



Università degli studi di Napoli "Federico II"  
Facoltà di Scienze MM. FF. NN.

## Dottorato in Scienze Chimiche - XXI Ciclo

Indirizzo: Struttura e proprietà chimico-fisiche di molecole e sistemi di interesse biologico

### **Studio conformazionale di domini funzionali proteici**

Tutore Prof. Lucia Falcigno

Dottoranda:

Dott. Luisa Calvanese

Co-tutore Dott. Menotti Ruvo

Triennio 2005-08



Università degli studi di Napoli "Federico II"  
Facoltà di Scienze MM. FF. NN.

Dottorato in Scienze Chimiche - XXI Ciclo

PhD Thesis Title:

**Conformational study of protein functional domains**

Tutore Prof. Lucia Falcigno

Dottoranda:

Dott. Luisa Calvanese

Co-tutore Dott. Menotti Ruvo

Triennio 2005-08

## Indice

Riassunto	pag. 1
Abstract	pag. 4
CAPITOLO 1	
Introduzione	pag. 6
1.1 La superfamiglia dei TGF- $\beta$	pag. 7
1.2 I recettori dei TGF- $\beta$	pag. 17
1.3 Meccanismo di attivazione dei recettori dei TGF- $\beta$	pag. 28
1.4 Regolazione dell'attivazione dei recettori dei TGF- $\beta$	pag. 31
1.5 La proteina Cripto	pag. 32
1.5.1 Le funzioni biologiche di Cripto	pag. 33
1.5.2 Cripto e la crescita tumorale	pag. 34
1.5.3 Domini funzionali di Cripto	pag. 37
1.6 Oggetto della tesi	pag. 43
CAPITOLO 2 - Metodiche Sperimentali	
2.1 Sintesi chimica	pag. 44
2.2 Esperimenti di dicroismo circolare	pag. 47
2.3 Esperimenti NMR	pag. 47
2.3.1 Constraint sperimentali da misure NMR	pag. 49
2.4 Calcolo dei modelli molecolari	pag. 50
2.4.1 CYANA: Torsion Angle Dynamics	pag. 50
2.4.2 AMBER: Meccanica Molecolare e Minimizzazione energetica	pag. 52
2.5 Homology Modelling	pag. 53
2.6 Esperimenti di docking	pag. 56
2.7 Esperimenti di binding SPR	pag. 59
CAPITOLO 3	
Risultati e Discussione	pag. 65
3.1 Sintesi in fase solida dei domini CFC ed EGF-like di Cripto	pag. 65

3.2 Analisi conformazionale di hCFC a pH 7 in tampone fosfato	pag. 68
3.2.1 Analisi NMR di hCFC	pag. 69
3.2.2 Calcoli strutturali di hCFC	pag. 71
3.3 Analisi conformazionale di mCFC a pH 7 in tampone fosfato	pag. 73
3.3.1 Analisi NMR di mCFC	pag. 74
3.3.2 Calcoli strutturali di mCFC	pag. 78
3.4 Modello per omologia di ALK4-ECD e ALK7-ECD	pag. 79
3.5 Docking tra ALK4-ECD e hCFC	pag. 83
3.6 Studio di binding SPR tra hCFC e hEGF-like e i recettori ALK4/ActRIIB	pag. 86
3.7 Analisi conformazionale della proteina Nodal	pag. 88
3.7.1 Modello per omologia di Nodal	pag. 89
3.7.2 Analisi conformazionale del frammento Nodal53-69	pag. 89
3.7.3 Analisi conformazionale del frammento Nodal43-69	pag. 94
3.7.4 Analisi conformazionale del frammento Nodal66-90	pag. 98
3.8 Docking tra Nodal, ALK7-ECD e ALK4-ECD	pag. 102
 CAPITOLO 4 -Conclusioni	 pag. 108
 Bibliografia	 pag. 115
Appendice A	pag. 124
Appendice B	pag. 130
Appendice C	pag. 137
Appendice D	pag. 142
Appendice E	pag. 146

## Riassunto

Il progetto di questo dottorato di ricerca riguarda lo studio conformazionale in soluzione di domini funzionali proteici, mediante l'uso della spettroscopia NMR e di metodi computazionali. In particolare, l'attenzione è stata rivolta allo studio di alcuni fattori di crescita.

I fattori di crescita sono proteine coinvolte nella crescita, differenziazione e proliferazione cellulare. Queste proteine possono suddividersi in tre grandi gruppi: IGF (*Insulin Growth Factor*), FGF (*Fibroblast Growth Factor*), superfamiglia dei TGF- $\beta$  (*Transforming Growth Factor-beta*).

La superfamiglia dei TGF- $\beta$  comprende un grande numero di ligandi collegati strutturalmente che giocano un ruolo chiave nella progressione del ciclo cellulare. Un TGF- $\beta$  inizia la trasmissione del segnale (*signaling*) legandosi a recettori serina/treonina chinasi di tipo I e di tipo II presenti sulla superficie cellulare. In seguito a tale *binding*, c'è un avvicinamento dei recettori che permette al recettore di tipo II di fosforilare il dominio intracellulare del recettore di tipo I, che poi propaga il segnale attraverso la fosforilazione delle proteine Smad intracellulari. Si forma, quindi, un complesso di proteine Smad che viene traslocato nel nucleo dove va a regolare la trascrizione dei geni target dei TGF- $\beta$ .

Poiché la superfamiglia dei TGF- $\beta$  interviene nella crescita e proliferazione cellulare, molti dei suoi membri sono over-espressi, o il loro normale meccanismo è mal regolato, nei casi di crescita anomale di cellule quali i tumori. Proprio questo loro coinvolgimento nella genesi e sviluppo tumorale ha suscitato un grosso interesse scientifico verso lo studio funzionale e strutturale dei TGF- $\beta$ .

Tra i numerosi fattori di crescita, una famiglia con caratteristiche particolari è quella chiamata EGF-CFC, di cui la proteina Cripto è il capostipite. I fattori di crescita extracellulari EGF-CFC sono composti da due domini adiacenti ricchi di cisteine: EGF-like (Epidermal Growth Factor-like) e CFC (CRIPTO/FRL-1/Cryptic) di circa 40 residui ciascuno. La proteina Cripto è un elemento fondamentale per il *signaling* di Nodal, una proteina appartenente ai TGF- $\beta$ . In particolare, Cripto è il co-recettore di Nodal. Infatti, Nodal si lega al complesso recettoriale dell'attivina, formato dai recettori di tipo I (ALK4) e di tipo II (ActRIIB), solo in presenza di Cripto. In questo processo Cripto lega Nodal con il dominio EGF e lega ALK4 con il dominio CFC. Entrambi questi domini di Cripto interferiscono con l'attività onco-soppressiva delle attivine, o bloccando il recettore ALK4 o attraverso il legame all'attivina solubile. Sebbene entrambi i domini siano coinvolti nell'attività tumorigenica di Cripto, il dominio CFC ha un ruolo primario. Data l'importanza che la formazione di questa complessa rete di interazioni riveste, non solo nello sviluppo embrionale, ma anche nella proliferazione di cellule tumorali, è estremamente interessante ottenere informazioni strutturali sui vari fattori proteici coinvolti. Purtroppo, ad oggi, non sono note le strutture tridimensionali né in soluzione né allo stato solido di Cripto, di Nodal e del recettore ALK4. Quindi, nel corso di questo progetto di ricerca è stata intrapresa tale indagine strutturale. In particolare, siamo partiti dallo studio strutturale attraverso NMR in tampone fosfato del dominio CFC da topo e umano. I modelli molecolari ottenuti esibiscono una forma complessivamente ellissoidale e il

*folding* è globalmente esteso con tre *strand* collegati dai tre ponti disolfuro. E' importante notare che i residui ritenuti responsabili del *binding* con il recettore ALK4 (His<sup>9</sup> e Trp<sup>12</sup>) non sono localizzati nel *core* proteico, ma si ritrovano posizionati tra la fine del primo *strand* e l'inizio del primo *loop*, con le catene laterali esposte esternamente. Opposto al sito di *binding*, è localizzato un *cluster* di residui idrofobici che potrebbe svolgere una funzione stabilizzante dell'intera struttura proteica attraverso *binding* col dominio EGF-like, oppure potrebbe essere rilevante per la formazione d'interazioni intermolecolari.

Per approfondire lo studio strutturale sull'interazione del dominio CFC con il recettore ALK4, è stato costruito il modello per omologia del dominio extra-cellulare (ECD) di ALK4, basandosi sulla struttura cristallografica del recettore umano ALK3 (identità di sequenza 27 %) ed è stata modellata la struttura tridimensionale del complesso CFC/ALK4-ECD, usando *docking* macromolecolare assistito da dati sperimentali. Le soluzioni del *docking* sono state selezionate considerando i risultati degli studi di mutagenesi che indicano i residui H<sup>9</sup> e W<sup>12</sup> di CFC cruciali per il *binding* con il recettore ALK4. In base alle analisi condotte è stato scelto un modello rappresentativo del complesso CFC/ALK4-ECD che presenta buona complementarità di superficie, contatti elettrostatici e idrofobici favorevoli, basso valore dell'energia totale. Questo modello, consistente con i dati precedenti di mutagenesi, fornisce una base strutturale per la progettazione di analoghi modificati, che potranno confermare le ipotesi di interazione molecolare tra CFC ed ALK4 ed eventualmente funzionare come antagonisti.

Infine, abbiamo puntato la nostra attenzione su Nodal. La proteina Nodal è un omodimero di 220 amminoacidi con 14 cisteine, che formano 6 ponti disolfuro intra-catena e uno inter-catena che lega i due monomeri. E' stato costruito un modello per omologia sia del monomero sia del dimero della variante umana, utilizzando cinque template. Il modello di Nodal ottenuto per *homology modelling*, conserva il nodo di cisteina caratteristico della superfamiglia dei TGF- $\beta$  ed è formato da due  $\beta$ -*sheet* antiparalleli e un  $\alpha$ -elica perpendicolare ai  $\beta$ -*sheet*. I modelli ottenuti conservano le regioni caratteristiche di *binding* del recettore di tipo I ('*wrist epitope*') e del recettore di tipo II ('*knucle epitope*').

Contemporaneamente, per avere informazioni strutturali in soluzione di Nodal sono stati sintetizzati e caratterizzati tre frammenti, che riproducono tratti di sequenza presumibilmente coinvolti nel legame con i suoi *partner* proteici. L'analisi conformazionale NMR evidenzia, nelle condizioni sperimentali adoperate, una fedele riproduzione strutturale dei relativi tratti nella proteina parente. Dall'insieme dei risultati ottenuti, sia mediante NMR che mediante *homology modelling*, si possono evidenziare elementi strutturali potenzialmente rilevanti per il *binding* con i recettori o altri *partner* proteici.

In conclusione, questo lavoro di ricerca ha consentito l'acquisizione di informazioni strutturali, finora sconosciute, sulle componenti del complesso proteico che è alla base del meccanismo di *signaling* di Nodal via Cripto e ha permesso di formulare ipotesi sugli elementi strutturali rilevanti ai fini delle mutue interazioni.

Per confermare tali ipotesi è stato avviato uno studio di *binding* "Real Time Bia" con tecnica Surface Plasmon Resonance (SPR) utilizzando uno strumento BIACORE 3000. Sebbene diversi esperimenti siano

stati già condotti, occorrono tuttavia ulteriori prove per arrivare a chiarire il quadro completo di questa complessa rete d'interazioni

## Abstract

This PhD research project deals with the conformational study in solution of protein functional domains, by means of NMR spectroscopy and computational methods. In particular, attention has been paid to the study of the transforming growth factors.

The growth factors are protein involved in the cell growth, differentiation and proliferation. They can be divided in three major groups: IGF (Insulin Growth Factor), FGF (Fibroblast Growth Factor), the TGF- $\beta$  superfamily (Transforming Growth Factor-beta).

The TGF- $\beta$  superfamily is a large group of structurally related ligands that play a key role in the cell-cycle progression. A TGF- $\beta$  initiates the signalling by binding to type I and type II serine/threonine kinase receptors on the cell surface. Upon binding, the two receptors approach each other and the type II receptor can phosphorylate the intracellular domain of the type I receptor, which in turn phosphorylates a number of intracellular mediator proteins known as Smads. The activated Smad complexes are translocated into the nucleus and regulate the transcription of target genes.

Since the TGF- $\beta$  superfamily takes part in the cell growth and proliferation, many of its members are overexpressed or their normal mechanism is misregulated, in the case of anomalous cell growth as tumours. Such involvement into tumour genesis and development has stirred up a large scientific interest to the functional and structural study of TGF- $\beta$ .

Among the TGF- $\beta$ , a family with particular features is the so called EGF-CFC family whose the founding member is Cripto. The EGF-CFC extra-cellular growth factors are composed of two adjacent cystein-rich domains of about 40 residues: EGF-like (Epidermal Growth Factor-like) and CFC (CRIPTO/FRL-1/Cryptic).

Cripto is fundamental for the Nodal signalling. In particular, Cripto is the co-receptor of Nodal, a member of the TGF- $\beta$  superfamily. Indeed, only in the presence of Cripto, Nodal binds to the Activin receptor complex comprising the type I (ALK4) and the type II (ActRIIB) receptors. In this process, the EGF-like domain binds to Nodal while the CFC domain binds to ALK4. Both Cripto domains interfere with the Activin onco-suppressive activity or by blocking the ALK4 receptor or by direct binding to soluble Activin. Although both domains are involved in the Cripto tumorigenic activity, the CFC domain has a primary role. By considering the relevance of the formation of such complex network of interactions, not only in the embryonic development but also in the tumour progression, it is extremely interesting to gain structural information on the operating protein factors.

Unfortunately, up to now, the tridimensional structures, either in solution or at the solid state, of Cripto, Nodal and ALK4 receptor are unknown. So, in this research work this conformational investigation has been undertaken. In particular, we have started with the structural study by NMR in phosphate buffer of the mouse and human CFC domain. The obtained molecular models show an ellipsoidal compact shape and the folding is globally extended with three anti-parallel *strands*

connected by disulfide bridges. Significantly, in all the NMR molecular models, the side chains of residues H<sup>9</sup> and W<sup>12</sup>, reportedly involved in the binding with the ALK4 receptor are not buried in the protein core but are localized between the end of the first strand and at the beginning of the first loop with the side chains exposed to the solvent. In addition, a hydrophobic patch, that supposedly acts as an interface with the cell membrane or with the adjacent EGF-like domain was found opposite to the binding site.

In order to broaden our understanding of the CFC/ALK4 interaction, the homology model of the extra-cellular ALK4 domain (ALK4-ECD) has been built, using the crystal structure of the human ALK3 receptor (sequence identity 27%) as a template. Furthermore, a model of CFC/ALK4-ECD complex has been obtained by macromolecular docking assisted by experimental data. The docking solutions have been filtered by considering previous mutagenesis results indicating that the H<sup>9</sup> and W<sup>12</sup> residues of CFC have a crucial role in the binding to ALK4. The selected model of the CFC/ALK4-ECD complex with the lowest total energy shows a good surface complementarity, and favourable electrostatics and hydrophobic contacts. Such model, consistent with prior mutagenesis data, can be quite useful in designing modified analogues to help in a more thorough understanding of the CFC/ALK4 molecular interaction.

Finally, we have pointed our attention to Nodal. Nodal is an homodimer of 220 aminoacids with 14 cysteine residues forming 6 intra-chain disulfide bridges and 1 inter-chain that binds the two monomers. The homology model of both Nodal monomer and dimer has been built using five selected templates. The Nodal homology model keeps the "cysteine knot" motif, typical of TGF- $\beta$  superfamily, and shows two antiparallel  $\beta$ -sheets and an  $\alpha$ -helix perpendicular to the sheets. The models preserve the characteristic binding regions of the type I receptor (wrist epitope) and the type II receptor (knuckle epitope).

Concurrently, to obtain structural information in solution, three Nodal fragments have been synthesized and characterized. These fragments reproduce sequence regions probably involved in the binding to the receptors. The NMR analysis reveals, in the experimental explored conditions, a strict structural similarity with the corresponding segments in the native protein. On the grounds of all results, obtained either by NMR or by homology model, structural features, potentially relevant for the receptor binding, can be highlighted.

In conclusion, this research work has allowed structural information, so far unknown, of the components of the protein complex at the base of the Nodal signalling via Cripto and to formulate hypotheses on the structural elements important for the mutual interactions.

To confirm such hypotheses a "Real Time Bia" study by Surface Plasmon Resonance (SPR) with a BIACORE 3000 instrument has been undertaken. Although several experiments have been already carried out, further tests are necessary to clarify the complete picture of such a complicate interaction network.

# CAPITOLO 1

- 1.1 La superfamiglia dei TGF- $\beta$
- 1.2 I recettori dei TGF- $\beta$
- 1.3 Meccanismo di attivazione dei recettori dei TGF- $\beta$
- 1.4 Regolazione dell'attivazione dei recettori dei TGF- $\beta$
- 1.5 La proteina Cripto
- 1.6 Oggetto della tesi

## Introduzione

Lo sviluppo è un processo mediante il quale un organismo si forma a partire da una cellula mediante ripetute divisioni cellulari (mitosi) e differenziamento delle cellule che ne derivano, e acquisisce le strutture tipiche della sua specie. Lo sviluppo, fenomeno detto anche ontogenesi, è un processo tipico degli organismi pluricellulari. In particolare, con sviluppo embrionale, si intende il periodo in cui si formano i tessuti e gli organi necessari alla sopravvivenza dell'organismo (negli uccelli, ad esempio, è la fase di formazione del pulcino fino alla schiusa dell'uovo; nell'uomo, corrisponde alla fase in cui il feto si trova nell'utero materno). Una delle conclusioni più sorprendenti della moderna ricerca biologica è che il meccanismo di sviluppo è molto simile per tutti gli animali, uomo compreso.

Lo sviluppo dipende in larga misura dagli effetti di un gruppo di cellule su un altro. Questi effetti sono esercitati dall'induzione embrionale, che prevede il rilascio di un fattore inducente da un centro di segnalazione e la sua azione su un gruppo di cellule rispondenti. Esistono parecchie famiglie molecolari di tali importanti fattori che sono noti come fattori di crescita, citochine o ormoni.

Gli ormoni sono molecole di segnalazione ad ampio raggio (dette endocrine). Queste sostanze sono prodotte e secrete dalle cellule o dai tessuti e messe in circolazione attraverso il sangue e altri fluidi del corpo per influenzare l'attività di altre cellule o altri tessuti nel corpo. I fattori di crescita, invece, sono sintetizzati dalle cellule e influenzano la funzione cellulare della stessa cellula (autocrini) o di un'altra cellula vicina (paracrini). Queste molecole sono determinanti nella differenziazione, crescita, mobilità e espressione genica cellulare e nel definire lo sviluppo delle cellule in tessuti o organi. Normalmente i fattori di crescita sono efficaci in concentrazioni molto basse e hanno una forte affinità con i loro recettori presenti sulle cellule interessate. Un fattore di crescita può avere effetti biologici diversi a seconda del tipo di cellula con cui interagisce. La risposta di una cellula dipende fortemente dal tipo di recettori presenti su di essa. Molti fattori di crescita favoriscono o inibiscono la funzione cellulare e possono essere multifattoriali. In altre parole, possono essere necessarie due o più

sostanze per indurre una risposta cellulare specifica. La proliferazione, la crescita e lo sviluppo della maggior parte delle cellule richiede la combinazione specifica di più fattori di crescita piuttosto che di un singolo.

I fattori di crescita, le citochine e gli ormoni sono essenzialmente proteine, sebbene alcune siano piccole molecole lipidiche come l'acido retinoico. La maggior parte sono secrete da cellule segnalatrici, mentre alcune sono proteine integrali di membrana che possono solo interagire con cellule immediatamente adiacenti. Esse si legano a recettori specifici e attivano il *pathway* di trasduzione del segnale. Questo può portare all'attivazione o alla repressione di specifici geni o al cambiamento del comportamento cellulare mediato attraverso il citoscheletro o attraverso variazioni nel metabolismo cellulare. Il repertorio delle risposte delle cellule dipende dal tipo di recettori che esse possiedono, dal modo con cui essi sono accoppiati al *pathway* di trasduzione e da come questi *pathway* sono accoppiati.

### 1.1 La superfamiglia dei TGF- $\beta$ (fattori di crescita trasformanti-beta)

Il TGF- $\beta$ , originariamente scoperto come mitogeno, fu successivamente identificato come il prototipo di una vasta e diversificata superfamiglia di molecole segnale. La superfamiglia dei fattori di crescita trasformanti-beta (Transforming Growth Factors- $\beta$ , TGF- $\beta$ ) comprende un grande numero di proteine collegate strutturalmente che giocano un ruolo importante in diversi processi cellulari, quali la progressione del ciclo cellulare, la differenziazione cellulare, la funzione riproduttiva, lo sviluppo, la motilità, l'adesione, la crescita neuronale, la morfogenesi dell'osso, la cicatrizzazione delle ferite.<sup>1-5</sup> Questo ruolo è molto diversificato a seconda dei tessuti in cui vengono secreti e delle quantità in cui sono espressi; in alcuni casi possono anche agire da potenti inibitori della crescita come è stato osservato in una varietà di cellule epiteliali, endoteliali e linfoidi.<sup>6-13</sup>

Un TGF- $\beta$  inizia la trasmissione del segnale (*signaling*) legandosi a recettori serina/treonina chinasi di tipo I e di tipo II presenti sulla superficie cellulare. In seguito a tale *binding*, c'è un avvicinamento dei recettori che permette al recettore di tipo II di fosforilare il dominio intracellulare del recettore di tipo I, che poi propaga il segnale attraverso la fosforilazione delle proteine Smad intracellulari. Si forma, quindi, un complesso di proteine Smad che viene traslocato nel nucleo dove regola la trascrizione dei geni target dei TGF- $\beta$ . (Fig.1.1).

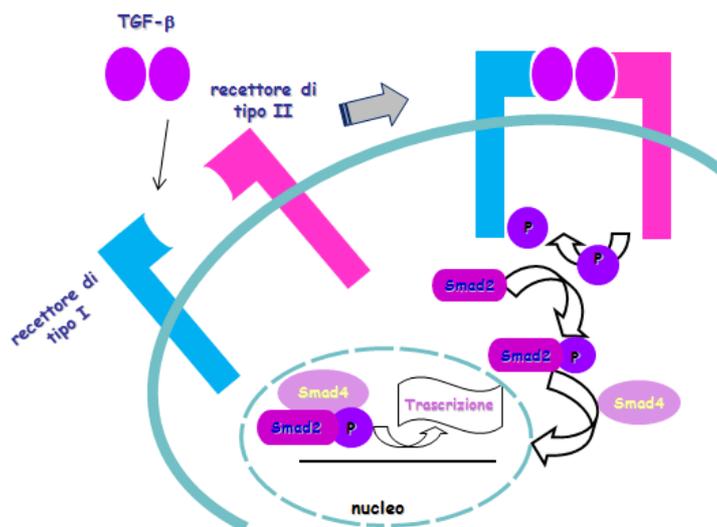


Fig.1.1: Pathway di trasduzione del segnale dei TGF-β

Poichè il *signaling* dei TGF-β è coinvolto nella crescita cellulare, la disfunzione di questo percorso contribuisce nella tumorigenesi. Mutazioni nei recettori della famiglia dei TGF-β o nelle proteine Smad sono state osservate in alcuni tumori. Il recettore di tipo II dei TGF-β, ad esempio, è inattivo a causa di mutazioni in molti tumori gastrointestinali umani<sup>14</sup> e la Smad4 è inattiva in quasi la metà di tutti i carcinoma pancreatici.<sup>15</sup>

La famiglia dei TGF-β può essere filogeneticamente divisa nei mammiferi in tre principali gruppi<sup>16,17</sup>:

- TGF-β
- Attivina/Inibina/Nodal
- BMP (proteina morfogenica dell'osso)/GDF(fattore di crescita e differenziazione)/MIS (ormone antimulleriano).

Questi tre gruppi possono suddiversi ulteriormente in relativi sottogruppi in base alla loro omologia di sequenza (Tabella 1).

**Tabella 1:** Membri della superfamiglia dei TGF-β<sup>18</sup>

Ligando	Nomi alternativi	Sottogruppi
<b>TGF-β</b>		TGF-β
TGF-β1		
TGF-β2		
TGF-β3		
<b>Attivina/Inibina/Nodal</b>		
Attivina A	Subunità Attivina β <sub>A</sub> -β <sub>A</sub>	Attivine
Attivina B	Subunità Attivina β <sub>B</sub> -β <sub>B</sub>	
Attivina AB	Subunità Attivina β <sub>A</sub> -β <sub>B</sub>	
Inibina A	Subunità Inibina α- Attivina β <sub>A</sub>	Inibine

Inibina B	Subunità Inibina $\alpha$ - Attivina $\beta_B$	
Nodal	Ndr	Nodal
<b>BMP/GDF/MIS</b>		
BMP2	BMP2A	BMP2/4
BMP4	BMP2B	
BMP3	Osteogenina	BMP3
BMP3b	GDF10/Sumitomo-BIP	
BMP5		BMP5/6/7
BMP6	Vgr1/DVR6	
BMP7	OP1	
BMP8A	OP2	
BMP8B	OP3/PC8	
BMP9	GDF2	BMP9/10
BMP10		
BMP11	GDF11	GDF8
BMP12	GDF7	GDF5/6/7
BMP13	GDF6	
BMP15	GDF9b	GDF9
GDF1		GDF1
GDF3	Vgr2	
GDF2	BMP9	BMP9/10
GDF5	CDMP1	GDF5/6/7
GDF6	CDMP1/BMP13	
GDF7	BMP12	
GDF8	Miostatina	GDF8
GDF11	BMP11	
GDF9		GDF9
GDF9b	BMP15	
GDF15	MIC1/PLAB/PTGFB/PDF	
GDF10	BMP3b	BMP3
MIS	MIF/AMH	MIS

I membri della superfamiglia dei TGF- $\beta$  si formano a partire da precursori secreti da cellule specifiche, in seguito alla loro maturazione proteolitica. I precursori contengono un peptide segnale aminoterminale, un pro-dominio ed un dominio maturo carbossiterminale, che costituisce il fattore di crescita attivo. La forma monomericamente di questi fattori di crescita

contiene 110-140 amminoacidi ed è caratterizzata da sei cisteine conservate. Quasi tutti i membri della superfamiglia dei TGF- $\beta$  hanno un settimo residuo di cisteina conservato, che è richiesto per formare le strutture dimeriche (~25kDa) covalentemente legate che interagiscono con i domini extracellulari dei loro rispettivi recettori.

Nel 1992, la struttura cristallina di TGF- $\beta$ 2 rivelò un esclusivo *fold* proteico<sup>19,20</sup> (Fig. 1.2):

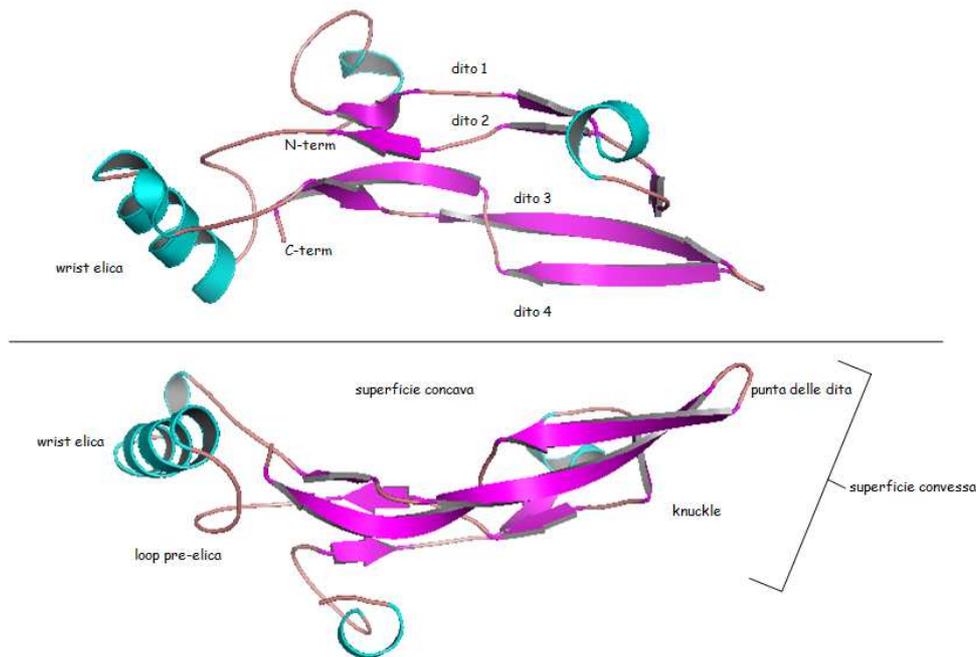


Fig.1.2: *Fold* dei TGF- $\beta$ <sup>20</sup> (pdbcode:1tfg). Un tipico monomero dei TGF- $\beta$  consiste di un nodo di cisteina con due coppie di  $\beta$ -*strand* antiparalleli e un  $\alpha$ -elica (polso). I  $\beta$ -*strand* formano una superficie concava e convessa per l'interazione con i recettori

Il monomero ha una forma sottile, allungata e ricurva che contiene un motivo strutturalmente conservato conosciuto come nodo di cisteina. Il nodo è costituito da un anello di otto residui amminoacidici, contenente due legami disolfuro intra-catena, con un terzo ponte disolfuro che passa attraverso l'anello. Oltre al nodo di cisteina il monomero è caratterizzato da due  $\beta$ -*sheet* antiparalleli e una  $\alpha$ -elica anfipatica (H3) con il suo asse perpendicolare agli *sheet* (Fig. 1.3).

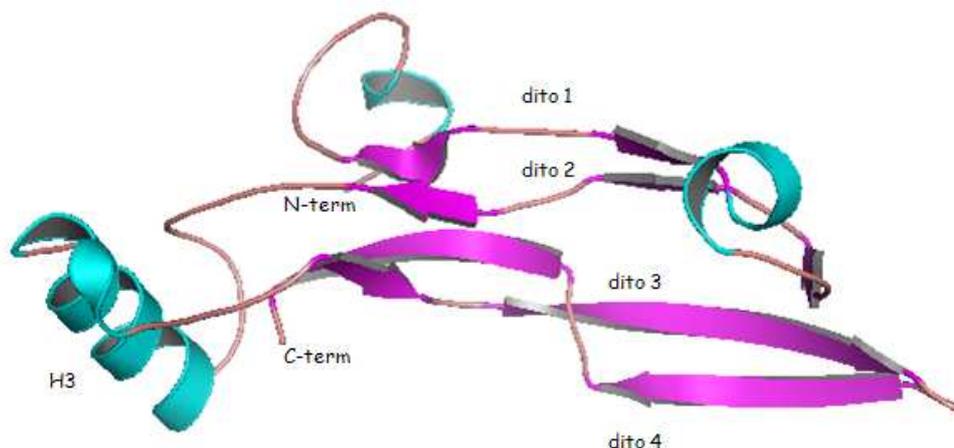


Fig.1.3: Rappresentazione con *ribbon* del monomero di TGF- $\beta$ 2<sup>20</sup> (1992; pdbcode:1tfg). I  $\beta$ -*sheet*, formati dal dito1-dito2 e dal dito3-dito4, sono colorati in rosa. L' $\alpha$ -elica (H3) è colorato in celeste

Il *fold* dei TGF- $\beta$  (Fig.1.2) ricorda una mano aperta, con i  $\beta$ -*strand* che fungono da dita, il nodo di cisteine da palmo, l'elica (H3) da polso e l'estremità amminotermiale da pollice. C'è una rete estesa di legami idrogeno tra gli *strand* di ciascun dito, i due  $\beta$ -*sheet* invece sono tra loro troppo lontani per formare interazioni non-covalenti. Essi sono invece legati attraverso i ponti disolfuro del nodo di cisteina e sono stabilizzati da contatti di van der Waals e interazioni idrofobiche. Uno sguardo più attento alle sequenze amminoacidiche delle strutture conosciute dei TGF- $\beta$  rivela l'assenza di un core idrofobico convenzionale nel monomero isolato. Sul monomero si identificano due regioni caratteristiche che sono coinvolte nel *binding* con i rispettivi recettori: il '*wrist epitope*' che lega il recettore di tipo I e il '*knucle epitope*' che lega il recettore di tipo II (Fig.1.2).<sup>21,22</sup>

Fino ad oggi, altri otto fattori di crescita della superfamiglia dei TGF- $\beta$  sono stati caratterizzati strutturalmente: le coordinate atomiche per TGF- $\beta$ 3,<sup>23,24</sup> BMP2,<sup>25-27</sup> BMP7,<sup>28,29</sup> BMP9,<sup>30</sup> GDF5<sup>31,32</sup> e Attivina<sup>33,34</sup> sono state determinate dalla cristallografia raggi-X; la struttura in soluzione di TGF- $\beta$ 1 è stata determinata mediante NMR,<sup>35</sup> mentre un modello teorico di BMP15 è stato ottenuto mediante *homology modelling*.<sup>36</sup> Il *fold* di questi TGF- $\beta$  è identico a quello di TGF- $\beta$ 2. La forma attiva dei TGF- $\beta$  è un dimero stabilizzato da interazioni idrofobiche e di solito rinforzato da un ponte disolfuro tra le subunità. Fanno eccezione GDF3 e GDF9, due fattori di crescita strutturalmente simili implicati nella maturazione dell'ovocita e nello sviluppo del follicolo.<sup>37</sup> Questi due ligandi non hanno la settima cisteina, normalmente usata nel formare il ponte disolfuro tra i due monomeri e possono formare omodimeri ed eterodimeri non-covalenti con altri fattori di crescita. Le subunità dell'Attivina,

inibina e poche BMP possono formare eterodimeri. Vari eterodimeri, così come BMP2/7<sup>38</sup> e BMP4/7<sup>39,40</sup> hanno un'attività biologica più potente dei corrispondenti omodimeri.<sup>41,42</sup> Di particolare interesse è l'inibina, un eterodimero consistente di un'unica  $\alpha$ -subunità e una  $\beta$ -subunità dell'Attivina, che è un inibitore del segnale delle BMP e delle Attivine.

