-*  
*atomus bernacchii*. *Fish Shellfish Immunol* 2000;10:343–57.
- [40] Near TJ. Estimating divergence times of neotethenid fishes using a fossil-  
calibrated molecular clock. *Antarctic Sc* 2004;16:37–44.
- [41] Coscia MR, Varriale S, De Santi C, Giacomelli S, Oreste U. Evolution of the  
Antarctic teleost immunoglobulin heavy chain gene. *Mol Phylogenet Evol*  
2010;55:226–33.
- [42] Merlino A, Varriale S, Coscia MR, Mazzarella L, Oreste U. Structure and  
dimerization of the teleost transmembrane immunoglobulin region. *J Mol*  
*Graph Mod* 2008;27:401–7.
- [43] Sanchez C, Alvarez A, Castillo A, Zapata A, Villena A, Dominguez J. Two  
different subpopulations of Ig-bearing cells in lymphoid organs of rainbow  
trout. *Dev Comp Immunol* 1995;19:79–86.
- [44] Haire R, Rast JP, Litman RT, Litman GW. Characterization of three isotypes of  
immunoglobulin light chains and T-cell antigen receptor a in zebrafish.  
*Immunogenetics* 2000;51:915–23.
- [45] Coscia MR, Giacomelli S, De Santi C, Varriale S, Oreste U. Immunoglobulin light  
chain isotypes in the teleost *Trematomus bernacchii*. *Mol Immunol*  
2008;45:3096–106.
- [46] Pilström L, Lundqvist ML, Wemmenstam NE. The immunoglobulin light chain  
in poikilothermic vertebrates. *Immunol Rev* 1998;166:123–32.
- [47] Stanfield RL, Zemla A, Wilson IA, Rupp B. Antibody elbow angles are influ-  
enced by their light chain class. *J Mol Biol* 2006;357:1566–74.
- [48] Coscia MR, Borriello A, Oreste U. Plasma and bile antibodies of the teleost  
*Trematomus bernacchii* specific for the nematode *Pseudoterranova decipiens*.  
*Dis Aquat Org* 2000;41:37–42.

## Ringraziamenti

Al termine di questo lungo percorso credo sia opportuno ringraziare tutti coloro che hanno contribuito al raggiungimento di questo traguardo così importante per me.

Desidero ringraziare innanzitutto il dott. Umberto Oreste e la dott.ssa Maria Rosaria Coscia, che mi hanno guidato e sostenuto con i loro utili consigli e con la loro grande esperienza e professionalità. Grazie a loro ho potuto comprendere cosa significhi realmente lavorare in un laboratorio di ricerca, ampliare le mie competenze e crescere a livello personale.

Ringrazio il prof. Lelio Mazzarella e il dott. Antonello Merlino per la loro assistenza con la loro immensa preparazione.

Un ringraziamento è inoltre rivolto alla prof.ssa Pompea Del Vecchio e al prof. Claudio Rivetti per la loro pazienza e disponibilità, che mi hanno permesso di effettuare misure molto utili per la mia ricerca.

Infine, ringrazio tutti i miei affetti, senza i quali non avrei mai potuto raggiungere i miei obiettivi, che mi hanno incoraggiato nei periodi bui e mi hanno incitato a tener duro nonostante le avversità e lo scoraggiamento. È a loro che va tutto il merito di questo lavoro. Grazie!