Nonostante il *fold* comune per tutti i monomeri, molte differenze possono essere osservate dal confronto della conformazione dei dimeri dei ligandi dei TGF- $\beta$ . Nella Fig.1.4 sono illustrate le strutture dimeriche di tre proteine appartenenti alla sottofamiglia dei TGF- $\beta$ : l'Attivina, BMP7 e TGF- $\beta$ 3.<sup>22</sup>

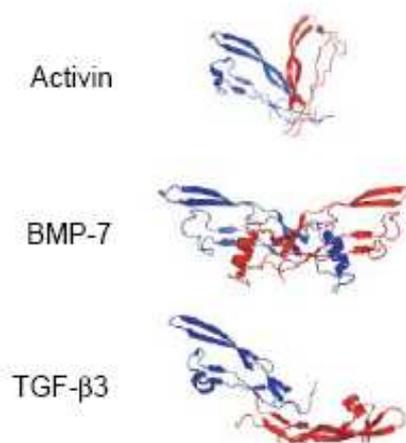


Fig.1.4: Confronto dei dimeri dell'Attivina (pdbcode:1nyu), BMP7 (pdbcode:1lxi), e TGF- $\beta$  (pdbcode:1ktz). I monomeri sono colorati rossi e blu

Nella formazione del dimero la parte non polare dell' $\alpha$ -elica di un monomero è in contatto con i residui idrofobici nello *sheet* dell'altro, con conseguente formazione di due core idrofobici legati da duplice simmetria. L'assemblaggio dei monomeri è molto meno compatto nella regione che circonda il ponte disolfuro intermolecolare, lasciando spazio per alcune molecole di acqua ordinate all'interfaccia del dimero. Tipicamente, circa il 20% di tutta la regione del monomero accessibile al solvente partecipa nell'interfaccia di dimerizzazione, suggerendo che la formazione dimerica è un processo energeticamente favorito.

Molti ligandi TGF- $\beta$  esibiscono l'arrangiamento simmetrico e allungato del dimero della BMP7<sup>29</sup> (pdbcode:1lxi), come ad esempio TGF- $\beta$ 1, TGF- $\beta$ 2 e BMP2.<sup>33</sup> Questo arrangiamento viene definito canonico. L'Attivina A e il TGF- $\beta$ 3 (Fig.1.4), invece, possono adattare conformazioni alternative, entrambi risultanti dalla mancanza della regione del 'polso-elica' (H3). Il

monomero dell'Attivina A (Fig.1.5) ha il *fold* canonico dei TGF- $\beta$  esclusa la regione dell' elica (H3), con piccole variazioni nei *loop* tra i  $\beta$ -*strand* che formano le dita.

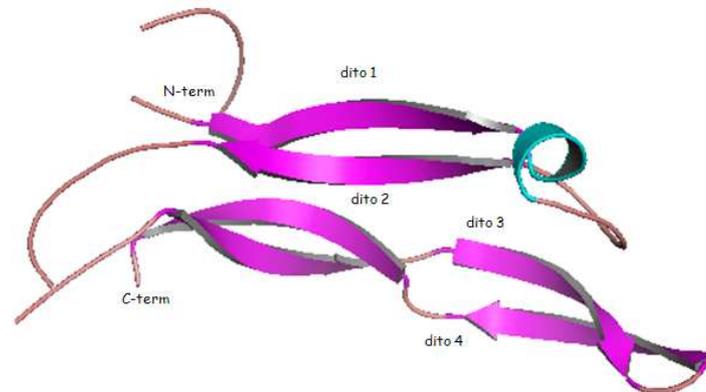


Fig.1.5: Rappresentazione con *ribbon* del monomero dell'Attivina A<sup>33</sup> (pdbcode:1nyu). I  $\beta$ -*sheet*, formati dal dito1-dito2 e dal dito3-dito4, sono colorati in rosa

Il valore dell'RMSD sui C $\alpha$  della sovrapposizione del monomero dell'Attivina A sui monomeri della BMP7 e TGF- $\beta$ 3 è rispettivamente di 1.25 Å<sup>2</sup> e 1.70 Å<sup>2</sup>.

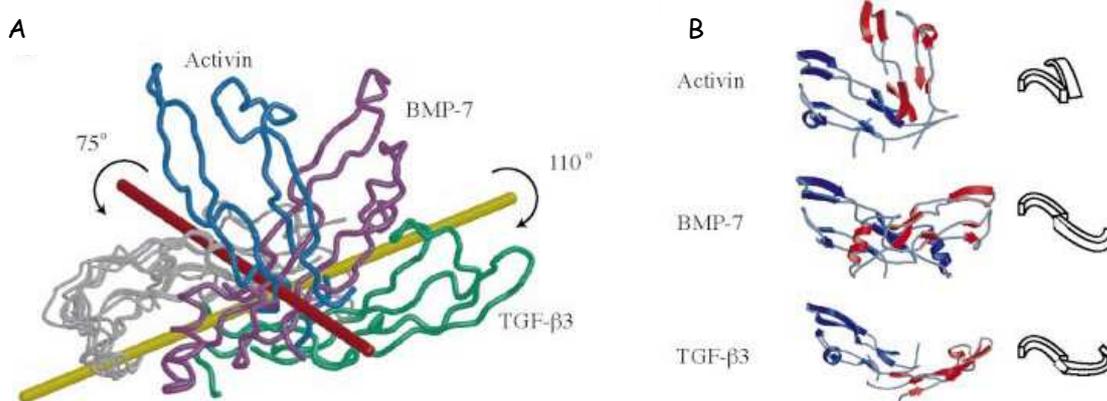


Fig.1.6:(A) Sovrapposizione dei dimeri dell'Attivina A (pdbcode:1nyu) (blu), BMP7 (pdbcode:1lxi) (viola) e TGF- $\beta$ 3 (pdbcode:1ktz) (verde) mantenendo la posizione di uno dei monomeri (grigio) fissa. (B) Confronto dei dimeri dell'Attivina A, BMP7 e TGF- $\beta$ 3. I monomeri sono colorati rossi e blu in ciascuna struttura e sono allineati ad un singolo monomero (blu). Uno schema del dimero è mostrato a destra

La sovrapposizione su un monomero dell'Attivina A, TGF- $\beta$ 3 e BMP7 mette in luce le grandi differenze conformazionali nella posizione del secondo monomero (Fig.1.6), che può essere descritta con movimenti del dominio della struttura della BMP7. Il cambiamento strutturale dal dimero esteso della BMP7 a quello compatto dell'Attivina A coinvolge una rotazione di 75°

di un monomero, su un asse di rotazione quasi perpendicolare all'asse longitudinale del dimero della BMP7. Al contrario, cambiamenti strutturali in TGF- $\beta$ 3 coinvolgono una rotazione di 110° sull' asse longitudinale del dimero della BMP7. Le rotazioni per l'Attivina A e TGF- $\beta$ 3 si verificano in direzioni opposte creando una forma 'chiusa' per l'Attivina e una forma 'aperta' per TGF- $\beta$ 3 (Fig.1.6).

Il dimero dell'Attivina, struttura ottenuta dai Raggi X in complesso con il recettore di tipo II,<sup>33</sup> (pdbcode:1nyu) è legato da un legame disolfuro intercatena (Cys<sup>80</sup>), che si trova vicino alla regione del 'polso-elica'. L'interfaccia del dimero è prevalentemente idrofobica e coinvolge due piccole parti adattate simmetricamente sui monomeri, che includono Cys<sup>80</sup>, Val<sup>82</sup>, Pro<sup>83</sup>, Leu<sup>86</sup>, Gln<sup>106</sup> e Asn<sup>107</sup> (Fig.1.7). Questi residui sono parzialmente conservati nella superfamiglia dei TGF- $\beta$ .<sup>43</sup> Cys<sup>80</sup> e Pro<sup>83</sup> sono completamente conservate nella Attivina B e C, mentre Val<sup>82</sup> è sostituita da una Ile nell'Attivina B, la Leu<sup>86</sup> da una Arg nell'Attivina C, la Gln<sup>106</sup> da una Pro in entrambi le Attivine, e l'Asn<sup>107</sup> da un Asp nell'Attivina C. Le differenze nei residui tra l'Attivina e altri membri della famiglia TGF- $\beta$  all' interfaccia dimerica canonica dei TGF- $\beta$  può avere effetti sulla flessibilità conformazionale dell'Attivina A. Tra BMP7 e TGF- $\beta$ 2, c'è una buona corrispondenza tra i residui coinvolti nell'interfaccia di dimerizzazione.<sup>28</sup> Dei 22 residui nascosti all'interfaccia dimerica canonica, 17 sono identici o sostituiti in modo conservativo nell'Attivina. Solo cinque residui sono differenti in dimensione e/o polarità. La Phe<sup>16</sup> dell'Attivina A è sostituita da una Leu sia in BMP7 sia in TGF- $\beta$ 2, la Phe<sup>58</sup> è sostituita da Asp e Glu, la Gln<sup>65</sup> da Leu in entrambi, l'Asn<sup>76</sup> è sostituita da Val e Ser, e la Leu<sup>77</sup> da Pro e Ala. La Leu<sup>86</sup> è sempre presente all'interfaccia dimerica sia nell'arrangiamento canonico (es. BMP7, BMP2, TGF- $\beta$ 2 etc.) sia nell'altro (Attivina A). La maggiore flessibilità nella regione del '*wrist epitope*' dell'Attivina A può quindi essere dovuta sia ai differenti residui presenti sia a differenze strutturali, quali la mancanza dell' $\alpha$ -elica.

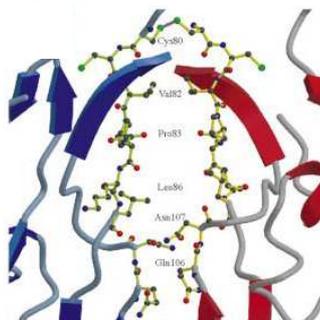


Fig.1.7: Rappresentazioni dei residui all'interfaccia del dimero dell'Attivina A. I *ribbon* dei monomeri sono colorati in blu e rosso

In altre strutture Raggi X in cui l'Attivina A si trova legata al recettore di tipo II<sup>34</sup> (pdbcode:1s4y) o al suo antagonista la Follistatina<sup>44</sup> (pdbcode:2b0u), si osserva un'ulteriore orientazione del dimero dell'Attivina A (Fig.1.8).

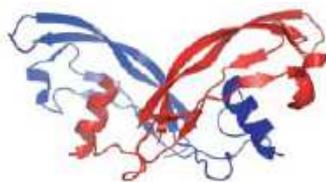


Fig.1.8: Conformazione aperta e estesa del dimero dell'Attivina A (pdbcode:1s4y,<sup>34</sup> pdbcode:2b0u<sup>44</sup>)

L'Attivina ha una conformazione aperta e estesa ed è moderatamente ruotata rispetto all'orientazione canonica dei TGF- $\beta$ ; all'interfaccia del dimero conserva elementi canonici di struttura secondaria, come una parte dell'  $\alpha$ -elica ( H3) (Fig.1.9).

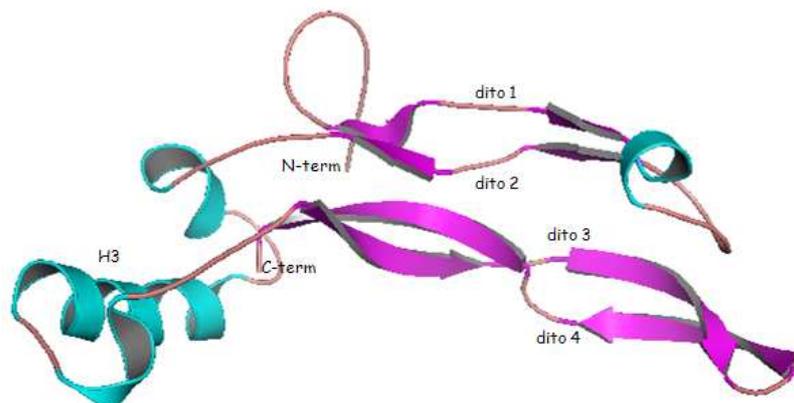


Fig.1.9: Rappresentazione con *ribbon* del monomero dell'Attivina A<sup>44</sup> (pdbcode:2b0u). I  $\beta$ -*sheet*, formati dal dito1-dito2 e dal dito3-dito4, sono colorati in rosa. L' $\alpha$ -elica (H3) è colorato in celeste

I due monomeri di TGF- $\beta$ 3 (Fig.1.4 e Fig.1.6B), invece, che formano l'omodimero biologicamente attivo hanno un arrangiamento testa-coda e sono uniti da un legame disolfuro alla posizione 77 sia nella forma libera<sup>23</sup> (pdbcode:1tgk) sia nella forma legata al recettore di tipo II<sup>24</sup> (pdbcode:1ktz).

Il *fold* del monomero di TGF- $\beta$ 3 libero ricorda quello canonico dei TGF- $\beta$ ; infatti è formato dai due  $\beta$ -*sheet* antiparalleli e l' $\alpha$ -elica perpendicolare agli *sheet* (H3). Il confronto del monomero del TGF- $\beta$ 3 libero e legato al recettore di tipo II rivela un piccolo movimento del  $\beta$ -turn che unisce i due *sheet*; nella forma legata, inoltre, mancano il segmento N-terminale che include il *turn* di elica (H1) e la terza  $\alpha$ -elica centrale (H3) (Fig.1.10).

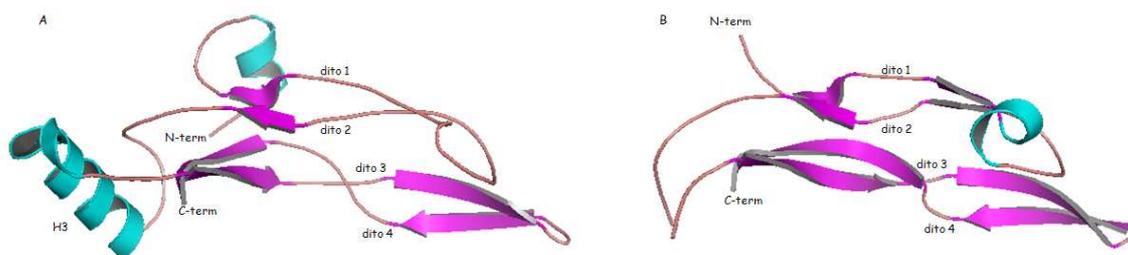


Fig.1.10: Rappresentazione con *ribbon* del monomero di TGF- $\beta$ 3 nella forma libera (A) (pdbcode:1tqk<sup>23</sup>) e legato al recettore di tipo II (B) (pdbcode:1ktz<sup>24</sup>). I  $\beta$ -*sheet*, formati dal dito1-dito2 e dal dito3-dito4, sono colorati in rosa. L' $\alpha$ -elica (H3) è colorato in celeste

Oltre alla variazione di questi elementi strutturali, vi è un riarrangiamento dei due monomeri di TGF- $\beta$ 3. Uno dei monomeri nella forma legata subisce una rotazione di 101° intorno ad un'asse approssimativamente collineare con la direzione dei relativi  $\beta$ -*strand* dell'altro monomero (Fig.1.11).

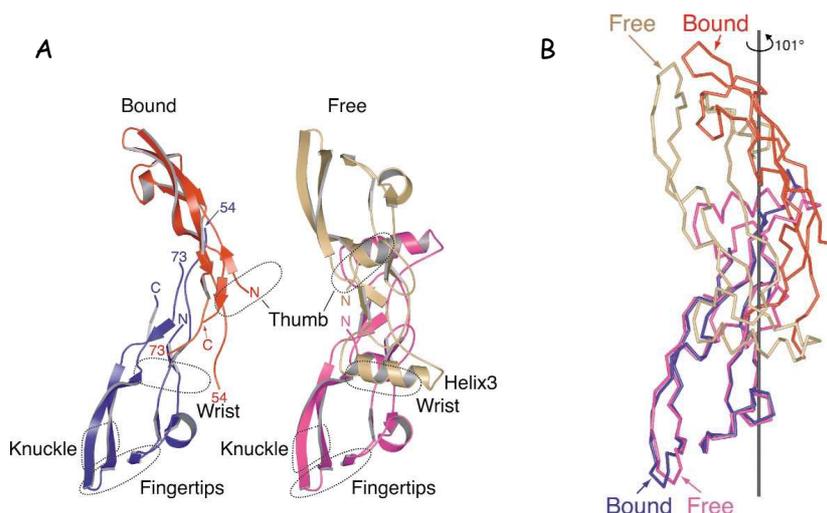


Fig.1.11: (A): Differenze strutturali tra il TGF- $\beta$ 3 legato (monomeri colorati blu e rosso) e libero (monomeri colorati rosa e marrone). (B): Confronto tra le strutture dei dimeri di TGF- $\beta$ 3 legato e libero, nel quale le strutture sono allineate tra il monomero blu della forma legata e il monomero rosa della forma libera (l'RMSD è di 1.1 Å)

La transizione strutturale osservata dell'omodimero è probabilmente legata alla perturbazione dell' $\alpha$ -elica H3, poichè l'elica H3 nella forma libera di TGF- $\beta$ 3 contribuisce nel 58% della superficie nascosta (680 Å<sup>2</sup> per monomero) e nella maggior parte dei contatti idrofobici all'interfaccia dell'omodimero. Ulteriori studi strutturali sia di dicroismo circolare e sia NMR del ligando libero indicano che TGF- $\beta$ 3 consiste di una regione stabile con ponti disolfuro, che

comprende i due  $\beta$ -sheet anti-paralleli e una regione 'palmo/polso' instabile, che è formata dall' $\alpha$ -elica H3 e i loop collegati ai residui N- e C-terminali.

### 1.2 I recettori dei TGF- $\beta$

I fattori della famiglia dei TGF- $\beta$  sono espressi in pattern tessuto-specifici e possono funzionare in maniera:

- endocrina, in cui il fattore di crescita influenza una cellula differente in fenotipo dalla sua cellula d'origine e localizzata in un luogo anatomicamente lontano;
- paracrina, in cui il fattore di crescita influenza una cellula adiacente o vicina ma differente come fenotipo dalla sua cellula d'origine;
- autocrina, in cui il fattore di crescita influenza la sua cellula d'origine o altre cellule identiche a quella nel fenotipo.

La specificità recettoriale, la distribuzione nei tessuti, e i livelli di espressione possono tutti interessare le risultanti risposte cellulari. I TGF- $\beta$  trasmettono il segnale legandosi a recettori di membrana serina-treonina chinasi (Tabella2), chiamati recettori di tipo I e II.<sup>45</sup>

**Tabella 2:** Recettori della superfamiglia dei TGF- $\beta$ <sup>18,46</sup>

Recettori	Nomi alternativi	Ligandi
<b>Recettori di tipo I</b>		
Gruppo ALK1		
ALK1	TSR1/SKR3/ACVRL1	TGF- $\beta$ , Attivina A
ALK2	TSK7L/ActRIA/SKR1	TGF- $\beta$ , Attivina A, MIS, BMP6/7
Gruppo ALK3		
ALK3	BMPRI/BMPRIA/BRK1/Tfr11/ ACVRLK3	BMP2/4, BMP6/7, GDF1, GDF3
ALK6	BMPRII/BRK2	BMP2/4, GDF5/6, GDF9b, MIS, BMP6/7
Gruppo ALK5		
ALK4	ActRIB/ACVR1B/SKR2	Attivina A, GDF1, GDF3, Vgr1 e Nodal (con EGF-CFC), GDF11
ALK5	TGF- $\beta$ RI/SKR4	TGF- $\beta$ , GDF11
ALK7		Nodal, Attivina B, Attivina AB, GDF3

GDF11

**Recettori di tipo II**

TGF- $\beta$ RII		TGF- $\beta$
BMPRII	BRK3/T-Alk	Inibina A (con TGFRIII), BMP2/4, BMP6/7, GDF5/6, GDF9b
ActRIIA	ActRII, ACVR2	Attivina A, Inibina A/B, GDF1, GDF3 e Nodal (con EGF-CFC), BMP2, BMP6/7, GDF5, GDF9b, GDF8
ActRIIB	ACVR2B	Attivina A, Inibina A/B, GDF1, GDF3, Vgr1 e Nodal (con EGF-CFC), BMP2, BMP6/7, GDF5, GDF8/11
MISRII	AMHR2	MIS

I recettori di tipo I e tipo II sono glicoproteine di circa 55 e 70 kDa, rispettivamente, che interagiscono dopo il legame al ligando. Entrambi i tipi di recettori serina/treonina chinasi consistono di circa 500 amminoacidi, organizzati sequenzialmente in un dominio extracellulare N-terminale con cui lega il ligando, una regione transmembranica e un dominio intracellulare C-terminale serina/treonina chinasi (Fig.1.12).<sup>17</sup> Nel dominio extracellulare di entrambi recettori è presente una regione di nove amminoacidi (CCX<sub>4-5</sub>CN), ricca di cisteine (box di cisteine). I recettori di tipo I differiscono dai recettori di tipo II nella sequenza del dominio chinasi, nella presenza di una piccola serie di residui tandem Ser/Gly e altri amminoacidi conservati immediatamente prima del dominio chinasi (Type I box), e un pattern caratteristico di sette residui di cisteina extracellulari che precedono il box di cisteine (CXCX<sub>3-5</sub>C<sub>4-25</sub>C<sub>4-5</sub>C<sub>13-16</sub>GCX<sub>7-19</sub>CX<sub>11-13</sub>).

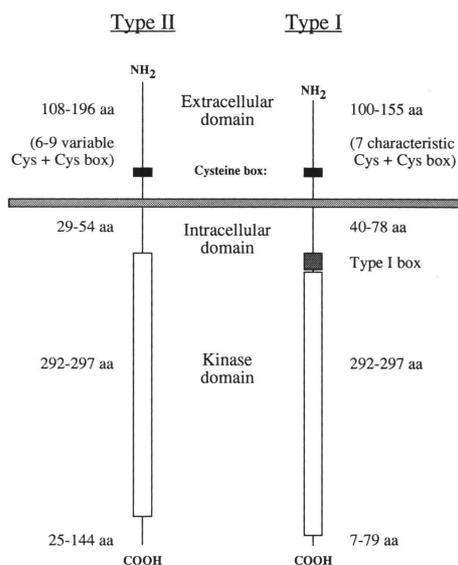


Fig.1.12 Schema dei recettori di tipo I e di tipo II della superfamiglia dei TGF- $\beta$ <sup>1</sup>

Il primo recettore identificato della superfamiglia dei TGF- $\beta$  è stato il recettore di tipo II dell'Attivina (ActRII).<sup>47</sup> Poco dopo, una grande classe di recettori serina-treonina sono stati identificati, con caratteristiche strutturali simili al recettore dell'Attivina.<sup>22</sup> Ad oggi, quattro altri recettori di tipo II sono stati identificati nei mammiferi: ActRIIB,<sup>48</sup> AMHR-II,<sup>49,50</sup> T $\beta$ RII,<sup>51</sup> and BMPRII.<sup>52-55</sup> In aggiunta sono stati clonati sette recettori di tipo I chiamati recettori dell'Attivina con attività chinasicca: da ALK1 ad ALK7.<sup>56-61</sup>

La prima struttura RX dei recettori dei TGF- $\beta$  che è stata risolta è stata quella del dominio extracellulare di ActRIIA<sup>62</sup>. Ad oggi altre strutture RX sono note: ActRIIB,<sup>33</sup> T $\beta$ RII,<sup>24</sup> ALK3<sup>26,27</sup> e T $\beta$ RI.<sup>63</sup>

Tutte le strutture dei domini extracellulari esibiscono un *fold* simile a tre-dita (three-finger toxin *fold*), con ciascuno dito formato da coppie di  $\beta$ -*strand* anti-paralleli, stabilizzati da cinque ponti disolfuro formati da dieci residui di cisteina conservati (Fig.1.13). Le tre coppie di  $\beta$ -*strand* antiparalleli si curvano per generare una superficie concava per legare il ligando.<sup>22</sup> Nonostante il *fold* comune e i residui di cisteina conservati nei  $\beta$ -*sheet*, non esiste né un'identità di sequenza né sovrapposizioni funzionali tra i due tipi di recettori.

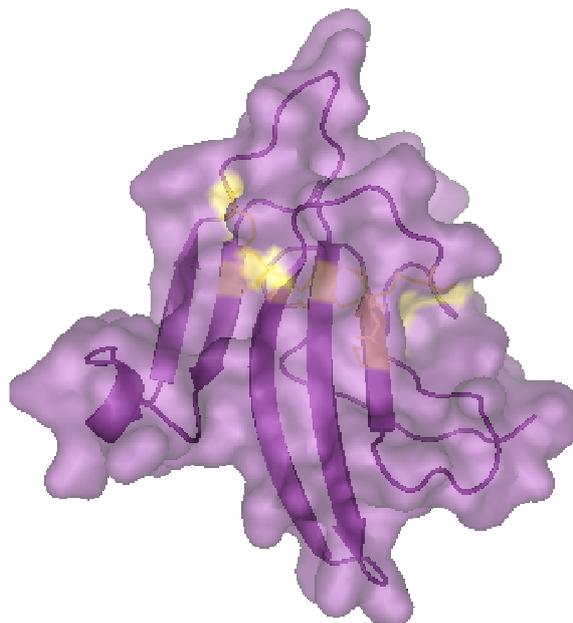


Fig.1.13: Struttura del dominio extracellulare del recettore ActRIIA<sup>62</sup> (pdbcode:1bte)

Una caratteristica comune di tutti recettori di tipo I, ma non dei recettori di tipo II, è la presenza del motivo TTSGSGSG altamente conservato nella loro regione citoplasmatica, chiamato il dominio GS, che gioca un ruolo chiave nella regolazione dell'attività chinasicca del recettore di tipo I.

I recettori di tipo I e tipo II differiscono nell'affinità e specificità, e possono essere espressi diversamente per permettere la regolazione della cascata di segnali intracellulari.

Un TGF- $\beta$  inizia quindi la trasmissione del segnale legandosi a recettori serina/treonina chinasi di tipo I e di tipo II presenti sulla superficie cellulare. In seguito a tale *binding*, c'è un avvicinamento dei recettori che permette al recettore di tipo II di fosforilare il dominio intracellulare del recettore di tipo I. Si forma un olo-complesso ternario tra il ligando TGF- $\beta$  e i recettori di tipo I e di tipo II.<sup>64,65</sup>

L'accesso dei ligandi al recettore è regolato da molte famiglie di proteine globalmente conosciute come ligandi trappola (Fig.1.14).

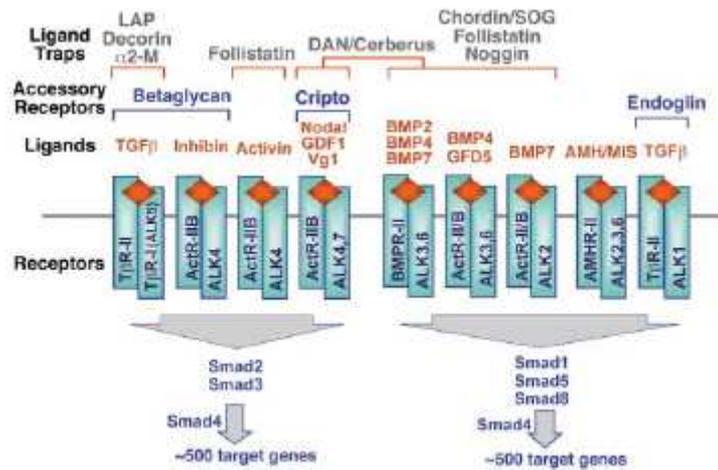


Fig.1.14: Una descrizione schematica della relazione tra i ligandi dei TGF- $\beta$ , i ligandi trappola, recettori accessori, e i recettori di tipo I e di tipo II. Gli R-Smad 1,2,3,5, ed 8 sono raggruppati in base alla loro specificità di segnale

Il complesso della struttura cristallina Noggin-BMP7<sup>66</sup> (pdbcode:1m4u) rivela che il ligando trappola Noggin inibisce la BMP7 bloccando le superfici che sono richieste per interagire con i recettori di tipo I e di tipo II della BMP<sup>66</sup> (Fig.1.15):

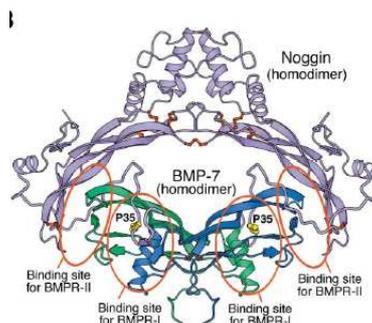


Fig.1.15: Struttura di un dimero di Noggin (colorato in viola) in complesso con un omodimero della BMP7<sup>66</sup> (pdbcode:1m4u). Le regioni N-terminali di Noggin interagiscono con la stessa area superficiale che è richiesta per il *binding* ai recettori di tipo I e II. Un residuo cruciale Pro35 di Noggin è colorato in giallo. Questo residuo occupa la stessa tasca idrofobica di BMP7 che è richiesta per il *binding* del recettore di tipo I della BMP

L'assemblamento dinamico che porta alla formazione dell' olo-complesso ligando-recettore può differire tra i ligandi. Esistono due distinti modi di interazione ligando-recettore, uno coinvolge la sottofamiglia delle BMP e l'altro è rappresentato dai TGF- $\beta$  e le Attivine. Le BMP, come la BMP2 e BMP4, hanno un'alta affinità per i domini extracellulari dei recettori di tipo I e una bassa affinità per i recettori di tipo II <sup>17</sup>. La struttura dell'omodimero della BMP2 in complesso con i due ectodomini del recettore di tipo I (BMPRIA-ECD) rivela che il recettore lega il '*wrist epitope*' del dimero della BMP2, con contatti che coinvolgono entrambi i monomeri della BMP2<sup>26</sup> (pdbcode:1es7) (Fig.1.16).

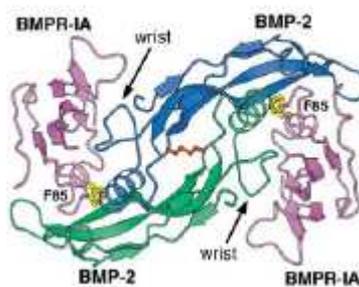


Fig.1.16: Complesso BMP2:BMPRIA-ECD<sup>26</sup> (pdbcode:1es7). Il residuo altamente conservato nei recettori di tipo I (Phe<sup>85</sup> in BMPRIA-ECD) è mostrato in giallo

Queste interazioni sono prevalentemente idrofobiche, con la Phe85 di BMPRIA-ECD che gioca un ruolo chiave. La catena laterale della Phe<sup>85</sup>, secondo il modello chiave-serratura, punta verso una tasca idrofobica formata all'interfaccia dei due monomeri della BMP2. La Phe<sup>85</sup> è conservata o sostituita da un residuo simile in tutti i recettori di tipo I eccetto ALK1; tutti i

residui che formano la tasca idrofobica nel dimero della BMP2 sono altamente conservati in tutti gli altri ligandi dei TGF- $\beta$ . Questo suggerisce che la tasca idrofobica di legame è essenziale per le interazioni tra i ligandi dei TGF- $\beta$  e i recettori di tipo I.<sup>26</sup> Successivamente, la determinazione delle strutture cristallografiche dei complessi ternari BMP2: BMPRIA-ECD:ActRIIA-ECD<sup>27,67</sup> ha rivelato che i domini extracellulari dei due recettori non interagiscono e non inducono significativi cambi conformazionali (Fig.1.17).

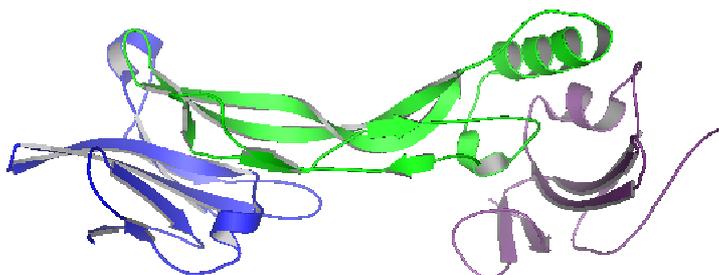


Fig.1.17: Rappresentazione del complesso BMP2 (verde):BMPR-IA (viola):ActRIIA (blu). (pdbcode:2goo)<sup>27</sup>

Il *binding* tra la BMP2 e ActRIIA-ECD avviene tra la superficie concava del recettore e il 'knuckle epitope' del dimero del ligando. L'interfaccia di legame è altamente idrofobica ed è formata da 12 residui della BMP2 e 10 residui di ActRIIA-ECD. Tre residui di ActRIIA-ECD (Phe<sup>42</sup>, Trp<sup>60</sup>, Phe<sup>83</sup>) sono importanti nel legame alla BMP2. Questi residui formano un core idrofobico all'interfaccia BMP2:ActRIIA-ECD in contatto con i residui Ala<sup>34</sup>, Pro<sup>35</sup>, Ser<sup>88</sup>, Met<sup>89</sup> e Leu<sup>90</sup> della BMP2 (pdbcode:2goo).<sup>27</sup>

Al contrario delle BMP, i TGF- $\beta$  e le Attivine hanno una grande affinità per i recettori di tipo II e non interagiscono con i recettori di tipo I isolati.<sup>68</sup> In questi casi il ligando lega prima fortemente il dominio extracellulare del recettore di tipo II; questo legame permette l'incorporazione del recettore di tipo I, con la formazione di un grande complesso ligando-recettore che coinvolge il ligando dimerico e quattro molecole di recettore.

L'analisi strutturale del dominio extracellulare umano del recettore di tipo II dei TGF- $\beta$  (T $\beta$ R<sub>II</sub>-ECD) in complesso con TGF- $\beta$ 3 rivela che il *binding* avviene all'estremità degli *sheet* (punta delle dita; 'fingertip') del dimero dei ligandi (Fig.1.18).

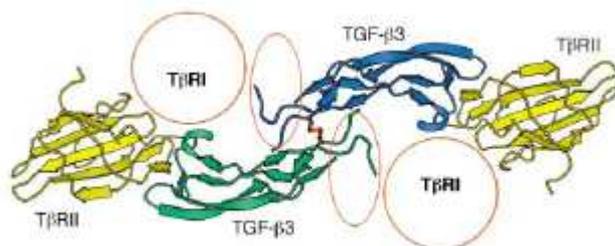


Fig.1.18: Struttura del dominio extracellulare umano del recettore di tipo II dei TGF- $\beta$  (T $\beta$ RII-ECD) legato a TGF- $\beta$ .<sup>24</sup> Le regioni disordinate del 'wrist' dei dimeri della BMP2 sono indicati da ovali rossi, mentre i siti di legame predetti per T $\beta$ RI sono mostrati da cerchi rossi. L'orientazione di TGF- $\beta$  è la stessa della BMP2 nella Fig.1.16

Ciascun recettore lega un monomero del dimerico di TGF- $\beta$ . Questa interazione crea due *patch* di superfici concave posizionate simmetricamente, che sono considerate i siti di *binding* per il dominio extracellulare del recettore di tipo I (Fig.1.18).

Contrariamente quindi all'interazione indipendente dei recettori della BMP, la struttura RX del complesso TGF $\beta$ :T $\beta$ RII-ECD:T $\beta$ RI-ECD<sup>63</sup> rivela che l'assemblaggio del complesso è sequenziale e cooperativo (Fig.1.19).

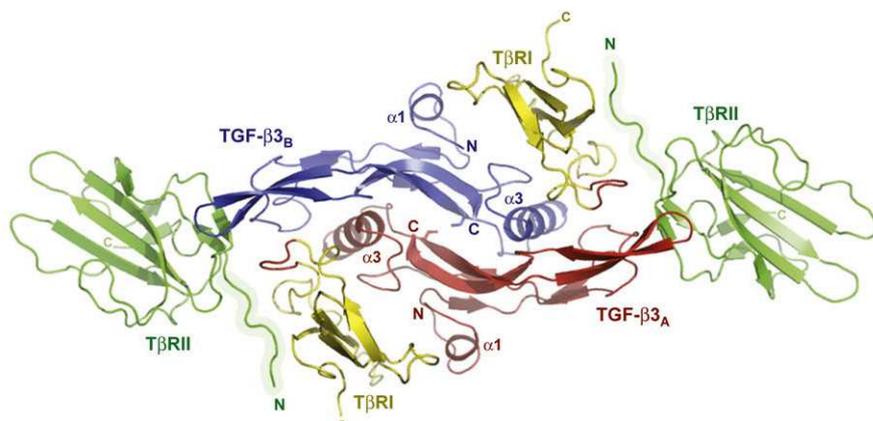


Fig.1.19: Rappresentazione del complesso TGF $\beta$ (rosso-blu):T $\beta$ RII-ECD (giallo):T $\beta$ RI-ECD (verde) (pdbcode:2pyj)<sup>63</sup>

In questo complesso il ligando TGF- $\beta$  non adotta la conformazione 'aperta' osservata nel complesso con T $\beta$ RII-ECD,<sup>24</sup> ma adotta la conformazione 'chiusa' vista nelle strutture cristalline del ligando libero.<sup>23</sup> L'impedimento sterico o contatti di van der Waals tra T $\beta$ RI e l'elica H3 ( $\alpha$ 3) di TGF- $\beta$  può portare il ligando alla conformazione 'chiusa' nel complesso ternario. E' stato proposto che il *loop* flessibile precedente l'elica  $\alpha$ 3 dei ligandi TGF- $\beta$ , il *loop* pre-elica, è il motivo conservato di riconoscimento di tutti i recettori di tipo I.<sup>69</sup> L'allineamento sequenziale di TGF- $\beta$  su BMP2 nella struttura del complesso BMP2:BMPRIA-

ECD<sup>25</sup> mostra che i TGF- $\beta$  non hanno tre residui che appartengono al *loop* pre-elica della BMP2 (Asp<sup>53</sup>, His<sup>54</sup>, e Leu<sup>55</sup>) (Fig.1.20A). Questi residui adottano una conformazione  $\alpha$ -elicoidale complementare alla superficie concava di BMPRIA. Il piccolo *loop* pre-elica delle isoforme dei TGF- $\beta$  preclude un'interazione ligando-recettore di tipo I simile alla BMP2. In aggiunta, piccoli ma significativi ingombri sterici con il recettore di tipo I in quella posizione e un'elica unica ( $\alpha$ 1) N-terminale, comportano il modo alternativo di TGF- $\beta$ 3 di legarsi a T $\beta$ RI.

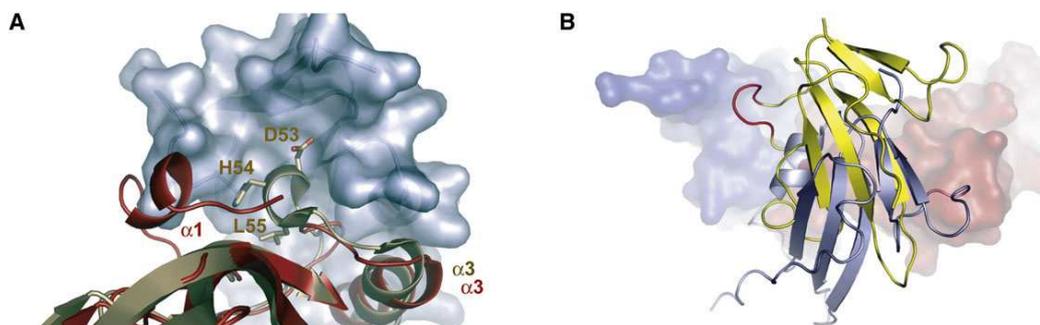


Fig.1.20: **(A)** Sovrapposizione di TGF- $\beta$ 3<sub>A</sub> (rosso) su BMP2<sub>A</sub> (oro) nel complesso binario BMP2:BMPRIA-ECD. La superficie di BMPRIA-ECD è rappresentata in grigio (le catene dei monomeri TGF- $\beta$ 3<sub>B</sub> e BMP2<sub>B</sub> non sono mostrate per semplicità). **(B)** Sovrapposizione del complesso binario della BMP2 con il complesso ternario di TGF- $\beta$ 3. TGF- $\beta$ 3<sub>A</sub> e TGF- $\beta$ 3<sub>B</sub> sono mostrati con le superfici rosse e blu, rispettivamente, BMPRIA-ECD e T $\beta$ RI-ECD sono mostrati con il *ribbon* viola e giallo, rispettivamente. (BMP2 e T $\beta$ RRII non sono mostrati)

La sovrapposizione dei ligandi nelle strutture cristalline del complesso ternario di TGF $\beta$ :T $\beta$ RRII-ECD:T $\beta$ RI-ECD e del complesso binario BMP2:BMPRIA-ECD mostra che T $\beta$ RI-ECD è ruotato di 45° rispetto a BMPRIA-ECD lungo l'asse longitudinale del ligando, consentendo l'avvicinamento dell'estensione pre-elica di T $\beta$ RI-ECD alle dita di TGF- $\beta$ 3 (Fig.1.19 e Fig.1.20B). Questa rotazione di T $\beta$ RI-ECD, combinata alla compattezza del *loop* pre-elica di TGF- $\beta$ 3, riduce drasticamente l'interazione tra TGF- $\beta$ 3<sub>A</sub> e T $\beta$ RI-ECD, che sono separati in gran parte da un canale riempito di solvente. L'interazione dell'estensione pre-elica di T $\beta$ RI-ECD con TGF- $\beta$ 3 è prevalentemente idrofobica e coinvolge anche altri residui vicini: Ile<sup>54</sup> e Phe<sup>60</sup>. A causa delle sue due proline (Pro<sup>55</sup> e Pro<sup>59</sup>), l'estensione si curva bruscamente, creando un rigonfiamento perpendicolare che si inserisce nell'ampio solco che separa le dita del ligando (Fig.1.21). Le catene laterali idrofobiche di tre residui del recettore (Ile<sup>54</sup>, Pro<sup>55</sup> e Phe<sup>60</sup>) interagiscono con quattro residui idrofobici del ligando (Trp<sup>30</sup>, Trp<sup>32</sup>, Tyr<sup>90</sup>, e Leu<sup>101</sup>), che non variano nelle isoforme di TGF $\beta$ . Il residuo Phe<sup>60</sup> di T $\beta$ RI non interagisce con la cavità idrofobica dell'interfaccia del dimero del ligando, come il corrispondente residuo di BMPRIA-

ECD (Phe<sup>85</sup>). L'anello fenolico di Phe<sup>60</sup> si pone lungo la superficie del ligando, avvicinandosi per *stacking* all'anello indolico del Trp30 di TGF- $\beta$ 3.

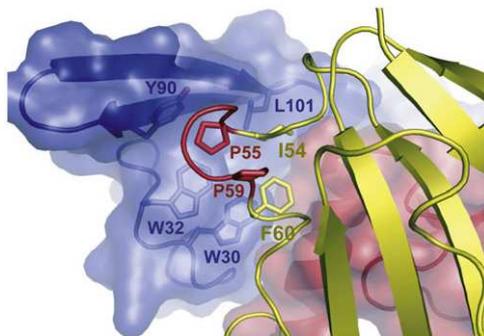


Fig.1.21: Interazione tra l' estensione pre-elica (rosso) di T $\beta$ RI-ECD (giallo) e le dita del monomero di TGF- $\beta$ 3 (blu)

L'analisi strutturale del dominio extracellulare umano del recettore di tipo II dell'Attivina (ActRIIB-ECD) in complesso con il ligando rivela che il recettore si lega all'estremità degli *sheet* dell'Attivina (Fig.1.22), in un modo che differisce dal legame di TGF- $\beta$ 3 a T $\beta$ RII.

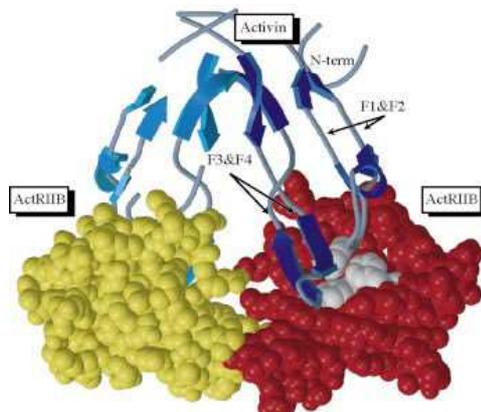


Fig.1.22: Rappresentazione del complesso ActRIIB-ECD:ActivinA. Il dimero dell'Attivina è mostrato con il *ribbon* blu chiaro o scuro per ciascun monomero e le dita sono indicate con F1-F4; i modelli di ActRIIB-ECD sono rappresentati con CPK gialli e rossi. Il cluster idrofobico (Y<sup>60</sup>, W<sup>78</sup>, F<sup>101</sup>) è mostrato in grigio

T $\beta$ RII-ECD lega TGF- $\beta$ 3 esclusivamente attraverso il primo dito del *fold* (Fig.1.23B). Al contrario ActRIIB-ECD lega l'Attivina A attraverso la superficie concava formata dalle dita 2 e 3 (L23) (Fig.1.23 A). La lunghezza e la posizione di L23 differisce nei due recettori. In ActRIIB-ECD il *loop* L23 è piccolo, forma parte dell'interfaccia del complesso e consente l'accesso ai residui idrofobici importanti per il legame all'Attivina. In T $\beta$ RII, L23 è più lungo di sei amminoacidi ed è esteso verso la parte delle dita che comprende la superficie analoga

del cluster idrofobico di ActRIIB-ECD. Un'altra differenza dei recettori è la dimensione e l'orientazione del dito 1. In T $\beta$ RII-ECD il dito 1 si avvolge sul retro del recettore formando la superficie d'interazione, mentre in ActRIIB-ECD il dito 1 punta al di fuori del core del recettore e quindi non fa parte della superficie d'interazione (Fig.1.23).

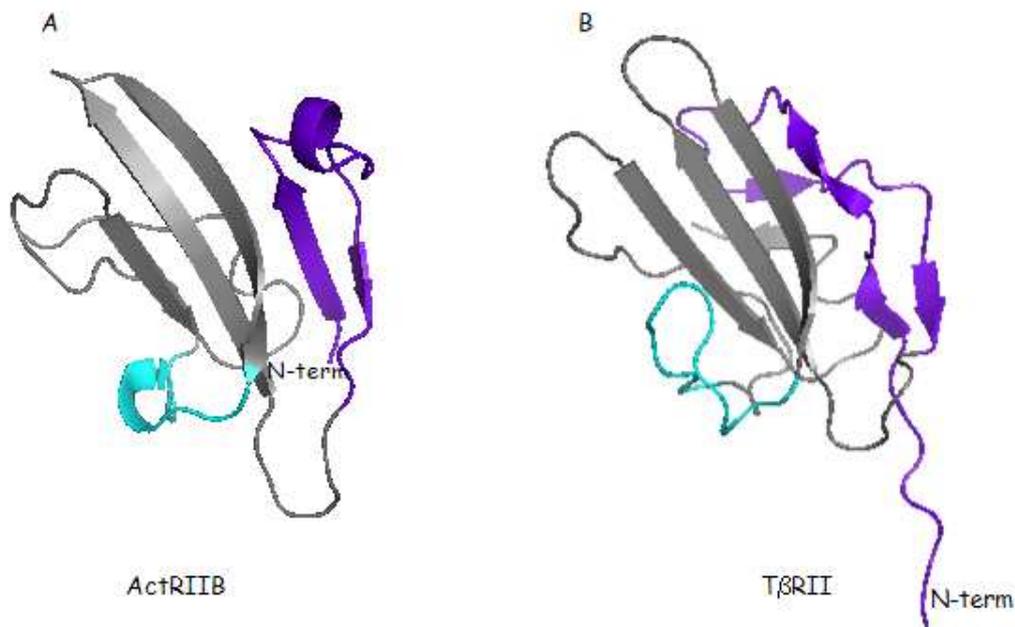


Fig.1.23: Rappresentazione con *ribbon* di (A) ActRIIB-ECD; (B) T $\beta$ RII-ECD. In celeste è evidenziato il *loop* L23, in viola il dito 1

L'interfaccia di legame ActRIIB-ECD:AttivinaA coinvolge residui idrofobici, carichi e polari (Fig.1.24). In ActRIIB-ECD, gli amminoacidi idrofobici (Val<sup>73</sup>, Val<sup>99</sup>, Tyr<sup>60</sup>, Trp<sup>78</sup>, Leu<sup>79</sup>, Phe<sup>82</sup>, Phe<sup>101</sup>) sulla superficie concava generano il core dell'interfaccia che interagisce con i residui idrofobici del '*knuckle epitope*' dell'Attivina (Phe<sup>17</sup>, Ile<sup>30</sup>, Ala<sup>31</sup>, Pro<sup>32</sup>, Pro<sup>88</sup>, Leu<sup>92</sup>, Tyr<sup>94</sup>, Ile<sup>100</sup>). All'interfaccia di *binding* sono osservate interazioni ioniche sia inter che intra-molecolari.

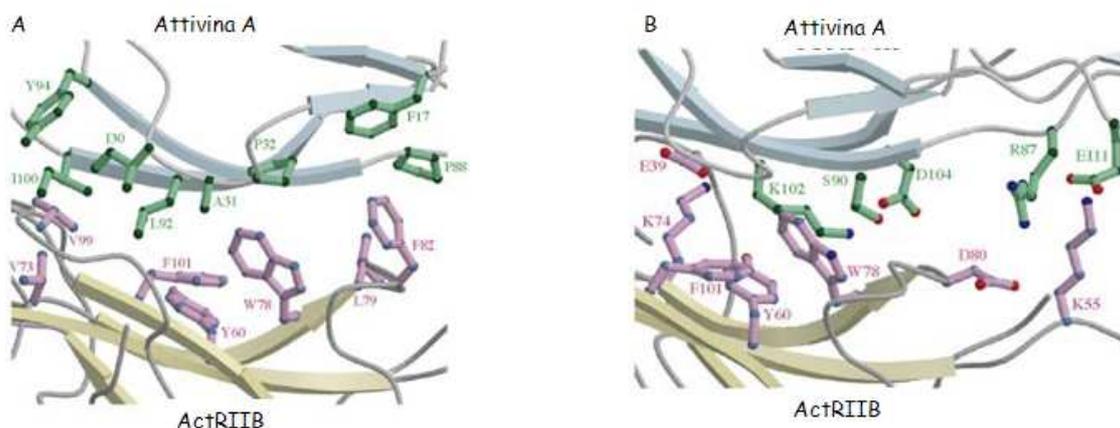


Fig.1.24:(A) Interazioni idrofobiche all'interfaccia ActRIIB-ECD:Attivina A. (B) Interazioni idrofiliche all'interfaccia ActRIIB-ECD:Attivina A

L'analisi delle sequenze suggerisce che MIS e AMHR-II possono usare un'altra interfaccia di *binding*,<sup>29</sup> evidenziando ulteriormente la complessità nel legame del ligando al recettore di tipo II. Queste osservazioni identificano modi differenti di assemblamento di complessi recettoriali e supportano l'idea che i recettori di tipo II giocano un ruolo più importante nella determinazione della specificità per l'assemblaggio del complesso recettoriale.

Oltre alle strutture cristallografiche, la conoscenza sui precisi siti di interazione del recettore di tipo I e di tipo II è stata completata con studi di mutagenesi ed analisi *in silico*. L'approccio attraverso gli studi di mutagenesi ha fornito predizioni molto accurate. Per esempio, le mutazioni nella 'wrist elica' e nella regione concava delle dita della BMP2 conferma pienamente le osservazioni ottenute dalla struttura cristallina BMP2:BMPRIA-ECD (Fig.1.16). Usando lo stesso approccio di mutagenesi è stato localizzato il sito di *binding* per il recettore di tipo II nella regione convessa di BMP2.<sup>26</sup>

Studi *in silico* hanno evidenziato posizioni di anelli aromatici e fenolici come determinanti per il *binding* al recettore di tipo I.<sup>43</sup> Residui di triptofano sono altamente conservati in tutta la superfamiglia dei TGF- $\beta$  nella regione concava delle dita. L'Attivina e le BMP hanno residui amminoacidici identici o simili in questa regione (Fig.1.25).

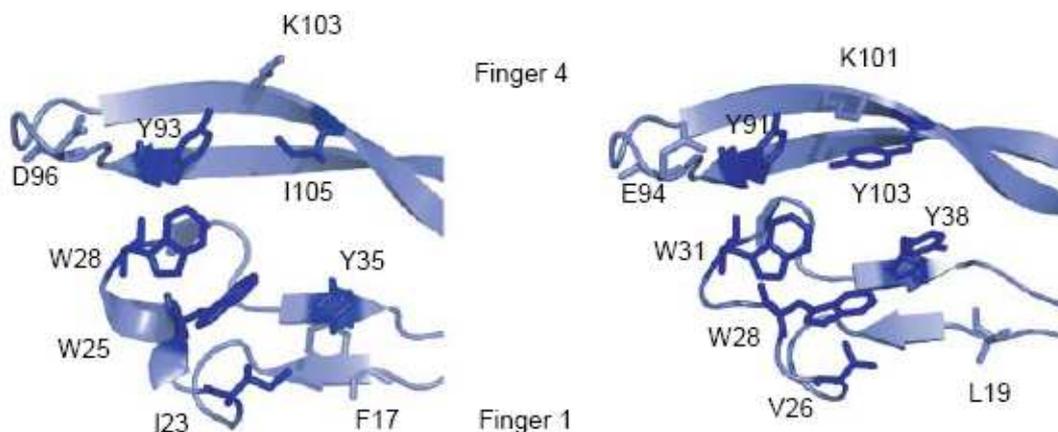


Fig.1.25: Conservazione del sito di *binding* al recettore di tipo I. La regione concava delle dita dell'Attivina (sinistra) e della BMP2 (destra) è mostrata. I residui della BMP2 che interagiscono con BMPRIA-ECD nella struttura cristallina e i corrispondenti residui sull'Attivina determinati da studi di mutagenesi sono indicati. Le catene laterali mostrate in blu scuro rappresentano i residui che contribuiscono al *patch* idrofobico responsabile del *binding* al recettore di tipo I

La mutazione nell'Attivina di uno dei residui di triptofano conservati produce un vero antagonista capace di legare solo il recettore di tipo II. Operando ulteriori mutazioni sui residui amminoacidici che danno luogo ad un *patch* idrofobico sulla superficie convessa dell'Attivina<sup>70</sup> e mutando anche la '*wrist elica*' si osserva interruzione del *signaling*, il che lascia ipotizzare che il recettore di tipo I si leghi contemporaneamente a due monomeri del ligando utilizzando queste due regioni.<sup>71</sup>

### 1.3 Meccanismo di attivazione dei recettori dei TGF- $\beta$

Il legame del fattore dimerico ai domini extracellulari di entrambi i tipi di recettori induce il loro avvicinamento e una conformazione vantaggiosa per i domini chinasi intracellulari dei recettori, facilitando la fosforilazione e la seguente attivazione del recettore di tipo I (Fig.1.26).

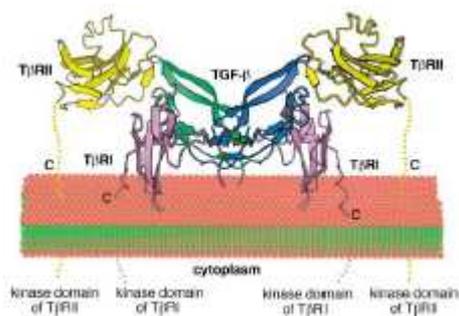


Fig.1.26 Modello strutturale di un complesso recettoriale completo dei TGF- $\beta$ , visto lungo la membrana cellulare. Le estremità C-terminali dei recettori sono rappresentate da linee tratteggiate che entrano nella membrana

Il domino attivo chinasi del recettore di tipo II fosforila i residui multipli di serina e treonina nella sequenza TTSGSGSG della regione citoplasmatica del dominio GS del recettore di tipo I, portando alla sua attivazione (Fig.1.27).

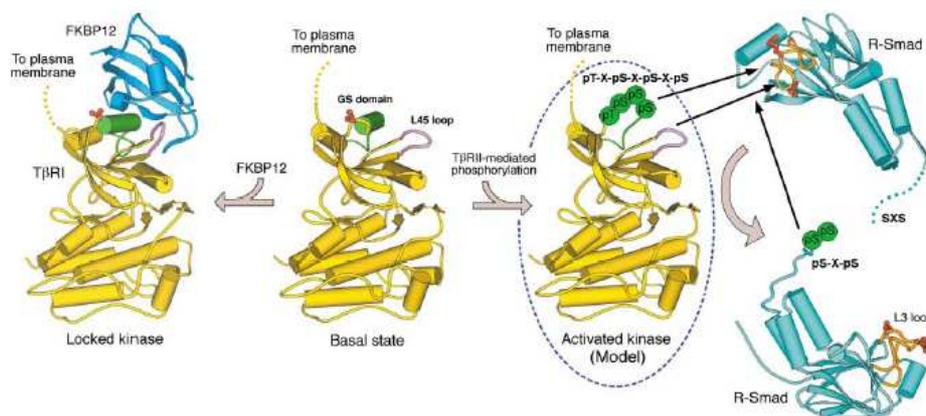


Fig.1.27:Attivazione del recettore chinasi di tipo I e riconoscimento di R-Smad. Nello stato fondamentale (basal state), il recettore di tipo I rimane non fosforilato.<sup>72</sup> Questa conformazione può essere riconosciuta sia da FKBP12<sup>72</sup> per sequestrarla in uno stato inattivo (freccia a sinistra) sia da T $\beta$ R<sub>II</sub> per la fosforilazione nel dominio GS (freccia a destra). Dopo la fosforilazione, T $\beta$ R<sub>I</sub> usa il suo dominio GS e il loop L45 per l'interazione con la tasca di residui basici e il loop L3 di un R-Smad, che viene fosforilato nel suo motivo SXS C-terminale

A causa del suo ruolo cruciale nell' attivazione del recettore, il dominio GS serve come un importante regolatore nel *signaling* dei TGF- $\beta$ . Per esempio, l' immunofillina FKBP12 inibisce il segnale dei TGF- $\beta$  attraverso il legame al dominio GS non fosforilato del recettore di tipo I.<sup>72</sup> Questa interazione blocca il centro catalitico chinasi dei recettori di tipo I in una conformazione non produttiva (Fig.1.27). Dettagli significativi sul meccanismo di attivazione del recettore sono stati rivelati da uno studio nel quale una proteina T $\beta$ R<sub>I</sub> tetrafosforilata

nel dominio GS è stata generata usando una strategia di 'ligation protein' in vitro.<sup>72</sup> La proteina T $\beta$ RI fosforilata lega efficacemente la Smad2 in vitro ed esibisce una forte specificità di fosforilare i residui di serina del dominio C-terminale di Smad2.<sup>72</sup> In aggiunta, il T $\beta$ RI tetra-fosforilato non può essere riconosciuto da FKBP12. Così, la fosforilazione attiva T $\beta$ RI attraverso il cambiamento del dominio GS da sito preferito di *binding* per un inibitore in superficie di *binding* per i substrati R-Smad.

Le Smad sono una famiglia di molecole che rappresentano il bersaglio immediatamente a valle del recettore di tipo I, e che regolano la segnalazione dei TGF- $\beta$ . Nelle cellule di mammifero sono state finora identificate otto molecole Smad,<sup>73</sup> costituite da tre classi funzionali: le Smad recettore-regolante (R-Smad), le Smad Co-mediatori (Co-Smad), e le inibitori Smad (I-Smad). Il termine Smad deriva dall'unione dei nomi dei geni correlati (Mad, gene mother against decapentaplegic, ed il gene Sma), identificati rispettivamente in *Drosophila* ed in *Caenorhabditis elegans*.<sup>74,75</sup>

Il recettore di tipo I attivato trasmette il segnale tramite le R-Smad (Smad 1, 2, 3, 5 e 8). Le Smad 1, 5 ed 8 sembrano regolare la segnalazione delle BMP.<sup>68,73</sup> Le Smad 2 e 3 regolano la segnalazione dei TGF- $\beta$  e dell' Attivina.<sup>68,76</sup> Il recettore di tipo I interagisce direttamente con le Smad e le fosforila nella sequenza di consenso conservata SXS.<sup>68</sup> Le R-Smad sono presentate al recettore di tipo I attraverso le proteine scaffold come SARA (ancora SMAD per l'attivazione del recettore), che facilitano il loro riconoscimento e la fosforilazione.<sup>77</sup> Una volta che le R-Smad sono fosforilate, si dissociano da SARA e interagiscono con Smad 4 (Co-Smad), determinando la traslocazione del complesso nel nucleo, dove interagiscono con altri regolatori trascrizionali per reprimere o attivare l'espressione dei geni (Fig.1.28).

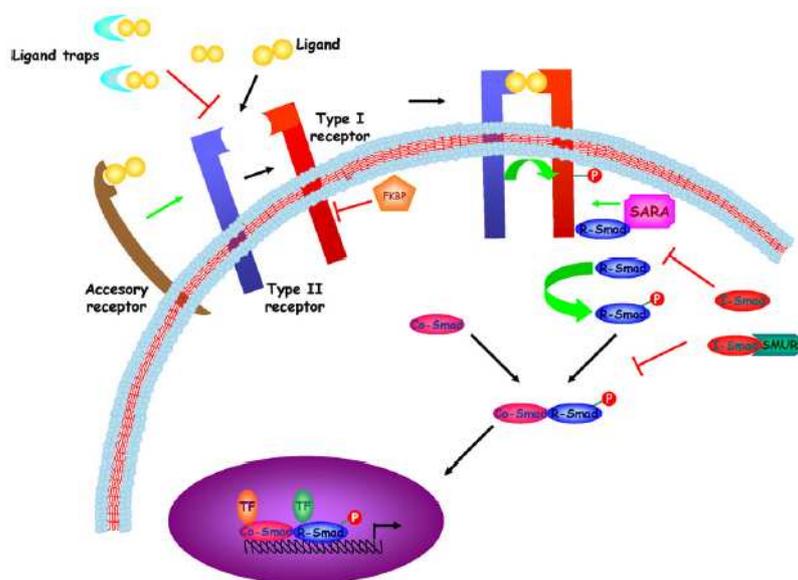


Fig.1.28: *Signaling* della superfamiglia TGF- $\beta$  <sup>78</sup>

Il meccanismo inibitorio citoplasmatico coinvolge le proteine appartenenti alle proteine Smad conosciute come I-Smad (Smad6 e 7); queste molecole inibiscono il segnale indotto dal complesso R-Smad/Co-Smad, interferendo nel *binding* di R-Smad ai recettori di tipo I<sup>79,80</sup> o competendo nel *binding* a Co-Smad.<sup>81-84</sup>

#### 1.4 Regolazione dell'attivazione dei recettori dei TGF- $\beta$

Gli *step* principali dell'attivazione del recettore sono fortemente regolati. L'accesso dei ligandi TGF- $\beta$  ai loro recettori è controllato da due classi di molecole con funzioni opposte (Fig.1.14). Una classe comprende un gruppo di proteine solubili che agiscono come trappole che legano il ligando, sequestrandolo e bloccando il suo accesso ai recettori di membrana. Esse includono le pro regioni dei precursori dei TGF- $\beta$ , che dopo la scissione proteolitica rimangono non covalentemente legate al dominio bioattivo come un 'polipeptide latente-associato' (LAP); le piccole decorine e le proteine  $\alpha$ 2-macroglobuline, che si legano ai TGF- $\beta$  liberi; la follistatina, che lega le Attivine e le BMP; e tre distinte famiglie di proteine - Noggin, Chordin/SOG, e DAN/Cerberus - i cui membri legano le BMP. La risoluzione della struttura cristallina del complesso Noggin-BMP7 (Fig.1.15) ha rivelato per la prima volta come uno di questi fattori lega la sua molecola target, impedendo il contatto tra il ligando e i recettori di membrana.<sup>66</sup> Noggin è un regolatore cruciale dell'attività delle BMP durante il modellamento dorsale-ventrale dei vertebrati, l'induzione e differenziazione dei neuroni, la scheletrogenesi e la formazione delle articolazioni.<sup>85,86</sup> Noggin agisce come antagonista di molte BMP, inclusa la BMP7. La struttura del complesso Noggin-BMP7 rivela che Noggin inibisce la BMP7 bloccando le superfici che sono richieste per l'interazione con i recettori di tipo I e di tipo II della BMP (Fig.1.15).<sup>66</sup> La parte N-terminale di ciascun monomero di Noggin adotta una conformazione estesa e si avvolge intorno ad un monomero della BMP7, occupando direttamente i siti di contatto dei recettori. Un residuo di prolina in questa parte, Pro<sup>35</sup>, riempie la tasca idrofobica della BMP7 che normalmente ospita la Phe<sup>85</sup> del recettore di tipo I. E' interessante notare che il *core* della struttura del monomero di Noggin contiene un nodo di cisteine. Questo, e la configurazione strutturale suggeriscono che il ligando e l'antagonista si sono evoluti da un progenitore comune.

L'altra classe di molecole che controlla l'accesso dei ligandi ai recettori include le proteine ancorate alla membrana che agiscono come recettori accessori, o corecettori, promuovendo il legame del ligando ai recettori. Il recettore di tipo III dei TGF- $\beta$  (un proteoglicano ancorato alla membrana) interviene nel mediare il *binding* dei TGF- $\beta$  al recettore di tipo II.<sup>68,87</sup> I

betaglicani, invece, legano l' inibina e facilitano il suo accesso ai recettori dell'Attivina, anticipando il *binding* dell'Attivina e bloccando la sua attività biologica.<sup>88</sup> Studi genetici e biochimici suggeriscono che una proteina, relazionata ai betaglicani, l'endogлина facilita il *binding* dei membri della famiglia dei TGF- $\beta$  al recettore di tipo I (ALK1) nelle cellule endoteliali, e questo ruolo è critico per l'omeostasi vascolare.<sup>3,89</sup>

La proteina BAMBI ( BMP and Activin receptor membrane bound inhibitor, conosciuta anche come Nma), rappresenta un tipo molto differente di regolatore negativo dell'attivazione del recettore.<sup>90</sup> BAMBI ha le caratteristiche strutturali di un recettore di richiamo (decoy), con un dominio extracellulare e una piccola regione citoplasmatica con sequenza simile ai recettori di tipo I. BAMBI compete con il recettore di tipo I per l'incorporazione nei complessi recettoriali indotti dal ligando, inibendo l'attivazione del recettore.

La proteina cripto ed altri membri della famiglia degli EGF-CFC, prodotti sia come fattori secreti sia come componenti di superficie cellulare, sono i co-recettori richiesti per il *signaling* di Nodal, Vg1, GDF1 e GDF3.<sup>91-94</sup> I ligandi che richiedono le proteine EGF-CFC come co-recettori segnalano attraverso i recettori dell'attivina ActRII/ActRIIB e ALK4.

### 1.5 La proteina Cripto

La proteina Cripto è il capostipite dei fattori di crescita cellulare chiamati EGF-CFC.<sup>95</sup> Nel 1978, anno in cui questa proteina fu clonata, le venne attribuito il nome 'Cripto' in quanto non mostrava alcuna relazione con proteine conosciute.<sup>96</sup> Solo alcuni anni più tardi apparve evidente che presentava le caratteristiche funzionali dei fattori di crescita cellulare.<sup>93</sup>

Cripto è una proteina extracellulare trovata nei vertebrati, come ad esempio nell'uomo, nei polli, nei pesci e nelle rane. Il suo nome varia in base alla specie d'appartenenza: Cripto e Criptic nell'uomo e nei topi, FRL-1 nelle rane, Oep (one-eyed pinhead) nei pesci e Cripto nei polli.<sup>97</sup> Le varianti di 'Cripto' presentano una bassa omologia, stimabile tra il 20 ed il 30%.

Queste proteine hanno delle caratteristiche peculiari, infatti presentano (Fig.1.29):

- a. un peptide segnale N-terminale;
- b. una parte centrale formata da due domini ricchi di cisteine, noti come EGF-like (E

p

idermal G

g

rowth F

f

actor-like) e CFC (C

ri

pto/FRL-1/C

ry

ptic), di circa 40 residui, ciascuno con tre ponti disolfuro;<sup>95</sup>
- c. una regione idrofobica C-terminale che è legata alla membrana cellulare attraverso un glicosil-fosfatil-inositolo (GPI).

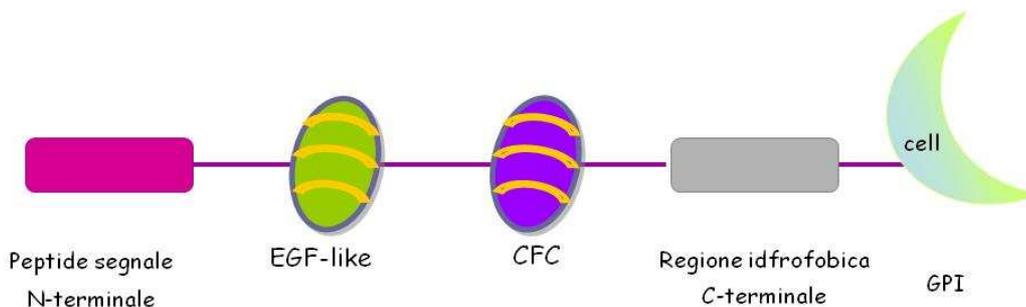


Fig.1.29: Schema rappresentativo della proteina Cripto

Le proteine EGF-CFC sono composte da circa 200 amminoacidi e il peso molecolare della regione conservata tra le varie specie animali è compreso tra 18 e 21 kDa (Fig.1.30).

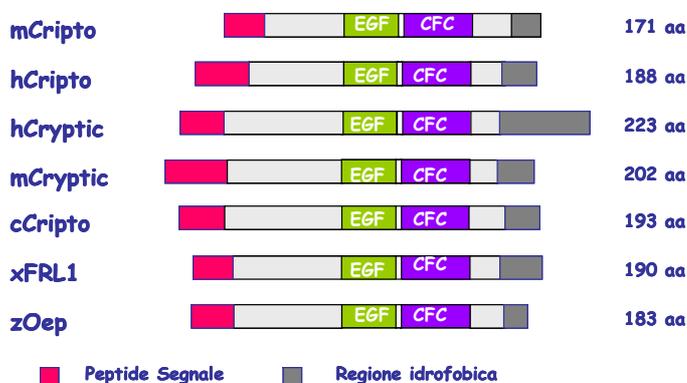


Fig.1.30: Schema della famiglia di proteine EGF-CFC

### 1.5.1 Le funzioni biologiche di Cripto

La proteina Cripto è coinvolta in vari processi fisiologici come l'embriogenesi, la morfogenesi e la differenziazione delle cellule staminali embrionali (ES).<sup>98,99</sup> In particolare nell'embriogenesi è coinvolta in processi importanti come il modellamento dell'asse anteriore-posteriore e nella morfogenesi della ghiandola mammaria e del cuore. Inoltre regola la differenziazione delle cellule staminali embrionali in neuroni e in cardiomiociti.<sup>100</sup> La proteina Cripto risulta però assente nei tessuti normali adulti.

Studi genetici molecolari sui pesci e sui topi hanno rivelato che le proteine EGF-CFC sono essenziali nelle prime fasi dello sviluppo embrionale, nello specificare la formazione degli assi anteriore-posteriore e destro-sinistro del corpo, e l'accumulo dei primi agglomerati cellulari durante la gastrulazione.

Il *pathway* principale identificato per il *signaling* delle proteine Cripto, che supporta la sopravvivenza delle cellule, coinvolge alcune proteine che appartengono alla superfamiglia dei TGF- $\beta$ , includendo Nodal,<sup>99-103</sup> Vg1/GDF1,<sup>94</sup> GDF3,<sup>91</sup> Attivina A,<sup>104</sup> e Attivina B.<sup>105</sup> In

particolare, le proteine Cripto sono state identificate come co-recettori della proteina Nodal, GDF1 e GDF3. Infatti il legame di Nodal/GDF1/GDF3 al complesso recettoriale dell'Attivina formato dal recettore transmembrana serina-treonina chinasi di tipo I, ALK4, e dal recettore transmembrana serina-treonina chinasi di tipo II, ActRIIB, avviene solo in presenza della proteina Cripto.

Cripto, che è legato alla membrana cellulare, è in grado di reclutare Nodal/GDF1/GDF3 e di renderli conformazionalmente adatti ad intascarsi nel complesso recettoriale dell'Attivina, dando così inizio alla trasmissione dei segnali di questi TGF- $\beta$  (Fig. 1.31).

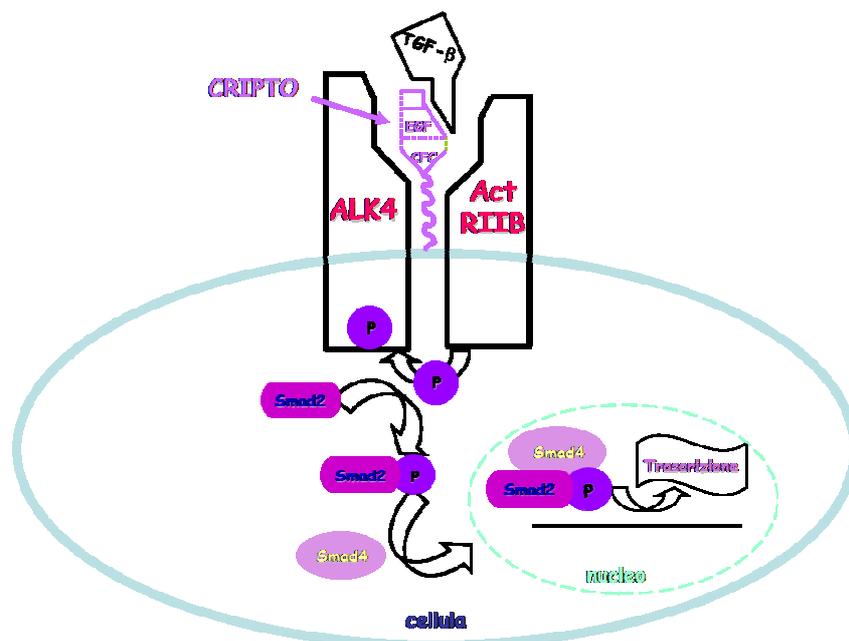


Fig.1.31: Schema rappresentativo della trasmissione dei segnali di Nodal

Il primo *step* del *signaling* intracellulare è la fosforilazione di residui di serina e treonina presenti nella sequenza TTSGSGSG del dominio GS citoplasmatico di ALK4.<sup>3,99</sup> A questo punto le chinasi attivate richiamano uno specifico fattore Smad e fosforilano il motivo SSXS C-terminale di questo,<sup>106,107</sup> consentendo l'accumulo di proteine Smad nel nucleo dove regolano molteplici segnali trascrizionali.<sup>68</sup>

### 1.5.2 Cripto e la crescita tumorale

Il crescente interesse scientifico destato dalla proteina Cripto, oltre al ruolo fondamentale che svolge nell'embriogenesi, è dovuto principalmente al suo coinvolgimento nei processi tumorali. Infatti, mentre nei tessuti adulti normali questa proteina risulta assente, essa è invece altamente sovraespressa in tumori solidi e particolarmente per il 75-80% dei casi di

carcinoma alla mammella, al colon e al polmone e per il 50-60% dei casi di carcinoma ai testicoli, alle ovaie, al pancreas e allo stomaco.<sup>105</sup>

Inoltre, il livello d'espressione di Cripto aumenta con il grado di avanzamento della displasia.<sup>105</sup>

In vitro, la proteina Cripto presenta molte proprietà oncogeniche, come la trasformazione di cellule immortalizzate, l'induzione della migrazione delle cellule e la stimolazione di morfogenesi ramificata.<sup>108</sup>

E' stato proposto che Cripto può favorire la crescita tumorale bloccando l'attività onco-soppressiva dell'Attivina B, un altro membro della famiglia dei fattori di crescita cellulare TGF- $\beta$ .<sup>95,105</sup>

L'Attivina è sintetizzata come un omo- o etero-dimero di subunità  $\beta$  risultando in una varietà di possibili specie molecolari. Le subunità  $\beta$  sono simili in tutte le strutture della superfamiglia dei fattori di crescita e differenziazione TGF- $\beta$ . La nomenclatura dell'Attivina riflette il tipo di  $\beta$  subunità nell'eterodimero: Attivina A ( $\beta_A$ - $\beta_A$ ); Attivina B ( $\beta_B$ - $\beta_B$ ); Attivina AB ( $\beta_A$ - $\beta_B$ ).<sup>109</sup>

L'Attivina è stata classificata come una citochina<sup>♥</sup>, come fattore di crescita e di differenziazione e come ormone endocrino (ormoni che, dopo essere stati sintetizzati, vengono riversati nel sangue e da questo distribuiti anche nelle zone corporee lontane dalle ghiandole endocrine).<sup>109</sup> E' prodotta da una grande varietà di tessuti inclusi la ghiandola pituitaria, il cervello e il midollo osseo.

L'Attivina stimola la produzione dell'ormone FSH (**F**ollicle **S**timulating **H**ormone) da parte dell'ipofisi, stimola la differenziazione cellulare degli eritrociti, regola i cicli cellulari dei geni e la morte di cellule del fegato e inoltre agisce come fattore di crescita di cellule nervose.<sup>109</sup>

Infine l'Attivina, come tutte proteine che appartengono alla famiglia TGF- $\beta$ , ha un'attività onco-soppressiva nelle cellule della mammella, del fegato e dei reni.<sup>105</sup> La normale attività biologica dell'Attivina si sviluppa mediante il legame al complesso recettoriale ActRII/ALK4 con successiva trasmissione dei segnali. In tumori alla prostata, l'espressione ridotta di ActRII (recettore dell'Attivina di tipo II) è collegata con la crescita del tumore, e in alcuni tumori al pancreas, sono state ritrovate mutazioni di ALK4 (recettore dimerico di tipo I).<sup>110,111</sup>

---

<sup>♥</sup> Le **citochine** sono una famiglia di proteine che svolgono un'importante funzione di mediazione nella comunicazione delle cellule. Le citochine intervengono nella comunicazione sia fra cellule contigue all'interno dello stesso tessuto, sia fra cellule di tessuti differenti e distanti. Queste proteine sono state studiate anche per altre funzioni che esercitano nell'organismo, in particolare per il ruolo di stimolatrici della crescita cellulare e di inibitrici della replicazione virale.

Le citochine vengono solitamente sintetizzate e rilasciate dalle cellule in piccole quantità, generalmente in modo controllato e in risposta a stimoli locali. Quindi, esse agiscono in modo specifico su cellule-bersaglio, che le riconoscono mediante particolari recettori poste sulla superficie esterna della membrana plasmatica. Il riconoscimento tra citochina e recettore attiva, all'interno della cellula, reazioni chimiche che costituiscono la risposta cellulare al messaggio portato dalla citochina.

Quindi alterazioni dell'equilibrio di questo complesso recettoriale influenzano l'attività onco-soppressiva dell'Attivina e possono quindi costituire un fattore di innesco della crescita del tumore. Pertanto, è stato proposto da due diversi gruppi di ricerca che la proteina Cripto può favorire la crescita tumorale interferendo nel regolare meccanismo biologico dell'Attivina.<sup>104,105</sup>

A livello molecolare, Cripto inibisce la formazione del complesso ALK4/ActRIIB/Attivina necessario al *signaling* di quest'ultima.

Gli studi condotti, tuttavia, discordano su un punto chiave, ossia se l'Attivina si lega direttamente a Cripto o se la loro interazione è mediata dai recettori dell'Attivina.

E' stato osservato<sup>104</sup> che Cripto lega l'Attivina A in presenza di ActRIIB, concludendo che Cripto inibisce il legame dell'Attivina A con ALK4 attraverso la formazione di un complesso inattivo Attivina/ActRIIB/Cripto (Fig.1.32). E' stato inoltre dimostrato che mutanti con delezioni nel dominio EGF-like non sono in grado di legare l'Attivina A, concludendo che Attivina A e Nodal legano Cripto nello stesso sito.

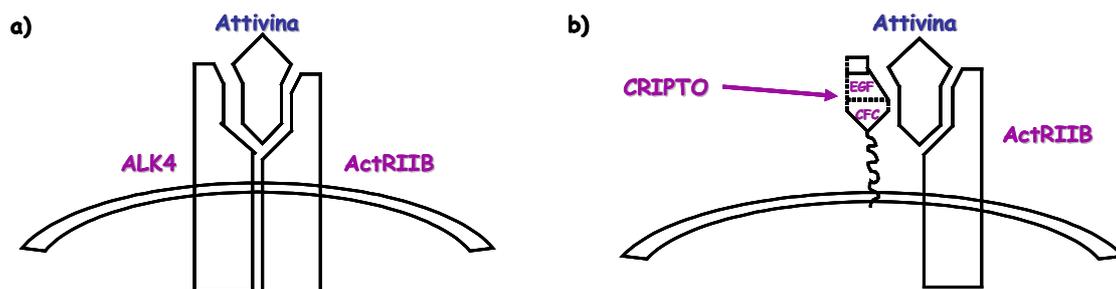


Fig.1.32: Modello schematico proposto da Gray et al., 2003<sup>104</sup>: a) Regolare *binding* dell'Attivina nel complesso recettoriale; b) Interazione della proteina Cripto con l'Attivina

Invece, Adkins e colleghi<sup>105</sup> hanno dimostrato che esiste un'interazione diretta tra la forma solubile di Cripto e l'Attivina B *in vitro*, senza riuscire però a rilevare un'analogha interazione tra Cripto e l'Attivina A.

Il meccanismo proposto da Adkins e colleghi è il seguente: l'Attivina B per svolgere la sua regolare attività biologica necessita del legame con il suo complesso recettoriale, formato da ActRII e ALK4 (Fig.1.33a); Cripto, però, è capace di legare, attraverso il suo dominio CFC, ALK4 e l'Attivina B interferendo nel regolare meccanismo di quest'ultima (Fig.1.33b).<sup>95,105</sup>

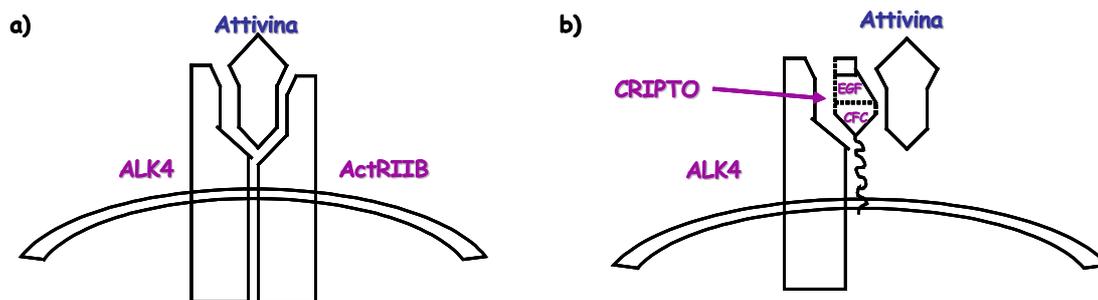


Fig.1.33: Modello schematico proposto da Adkins et al., 2003<sup>105</sup>: a) Regolare *binding* dell'Attivina nel complesso recettoriale; b) Interazione della proteina Cripto con l'Attivina

Inoltre, questi autori hanno osservato che proteine Cripto mutate al sito O-fucosilato del dominio EGF-like non interagiscono con Nodal, ma sono ancora in grado di legare l'Attivina B, stabilendo così che Attivina B e Nodal legano regioni diverse di Cripto. Poiché anticorpi anti-CFC sono in grado di bloccare il legame tra Attivina B e Cripto gli autori hanno dedotto che il dominio CFC è critico per l'interazione con l'Attivina B.

Il contributo di questi studi, nonostante le divergenze, ha segnato il punto di partenza per la comprensione della funzione di Cripto nello sviluppo dei tumori.

E' ipotizzabile, quindi, che Cripto possa promuovere la crescita tumorale sia inibendo l'attività onco-soppressiva dell'Attivina sia favorendo il *signaling* di Nodal o di altri fattori di crescita TGF- $\beta$ .

Ulteriori indagini sono quindi necessarie per giungere alla completa descrizione, su base molecolare, del meccanismo d'azione di Cripto. Lo studio di questo tema è di particolare interesse, dal momento che Cripto si profila ad oggi come uno dei più importanti target per farmaci antitumorali.

### 1.5.3 Domini funzionali di Cripto

La struttura tridimensionale della proteina Cripto umana o di altri membri della famiglia EGF-CFC non è nota, nonostante il notevole e crescente interesse verso la comprensione della loro funzione biologica.

Le proteine della famiglia EGF-CFC contengono due domini conservati: una variante dei fattori EGF (chiamato EGF-like) e un dominio caratteristico di questa famiglia proteica, ricco di cisteine, chiamato CFC. Il dominio EGF-like, così definito per la somiglianza ai fattori EGF, mostra un'omologia di sequenza pari al 60-70% tra le varianti di Cripto nei diversi vertebrati. L'omologia di sequenza tra i vari domini CFC varia tra il 35 ed il 48%. Un'omologia di sequenza più alta (80%) si osserva per il dominio CFC nella proteina Cripto umana e da topo.<sup>97</sup> La

proteina Cripto umana e quella da topo sono formate da 188 e 171 amminoacidi, rispettivamente. E' interessante notare l'elevata omologia di sequenza per i domini EGF-like (Fig.1.34: evidenziato in verde) e CFC (Fig.1.34: evidenziato in viola), presenti nella parte centrale della proteina.

```

mCripto  MGY-----FSSVLLVAISSAFE----FGVAGRD-----LATRDNSTWDQKEPAVRDRSFQFVPSV-GIQNSKSLNKT 65
hCripto  MDC-----RKMARFSYSVIWIMAIKVF-----LGLVAGLGHQEFARPSRGLAFRDDSTWPGEEPATIPRSSQRVPPM--GIQHSKELNRT 81

mCripto  CCLNGGTCILGSFCACPPSFYGRNCEHDVRCGSSILHGTWLPKKSLRCWHEQLHLPQTFLPGCDGHVMD QDLKASRTPOQTSPV---TTTFMLA 161
hCripto  CCLNGGTCMLGSFCACPPSFYGRNCEHDVRCENCGSVPHDTWLPKKSLRCWHEQLRCFPQAFLPGCDGHVMD EHLVASRTPPELPSAR--TTTFMLV 178

mCripto  GAQF-----LDMKV 171
hCripto  GQLS-----IQ-SYY 188
    
```

Fig.1.34: Allineamento delle sequenze amminoacidiche della proteina Cripto da topo (mCripto) e umana (hCripto). I domini EGF-like sono evidenziati in verde, mentre i domini CFC in viola

Questi due domini sono formati da circa 40 amminoacidi e sono ricchi di cisteine con tre ponti disolfuro ciascuno. Lo schema di ponti disolfuro, identico a quello riscontrato per la proteina umana, è rappresentato in Fig.1.35. Per il dominio EGF-Like lo schema è: 1-3, 2-4,5-6, mentre per il dominio CFC è: 1-4, 2-6,3-5.



Fig.1.35: Schema rappresentativo dei ponti disolfuro presenti nei domini EGF-like e CFC

Nella proteina Cripto umana i primi trenta costituiscono il peptide segnale. La proteina matura inizia, dal residuo di Leu<sup>31</sup>. I domini EGF-like (K<sup>76</sup>-R<sup>111</sup>) e CFC (K<sup>112</sup>-D<sup>150</sup>) sono colorati in verde e in viola, rispettivamente. Sono inoltre evidenziati in Fig.2.1 i siti di glicosilazione: Asn<sup>79</sup> N-glicosilato; Ser<sup>40</sup> e Ser<sup>161</sup> O-glicosilate; Thr<sup>88</sup> O-fucosilata. La Ser<sup>161</sup> costituisce il sito di aggancio del gruppo GPI e la sequenza segnale per questo ancoraggio è sottolineata (M<sup>170</sup>-Y<sup>188</sup>).

```

1   MDCRKMARFS YSVIWI MAIS KVFELGLVAG LGHQEFARPS RGYLAFRDDS
51  IWPQEEPAIR PRSSQRVPPM GIQHSKELNR TCCLNGGTCM LGSFCACPPS
101 FYGRNCEHDV RKENCGSVPH DTWLPKKCSL CKCWHGQLRC FPQAFLPGCD
151 GHVMDEHLVA SRTPELPPSA RTTTFMLVGI CLSIQSYY

```

Fig.1.36: Struttura primaria di hCripto. La sequenza segnale è riportata in corsivo. I domini EGF-like e CFC sono riportati in verde e viola, rispettivamente. I siti di glicosilazione sono in grassetto fucsia. I residui segnale per l'ancoraggio di GPI sono sottolineati

Sulla base di studi di mutagenesi è stato proposto che i due domini di Cripto sono strutturalmente distinti e possono interagire in modo diverso: Cripto lega Nodal con il dominio EGF-like<sup>112</sup> e ALK4 con il dominio CFC.<sup>103,113,114</sup>

Alcune mutazioni puntiformi nel dominio EGF-like sono cruciali nel *signaling* della proteina Cripto.<sup>112</sup> La sostituzione della Thr<sup>72</sup> (Fig.1.34) della proteina Cripto topo con un Ala o Ser porta a una diminuzione del *signaling*. Il residuo di Thr della proteina Cripto è O-fucosilato e fa parte di una sequenza *consensus*, C-X<sub>4-5</sub>-(T/S)-C, osservabile tra la seconda e la terza cisteina del dominio EGF-like. Il mutante T72S della proteina Cripto è O-fucosilato ma ha una attività bassa, simile al mutante T72A, nei saggi di *signaling* della proteina Nodal. Il fatto che i due mutanti hanno attività simile suggerisce che il gruppo metile della Thr<sup>72</sup> gioca un ruolo cruciale nell'abilità di Cripto di supportare il *signaling* di Nodal. La Thr è un residuo conservato nel dominio EGF-like per tutti i membri della famiglia EGF-CFC.<sup>99</sup> Anche il mutante T88A della proteina Cripto umana esibisce un livello di attività basso nei saggi di *signaling* della proteina Nodal.<sup>98</sup>

Il residuo di Thr non è mai sostituito da una Ser (altro sito possibile di ancoraggio del fucosio), consistente con l'idea che l'O-fucosio non è essenziale per la funzione biologica del dominio EGF-like.<sup>112</sup> Un'altra mutazione puntiforme del dominio EGF-like da topo di un residuo che è immediatamente adiacente alla Thr<sup>72</sup>, G71N, dà luogo ad una perdita completa dell'attività di Cripto. Un altro residuo cruciale è la Phe<sup>78</sup>, la mutazione di questo residuo con un'Ala inattiva completamente Cripto.

Mutazioni puntiformi sul dominio CFC influenzano, invece, l'interazione della proteina Cripto con il recettore ALK4. Saggi ELISA effettuati con i domini sintetici hanno mostrato che la sostituzione del Trp<sup>107</sup> con un'alanina nel dominio CFC da topo (corrispondente al Trp<sup>123</sup> nella sequenza umana) abolisce totalmente il *binding* ad ALK4, mentre la sostituzione dell'His<sup>104</sup> (corrispondente a His<sup>120</sup> nella variante umana) abolisce solo parzialmente il riconoscimento al recettore.<sup>114</sup> Inoltre la mutazione Pro125Leu nel dominio CFC è stata correlata con anomalie dello sviluppo in feti umani.<sup>115</sup>

In letteratura sono disponibili informazioni strutturali sui singoli domini. Del dominio CFC e di un suo mutante (CFC-Mut) puntiforme (W107A) è stata determinata la struttura tridimensionale mediante NMR della proteina Cripto da topo.<sup>116</sup> Del dominio EGF-like<sup>117</sup> e del dominio CFC umano<sup>118</sup> sono stati costruiti modelli per omologia.

I modelli molecolari di CFC e CFC-Mut sono stati determinati mediante NMR a pH acidi e a pH quasi neutro (pdbcode:2j5h).<sup>116</sup> Tutte le strutture determinate mostrano un *folding* globalmente esteso, esse appaiono più strutturate a pH fisiologico rispetto al pH acido. Il *folding* è globalmente esteso in tutti i casi con tre *strand* collegati dai ponti disolfuro Cys<sup>99</sup>-Cys<sup>117</sup>, Cys<sup>115</sup>-Cys<sup>124</sup> e Cys<sup>112</sup>-Cys<sup>133</sup>. Non si osserva una regolarità nella conformazione dei tre ponti disolfuro che mostrano in tutti i casi un' uguale distribuzione tra conformazioni "*left-handed*" ( $\chi_3 = -90^\circ \pm 30^\circ$ ) e "*right-handed*" ( $\chi_3 = +90^\circ \pm 30^\circ$ ). La presenza di conformazioni multiple dei ponti disolfuro, d'altro canto, non altera sostanzialmente la struttura del backbone.<sup>119</sup>

Ciascun modello esibisce una forma complessivamente ellissoidale (Fig.1.37) che si rivela più compatta per CFC a pH 3 ( $24 \times 20 \times 18 \text{ \AA}^3$ ) rispetto a quella prolata di CFC a pH 6 ( $32 \times 12 \times 17 \text{ \AA}^3$ ) e CFC-Mut ad entrambi i valori di pH ( $30 \times 14 \times 16 \text{ \AA}^3$  a pH 3 e  $30 \times 14 \times 10 \text{ \AA}^3$  a pH 6).

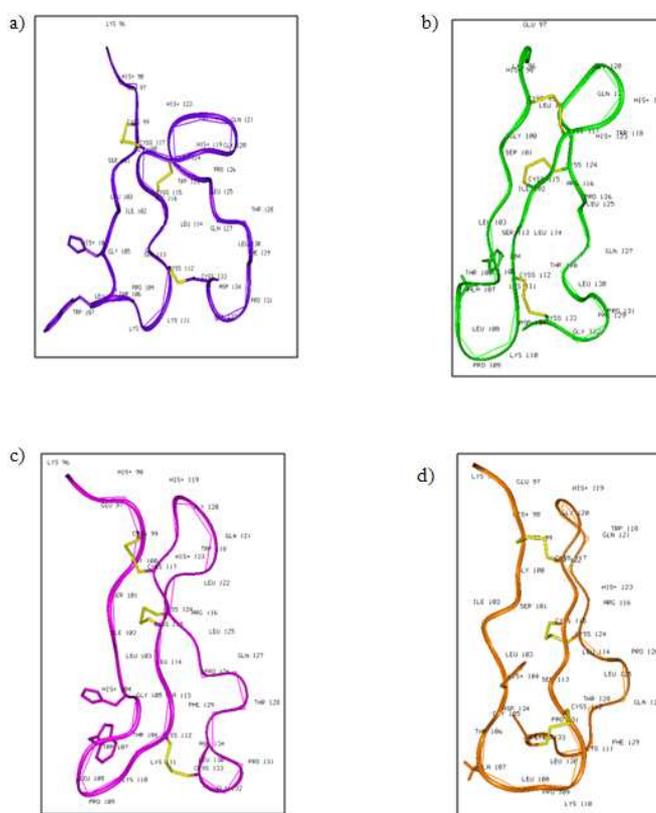


Fig.1.37: Rappresentazione con *ribbon* dei modelli molecolari medi: (a) CFC pH 3 in viola; (b) CFC-Mut pH 3 in verde; (c) CFC pH 6 in fucsia; (d) CFC-Mut pH 6 in arancione. Le catene laterali dell'His104 e del Trp107 sono mostrate (pdbcode:2j5h)

Dal punto di vista delle implicazioni funzionali, i modelli sperimentali di CFC rivelano che i residui ritenuti responsabili del *binding* con il recettore ALK4, His<sup>104</sup> e Trp<sup>107</sup>, non sono mai localizzati nel core proteico, ma si ritrovano posizionati tra la fine del primo *strand* e l'inizio del primo *loop*, con le catene laterali esposte esternamente. Opposto al sito di *binding* si ritrova un *patch* idrofobico (Trp<sup>118</sup>, Leu<sup>122</sup>, Leu<sup>125</sup>, Pro<sup>126</sup>, Phe<sup>129</sup>, Leu<sup>130</sup>) (Fig.1.38). Questo *patch* è presente in più del 60% delle strutture finali selezionate, suggerendo che può avere un ruolo nel riconoscimento di altre superfici idrofobiche derivate dal dominio adiacente EGF-like o da altri partner proteici solubili. Nei modelli molecolari del dominio CFC sono stati identificati anche *patch* idrofilici sopra (Lys<sup>96</sup>, His<sup>98</sup>, His<sup>119</sup>) e sotto (Lys<sup>110</sup>, Lys<sup>111</sup>) il sito di *binding*. Sorprendentemente, un simile assemblaggio di residui idrofilici/idrofobici non è osservato nelle superfici di Connolly di CFC-Mut (a pH 3 e a pH 6), suggerendo un possibile coinvolgimento di questi residui nel modulare il *binding* ad ALK4 (Fig.1.38).

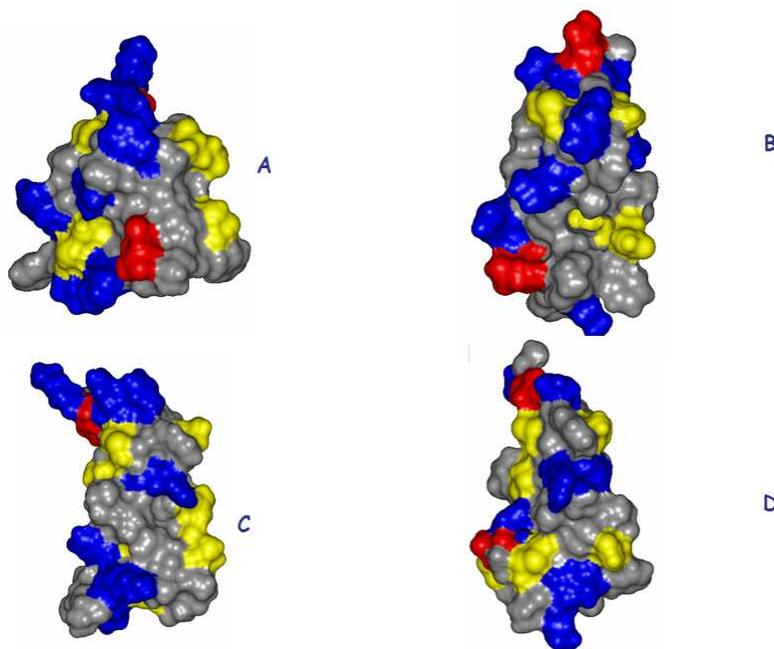


Fig.1.38: Confronto delle superfici di Connolly dei modelli molecolari di CFC e CFC-Mut. (A) CFC pH 3; (B) CFC-Mut pH 3; (C) CFC pH 6; (D) CFC-Mut pH 6. I residui sono colorati per tipo: acidi in rosso, basici in blu, polari in giallo e idrofobici in grigio

Per la costruzione del modello per omologia del dominio EGF-like umano è stato usato come *template* la struttura NMR del dominio EGF del fattore IX umano,<sup>120</sup> (pdbcode:1ixa), che ha un'omologia di sequenza con EGF-like del 54%. Il dominio EGF-like (Fig.1.39) ottenuto è formato da due  $\beta$ -*hairpin* di sei residui nelle regioni N- e C-terminale; ed è caratterizzato,

inoltre, da un *patch* idrofobico e una distribuzione di residui carichi che sono centrati sul sito opposto dei  $\beta$ -hairpin.

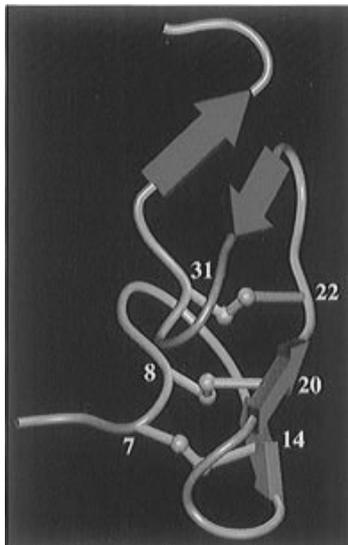


Fig.1.39: Modello per omologia del EGF-like umano<sup>117</sup>

Il modello molecolare<sup>118</sup> del dominio CFC umano (Fig.1.40) è stato ottenuto sulla base della struttura NMR di PMP-C<sup>121</sup> (un inibitore serina proteasi) (pdbcode:1pmc). Il modello molecolare rivela un'alta concentrazione di residui idrofobici (Trp<sup>134</sup>, Leu<sup>138</sup>, Phe<sup>141</sup>, Pro<sup>142</sup>, Phe<sup>145</sup> e Leu<sup>146</sup>) in una regione opposta a quella definita dai residui di His e Trp<sup>♥</sup> ritenuti coinvolti nel legame di Cripto con il recettore ALK4,<sup>103</sup> analogamente a quanto osservato nei modelli molecolari del dominio CFC da topo, ottenuti mediante NMR.<sup>116</sup>

Foley e colleghi<sup>118</sup> hanno costruito un modello molecolare del dominio EGF-like umano, basandosi sulla struttura NMR del dominio EGF (pdbcode:1epi) da topo, utilizzando lo stesso allineamento dei residui di cisteina descritto da Lohmeyer e colleghi.<sup>117</sup> Sono stati costruiti due modelli teorici dell'intera proteina Cripto unendo i due modelli per omologia ottenuti per i due domini. Nel primo modello (Fig.1.40A) i domini EGF-like e CFC sono disposti in un'arrangiamento prevalentemente esteso, mentre il secondo modello (Fig.1.40B) ipotizza una struttura globulare, nella quale i due domini presentano molti contatti non-covalenti.

<sup>♥</sup>Questi residui corrispondono a His<sup>120</sup> e Trp<sup>123</sup> ovvero His<sup>104</sup> e Trp<sup>107</sup> della proteina Cripto umana e da topo, rispettivamente.

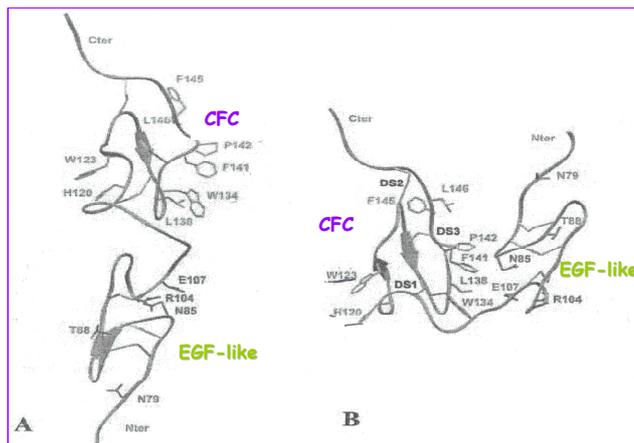


Fig.1.40: Modelli molecolari per i domini EGF-CFC della proteina Cripto umana: **A)** arrangiamento esteso; **B)** arrangiamento globulare<sup>118</sup>

### 1.6 Oggetto della tesi

Lo scopo di questo lavoro di dottorato è stato quello di ottenere informazioni strutturali sui diversi fattori proteici che intervengono nel processo di *signaling* Nodal/Cripto.

E' stata effettuata la sintesi in fase solida presso l'Istituto di Biostrutture e Bioimmagine (IBB) del CNR del dominio CFC umano (hCFC). E' stata risolta la struttura 3D del dominio CFC della proteina Cripto da topo e umana attraverso l'analisi NMR in tampone fosfato. Per avere informazioni strutturali in soluzione della proteina umana di Nodal, presso l'IBB sono stati sintetizzati tre frammenti di Nodal, che riproducono tratti di sequenza presumibilmente coinvolti nel legame con i suoi partner proteici. Di questi frammenti è stata fatta un' analisi conformazionale in una miscela TFE/H<sub>2</sub>O 30/70 (v/v) mediante NMR. Sono stati poi ottenuti mediante *homology modelling* i modelli molecolari dei recettori di tipo I, ALK4 e ALK7 e di Nodal con i quali sono stati eseguiti esperimenti di *docking* con le strutture NMR dei rispettivi partner proteici. Dai complessi selezionati è stato possibile formulare una serie di ipotesi sugli elementi strutturali potenzialmente rilevanti ai fini delle mutue interazioni.

## CAPITOLO 2

2.1 Sintesi Chimica
2.2 Esperimenti di dicroismo circolare
2.3 Esperimenti NMR
2.4 Calcolo dei modelli molecolari
2.5 Homology Modelling
2.6 Esperimenti di docking
2.7 Esperimenti di binding SPR

### Metodiche sperimentali

#### 2.1 Sintesi chimica

La tecnica SPPS (Solid Phase Peptide Synthesis) messa a punto nel 1963 ad opera di Bruce Merrifield<sup>122</sup>, è attualmente il metodo più vantaggioso per la sintesi peptidica.

La sintesi di un peptide con tale tecnica richiede in genere cinque passaggi:

- assemblaggio della catena;
- sblocco dalla resina e rimozione dei gruppi protettori in catena laterale;
- purificazione;
- eventuali modifiche chimiche post-sintesi;
- caratterizzazione.

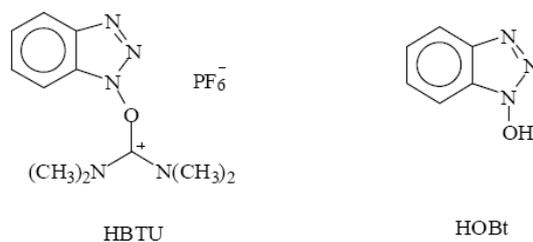
La SPPS si basa sull'addizione sequenziale ad un supporto polimerico insolubile (resina) di residui amminoacidi, protetti sia in catena laterale che sul gruppo  $\alpha$ -ammino,.

Il peptide è assemblato dall'estremità C-terminale a quella N-terminale con il gruppo  $\alpha$ -carbossilico dell'ultimo amminoacido della sequenza ancorato ad un supporto solido tramite un linker: si ha così la crescita della catena polimerica in fase eterogenea. L'ancoraggio ad una matrice insolubile costituisce il vantaggio principale di questa tecnica in quanto permette di condurre l'intera sintesi nello stesso contenitore (reattore); il peptide in crescita, infatti, può essere isolato dopo ogni passaggio dai reagenti in eccesso o da sottoprodotti solubili attraverso semplici lavaggi e filtrazioni. Questo comporta un minor spreco di materiale dovuto ad una minore manipolazione ed una sensibile riduzione dei tempi di lavoro in quanto la procedura può essere automatizzata.

In base alla scelta del linker per ancorare il peptide in crescita al supporto solido si può ottenere, in seguito allo sblocco (*cleavage*), un peptide col gruppo COOH o amidato (CONH<sub>2</sub>). Nel caso si vogliano condurre reazioni che coinvolgono selettivamente il terminale amminico o carbossilico, inoltre, è possibile sbloccare il peptide dalla resina lasciando inalterate le protezioni in catena laterale scegliendo un *linker* opportuno.



reazioni secondarie. Nel presente lavoro di dottorato è stata utilizzata l'attivazione *in situ* con HBTU/HOBt/DIEA (1:1:2).



**Fig.2.2:** Strutture dei reagenti HBTU e HOBt

I tempi di *coupling* e deprotezione sono stati 15 e 25 minuti, rispettivamente. Il distacco del peptide dalla resina è stato ottenuto per trattamento con una miscela di TFA/H<sub>2</sub>O/TIS (94:3:3 v/v/v).

Per la sintesi del dominio CFC, i sei residui di cisteina presenti in sequenza, introdotti come derivati tritile (Trifenilmetile-Trt), sono prodotti con gruppi tiolici liberi dopo il distacco dalla resina. Nella sintesi dei frammenti di Nodal, invece, le catene laterali delle Cys sono protette con gruppi acetamidometile (Acm). I peptidi sono stati assemblati utilizzando un sintetizzatore automatico di peptidi 433A (Applied Biosystems, Foster City, USA), utilizzando residui aminoacidici con protezioni standard che sono rimosse durante la reazione di deprotezione con TFA, su una resina che produce peptidi con gruppi ammidici al C-terminale. Le reazioni di accoppiamento sono state condotte per 30 minuti e le reazioni di deprotezione per 20 minuti. Il polipeptide assemblato su resina è stato acetilato all'N-terminale mediante trattamento con anidride acetica e DIEA. Il polipeptide è stato staccato dalla resina mediante trattamento con una miscela di TFA-TIS-H<sub>2</sub>O (94:3:3) per 5 ore a temperatura ambiente. Il prodotto è stato precipitato da etere etilico freddo e separato dal solvente mediante centrifugazione. Infine il prodotto è stato liofilizzato. Dopo liofilizzazione, questo è stato ridotto in 100mM DTT in acqua per 30 min a temperatura ambiente, purificato attraverso HPLC in fase inversa e rifoldato attraverso ossidazione spontanea in di tampone carbonato, 100 mM, pH 8.5 contenente 0.5 mM di EDTA introdotto per impedire la complessazione dei tioli liberi da parte di cationi metallici. Durante il *refolding*, il peptide viene lasciato alla concentrazione di circa 4  $\mu$ M nel tampone descritto, sotto agitazione ed esposto all'aria. Il peptide rifolda rapidamente e quantitativamente dopo circa 1.5 h perdendo 6 uma alla fine del processo spontaneo di ossidazione delle cisteine e non subisce alcuna trasformazione fino a 12 h. Dopo il *refolding*, il dominio è stato purificato attraverso HPLC in fase inversa e spettrometria di massa.

Il polipeptide sintetico corrispondente al dominio EGF-like umano (75-113) con la treonina 88 non derivatizzata con fucosio come succede nella proteina nativa, è stato preparato nelle stesse condizioni sperimentali. Tutti gli altri peptidi sintetici sono stati preparati analogamente ma non sono stati sottoposti a processi di *refolding*, sono stati successivamente purificati e caratterizzati mediante LC-MS.

## 2.2 Esperimenti di dicroismo circolare

Le misure di dicroismo circolare sono state ottenute con uno spettropolarimetro Jasco modello J-715, calibrato a 290 nm con una soluzione acquosa di acido solfonico D(+)-10-canfora.

Le misure CD sono state condotte a temperatura ambiente adottando i seguenti parametri: *range*, 190-260 nm; *step resolution*, 0.2 nm; *scan speed*, 20 nm/min; *response*, 16 sec; *band width*, 2.0 nm. Sono state utilizzate celle cilindriche in quarzo fuso con cammino ottico di 0.1 mm.

Ogni peptide è stato analizzato mediante dicroismo circolare utilizzando diversi sistemi solvente. Generalmente gli spettri sono stati registrati utilizzando soluzioni con concentrazioni di peptide ( $C_{pep}$ ) dell'ordine di  $10^{-3}$ - $10^{-4}$ M in tampone fosfato 10 mM a pH 7.0, con diverse percentuali (da 0 a 70 %) di TFE (v/v). Inoltre in alcuni casi sono state effettuate misure CD in tampone acetato 10 mM a pH 5.0 e in tampone citrato 10 mM a pH 3.0.

## 2.3 Esperimenti NMR

Il campione del dominio CFC umano (hCFC) per l'analisi NMR è stato preparato sciogliendo circa 4 mg di peptide in 540  $\mu$ L di tampone fosfato 100mM a cui sono stati aggiunti altri 60  $\mu$ L di D<sub>2</sub>O (99.96% D - Sigma Aldrich) e 120 $\mu$ L di CD<sub>3</sub>CN per un rapporto finale di tampone fosfato/D<sub>2</sub>O pari a 90/10 v/v e una concentrazione di circa 1.2 mM. Per il campione del dominio CFC da topo (mCFC) circa 4 mg di peptide sono stati sciolti in 630  $\mu$ L di tampone fosfato 100mM e 70  $\mu$ L di D<sub>2</sub>O (99.96% D - Sigma Aldrich) per un rapporto di tampone fosfato/D<sub>2</sub>O pari a 90/10 v/v e una concentrazione di circa 1.3 mM. Il pH misurato per le soluzioni era uguale a 6.8 per hCFC e 7.2 per mCFC.

Per il frammento di Nodal43-69, il campione è stato preparato sciogliendo 3 mg di peptide in 420  $\mu$ L di H<sub>2</sub>O (purezza per HPLC senza TFA) a cui sono stati aggiunti 180  $\mu$ L di TFE (TFE-d<sub>3</sub> - Sigma Aldrich) per un rapporto finale di TFE/H<sub>2</sub>O pari a 30/70 v/v e una concentrazione di

circa 1.5 mM. I campioni per gli altri due frammenti di Nodal sono stati preparati utilizzando la stessa miscela solvente, sciogliendo 1.5 mg del peptide Nodal53-69 e 2.5 mg di Nodal66-90, ottenendo soluzioni con una concentrazione di 1.2 mM per Nodal53-69 e 1.4 mM per Nodal66-90.

Per tutti i campioni come standard di riferimento interno è stato utilizzato il 3-trimetilsililpropionato-2,2',3,3'-d<sub>4</sub> sodico (TSP).

Gli spettri NMR mono- e bi-dimensionali per il campione di hCFC e Nodal53-69 sono stati acquisiti su uno spettrometro Varian Inova, operante alla frequenza protonica di 600 MHz, sito presso l'Istituto di Biostrutture e Bioimmagini (IBB) del CNR di Napoli. Gli spettri per gli altri campioni sono stati acquisiti su uno spettrometro INCA (Consorzio Interuniversitario Chimica per l'Ambiente) Varian Inova, operante alla frequenza protonica di 500 MHz, sito presso il "Centro Interdipartimentale di Metodologie Chimico Fisiche", Università "Federico II" di Napoli. La caratterizzazione NMR è stata condotta alla temperatura di 298 K.

Esperimenti bidimensionali omonucleari protonici, quali *total correlation spectroscopy* (TOCSY),<sup>123</sup> *nuclear Overhauser effect spectroscopy* (NOESY),<sup>124</sup> *double quantum-filtered correlated spectroscopy* (DQF-COSY),<sup>125</sup> sono stati acquisiti in modalità *phase-sensitive* con il metodo di States-Haberhorn.<sup>126</sup> Gli esperimenti sono stati registrati con 2048 punti in F2 (4096 per i DQF-COSY) e 512 punti, portati a 1024 mediante *zero-filling*, in F1.

Il *mixing time* usato negli esperimenti TOCSY è stato di 70 ms. Gli esperimenti NOESY sono stati acquisiti con 100 ms, 200 ms e 300 ms di *mixing time*. Il segnale dell' H<sub>2</sub>O è stato soppresso adoperando sequenze con gradienti.<sup>127</sup>

I FID sono stati moltiplicati, in entrambe le dimensioni, con funzioni di pesatura di tipo *sine-bello gaussiana*, prima della trasformata di Fourier.

Gli spettri sono stati analizzati usando i programmi NMRPipe e NMRView.<sup>128,129</sup> L'identificazione dei sistemi di spin degli amminoacidi è stata effettuata dall'analisi degli spettri, TOCSY e DQCOSY, mentre l'assegnazione sequenziale dei diversi sistemi di spin è stata realizzata dall'analisi dello spettro NOESY, secondo il metodo di *Wuthrich (1986)*.<sup>130</sup>

I volumi dei cross-peak presenti negli spettri NOESY 300 ms, integrati con il programma NMRView,<sup>129</sup> sono stati convertiti in distanze con il programma CALIBA<sup>131</sup> del pacchetto CYANA.<sup>132</sup>

### 2.3.1 Constraint sperimentali da misure NMR

Il calcolo delle distanze interprotoniche è stato realizzato in modalità semiautomatica da CALIBA, una macro del programma CYANA.

CALIBA (*CALIBrAtion of NOE intensity versus distance constrain*) è un programma che considera come input l'intensità degli effetti NOE sperimentali e fornisce come output le distanze interprotoniche.<sup>131</sup> In particolare, CALIBA converte il volume di un picco NOE, misurato in un esperimento NOESY, in un limite superiore di distanza (upl = upper limit); in questo modo prepara l'input principale per il programma DYANA, cioè la lista degli upl.

CALIBA è basato, in particolare, sulla seguente assunzione: il volume  $V$  di un cross-peak di uno spettro NOESY, è inversamente proporzionale alla potenza "  $n$  " della corrispondente distanza di legame ( $d$ ):

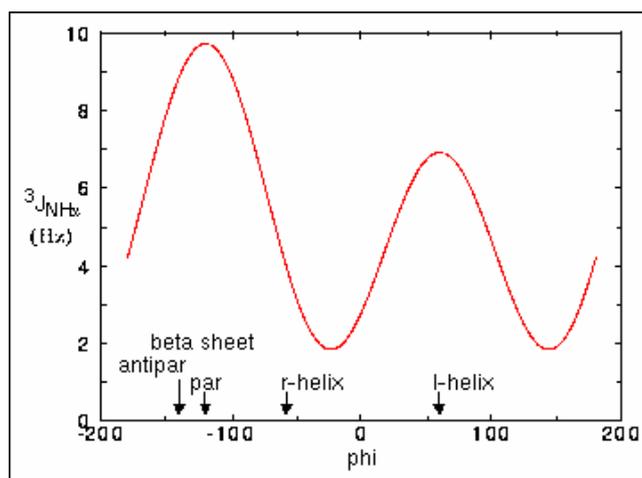
$$V = C/d^n \Rightarrow d = (C/V)^{1/n} \quad (1)$$

dove  $C$  è una costante ed i valori di  $n$  sono compresi tra 4 e 6; chiaramente il valore  $n=6$  è ottenuto teoricamente, assumendo una struttura rigida mentre i valori di  $n < 6$  sono calcolati empiricamente allo scopo di migliorare l'approssimazione della relazione (1) per picchi che coinvolgono i protoni delle catene laterali.

Fissando un valore di riferimento per una distanza nota tra due protoni (ad esempio due protoni geminali), tale macro esegue una calibrazione ricavando i rispettivi valori delle distanze mediante un algoritmo.

Ad esempio, per l'analisi di hCFC in H<sub>2</sub>O/D<sub>2</sub>O 90/10 (v/v) a pH 7 è stata utilizzata, come distanza di riferimento per la calibrazione, quella tra i protoni  $\alpha$  e  $\alpha'$ CH<sub>2</sub> della Gly<sup>37</sup> (2.2 Å); mentre per l'analisi di Nodal53-69 in TFE/H<sub>2</sub>O 30/70 (v/v) è stata utilizzata la distanza tra i protoni  $\beta$  e  $\beta'$ CH<sub>2</sub> della Pro<sup>16</sup> (2.2Å). L'elenco completo delle distanze, corrispondenti ai volumi dei NOE misurati, viene prodotto da CALIBA in formato upl e può essere letto da CYANA, passando allo stadio successivo dell'analisi.

Altri constraint utilizzati nel programma CYANA possono essere gli angoli di torsione  $\phi$ , stimabili dalle costanti d'accoppiamento  $^3J_{\text{NH-}\alpha\text{CH}}$ . Infatti, per un legame peptidico *trans* le  $^3J_{\text{NH-}\alpha\text{CH}}$  sono correlate all'angolo di torsione  $\phi$  attraverso una relazione<sup>133</sup> che ha la forma di una curva di Karplus<sup>134</sup>(Fig.2.3).



**Fig.2.3:** Curva di Karplus derivata empiricamente tra  ${}^3J_{\text{NH}-\alpha\text{CH}}$  vicinale coupling constant e angolo torsionale  $\phi$

Le costanti d'accoppiamento  ${}^3J_{\text{NH}-\alpha\text{CH}}$  sono state stimate dallo spettro DQF-COSY, misurando la separazione picco-picco delle componenti antifase dei cross peak NH- $\alpha$ CH.

Solo per il frammento Nodal43-69 per un certo numero di residui sono state misurate costanti d'accoppiamento  ${}^3J_{\text{NH}-\alpha\text{CH}}$  con valori di circa 6 Hz. In questi casi, in base alla curva di Karplus riportata in Fig.2.3, i rispettivi angoli diedri  $\phi$  sono stati impostati a  $-57^\circ$  con una tolleranza di  $\pm 10^\circ$ .

## 2.4 Calcolo dei modelli molecolari

Gli upl e gli angoli di torsione vengono utilizzati come vincoli sperimentali NMR per il calcolo del modello molecolare attraverso l'utilizzo della procedura TAD del programma CYANA. I modelli molecolari vengono poi minimizzati energeticamente attraverso l'utilizzo del programma AMBER.

### 2.4.1 CYANA: Torsion Angle Dynamics

Lo scopo principale del programma CYANA (*Combined assignment and Dynamics Algorithm for Nmr Applications*) è quello di riuscire a valutare e soprattutto a calcolare, sulla base di vincoli conformazionali derivanti da esperimenti NMR (NOE, costanti d'accoppiamento J) ottenuti sperimentalmente, le strutture tridimensionali di acidi nucleici e proteine. CYANA sostituisce il precedente programma di calcolo computazionale chiamato DIANA,<sup>131,135</sup> incorporando tutte le precedenti funzionalità di DIANA, e in più, diverse estensioni, tra cui un nuovo efficiente metodo di calcolo strutturale: la dinamica nello spazio degli angoli di torsione (TAD).

CYANA utilizza un metodo di calcolo strutturale basato sulla dinamica nello spazio degli angoli di torsione, mantenendo fissi, a valori 'natural', i parametri covalenti della struttura, cioè distanze di legame, angoli di legame, chiralità, planarità, con una notevole riduzione dei tempi di calcolo.<sup>132</sup> Questo programma, sfrutta i dati sperimentali forniti dalla spettroscopia NMR, attuando un *annealing* simulato combinato con le dinamiche molecolari nello spazio degli angoli di torsione, che porta al calcolo della soluzione numerica dell'equazione del moto della meccanica classica (equazioni di Lagrange) dove le coordinate generalizzate sono gli angoli di torsione. Il calcolo eseguito in CYANA consiste nella minimizzazione di una funzione *target* variabile con lo scopo di trovare strutture tridimensionali che soddisfino i vincoli conformazionali e le proprietà covalenti della molecola. La funzione *target*, strettamente connessa all'energia potenziale ( $V$ ), è definita in modo tale che assume valori solo positivi ed è uguale a zero solo se tutti i vincoli sperimentali sono soddisfatti. Il sistema è accoppiato ad un bagno termostatico che viene raffreddato lentamente partendo dalla sua temperatura massima iniziale, così da permettere al sistema di superare le barriere tra i diversi minimi locali della funzione *target*.

La qualità delle strutture ottenute dal calcolo può essere giudicata in maniera quantitativa dal valore del *Root Mean Square Deviation* o RMSD.<sup>131</sup> L'RMSD misura la differenza atomica media tra le coordinate atomiche cartesiane di coppie di strutture (ad esempio ogni singola struttura e la media) e permette, quindi, di valutare la convergenza del calcolo.

La selezione di strutture con valori minimi di funzione *target*, la quale misura l'accordo con i vincoli sperimentali imposti, garantisce la rappresentazione realistica della conformazione molecolare in soluzione.

Per i nostri sistemi è stato generato un primo insieme di quaranta conformeri usando il programma CYANA. La libreria del programma è stata opportunamente modificata per ottenere i residui N- e C- terminali delle sequenze rispettivamente nella forma acetilata e amidata ed è stata inserita la Cys modificata con il gruppo Acm. Il calcolo è stato condotto utilizzando come vincoli sperimentali le distanze interprotoniche ricavate via CALIBA dalle intensità degli effetti NOE. Le distanze sperimentali sono state in particolare utilizzate come limiti superiori (upl) alle distanze interprotoniche dei conformeri prodotti nel calcolo. Un singolo conformero presenterà energia tanto più bassa, ovvero un tanto più basso valore della funzione *target*, quanto minore è il numero complessivo di violazioni rispetto alle distanze massime (sperimentali) impostate. Sono stati inoltre introdotti vincoli per alcuni angoli diedri del *backbone* per il frammento Nodal43-69. Gli angoli  $\Phi$ , relativi a residui con

una costante d'accoppiamento  $^3J_{\text{NH}\alpha\text{CH}}$  minore di 6 Hz, sono stati imposti a  $-57^\circ$ , con una tolleranza di  $\pm 10^\circ$ , in base alla curva di Karplus.<sup>134</sup> Per i calcoli sono stati adoperati i parametri standard forniti dal programma CYANA.

#### 2.4.2 AMBER: Meccanica Molecolare e Minimizzazione energetica

I programmi di meccanica molecolare (MM), che sono stati sviluppati nel corso degli anni, vengono attualmente utilizzati soprattutto per simulare e "predire" le proprietà dinamiche e termodinamiche delle molecole.<sup>136-138</sup>

Tali programmi, inoltre, vengono utilizzati per raffinare le strutture e simulare il loro comportamento in soluzione, nel rispetto dei dati sperimentali.

L'approssimazione sulla quale tutti i metodi di MM si basano, è la possibilità di ottenere una funzione univoca, continua e differenziabile, che descriva l'energia dello stato elettronico di un sistema, esclusivamente sulla base delle coordinate dei nuclei. Tale funzione è detta 'Campo di Forza' ed è descritta da una sommatoria di quattro termini: lo "stretching" tra due atomi legati, il "bending" tra tre atomi legati, la torsione tra angoli diedri e le interazioni di non legame tra atomi non legati covalentemente (Van der Waals, interazioni elettrostatiche). Nella costruzione di tali campi di forza si effettua un'ulteriore assunzione: esistono dei valori 'natural' per lunghezze di legame, per angoli tra legami, per angoli di torsione e così via, che dipendono esclusivamente del tipo di atomi coinvolti. I programmi di meccanica molecolare possono implementare campi di forza diversi, ciascuno dei quali è stato sviluppato e "parametrizzato" per meglio descrivere specifiche classi di molecole. Una volta definito il campo di forza, risulta definibile l'energia della struttura molecolare .

Il programma CYANA, che non è definibile a tutti gli effetti un come programma di MM, attua una minimizzazione della funzione target variando i parametri relativi agli angoli di torsione ed alla sovrapposizione delle superfici di Van der Waals. Dal momento che questa procedura non è ritenuta sufficiente all'ottenimento di strutture 'realistiche', le migliori strutture ottenute da CYANA vengono ulteriormente raffinate mediante l'impiego di un programma di meccanica molecolare più completo, come ad esempio AMBER 6.0 (*Assisted Model Building with Energy Refinement*).<sup>139</sup>

Un protocollo di calcolo ritenuto affidabile prevede che le strutture CYANA, con il più basso valore della funzione target, siano sottoposte a minimizzazione energetica usando il modulo SANDER (*Simulated Annealing with NMR-Derived Energy Restraints*) del pacchetto AMBER 6.0. Le nostre minimizzazioni energetiche sono state condotte *in vacuo* utilizzando un campo di forza AMBER, indicato con la sigla "ff 91", in cui la costante dielettrica viene considerata

dipendente dalla distanza  $\epsilon_r$ .<sup>140</sup> Un *cut-off* sulle distanze di 12 Å è usato per valutare le interazioni di non legame. La minimizzazione energetica vincolata è stata condotta nei primi 200 *step* utilizzando il metodo *steepest descent* e nei successivi 2000 *step* con il metodo del gradiente coniugato. Per ognuno dei peptidi oggetto di studio sono state selezionate le migliori strutture, che in accordo con i dati sperimentali, mostrano i più bassi valori di energia potenziale. Per la successiva analisi statistica è stato poi utilizzato il programma di grafica molecolare MOLMOL.<sup>141</sup>

## 2.5 Homology Modelling

L'*homology modelling* si basa sull'osservazione che proteine che presentano un buon livello di similarità di sequenza, mostrano in genere un buon livello di similarità di struttura. Quando la proteina in esame (*target*) presenta un significativo grado di similarità (>30%) con una proteina a struttura nota (o *templato*),<sup>142</sup> si può utilizzare la struttura del templato come sistema di riferimento per costruire il modello della proteina data. La prima operazione da compiere è quella di verificare se esiste una (o più di una) proteina di struttura nota, che abbia una sequenza amminoacidica simile a quella della proteina *target*. Dato che la similarità della conformazione della catena principale nel *core* di due proteine aumenta all'aumentare della somiglianza tra le loro sequenze ne consegue che il miglior templato per costruire un modello è quello che ha la maggiore similarità in sequenza con la proteina *target*.<sup>143</sup> Numerosi algoritmi permettono di misurare l'identità o la similarità fra due sequenze in modo sufficientemente accurato da valutare se una proteina è adatta a fare da templato per la costruzione di un modello per omologia.

I programmi di allineamento identificano le regioni strutturali conservate in famiglie di proteine e valutano i gradi di omologia. I programmi da noi utilizzati per ottenere l'allineamento di sequenza sono stati: PSIBLAST, HHpred e mGentHreader.

PSI-BLAST<sup>144</sup> (*Position Specific Iterative BLAST*) rappresenta un'estensione del programma di base BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool), ricerca la similarità della proteina *target* all'interno di un database e produce un output in forma di allineamento della sequenza della proteina *target* con le sequenze del database con cui hanno mostrato una similarità significativa. La bontà dell'allineamento si può determinare dal valore dell'identità di sequenza ( $\%seq = N_{ident}/N_{align}$ ), che è calcolata come frazione di coppie di residui identici ( $N_{ident}$ ) fra tutti quelli allineati ( $N_{align}$ ), e dal valore dell' *Evalue*, che rappresenta la

probabilità che uno score alto quanto quello osservato tra le due sequenze possa essere ottenuto per puro caso in una ricerca di similarità in un database di dimensioni  $D$ .

Il programma mGenTHREADER<sup>145</sup> si basa sul riconoscimento di *fold*. In questo programma si fa l'assunzione che la struttura nativa corrisponde alla conformazione a minima energia tra quelle accessibili alla proteina all'equilibrio. La sequenza target è infilata in tutti i possibili allineamenti in ogni struttura del database di strutture note e una funzione energia è usata per valutare gli allineamenti. Gli allineamenti con un basso valore di energia corrispondono al fold di tipo nativo per la sequenza target. Per valutare la compatibilità tra la sequenza-struttura si utilizza il parametro Score:

$$S = (E - E_m)/\sigma$$

dove  $E$  è l'energia di un dato allineamento,  $E_m$  è l'energia mediata su tutti gli allineamenti e  $\sigma$  la deviazione standard. Alti valori dello Score indicano che la sequenza target è probabilmente compatibile con la struttura. Un altro parametro di cui bisogna tener conto per valutare la bontà di un allineamento è il P-value, che rappresenta la probabilità di avere un falso positivo; bassi valori P-value sono indicativi del fatto che l'allineamento che si osserva non è dovuto al caso.

Dal programma HHpred<sup>146</sup> (Homology detection & structure prediction by HMM-HMM comparison) si ottiene un allineamento dal confronto della proteina target con gli HMM (Hidden Markov Models).<sup>147</sup> Gli Hidden Markov Models sono strutture computazionali utilizzate per descrivere pattern precisi che definiscono le famiglie di sequenze omologhe. In questo programma il valore dell'Evalue è dato dal rapporto tra il P-value e il numero di HMM presenti nel database.

L'allineamento si ottiene utilizzando non solo le sequenze amminoacidiche delle due proteine, ma anche la struttura della proteina template, la predizione della struttura secondaria e talvolta una sovrapposizione strutturale di proteine di struttura nota omologhe al template. La struttura del template può essere usata per verificare che eventuali inserzioni o delezioni non capitino in elementi di struttura secondaria o in regioni non esposte sulla superficie, mentre la sovrapposizione strutturale di proteine simili al template permette di individuare quali regioni siano più conservate strutturalmente nella famiglia, e che quindi saranno probabilmente conservate anche nella proteina target. Anche l'allineamento multiplo di sequenze della famiglia della proteina target contiene informazioni sulla conservazione e variabilità di ogni posizione.

Per la costruzione del modello per omologia dei domini extracellulari dei recettore di tipo I, ALK4 e ALK7, i programmi PSIBLAST,<sup>144</sup> HHpred<sup>146</sup> e mGentHreader<sup>145</sup> ci hanno fornito un unico template possibile per effettuare l'allineamento.

Il programma PSIBLAST dopo tre iterazioni, con un Evalue di  $1.7 \times 10^{-13}$  e un'identità di sequenza del 27%, ci ha fornito la struttura del recettore umano ALK3 (pdbcode: 1rew; catena A, risoluzione 1.86 Å) come miglior template per la costruzione del modello per omologia di ALK4. 1rew è stata, anche, la prima struttura identificata, nel database PDB70\_15Apr07, dalla ricerca per confronto con i profili di Markov (Hidden Markov Model, HMM) effettuata dal programma HHpred, con un Evalue di  $3 \times 10^{-27}$  e dal programma mGenThreader con un P-value di  $10^{-27}$ . Per la costruzione del modello per omologia di ALK7 il programma PSIBLAST dopo due iterazioni, con un Evalue di  $2 \times 10^{-14}$  e un'identità di sequenza del 39%, ci ha fornito come template la struttura del recettore umano ALK5 (pdbcode: 2pjj, catena C, risoluzione 3.00 Å). 2pjj è stata, anche, la prima struttura identificata, nel database PDB70\_20Mag08, dalla ricerca per confronto con i profili di Markov (Hidden Markov Model, HMM) effettuata dal programma HHpred, con un Evalue di  $3.2 \times 10^{-30}$  e dal programma mGenThreader con un P-value di  $2 \times 10^{-5}$ . Gli allineamenti target-template ottenuti dai programmi PSIBLAST, HHpred e mGentHreader sono stati utilizzati per la costruzione dei modelli molecolari dei domini extracellulari di ALK4 e ALK7, utilizzando il programma Modeller6v2.<sup>148</sup>

Per la costruzione del modello per omologia del monomero e dimero della proteina Nodal umana il programma HHpred ci ha fornito cinque template possibili, quattro appartenenti alle BMP (BMP6, BMP9, BMP2 e BMP3) e uno appartenente alle GDF (GDF5). Tutti i template mostrano valori di identità di sequenza e Evalue di seguito riportati (Tabella2.1).

**Tabella2.1:** Risultati della ricerca nel database PDB70\_14Apr08 effettuata dal programma HHpred

Hit	pdb_code			Ident	Prob	E-value	P-value	Score	Risoluz
<u>1</u>	<a href="#">1lxi_A</a>	BMP6	monomero	41 %	100.0	0	0	269.6	2.00Å
<u>2</u>	<a href="#">1zkz_A</a>	BMP9	monomero	43 %	100.0	2.8E-45	0	266.3	2.33Å
<u>3</u>	<a href="#">1waq_A</a>	GDF5	monomero	38 %	100.0	4.2E-45	0	265.1	2.28Å
<u>4</u>	<a href="#">2h62_A</a>	BMP2	Complex recettore	40 %	100.0	8.5E-45	0	257.8	1.85Å
<u>5</u>	<a href="#">2qcg_A</a>	BMP3	dimero	39 %	100.0	5.2E-45	0	253.5	2.21Å

Dei cinque templati è stato effettuato un allineamento di struttura secondaria, utilizzando il programma *SSM*<sup>149</sup> (**S**econdary **S**tructure **M**atching) ed è stato costruito un profilo HMM, utilizzando il programma *HMMer*.<sup>150</sup> A questo profilo è stata allineata la sequenza umana di Nodal. Per la costruzione dell'HM è stato utilizzato il programma *Modeller6v2*.<sup>148</sup>

La qualità dei modelli ottenuti è stata appurata utilizzando alcuni programmi di controllo qualità, come *WhatCheck*,<sup>151,152</sup> *Verify3D*<sup>153,154</sup> e *ProQ*.<sup>155,156</sup>

## 2.6 Esperimenti di docking

Gli studi di *docking* sono stati effettuati con il pacchetto 3D-Dock del Biomolecular Modelling Laboratory of Cancer Research, UK (<http://www.bmm.icnet.uk/docking>). I complessi sono stati generati con *FTDock*,<sup>157</sup> che classifica e ordina le soluzioni del *docking* in base alla complementarità di superficie (*SCscore*). *FTDock* per effettuare il *docking* utilizza l'algoritmo di Katchalski-Katzir et al. (1992),<sup>158</sup> si parte dalla descrizione geometrica del recettore (A) e del ligando (B) conoscendo le loro coordinate atomiche (Fig.2.4).

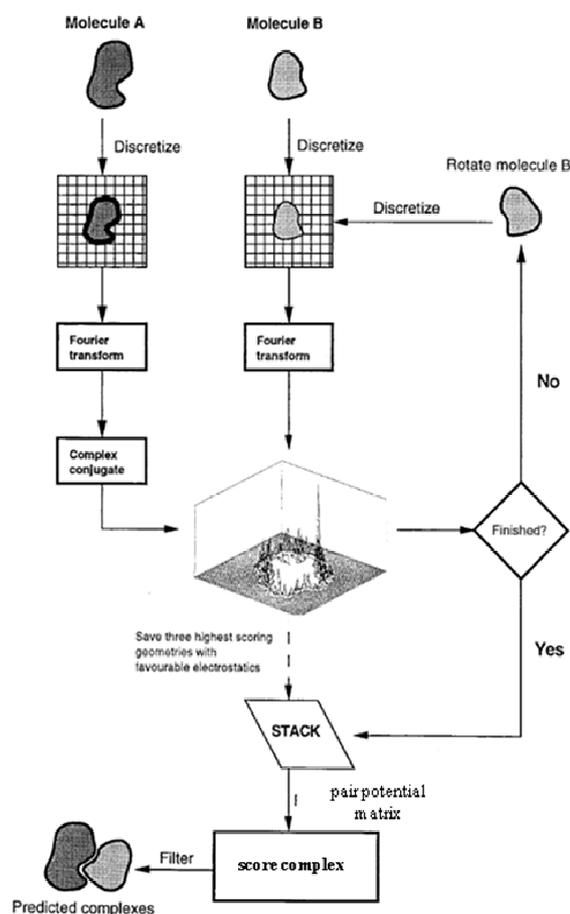


Fig.2.4: Schema del protocollo di lavoro dell'algoritmo di Katchalski-Katzir utilizzato dal programma *FTDock*

Le due molecole vengono proiettata su una griglia a tre dimensioni (l,m,n) e sono rappresentate da funzioni discrete ( $f_{A,l,m,n}$  e  $f_{B,l,m,n}$ ), che sono uguali:

- 1: rappresentano punti della griglia sulla superficie della molecola
- 0: rappresentano punti della griglia all'esterno della molecola
- $\rho$ : rappresentano punti della griglia all'interno della molecola

La funzione di correlazione che descrive la complementarità di superficie (SCscore) delle molecole A e B lungo ciascun vettore di traslazione è uguale ad:

$$f_{C\alpha,\beta,\gamma} = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N f_{A,l,m,n} \times f_{B,l,m,n}$$

dove  $\alpha, \beta, \gamma$  sono i vettori di traslazioni della molecola B rispetto ad A. Questa funzione assume valori:

- **positivi**: buona complementarità di superficie
- **negativi**: penetrazioni delle superfici
- **0**: le proteine non sono in contatto.

FTDOCK utilizza un'unità di griglia di 0.7 Å, uno spessore della superficie di 1.3 Å, un angolo di rotazione di 12°.

Nel *docking* tra ALK4-ECD e hCFC, il modello per omologia del dominio extracellulare del recettore ALK4 è stato considerato come la parte statica mentre ciascuna delle dieci strutture NMR a più bassa energia di hCFC (chiamate hCFC1-10) è stata considerata come la parte mobile. Un totale di 9240 rotazioni sono state valutate, ottenendo 10000 complessi per ogni struttura NMR. Queste soluzioni del *docking* sono state classificate e ordinate usando RPScore.<sup>159</sup> Questo programma classifica le soluzioni del *docking* usando una matrice empirica di potenziali di coppie di residui. In questo lavoro di dottorato è stata usata la matrice i90\_p05\_d4.5\_2dp.matrix. Queste soluzioni sono state poi filtrate applicando un *constraint* sulla distanza di 4.5 Å ai residui che sono noti dai dati di letteratura, fare parte dell'interfaccia proteina-proteina.

Per il *docking* tra hCFC e ALK4-ECD sono stati accettate solo le soluzioni che esibivano l'H<sup>9</sup> e il W<sup>12</sup> di hCFC ben orientate verso l'interfaccia idrofobica di legame di ALK4-ECD.

La procedura sopra descritta ha portato a nove soluzioni per la struttura del ligando hCFC1, due soluzioni per hCFC2, venti soluzioni per hCFC3, nessuna soluzione per hCFC4, sedici

soluzioni per hCFC5, otto soluzioni per hCFC6, quattordici soluzioni per hCFC7, cinque soluzioni per hCFC8, sei soluzioni per hCFC9 e quattro soluzioni per hCFC10. L'ispezione visiva ci ha portato a scartare le soluzioni per hCFC2, hCFC6, hCFC8, hCFC9, hCFC10, così come sette per hCFC1, dieci per hCFC3 e hCFC4, e sei per hCFC5. Le soluzioni rimanenti (Tabella2.2) sono state rifinite utilizzando il programma MULTIDOCK, che effettua una minimizzazione energetica utilizzando una libreria di rotameri delle catene laterali.<sup>160</sup>

**Tabella2.2:** Soluzioni del *docking* di ALK4/hCFC selezionati dopo il filtro

NMR structures	Multidock				
	SCscore	RPScore	Total Energy	Qenergy	bumps
<b>hCFC1</b>					
1	149	2.77	-38.98	-12.76	0
2	197	2.06	-36.08	-14.66	3
<b>hCFC3</b>					
1	146	2.98	-30.56	-3.85	2
2	154	2.71	-28.65	-2.87	6
3	148	2.57	-36.87	-8.14	1
4	151	2.48	-40.33	-8.64	1
5	149	2.39	-22.40	-4.64	7
6	154	2.31	-30.79	-12.8	7
7	156	1.73	-26.95	-5.44	1
8	156	1.61	-1.9	-7.8	15
9	181	1.57	-37.00	-6.2	1
<b>10</b>	<b>161</b>	<b>1.49</b>	<b>-41.67</b>	<b>-14.9</b>	<b>1</b>
<b>hCFC5</b>					
1	149	4.28	-37.47	-5.63	2
2	147	3.58	-32.51	+5.98	8
3	188	2.99	-54.58	-9.89	1
4	169	2.56	-49.92	-7.84	2
5	228	2.33	-41.90	-14.63	6
6	166	2.29	-59.30	-19.62	11
7	181	1.81	-18.99	-11.29	8
8	162	1.80	-35.46	-24.15	6
9	156	1.49	-33.89	-13.88	9

10	214	1.35	-43.48	-6.35	3
11	157	1.03	-41.97	-20.05	1
<b>hCFC6</b>					
1	151	1.81	-33.38	-8.82	0
2	151	1.71	-34.64	-9.89	0
3	155	1.64	-29.27	-4.95	2
4	148	1.58	-30.24	-4.59	0

Di queste, la soluzione migliore (il complesso 10 di hCFC3) è stata scelta guardando l'RP Score, l'energia, e le caratteristiche geometriche all'interfaccia.

Nel *docking* tra i recettori ALK4-ECD, ALK7-ECD e il dimero di Nodal, il modello per omologia del dominio extracellulare dei recettori è stato considerato come la parte statica, mentre il dimero di Nodal è stato considerato come la parte mobile. Sono state accettate solo le soluzioni, che presentavano il '*wrist epitope*' di Nodal in prossimità della superficie concava di *binding* dei recettori. Dopo l'applicazione di questo filtro dalle 10000 soluzioni iniziali abbiamo ottenuto 1132 complessi per Nodal:ALK4-ECD e 1112 per Nodal:ALK7-ECD. Di questi mettiamo a confronto il primo complesso ottenuto dopo la procedura di filtro, minimizzato energeticamente con il programma MULTIDOCK (Tabella2.3). È interessante notare che il primo complesso in ordine di RP Score ottenuto dopo il filtro è nell'ordine il primo dei 10000 conformeri calcolati inizialmente per ALK7 e il secondo per ALK4.

**Tabella2.3:** Primo complesso del *docking* di ALK4/hCFC e ALK7/hCFC selezionato dopo il filtro

	Multidock				
	SCscore	RPscore	Total Energy	Qenergy	bumps
Nodal:ALK4-ECD	251	6.75	-34.06	-24.29	6
Nodal:ALK7-ECD	167	5.46	-40.78	-6.64	1

## 2.7 Esperimenti di binding SPR

Per stimare il riconoscimento dei domini hCFC ed hEGF-like verso i recettori ALK4 e ActRIIB sono stati effettuati esperimenti di *binding* "Real Time Bia" con tecnica Surface Plasmon Resonance (SPR) immobilizzando il recettore di tipo I, ALK4, e il recettore di tipo II, ActRIIB covalentemente su *biochips*. Le proteine ricombinanti umane Activin RIB/ALK-4- e ActRIIB fuse a -Fc sono state comprate dalla R&D Systems (Minneapolis, US).

La tecnica 'Real Time Bia' (Biomolecular Interaction Analysis) è una tecnica innovativa per lo studio delle interazioni biomolecolari, che si basa su un fenomeno ottico detto Surface Plasmon Resonance (SPR).

In questa tecnica una specie interagente (il recettore) si trova immobilizzata sulla superficie di un *chip*, mentre l'altra (il ligando) è in soluzione. Ciò che avviene nell'SPR è legato ad un cambiamento dell'indice di rifrazione sulla superficie del *sensor chip* provocato da una alterazione correlata al legame tra ligando e recettore. Il segnale SPR è monitorato in continuo e le interazioni tra biomolecole possono essere studiate in tempo reale.

La strumentazione (Fig.2.5) è costituita da diverse parti fondamentali che sono:

- UNITA' DI RIVELAZIONE: include componenti ottici ed elettronici per generare e misurare la risposta SPR (sorgente luminosa, prisma, e rivelatore);
- SENSOR CHIP: presenta una superficie biospecifica dove hanno luogo le interazioni tra analita e ligando;
- INTEGRATED MICROFLUIDICS CARTRIDGE (IFC): contiene i microtubi che trasportano il tampone, il *loop* per il caricamento del campione e le valvole;

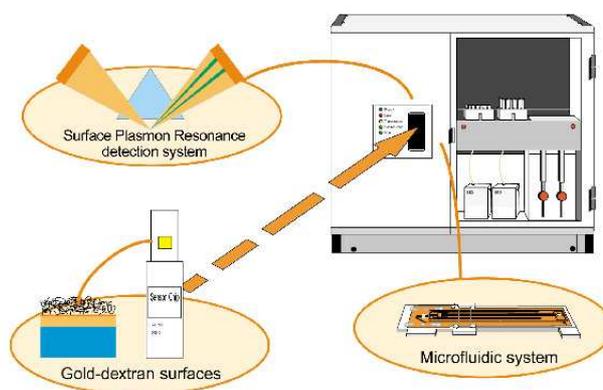


Fig.2.5:Strumentazione per la tecnica Real-Time Bia

Il *sensor chip* (Fig.2.6) è costituito da un supporto in vetro su cui è stato depositato un sottilissimo strato in oro. L'oro presenta infatti inerzia chimica verso soluti e solventi e fornisce una buona risposta in SPR. Lo strato in oro a sua volta è ricoperto da una matrice legata covalentemente ad esso sulla quale vengono immobilizzate le biomolecole. Il *chip* più utilizzato è il CM5 la cui matrice è costituita da destrano carbossimetilato non cross linkato,

che fornisce un intorno adatto per lo studio delle interazioni biomolecolari. Il destrano è costituito da un polimero lineare di unità di glucosio carbossimetilato.

I due recettori ricombinanti (collegati a Fc) sono stati immobilizzati sul *chip*, in due distinti canali.

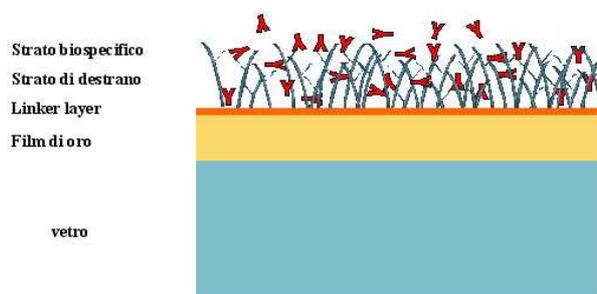


Fig.2.6: *Sensor chip*

La fase in cui il recettore viene immobilizzato sulla superficie del *chip* viene detta *coupling*. Le reazioni chimiche coinvolte nel processo di immobilizzazione del recettore sul *chip* CM5, che è rappresentato da tre eventi fondamentali:

1. ATTIVAZIONE: i gruppi carbossilici vengono attivati chimicamente tramite N-etil-N'-(3-dimetilaminopropil) carbodimide cloridrato (EDC) e N-idrossisuccinimide (NHS).
2. COUPLING: il recettore reagisce coi gruppi carbossilici attivati della matrice. I gruppi funzionali della molecola di ligando coinvolto in questa reazione possono essere diversi, nel caso delle proteine si forma un legame ammidico con i gruppi amminici degli amminoacidi basici (*amine coupling*).
3. INATTIVAZIONE: la succinimide che non ha reagito con il recettore viene sostituita con etanolamina.

Nel nostro caso specifico la superficie di destrano carbossimetilato è stata attivata attraverso un'iniezione di sette minuti di una soluzione contenente 200mM di EDC e 50 mM di NHS. Successivamente, nella fase di *coupling* vengono iniettate le soluzioni di ALK4 o ActRIIB (soluzioni 50 µg/ml in 10mM di acetato di sodio, pH 4.0). Un flusso continuo di HBS (Hepes 10 mM, NaCl 150 mM, EDTA 3.4 mM) a 20 µl/min è stato mantenuto, per tre minuti e la succinimide che non ha reagito con il recettore è stata sostituita tramite l'iniezione di una soluzione 1M di etano lamina, pH 8.5, per 7 minuti. I reagenti per l'immobilizzazione delle proteine sono stati ottenuti da GE Healthcare.

Una volta che il recettore è stato immobilizzato sul *sensor chip* si può procedere con la seconda fase che è rappresentata dall'analisi dell'interazione recettore-ligando.

Il campione viene caricato manualmente o attraverso un autocampionatore e viene trasportato attraverso un sistema di pompe pneumatiche a livello della cella d'analisi dove può interagire con la biomolecola immobilizzata sul *sensor chip*. Se avviene l'interazione si crea un segnale SPR che viene registrato nel tempo. L'unità di misura dell'SPR è l'unità di risonanza (RU).

Monitorando in modo continuato il segnale SPR si ottiene un diagramma detto **SENSORGRAMMA** nel quale si distinguono le diverse fasi dell'interazione: (Fig. 2.7)

- la fase di associazione: il ligando si lega al recettore fino al raggiungimento dell'equilibrio;
- la fase di dissociazione: coincide col termine dell'iniezione e fornisce dati utili sulla stabilità del complesso recettore-ligando;
- la fase di rigenerazione: il ligando viene completamente rimosso dalla superficie del *chip*.

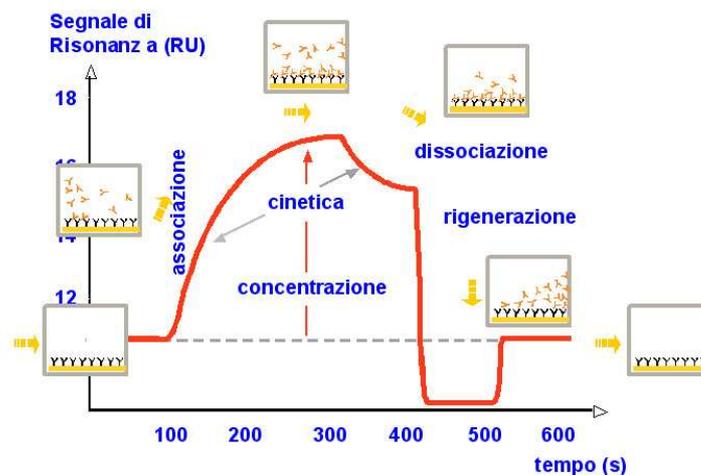


Fig.2.7: Sensorgramma

In genere per ogni esperimento vengono registrati almeno 6-8 sensorgrammi caricando diverse concentrazioni di ligando. Questi vengono poi rielaborati da un software che, tramite l'utilizzo di opportuni modelli matematici, fornisce informazioni sulle costanti cinetiche e di affinità del complesso recettore-ligando. L'impiego di questa tecnica permette il monitoraggio diretto delle fasi di associazione e dissociazione del complesso. La rielaborazione dei sensorgrammi mediante un opportuno software consente di ottenere dei

valori delle costanti di velocità di associazione e dissociazione e delle costanti d'equilibrio di affinità e dissociazione.

$K_a$  o  $K_{on}$  = costante di velocità di associazione ( $\text{sec}^{-1}\text{M}^{-1}$ )

$K_d$  o  $K_{off}$  = costante di velocità di dissociazione ( $\text{sec}^{-1}$ ).

$K_D$  = costante di dissociazione (M) =  $K_{off}/K_{on}$

$K_A$  = costante di affinità ( $\text{M}^{-1}$ ) =  $K_{on}/K_{off}$ .

I nostri sensorgrammi sono stati registrati su uno strumento BIACORE 3000 (Biacore AB, Uppsala, Sweden).

I saggi di *binding* sono stati effettuati attraverso l'iniezione di soluzioni dei domini hCFC ed hEGF-like a concentrazioni crescenti, nel range 0-40  $\mu\text{M}$  e usando come riferimento il segnale prodotto da un *binding* non-specifico nel canale del riferimento su cui è stato legato l'Fc. L'Fc (frazione costante degli anticorpi) è stato immobilizzato utilizzando un campione di proteina a 50  $\mu\text{g}/\text{ml}$  preparato presso l'IBB.

I tempi di *binding* e dissociazione del peptide sono stati 3 minuti e 2 minuti, rispettivamente. Dati cinetici triplicati sono stati collezionati per ciascuna concentrazione delle soluzioni. I dati sono stati correttamente manipolati per sottrarre il segnale di riferimento e sono stati rielaborati utilizzando il software BIAevaluation4.1 (Biacore AB, Uppsala, Sweden).

A livello teorico il principio di funzionamento del 'Real Time Bia' (Biomolecular Interaction Analysis) si basa, come già accennato, su un fenomeno ottico detto Surface Plasmon Resonance (SPR). In corrispondenza dell'interfaccia tra due mezzi trasparenti con diverso indice di rifrazione (ad esempio vetro e soluzione) la luce proveniente dal lato del mezzo con più alto indice di rifrazione viene in parte riflessa ed in parte rifratta. Al di sopra di un angolo di incidenza detto critico non viene rifratta attraverso l'interfaccia alcuna luce e si ha riflessione totale. In queste condizioni però una componente del campo elettromagnetico della radiazione incidente, detta onda evanescente (Fig.2.8), si propaga ad una certa distanza nel mezzo dotato di più basso indice di rifrazione.

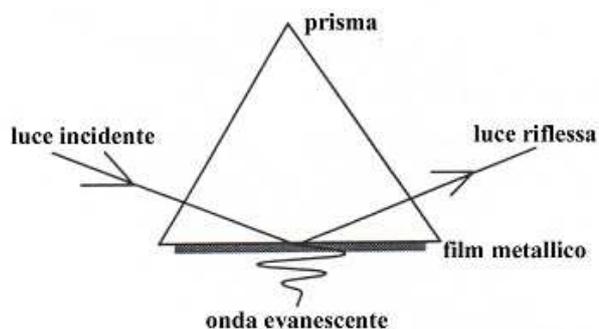


Fig.2.8: Schema dell'onda evanescente

In questo caso si lavora proprio in condizioni di riflessione totale. Nel *sensor chip* i due mezzi con diverso indice di rifrazione sono il vetro e la soluzione tampone; all'interfaccia tra di essi si trova il sottilissimo strato in oro. In queste condizioni l'onda evanescente che si genera interagisce con gli elettroni di superficie delocalizzati dell'oro, detti Plasmoni, presenti verso la soluzione tampone ed in seguito a questa interazione si verifica il fenomeno di risonanza di tali elettroni (Surface Plasmon Resonance) con conseguente riduzione dell'intensità della radiazione riflessa. L'angolo al quale si osserva questo fenomeno è detto angolo SPR ed il sistema che misura questo angolo è costituito da una serie di diodi fissati alla stessa lunghezza d'onda ma che registrano il segnale ad angoli SPR differenti. L'angolo SPR è influenzato da tre parametri:

- caratteristiche del film metallico;
- lunghezza d'onda della radiazione incidente;
- indice di rifrazione dei mezzi presenti a livello dei due lati del metallo (vetro e soluzione).

Le proprietà del film di metallo, la lunghezza d'onda e l'indice di rifrazione del vetro (mezzo più denso) sono mantenuti costanti, il fenomeno SPR viene quindi utilizzato per valutare l'indice di rifrazione dello strato acquoso immediatamente adiacente alla superficie di metallo del *chip* (oro). Dal momento che il recettore è immobilizzato al *chip* tale indice di rifrazione è influenzato principalmente dalla concentrazione oltre che dal peso molecolare del ligando a livello della superficie del *chip*. Semplificando: se il ligando interagisce con il recettore si verifica una variazione dell'indice di rifrazione dello strato acquoso e con conseguente cambiamento dell'angolo SPR da I a II; quest'ultimo effetto determina un segnale registrato dallo strumento (Fig.2.9). E' importante ricordare che il segnale dipende dal peso molecolare del ligando: tanto maggiore è il peso molecolare tanto maggiore sarà il segnale in SPR.

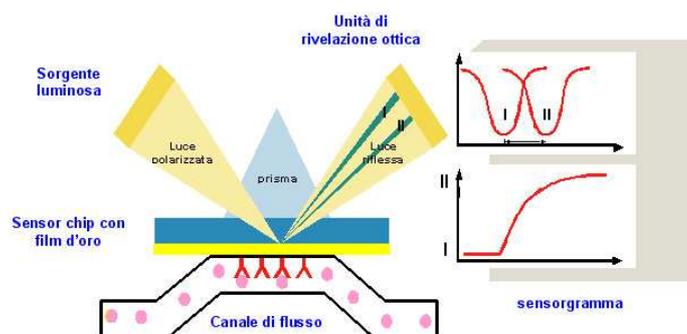


Fig.2.9: Angolo SPR

## CAPITOLO 3

- 3.1 Sintesi in fase solida dei domini CFC ed EGF-like di Cripto
- 3.2 Analisi conformazionale di hCFC a pH 7 in tampone fosfato
- 3.3 Analisi conformazionale di mCFC a pH 7 in tampone fosfato
- 3.4 Modello per omologia di ALK4-ECD e ALK7-ECD
- 3.5 Docking tra ALK4-ECD e hCFC
- 3.6 Studio di binding SPR tra hCFC e hEGF-like e i recettori ALK4/ActRIIB
- 3.7 Analisi conformazionale della proteina Nodal
- 3.8 Docking tra Nodal, ALK7-ECD e ALK4-ECD

### Risultati e Discussione

Il lavoro di ricerca si è sviluppato a partire dalla sintesi in fase solida dei domini CFC e EGF-like di Cripto e della proteina Nodal.

Si è scelta la strategia sintetica, e non l'espressione genica, per diverse ragioni. Innanzitutto i geni di Cripto e di Nodal possono essere espressi solo in linee cellulari di organismi superiori e, quindi, l'ottenimento di quantità sufficienti per l'analisi NMR è impedito dai notevoli tempi e costi di produzione. L'altra ragione è che non si riescono ad ottenere per espressione genica le proteine correttamente rinaturate (*refolded*).

Si anticipa che per sintesi chimica si è ottenuto il dominio CFC di Cripto, mentre, nonostante le molteplici e svariate strategie sintetiche adoperate, non si è riusciti ad ottenere la proteina Nodal; il dominio EGF-like, invece, è stato sintetizzato in modeste quantità non sufficienti per un'analisi NMR.

Parallelamente al lavoro di sintesi e analisi conformazionale in soluzione mediante CD ed NMR, si è sviluppato un lavoro computazionale mediante *homology modelling* ed esperimenti di *docking* per ricavare informazioni aggiuntive sulle caratteristiche strutturali dei sistemi interagenti.

#### 3.1 Sintesi in fase solida dei domini CFC e EGF-like di Cripto

Il dominio CFC 112-150, la cui sequenza è riportata in Fig.3.1, è costituito da 39 residui aminoacidici e rappresenta uno dei due domini funzionali ricchi di cisteine della proteina Cripto umana. Il polipeptide è stato sintetizzato presso l'IBB del CNR di Napoli, sotto forma di derivato acetilato all'N-terminale ed ammidato al C-terminale. Nella discussione che seguirà, per semplicità di descrizione, il primo residuo della sequenza (K<sup>112</sup>) sarà numerato come primo residuo (K<sup>1</sup>).

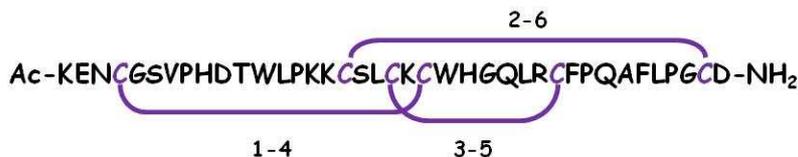


Fig.3.1 Sequenza amminoacidica del dominio CFC nella proteina Cripto umana, corrispondente ai residui 112-150

Il dominio CFC della proteina Cripto umana (hCFC) è stato ottenuto mediante sintesi in fase solida. Per i dettagli sulla sintesi chimica di hCFC si rimanda al paragrafo 2.1 delle Metodiche Sperimentali.

Il dominio hCFC è stato quindi sintetizzato su resina e il suo distacco è stato ottenuto per trattamento con una miscela di TFA/H<sub>2</sub>O/TIS (94:3:3 v/v/v). Dopo allontanamento dalla resina per filtrazione e dei solventi per evaporazione sotto vuoto, il peptide è stato precipitato con etere etilico freddo e liofilizzato. Il materiale solido recuperato, è stato ridotto in 100 mM di DTT per 30 min a temperatura ambiente, purificato attraverso HPLC in fase inversa e rifoldato attraverso ossidazione spontanea in 100 mM di tampone carbonato, pH 8.5 e 0.5 mM di EDTA.

Il peptide rifolda rapidamente e quantitativamente dopo circa 1.5 h e non subisce alcuna trasformazione fino a 12 h (Fig.3.2).

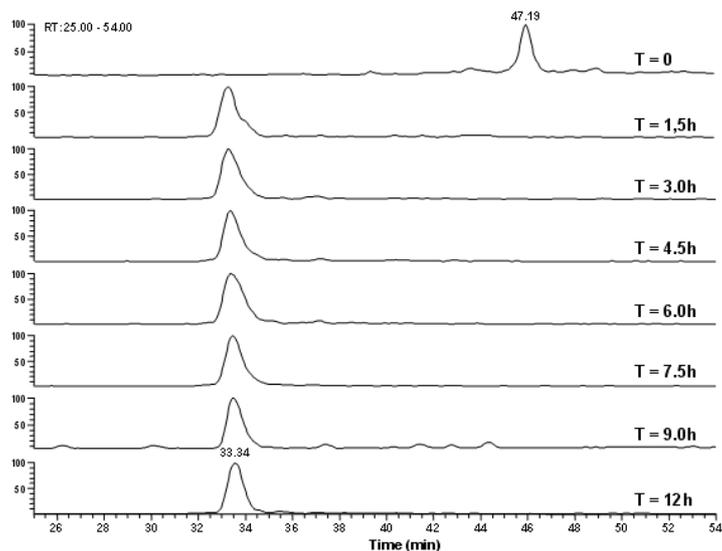


Fig.3.2 Refolding del dominio CFC di Cripto

Dopo il *refolding*, il dominio è stato purificato attraverso HPLC in fase inversa ottenendo un prodotto puro con un peso molecolare sperimentale in ottimo accordo con quello calcolato ( $PM_{teor/sper}=4468.9/4469.25$  uma) (Fig.3.3). La resa finale della sintesi del prodotto rifoldato

e purificato, calcolata sulla base della scala di sintesi iniziale e della quantità di prodotto recuperato, è stata del 2% circa.

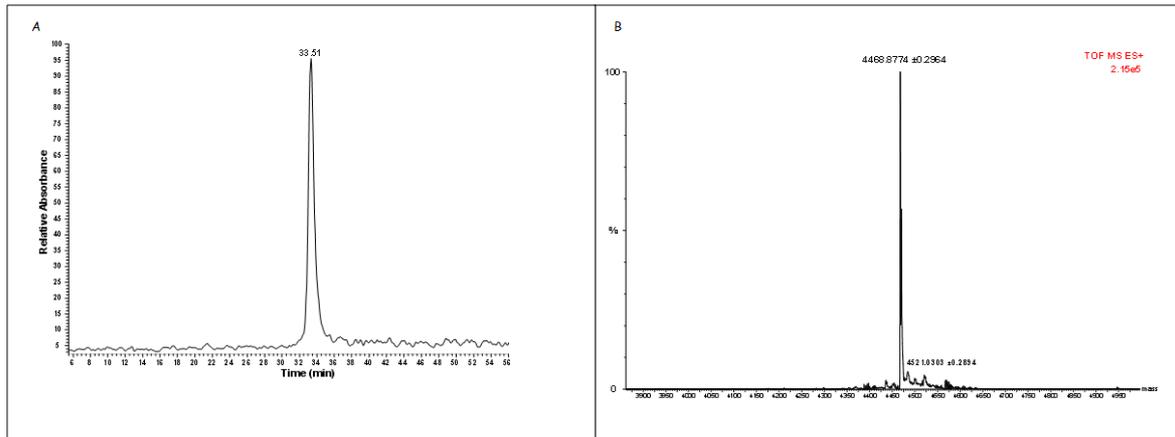


Fig.3.3: Il dominio CFC umano rifoldato e purificato è stato caratterizzato attraverso analisi HPLC (A) e spettrometria di massa (B) osservando un singolo piccolo, molto pulito a circa 33 min

La sintesi del dominio EGF-like, la cui sequenza e disposizione dei ponti disolfurici sono riportati nella successiva Fig.3.4, è stata condotta con modalità molto simili a quelle riportate per il dominio CFC, ma con rese molto più basse. Infatti il prodotto viene in buona parte perso durante la reazione di rinaturazione, perché da origine a numerosi sottoprodotti non correttamente foldati o solo parzialmente foldati. Nella successiva Fig.3.5, è riportata una sovrapposizione di corse cromatografiche relative al polipeptide durante il processo di rinaturazione insieme agli spettri di massa ESI prima del processo e dopo 9h.

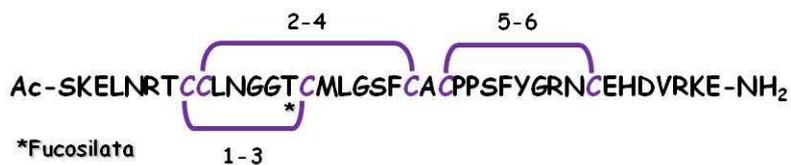


Fig.3.4: Sequenza e disposizione del dominio EGF-like di Cripto, corrispondente ai residui 75-113. La treonina 88 è fucosilata nella proteina matura. Quella sintetica è stata preparata senza la modifica post-traduzionale

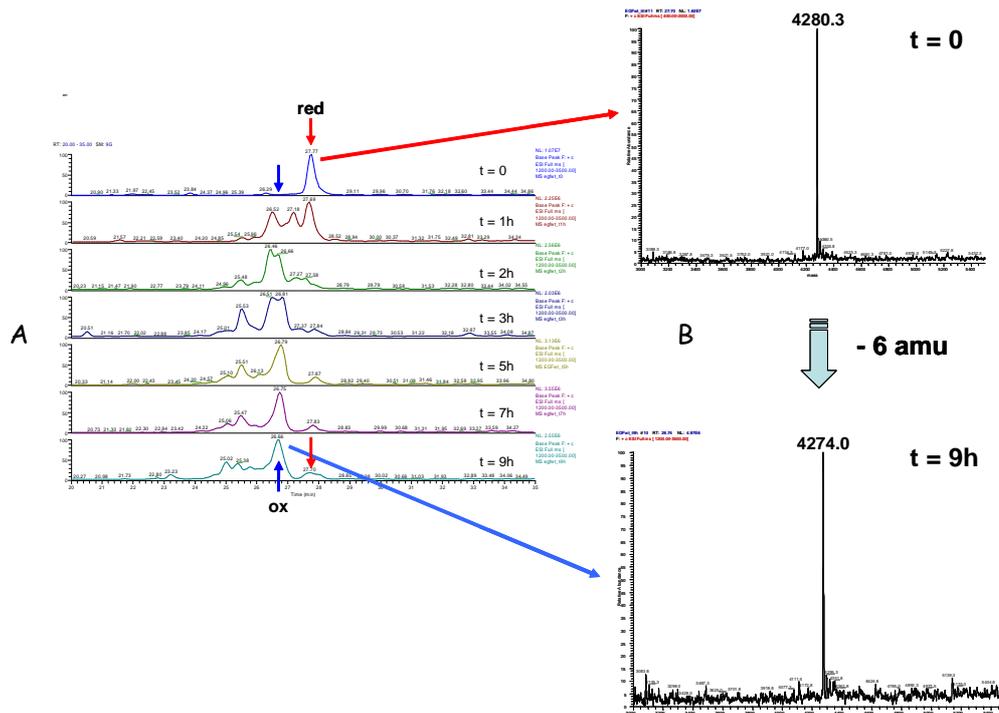


Fig.3.5: (A) *Refolding* del dominio EGF-like di Cripto. La rinaturazione è stata monitorata mediante analisi LC-MS per 9h. Il prodotto finale ha un peso molecolare di 6 uma inferiore a quello di partenza, indicativo della formazione dei 3 ponti disolfuro. (B) Spettri di massa dei prodotti ai tempi di ritenzione indicati

Il prodotto rinaturato è stato infine purificato su colonna a fase inversa, recuperando poco materiale (resa inferiore allo 0.1%). E' ragionevole assumere che le specie solo parzialmente rinaturate, abbiano successivamente reagito con il prodotto foldato, portando alla sua trasformazione in polimeri insolubili non separabili sulla colonna cromatografica.

### 3.2 Analisi conformazionale di hCFC a pH 7 in tampone fosfato

Per avere una prima diagnosi strutturale del dominio hCFC sono stati registrati spettri di dicroismo circolare presso l'IBB. Gli spettri CD del polipeptide rifoldato (Fig.3.6) sono stati registrati su soluzioni con una concentrazione di peptide pari a  $1.0 \times 10^{-4}$  M. Sono state preparate tre soluzioni separate: (i) 10 mM di tampone fosfato, pH 7.0; (ii) 10 mM di tampone fosfato, pH 7.0 con 10 % di TFE (v/v) e (iii) 10 mM di tampone fosfato, pH 7.0 con 20 % di TFE (v/v).

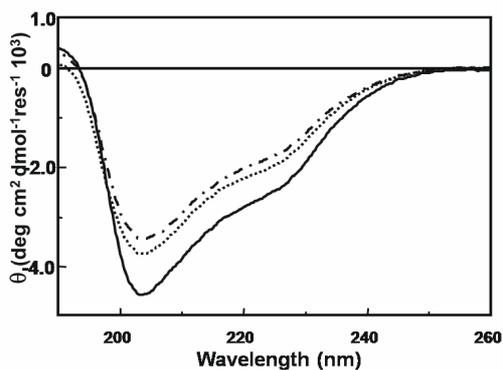


Fig.3.6: Spettri CD del dominio CFC umano rifoldato. La linea continua si riferisce al campione in 10 mM di tampone fosfato, pH 7.0; la linea a punti si riferisce al campione in 10 mM di tampone fosfato, pH 7.0 con 10 % di TFE (v/v) e la linea tratteggiata si riferisce al campione 10 mM di tampone fosfato, pH 7.0 con 20 % di TFE (v/v)

La struttura del peptide non cambia significativamente sotto l'aggiunta del TFE, suggerendo che le conformazioni sono ristrette dalla presenza dei ponti disolfuro.

### 3.2.1 Analisi NMR di hCFC

L'analisi conformazionale NMR di hCFC è stata condotta in tampone fosfato a pH 7, per avere informazioni strutturali in condizioni di pH simili a quelle fisiologiche.

L'assegnazione sequenziale di tutte le risonanze protoniche è stata ottenuta mediante l'esame comparato degli spettri TOCSY e NOESY, secondo il metodo classico di Wüthrich.<sup>130</sup>

In Fig.3.7A è riportata un'espansione dello spettro TOCSY 70 ms, dove sono messi in evidenza alcuni sistemi di spin.

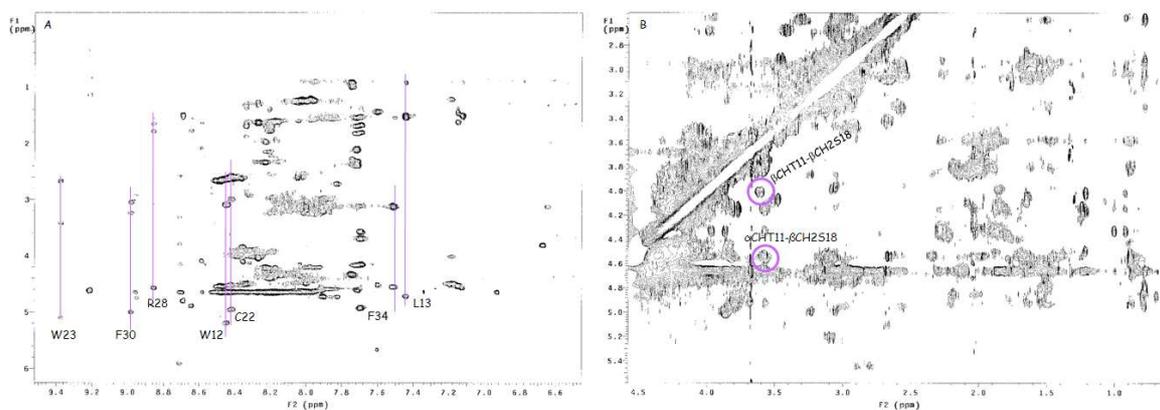


Fig.3.7: Spettri bidimensionali di hCFC in tampone fosfato/D2O 90/10 (v/v) a pH 7 a 298 K a 600 MHz: (A) Zona NH-alto campo dello spettro TOCSY 70 ms; (B) Zona alto campo-alto campo dello spettro NOESY 300 ms

Dall'analisi dello spettro DQF-COSY, è stato possibile confermare l'assegnazione dei sistemi di spin, sia delle glicine attraverso i contatti diretti NH- $\alpha$ CH sia dei protoni delle catene laterali. I *chemical shift* di tutti i sistemi di spin sono riportati in **Tabella3.1**.

**Tabella3.1:** *Chemical shift* (ppm) di hCFC in tampone fosfato/D<sub>2</sub>O 90/10 (v/v) pH 7, 298 K

Residuo	NH	$\alpha$ H	$\beta$ H	$\gamma$ H	altri
Lys <sup>1</sup>	8.27	4.04	1.64	1.29	$\epsilon$ CH <sub>2</sub> :2.94
Glu <sup>2</sup>	8.58	4.10	2.08, 1.64		
Asn <sup>3</sup>	7.83	4.73	2.73		$\gamma$ NH <sub>2</sub> :7.70, 6.90
Cys <sup>4</sup>	8.95	4.76	3.05, 2.94		
Gly <sup>5</sup>	8.56	4.16, 3.81			
Ser <sup>6</sup>	8.36	4.13	3.96, 3.86		
Val <sup>7</sup>	7.74	4.32	2.28	1.00, 0.93	
Pro <sup>8</sup>		4.93	2.19	2.00	$\delta$ H: 3.78
His <sup>9</sup>	8.20	4.24	3.22, 2.98		
Asp <sup>10</sup>	8.05	4.40	2.63		
Thr <sup>11</sup>	7.18	4.50	4.02	1.21	
Trp <sup>12</sup>	8.46	5.20	3.10, 2.64		2H: 7.33, 4H: 7.50, 5H:6.94 6H:7.40, 7H:7.27, NH:10.1
Leu <sup>13</sup>	7.44	4.70	1.54	1.54	
Pro <sup>14</sup>		4.85	2.20	2.03	
Lys <sup>15</sup>	8.33	4.50	1.87	1.37, 1.31	$\delta$ CH <sub>2</sub> :1.69; $\epsilon$ CH <sub>2</sub> :2.92
Lys <sup>16</sup>	8.20	4.39	1.79	1.64	$\delta$ CH <sub>2</sub> :1.71 $\epsilon$ CH <sub>2</sub> :2.92
Cys <sup>17</sup>		4.98	3.05, 2.95		
Ser <sup>18</sup>	7.69	4.93	3.69, 3.56		
Leu <sup>19</sup>	8.65	4.89	1.77	1.54	$\delta$ CH <sub>3</sub> :0.70
Cys <sup>20</sup>	8.71	5.92	3.30, 2.58		
Lys <sup>21</sup>	9.21	4.61	1.15		
Cys <sup>22</sup>	8.41	4.95	2.98, 2.61		
Trp <sup>23</sup>	9.37	5.11	3.43, 2.68		2H: 7.26, 4H: 7.65, 5H:7.33 6H:7.04, 7H:6.93, NH:10.4
His <sup>24</sup>	8.00	4.50	3.18, 3.12		
Gly <sup>25</sup>	8.32	4.13, 3.94			
Gln <sup>26</sup>	7.72	4.18	2.39, 2.33	2.11	
Leu <sup>27</sup>	8.69	4.80	1.56	1.51	$\delta$ CH <sub>3</sub> :0.81
Arg <sup>28</sup>	8.85	4.56	1.67, 1.54	1.79	$\delta$ CH <sub>2</sub> :3.08; $\epsilon$ NH:7.15, 6.70
Cys <sup>29</sup>	8.52	5.44	2.89, 2.66		
Phe <sup>30</sup>	8.99	5.00	3.24, 3.06		2, 6H: 7.27, 3, 5H: 7.45, 4H:7.20
Pro <sup>31</sup>		4.33	2.16, 2.03	2.36	$\delta$ H: 3.97, 3.85
Gln <sup>32</sup>	8.22	4.22	2.34, 2.16	1.98	
Ala <sup>33</sup>	8.03	4.13	1.25		
Phe <sup>34</sup>	7.51	4.55	3.12		2, 6H: 7.22, 3, 5H: 7.33, 4H:7.15
Leu <sup>35</sup>	8.17	4.34	1.93	1.61	$\delta$ CH <sub>3</sub> :0.85
Pro <sup>36</sup>		4.29	2.29, 2.04	1.89	3.83, 3.58
Gly <sup>37</sup>	8.71	4.15, 3.80			
Cys <sup>38</sup>	7.90	4.72	3.21, 3.04		
Asp <sup>39</sup>	8.49	4.54	2.66		

Per avere informazioni preliminari sulla struttura si considera il grafico delle deviazioni dei *chemical shift* dei protoni  $\alpha$ -CH rispetto ai loro valori di *random coil*<sup>161</sup> per ogni residuo (Fig.3.8).

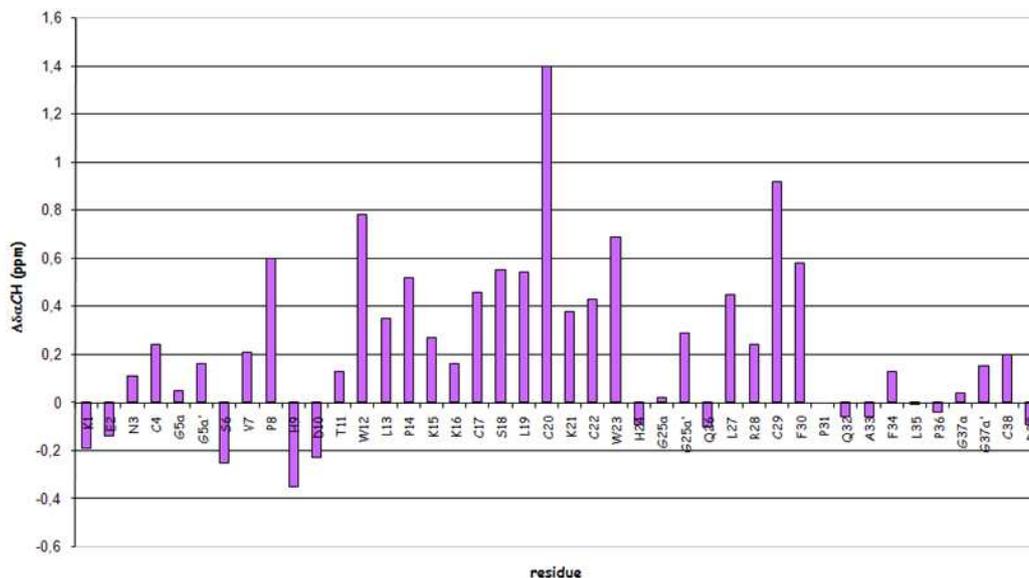


Fig.3.8: Deviazioni dei *chemical shift* (ppm) dei protoni  $\alpha$ -CH di hCFC a pH 7 dai valori *random coil*

Le deviazioni positive che si osservano sono indicative di una conformazione di tipo  $\beta$ -sheet.<sup>161</sup> Dall'analisi degli spettri NOESY 300ms è stato possibile effettuare l'assegnazione in sequenza dei sistemi di spin. Dall'analisi degli spettri NOESY nella zona NH-NH e nella zona NH-alto campo sono stati osservati alcuni contatti NOE *inter-strand*:  $\alpha$ CH C<sup>20</sup>- $\alpha$ CH C<sup>29</sup>,  $\beta$ CH<sub>2</sub> S<sup>6</sup>- $\beta$ CH<sub>2</sub> C<sup>20</sup>,  $\alpha$ CH T<sup>11</sup>- $\beta$ CH<sub>2</sub> S<sup>18</sup>,  $\beta$ CH T<sup>11</sup>- $\beta$ CH<sub>2</sub> S<sup>18</sup>,  $\alpha$ CH C<sup>20</sup>-NH F<sup>30</sup>,  $\alpha$ CH C<sup>29</sup>-NH K<sup>21</sup>. Sono stati misurati un totale di 146 contatti NOE (80 intraresiduo, 40 sequenziali, 26 *long-range*), che sono stati convertiti in altrettante distanze interprotoniche dal programma CALIBA (TabellaA1, Appendice A). Queste distanze sono state utilizzate come limiti superiori per la costruzione di modelli molecolari mediante l'uso del programma CYANA di *torsion angle dynamics*.

### 3.2.2 Calcoli strutturali di hCFC

Un primo insieme di quaranta conformeri è stato calcolato con il programma CYANA usando le distanze interprotoniche, ottenute dai NOE, come limiti superiori (upl). Per ottenere le strutture soddisfacenti i vincoli sperimentali sono state eseguite sette corse CYANA, usando la strategia TAD, modificando opportunamente i contatti di dubbia attribuzione che maggiormente violavano i limiti superiori imposti sulle distanze.

Il valore della *target function* calcolato su 100 strutture dell'ultima corsa CYANA è di  $1.54 \pm 0.70 \text{ \AA}^2$  e quello della deviazione quadratica media calcolata (RMSD) sul *backbone* è di  $4.13 \pm 0.95 \text{ \AA}$ . Dei 146 contatti NOE inseriti (80 intraresiduo, 40 sequenziali, 26 medio raggio) nessun contatto viene violato.

Delle strutture risultanti, le quaranta col più basso valore della *target function* (il valore medio della *target function* è di  $1.05 \pm 0.10 \text{ \AA}^2$ , il valore minimo è di  $0.85 \text{ \AA}^2$  e l' RMSD sul *backbone* è di  $4.21 \pm 0.90 \text{ \AA}$ ) sono state sottoposte a minimizzazione energetica usando il modulo SANDER del programma AMBER 6.0.<sup>139</sup> La statistica dell'analisi strutturale è riportata nella Tabella2A dell' Appendice A.

Le migliori dieci strutture sulle quaranta risultanti dalla minimizzazione energetica sono state selezionate in base alla loro energia e all'accordo con il dato sperimentale; in particolare, sono state scelte le strutture con energia non superiore a  $-157 \text{ kcal mol}^{-1}$ , il valore medio delle energie AMBER delle dieci strutture. Questi 10 conformeri, rappresentativi dello spazio conformazionale accessibile al peptide, sono stati analizzati con il programma di grafica molecolare MOLMOL.<sup>141</sup>

La sovrapposizione delle 10 strutture con la struttura media rivela un certo grado di flessibilità con un RMSD sull'intero *backbone* (1-39) pari a  $4.74 \pm 1.09 \text{ \AA}$ . Si osserva invece una minore flessibilità, e quindi un maggiore accordo, per alcuni tratti della sequenza: la sovrapposizione del *backbone* nei tratti 20-29 e 29-38 mostra un RMSD pari a  $1.80 \pm 0.56 \text{ \AA}$  e  $2.73 \pm 0.56 \text{ \AA}$ , rispettivamente.

Il modello molecolare medio di hCFC ottenuto dall'analisi conformazionale in tampone fosfato/D<sub>2</sub>O a pH 7, riportato in Fig.3.9, presenta una forma ellissoidale compatta con dimensioni approssimative di  $24 \text{ \AA}$  (h, altezza)  $\times$   $15 \text{ \AA}$  (l, larghezza)  $\times$   $12 \text{ \AA}$  (d, profondità).<sup>162</sup> Il *fold* della molecola è globalmente esteso con la presenza di tre *strand* antiparalleli collegati dai ponti disolfuro.

Nel modello molecolare sono evidenziati in rosa gli amminoacidi (H<sup>9</sup> e W<sup>12</sup>) ritenuti coinvolti nel *binding* con il recettore ALK4,<sup>118</sup> che mostrano le catene laterali esposte al solvente (Fig.3.9A).

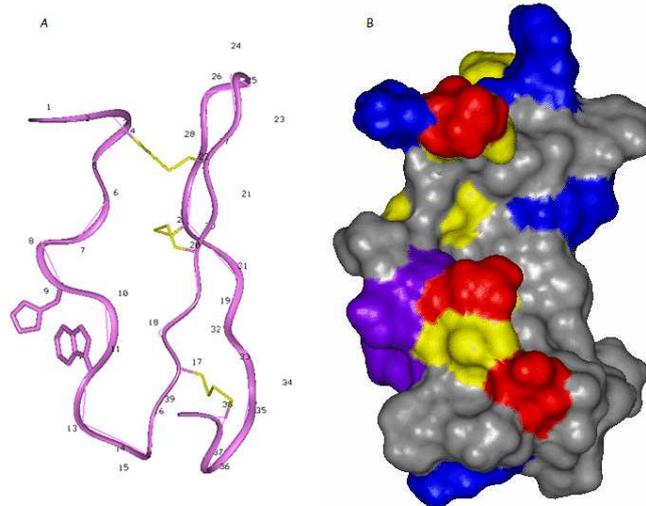


Fig.3.9: (A) Rappresentazione con *ribbon* del modello molecolare medio di hCFC dopo minimizzazione energetica AMBER. Le catene laterali dell'H<sup>9</sup> e W<sup>12</sup> sono evidenziate; (B) Superficie di *Connolly* del modello molecolare medio di hCFC. I residui sono colorati per tipo: acidi in rosso, basici in blu, polari in giallo e idrofobici in grigio. I residui del sito di *binding* H<sup>9</sup> e W<sup>12</sup> sono colorati in viola

Inoltre, in grigio è evidenziato il *patch* idrofobico (Fig.3.9B) dalla parte opposta al sito di *binding* (ad una distanza di circa 22 Å), sito ritenuto responsabile dell'interazione di Cripto con la membrana cellulare.<sup>118</sup>

### 3.3 Analisi conformazionale di mCFC a pH 7 in tampone fosfato

Il dominio CFC 96-134 della proteina Cripto da topo (mCFC), la cui sequenza è riportata in Fig.3.10, è costituito da 39 residui amminoacidici. Il polipeptide è stato ottenuto mediante la sintesi in fase solida presso l'IBB del CNR di Napoli, sotto forma di derivato acetilato all'N-terminale ed amidato al C-terminale.<sup>114</sup> Nella discussione che seguirà, per semplicità di descrizione, il primo residuo della sequenza (K<sup>96</sup>) sarà numerato come primo residuo (K<sup>1</sup>).

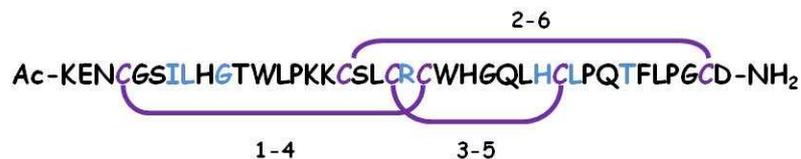


Fig.3.10: Sequenza amminoacidica del dominio CFC nella proteina Cripto da topo

Il dominio CFC da topo differisce da quello umano per sette residui amminoacidi, che nella sequenza sono evidenziati in blu (Fig.3.10).

Per il dominio CFC sono stati registrati spettri CD a diversi valori di pH (Fig.3.11) presso l'IBB. Gli spettri del polipeptide rifoldato sono stati registrati su soluzioni con una concentrazione di peptide di  $1.0 \times 10^{-3}$  M. Sono state preparate tre soluzioni separate: (i) 10 mM di tampone fosfato, pH 7.0; (ii) 10 mM di tampone acetato, pH 5.0 e (iii) 10 mM di tampone citrato, pH 3.0.<sup>114</sup>

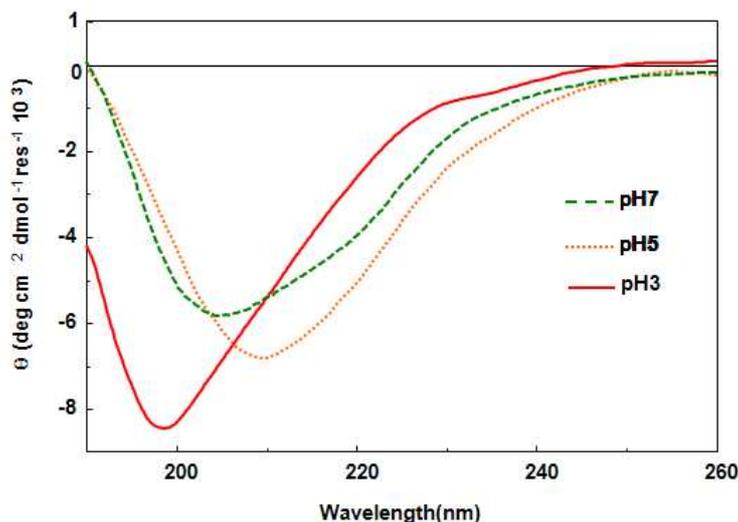


Fig.3.11: Spettri CD del dominio CFC da topo rifoldato. La linea tratteggiata verde si riferisce al campione in 10 mM di tampone fosfato, pH 7.0; la linea a punti arancione si riferisce al campione 10 mM di tampone acetato, pH 5.0 e la linea rossa si riferisce al campione 10 mM di tampone citrato, pH 3.0

L'analisi CD ha mostrato che la molecola non assume conformazioni canoniche, sebbene le curve sono indicative di strutture organizzate con contributi misti di struttura alfa e beta.

### 3.3.1 Analisi NMR di mCFC

L'analisi conformazionale NMR di mCFC è stata condotta in tampone fosfato a pH 7, come per hCFC. Il dominio mCFC era stato precedentemente caratterizzato mediante NMR sia a pH 3 sia a pH 6.<sup>116</sup> L'assegnazione sequenziale di tutte le risonanze protoniche è stata ottenuta mediante l'esame comparato degli spettri TOCSY, DQF-COSY e NOESY, secondo il metodo classico di Wüthrich.<sup>130</sup>

In Fig.3.12A è riportata un'espansione dello spettro TOCSY 70 ms, dove sono messi in evidenza alcuni sistemi di spin.

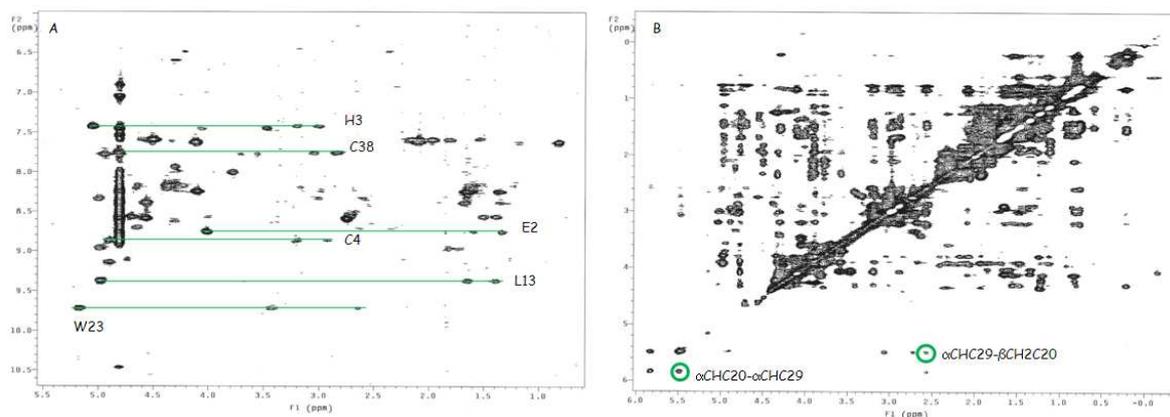


Fig.3.12: Spettri bidimensionali di mCFC in tampone fosfato/D<sub>2</sub>O 90/10 (v/v) a pH 7 a 298 K a 500 MHz: (A) Zona NH-alto campo dello spettro TOCSY 70 ms; (B) Zona alto campo-alto campo dello spettro NOESY 300 ms

Dall'analisi dello spettro DQF-COSY, è stato possibile confermare l'assegnazione dei sistemi di spin, sia delle glicine attraverso i contatti diretti NH- $\alpha$ CH sia dei protoni delle catene laterali. I *chemical shift* di tutti i sistemi di spin sono riportati in **Tabella3.2**.

**Tabella3.2:** *Chemical shift* (ppm) di mCFC in tampone fosfato/D<sub>2</sub>O 90/10 (v/v) pH 7, 298 K

Residuo	NH	$\alpha$ H	$\beta$ H	$\gamma$ H	altro
Lys <sup>1</sup>	8.24	4.09	1.34	1.69	$\epsilon$ CH <sub>2</sub> :2.98, $\epsilon$ NH:7.62, Ac.:2.01
Glu <sup>2</sup>	8.75	4.01	1.85,1.60	1.32	
His <sup>3</sup>	7.41	5.04	3.18,3.00		2H:7.00, 4H: 7.92
Cys <sup>4</sup>	8.87	4.90	3.19,2.92		
Gly <sup>5</sup>					
Ser <sup>6</sup>		4.25	3.95,3.86		
Ile <sup>7</sup>	7.61	4.11	2.09	1.17	$\delta$ CH <sub>3</sub> :0.81
Leu <sup>8</sup>	7.92	4.28	1.58	1.15	$\delta$ CH <sub>3</sub> : 0.82,0.21
His <sup>9</sup>		3.87	3.19, 2.68		4H: 6.62
Gly <sup>10</sup>	9.13	4.89			
Thr <sup>11</sup>	7.99	4.31	3.78	1.24	
Trp <sup>12</sup>		5.48	3.09,2.93		2H:7,15, 4H: 7.04, 5H:6.85; 6H:7.25, 7H:7.33
Leu <sup>13</sup>	9.36	4.97	1.64,1.46	1.38	$\delta$ CH <sub>3</sub> :0.92, 0.85
Pro <sup>14</sup>	--	4.83	2.40,2.10	2.30,2.20	$\delta$ H: 3.95, 3.80
Lys <sup>15</sup>	8.38	4.56	1.90,1.77	1.33	$\epsilon$ CH <sub>2</sub> :2.97, $\epsilon$ NH:7.62
Lys <sup>16</sup>		3.88	1.82,1.72	1.48	$\epsilon$ CH <sub>2</sub> :2.98, $\epsilon$ NH:7.62
Cys <sup>17</sup>		4.82	3.34,2.70		

<b>Ser<sup>18</sup></b>	7.77	4.93	3.70,3.55		
<b>Leu<sup>19</sup></b>	9.10	4.75	1.92	1.65	$\delta\text{CH}_3$ :0.91, 0.85
<b>Cys<sup>20</sup></b>	8.45	5.85	3.26,2.55		
<b>Arg<sup>21</sup></b>	9.16	4.96	1.46	1.38	
<b>Cys<sup>22</sup></b>	8.33	4.99	2.99,2.59		
<b>Trp<sup>23</sup></b>	9.70	5.16	3.42,2.64		2H: 7.41, 4H: 7.25, 5H:6.89 6H:7.33, 7H:7.64, NH:10.45
<b>His<sup>24</sup></b>	8.62	4.31	3.42		2H:6.88, 4H: 7.27
<b>Gly<sup>25</sup></b>	7.45	4.07,3.46			
<b>Gln<sup>26</sup></b>	7.58	4.49	2.11,1.96	1.79	CONH <sub>2</sub> : 7.79,6.52
<b>Leu<sup>27</sup></b>	8.55	4.69	1.50	1.38	$\delta\text{CH}_3$ :0.79
<b>His<sup>28</sup></b>		4.05	3.19,2.89		2H:6.33, 4H: 7.64
<b>Cys<sup>29</sup></b>		5.52	3.05,2.73		
<b>Leu<sup>30</sup></b>	8.96	4.98	1.78,1.70	1.88	$\delta\text{CH}_3$ :1.12
<b>Pro<sup>31</sup></b>			2.32,2.17	2.08	$\delta\text{H}$ : 3.70,3.59
<b>Gln<sup>32</sup></b>	8.70	4.65	2.64	2.33	CONH <sub>2</sub> : 7.50,6.92
<b>Thr<sup>33</sup></b>	8.18	4.37	4.22	1.20	
<b>Phe<sup>34</sup></b>	8.18	4.65	3.12,3.05		
<b>Leu<sup>35</sup></b>	7.57	4.63	1.52	1.52	$\delta\text{CH}_3$ :0.92
<b>Pro<sup>36</sup></b>		4.34	2.34,2.03	2.31,1.95	$\delta\text{H}$ :3.87,3.58
<b>Gly<sup>37</sup></b>	7.66	4.16,3.83			
<b>Cys<sup>38</sup></b>	7.75	4.81	3.03,2.74		
<b>Asp<sup>39</sup></b>	8.58	4.56	2.74		CONH <sub>2ter</sub> :7.44,7.06

In Fig.3.13 si riportano le deviazioni dei *chemical shift* dei protoni  $\alpha$ -CH di mCFC (verde), a confronto con le deviazioni osservate per hCFC (viola).

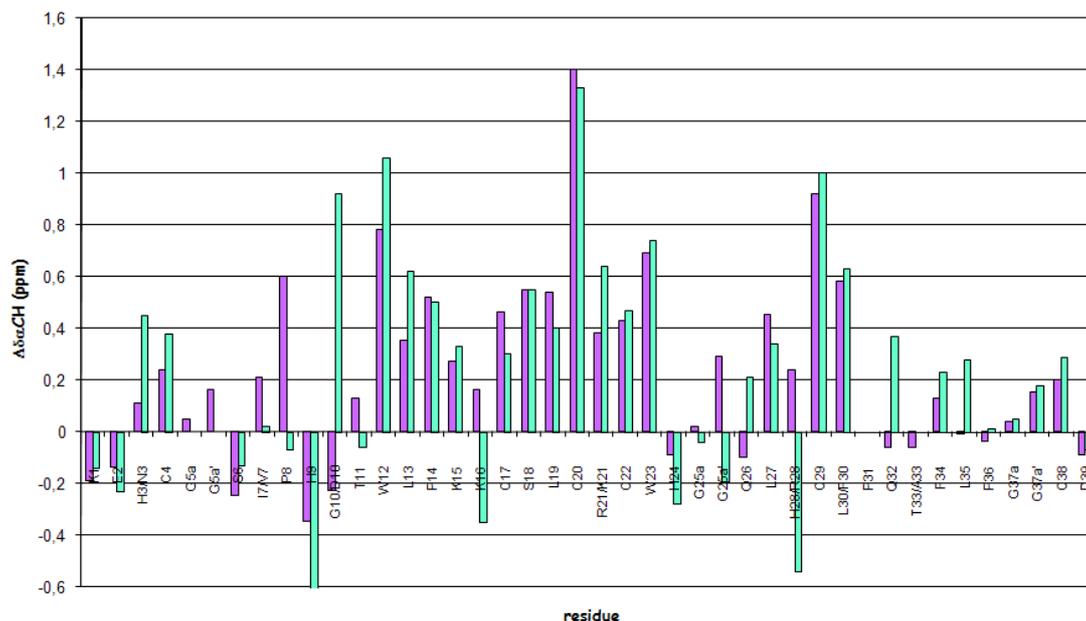


Fig.3.13: Deviazioni dei *chemical shift* (ppm) dei protoni  $\alpha$ -CH di mCFC (verde) e hCFC (viola) a pH 7 dai valori *random coil*

Per il dominio mCFC si osservano deviazioni positive, che sono indicative di una conformazione di tipo  $\beta$ -*sheet*<sup>61</sup> paragonabili a quelle ottenute per hCFC.

L'assegnazione di mCFC a pH 7 ha presentato notevoli difficoltà. A causa della maggiore velocità di scambio dei protoni ammidici con l' $H_2O$ , sette sistemi di spin sono assenti nella zona di correlazione NH-alto campo dello spettro TOCSY. È stato tuttavia possibile osservare sei sistemi di spin nella zona di correlazione scalare ad alto campo dello spettro TOCSY. Le ambiguità sull'assegnazione di alcuni di questi sistemi sono state risolte tramite l'osservazione dei contatti NOE  $NH_{i+1}-\alpha CH_i$  (come ad es.  $I^7-S^6$ ,  $L^{13}-W^{12}$ ,  $L^{30}-C^{29}$ ). Dall'analisi degli spettri NOESY 300ms è stato possibile effettuare l'assegnazione in sequenza dei sistemi di spin. Dall'analisi dello spettro NOESY nella zona ad alto campo, si è potuta confermare la presenza dei ponti disolfuro tra le cisteine, grazie all'osservazione dei contatti NOE, 4-90-2.98 ppm tra  $\alpha CH C^4-\beta CH_2 C^{22}$ , 5.85-5.52 ppm tra  $\alpha CH C^{20}-\alpha CH C^{29}$  e 5.52-2.57  $\alpha CH C^{29}-\beta CH_2 C^{20}$ . Il restante ponte disolfuro ( $C^{17}-C^{38}$ ) è stato assegnato per esclusione non potendosi osservare contatti NOE non ambigui tra le due cisteine per la vicinanza del *chemical shift* del protone  $\alpha CH$  di entrambi (4.82 ppm per la  $C^{17}$  e 4.81 ppm per la  $C^{38}$ ). Dall'analisi degli spettri NOESY nella zona NH-NH e nella zona NH-alto campo sono stati osservati alcuni contatti NOE *inter-strand*:  $NH C^4-\alpha CH H^{24}$ ,  $NH L^8-\beta CH_2 L^{19}$ ,  $NH L^{30}-\beta CH_2 S^{18}$ ,  $NH L^{19}-NH L^{30}$ ,  $NH L^{19}-NH L^{32}$ ,  $NH L^{30}-\beta CH_2 C^{20}$ .

Sono stati misurati un totale di 200 contatti NOE (109 intraresiduo, 54 sequenziali, 37 *long-range*), che sono stati convertiti in altrettante distanze interprotoniche dal programma CALIBA (TabellaB1, Appendice B). Queste distanze sono state utilizzate come limiti superiori per la costruzione di modelli molecolari mediante l'uso del programma CYANA di *torsion angle dynamics*.

### 3.3.2 Calcoli strutturali di mCFC

Un primo insieme di quaranta conformeri è stato calcolato con il programma DYANA usando le distanze interprotoniche, ottenute dai NOE, come limiti superiori (upl). Per ottenere le strutture soddisfacenti i vincoli sperimentali sono state eseguite otto corse CYANA, usando la strategia TAD, modificando opportunamente i contatti di dubbia attribuzione che maggiormente violavano i limiti superiori imposti sulle distanze.

Il valore della *target function* calcolato su 100 strutture dell'ultima corsa CYANA è di  $2.34 \pm 1.87 \text{ \AA}^2$  e quello della deviazione quadratica media calcolata (RMSD) sul *backbone* è di  $3.86 \pm 0.75 \text{ \AA}$ . Dei 200 contatti NOE (109 intraresiduo, 54 sequenziali, 37 *long-range*) nessun contatto viene violato.

Delle strutture risultanti, le quaranta col più basso valore della *target function* (il valore medio della *target function* è di  $1.03 \pm 0.12 \text{ \AA}^2$ , il valore minimo è di  $0.81 \text{ \AA}^2$  e l' RMSD sul *backbone* è di  $3.24 \pm 0.58 \text{ \AA}$ ) sono state sottoposte a minimizzazione energetica usando il modulo SANDER del programma AMBER 6.0.<sup>139</sup> La statistica dell'analisi strutturale è riportata nella TabellaB2 dell'Appendice B.

Le migliori dieci strutture sulle quaranta risultanti dalla minimizzazione energetica sono state selezionate in base alla loro energia e all'accordo con il dato sperimentale; in particolare, sono state scelte le strutture con energia non superiore a  $-74 \text{ kcal mol}^{-1}$ , il valore medio delle energie AMBER delle dieci strutture. Questi 10 conformeri, rappresentativi dello spazio conformazionale accessibile al peptide, sono stati analizzati con il programma di grafica molecolare MOLMOL.<sup>141</sup>

La sovrapposizione delle 10 strutture con la struttura media rivela un certo grado di flessibilità con un RMSD sull'intero *backbone* (1-39) pari a  $4.40 \pm 1.19 \text{ \AA}$ . Si osserva invece una minore flessibilità, e quindi un maggiore accordo, per il tratto 20-29 della sequenza dove la sovrapposizione del *backbone* mostra un RMSD pari a  $0.88 \pm 0.26 \text{ \AA}$ .

Il modello molecolare medio di mCFC ottenuto dall'analisi conformazionale in tampone fosfato/D<sub>2</sub>O a pH 7, riportato in Fig.3.14, presenta una forma ellissoidale compatta con

dimensioni approssimative di 22 Å (h, altezza) × 14 Å (l, larghezza) × 18 Å (d, profondità). Il *folding* della molecola è globalmente esteso con la presenza di tre *strand* antiparalleli collegati dai ponti disolfuro.

Nel modello molecolare sono evidenziati in verde gli amminoacidi (H<sup>9</sup> e W<sup>12</sup>) ritenuti coinvolti nel *binding* con il recettore ALK4,<sup>118</sup> che mostrano le catene laterali esposte al solvente (Fig.3.14A).

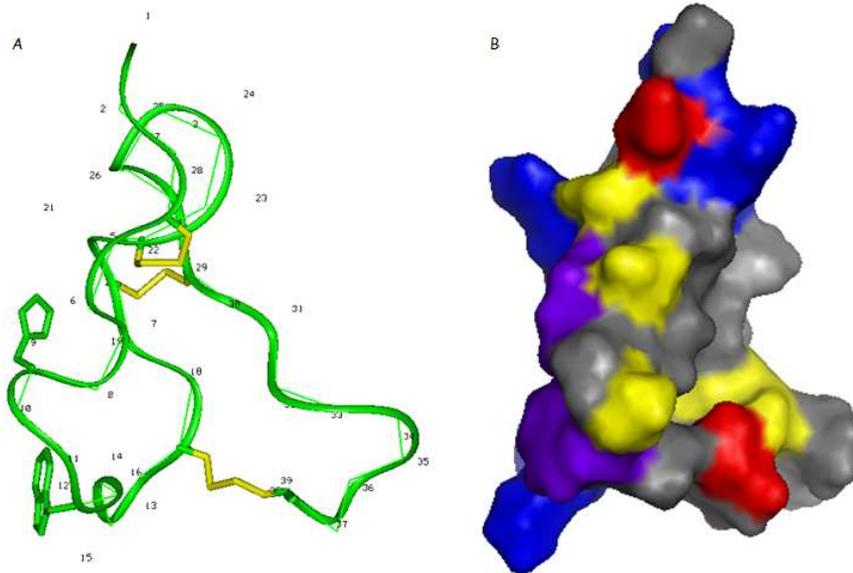


Fig.3.14: (A) Rappresentazione con *ribbon* del modello molecolare medio di mCFC dopo minimizzazione energetica AMBER. Le catene laterali dell'H<sup>9</sup> e W<sup>12</sup> sono evidenziate; (B) Superficie di *Connolly* del modello molecolare medio di mCFC. I residui sono colorati per tipo: acidi in rosso, basici in blu, polari in giallo e idrofobici in grigio. I residui del sito di *binding* H<sup>9</sup> e W<sup>12</sup> sono colorati in viola

Inoltre, in grigio è evidenziato il *patch* idrofobico (Fig. 3.14B) dalla parte opposta al sito di *binding* (ad una distanza di circa 18 Å) sito ritenuto responsabile dell'interazione di Cripto con la membrana cellulare.<sup>118</sup>

I modelli molecolari ottenuti per hCFC e mCFC sono molto simili a quelli precedentemente calcolati per il dominio CFC da topo a pH 3 e a pH 6.<sup>116</sup>

### 3.4 Modello per omologia di ALK4-ECD e ALK7-ECD

La struttura dei recettori di tipo I dell'Attivina, che sono coinvolti nel *signaling* della proteina Nodal, fino ad oggi non è stata determinata. Abbiamo costruito, quindi, un modello per omologia (HM) del dominio extracellulare del recettore umano di tipo I dell'Attivina, ALK4-ECD, che interagisce con il dominio CFC della proteina Cripto durante il *signaling*<sup>103,113,114</sup> e un

modello per il dominio extracellulare del recettore umano di tipo I dell'Attivina, ALK7-ECD, che interagisce direttamente con la proteina Nodal.<sup>163</sup> I recettori ALK4 e ALK7 intervengono entrambi nello sviluppo embrionale, ma ALK4 si trova nelle cellule dell'endoderma, che daranno vita ad intestino, sistema respiratorio, fegato, etc., mentre ALK7 si trova anche nel mesoderma, cellule che origineranno vasi sanguigni, cuore, etc.<sup>163</sup> Inoltre è stato proposto che ALK7 intervenga nella crescita cellulare degli adulti in combinazione con ligandi diversi da Nodal.<sup>46</sup>

#### HM di ALK4

Per la costruzione del modello per omologia di ALK4-ECD abbiamo usato come template la struttura cristallina del dominio extracellulare del recettore umano di tipo I delle BMP, ALK3-ECD (BRIA-ECD), in complesso con la BMP2 (pdbcode:1rew, risoluzione 1.86 Å).<sup>69</sup>

I domini extracellulari di ALK3 e ALK4 fanno parte della famiglia Pfam dei recettori dell'Attivina (PF01064) e presentano un'omologia di sequenza del 27%. Questi domini extracellulari sono idrofili e sono caratterizzati da una regione di nove amminoacidi (CCX<sub>4</sub>-<sub>5</sub>CN), ricca di cisteine (box di cisteine).<sup>17</sup> I programmi PSIBLAST,<sup>144</sup> HHpred<sup>146</sup> e mGentHreader<sup>145</sup> hanno identificato il recettore umano ALK3 (pdbcode:1rew) come miglior template per l'*homology modelling* per il dominio extracellulare del recettore umano ALK4 (E-values minore di 10<sup>-13</sup>). I tre programmi hanno fornito allineamenti *target*-template con le dieci cisteine coinvolte nei ponti disolfuro allineate correttamente e coincidenti in 72 posizioni su 91 dell'allineamento (Fig.3.15), permettendo di identificare un esteso core della proteina.

ALK4_ECD	5	GVQALLCACT	-SCLQ--ANY	TCETDGACMV	SIFNL-DGME	HHVRTCIPKV
ALK3_ECD	32	TLFFLKCYCS	GHCPPDAINN	TCITNGHCEA	IIEEDDQGET	TLASGCMKY-
		.	* * *	.	* ** * *	* . *
ALK4_ECD	51	ELVPAGKPFY	CLSSE---DL	RNTHCCYTDY	CNRIDLRVPS	G
ALK3_ECD	82	----EGSDFQ	CKDSPKAQLR	RTIECCRTNL	CNQYLQPTLP	P
		*	* * *	*	** * *	**.

Fig.3.15: Allineamento *target*-template di HHpred, basato sul confronto dei HMM-HMM. Le regioni con le box grigie sono conservate nel confronto degli allineamenti ottenuti da PSI-BLAST e mGentHreader

Tutti e tre gli allineamenti ottenuti da PSIBLAST, HHpred e mGenThreader sono stati adoperati per la costruzione del modello molecolare di ALK4-ECD, usando il programma Modeller6v2.<sup>148</sup> La qualità dei modelli ottenuti è stata appurata utilizzando alcuni programmi di controllo qualità, come WhatCheck,<sup>151,152</sup> Verify3D<sup>153,154</sup> e ProQ.<sup>155,156</sup> Il miglior modello in

accordo con gli *score* di qualità, è risultato quello costruito sulla base dell'allineamento di HHpred, basato sul confronto dei profili di Markov nascosti (Hidden Markov Models, HMM).<sup>147</sup>

Il modello di ALK4-ECD,<sup>162</sup> esibisce lo stesso *fold* a tre-dita di ALK3-ECD, è formato da due  $\beta$ -*sheet*, costituiti da due (residui 5-9 e 17-21) e tre (residui 24-32, 35-43 e 63-69) *strand*, rispettivamente; un piccolo segmento ad  $\alpha$ -elica tra i residui 54-58, e un' $\alpha$ -elica aggiuntiva tra i residui 75-78 (Fig.3.16A). Nel modello si osserva una faccia idrofobica, che coinvolge i residui 42-58, seguita da un *patch* di residui polari (S<sup>28</sup>, S<sup>59</sup>, S<sup>60</sup>, N<sup>65</sup>, T<sup>66</sup>, S<sup>83</sup>). Il sito opposto della molecola è altamente idrofilico, con una successione di *patch* di residui acidi (E<sup>20</sup>, D<sup>22</sup>, E<sup>47</sup>), polari (T<sup>10</sup>, S<sup>11</sup>, Q<sup>14</sup>, N<sup>16</sup>, T<sup>18</sup>) e basici (H<sup>37</sup>, H<sup>38</sup>, R<sup>40</sup>, R<sup>76</sup>) (Fig.3.16B).

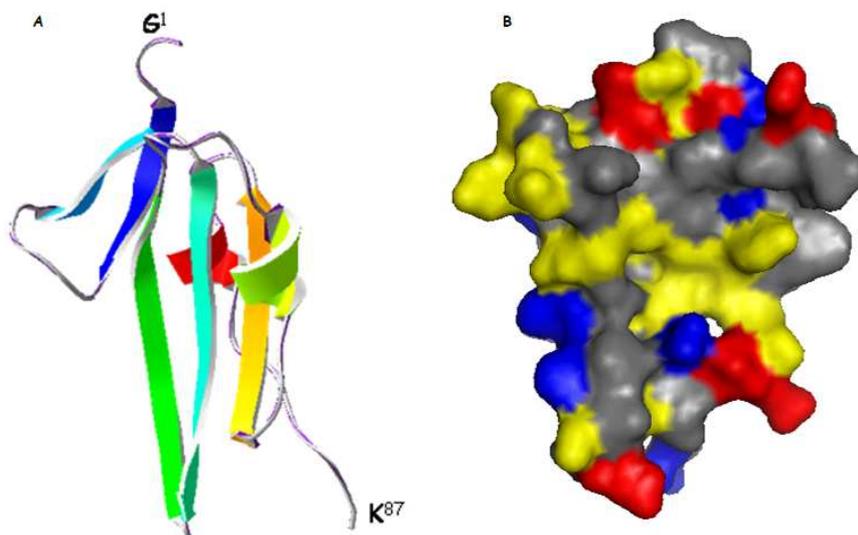


Fig.3.16: (A) Rappresentazione con *ribbon* del modello di ALK4-ECD. Il residuo G<sup>1</sup> corrisponde al residuo G<sup>5</sup> dell'allineamento (Fig.3.15); (B) Superficie di *Connolly* del modello di ALK4-ECD. I residui sono colorati per tipo: acidi in rosso, basici in blu, polari in giallo e idrofobici in grigio

Studi di mutagenesi, effettuati da Harrison et al. (2003)<sup>164</sup> su ALK4, hanno identificato cinque residui idrofobici (L<sup>13</sup>, I<sup>43</sup>, V<sup>46</sup>, L<sup>48</sup>, P<sup>50</sup>) come quelli rilevanti per il *binding* di ALK4 all'Attivina; questi residui insieme ad altri possono costituire l'interfaccia di *binding* per i ligandi. E' stato, anche, proposto che i ligandi TGF- $\beta$  hanno un'alta affinità di legame per la superficie concava idrofobica dei recettori di tipo I,<sup>22</sup> dove appunto i residui su menzionati si trovano. I residui V<sup>46</sup>, L<sup>48</sup>, e P<sup>50</sup>, tuttavia, si trovano nella regione di ALK4 che ha la più bassa omologia di sequenza con ALK3. La regione idrofobica di ALK4, formata dai residui 42-58, comprende un *loop* pre-elica (42-53) e una piccola  $\alpha$ -elica formata tra i residui 54-58. I residui V<sup>46</sup> e I<sup>43</sup>, ma non la L<sup>13</sup>, si trovano in prossimità del *patch* idrofobico e potrebbero

giocare un ruolo importante nel *binding* al ligando. Il *patch* idrofobico, inoltre, contiene i residui F<sup>55</sup> e la Y<sup>56</sup>, che hanno le catene laterali ben orientate e potrebbero interagire con il sito di *binding* (H<sup>9</sup> e W<sup>12</sup>) del dominio CFC. Il residuo F<sup>55</sup>, che corrisponde al residuo F<sup>85</sup> di ALK3, che è stato previsto essere un residuo fondamentale nell'interazione tra il recettore di tipo I e i ligandi TGF-β, non è assolutamente richiesto nel *binding* tra l'Attivina A e ALK4.<sup>26</sup> Studi di mutagenesi più recenti<sup>165</sup> hanno, però, dimostrato che l'Attivina e CFC hanno un diverso sito di *binding* sul recettore ALK4. Questi studi confermano, inoltre, che i residui ritenuti responsabili del *binding* con l'Attivina, ma non con Cripto, sono l'I<sup>43</sup>, la L<sup>48</sup> e la P<sup>50</sup>. Pertanto non essendoci prove certe sul mancato coinvolgimento del residuo F<sup>55</sup> di ALK4 nel *binding* a CFC, e considerata l'importanza di questo residuo nel *binding* di altri TGF-β, nella selezione dei complessi del *docking*, sono state considerate le possibili interazioni tra i residui F<sup>55</sup> e Y<sup>56</sup> con CFC. Questa scelta è supportata ulteriormente dalla considerazione della rilevanza cruciale degli anelli aromatici nel *binding* al recettore.<sup>43</sup>

#### HM di ALK7

Per la costruzione del modello per omologia di ALK7-ECD abbiamo usato come templatato la struttura cristallina del dominio extracellulare del recettore umano di tipo I dei TGF-β, ALK5-ECD (TβRI-ECD), nel complesso ternario con il ligando TGF-β3 e con il recettore di tipo II TβRII (pdbcode:2pjy, risoluzione 3.00 Å).<sup>63</sup>

I domini extracellulari di ALK5 e ALK7 fanno parte della famiglia Pfam dei recettori dell'Attivina (PF01064) e presentano un'omologia di sequenza del 39%. Questi domini extracellulari sono idrofilici e sono caratterizzati da una regione di nove amminoacidi (CCX<sub>4</sub>-<sub>5</sub>CN), ricca di cisteine (box di cisteine).<sup>17</sup> I programmi PSIBLAST,<sup>144</sup> HHpred<sup>146</sup> e mGentHreader<sup>145</sup> hanno identificato il recettore umano ALK5 (pdbcode: 2pjy) come miglior templatato per l'*homology modelling* per il dominio extracellulare del recettore umano ALK7 (E-value minore di 10<sup>-14</sup>). I tre programmi hanno fornito allineamenti *target*-templatato con le dieci cisteine coinvolte nei ponti disolfuro allineate correttamente e coincidenti in 74 posizioni su 80 dell'allineamento (Fig.3.17), permettendo di identificare un esteso core della proteina.

```

ALK7_ECD 4  GLKCVLLLCD  SSNFTCQTEG  ACWASVMLTN  GKEQVIKSCV  SLPEL---NA
ALK5_ECD 9  ALQCFCHLCT  KDNFTCVTDG  LCFVSVTETT  DKVIHNSSCI  AEIDLIPDR
          . * * * * *      * * * * * * * * * *      * * * * * * * * * *
          . * * * * *      * * * * * * * * * *      * * * * * * * * * *

ALK7_ECD 50 QVFC HSSN--  -NVTKTECCF  TDFCNNITL
ALK5_ECD 59 PFVCA PSSKT  GSVTTTYCCN  QDHCNKIEL
          * _ * _ _      * * * * *      * * * * * * * * * *
    
```

Fig.3.17: Allineamento *target*-templatato di HHPred, basato sul confronto dei HMM-HMM. Le regioni con le box grigie sono conservate nel confronto degli allineamenti ottenuti da PSI-BLAST e mGentHreader

Tutti e tre gli allineamenti ottenuti da PSIBLAST, HHpred e mGenThreader sono stati adoperati per la costruzione del modello molecolare di ALK7-ECD, usando il programma Modeller6v2.<sup>148</sup> La qualità dei modelli ottenuti è stata appurata utilizzando alcuni programmi di controllo qualità, come WhatCheck,<sup>151,152</sup> Verify3D<sup>153,154</sup> e ProQ.<sup>155,156</sup> Il miglior modello in accordo con gli *score* di qualità, è risultato quello costruito sulla base dell'allineamento di HHpred, basato sul confronto dei profili di Markov nascosti (Hidden Markov Models, HMM).<sup>147</sup>

Il modello di ALK7-ECD, esibisce lo stesso *fold* a tre-dita di ALK5-ECD, è formato da due  $\beta$ -*sheet*, costituiti da due (residui 2-5 e 15-18) e tre (residui 21-27, 34-40 e 59-62) *strand*, rispettivamente, e due piccoli segmenti ad  $\alpha$ -elica tra i residui 10-13 e tra i residui 42-44 (Fig.3.18A). Nel modello si osserva che la superficie concava di *binding* è caratterizzata da una distribuzione di residui polari e idrofobici (Fig.3.18B).

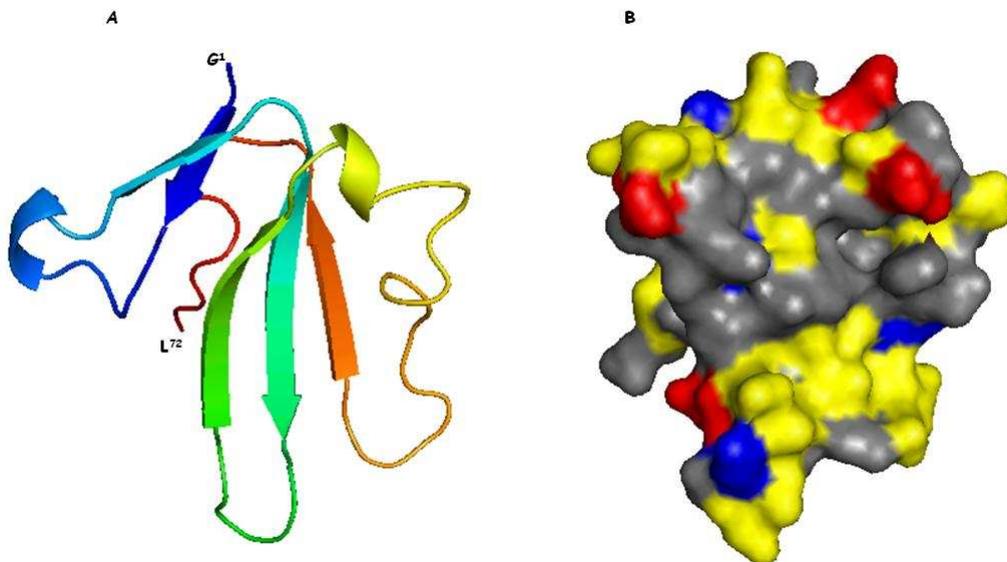


Fig.3.18: (A) Rappresentazione con *ribbon* del modello di ALK7-ECD. Il residuo  $G^1$  corrisponde al residuo  $G^{54}$  dell'allineamento (Fig.3.17); (B) Superficie di *Connolly* del modello di ALK7-ECD. I residui sono colorati per tipo: acidi in rosso, basici in blu, polari in giallo e idrofobici in grigio

Il modello per omologia di ALK7-ECD sarà utilizzato successivamente per prove di *docking* con la proteina Nodal (vedi par. 3.8).

### 3.5 Docking tra ALK4-ECD e hCFC

Per investigare l'interazione tra il dominio CFC di Cripto e il recettore ALK4, abbiamo effettuato il *docking* delle migliori dieci strutture NMR di hCFC (chiamate hCFC1-10) con il modello per omologia del dominio extracellulare di ALK4, utilizzando il programma di *docking*

rigido 3D-Dock.<sup>157,166</sup> Le soluzioni del *docking* sono state filtrate sulla base degli studi di mutagenesi, che indicano i residui H<sup>9</sup> e W<sup>12</sup> come cruciali per il *binding* al recettore ALK4. Come conseguenza di questa procedura di filtro, le 10000 soluzioni iniziali del *docking* sono state ridotte a circa dieci per ciascuna struttura NMR. Da un'ispezione visiva delle strutture filtrate, sono stati considerati solo quei complessi che esibivano H<sup>9</sup> e W<sup>12</sup> di hCFC ben orientate verso l'interfaccia concava idrofobica di legame di ALK4. Queste soluzioni del *docking*, ordinate dal loro valore dell'RPScore, sono state minimizzate energeticamente dal programma MULTIDOCK.

Per la nostra analisi sono state considerate solo le soluzioni del *docking* delle strutture NMR 3 e 5, poiché gli altri complessi sono stati considerati irrealistici per interazioni elettrostatiche sfavorevoli, sovrapposizioni steriche etc.<sup>162</sup>

Il complesso ottenuto dal *docking* della struttura NMR hCFC5 con ALK4 è stato considerato per primo per il suo valore più alto di RPScore (Fig.3.19).

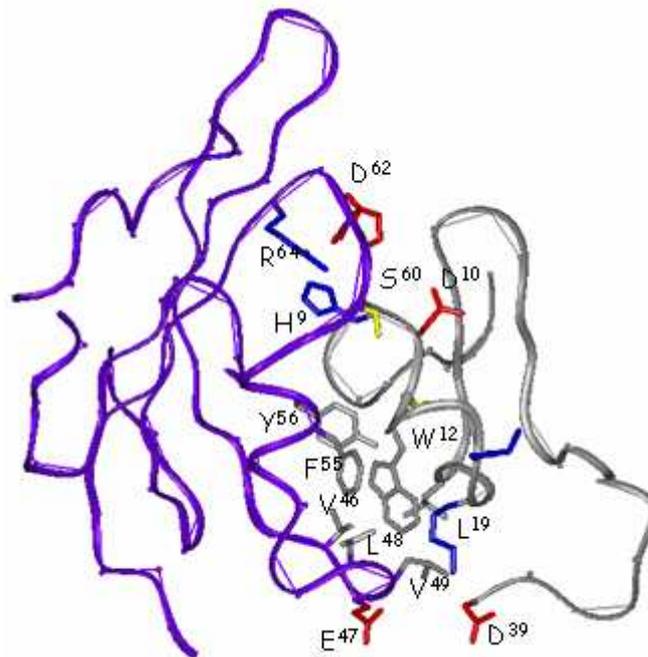


Fig.3.19: Rappresentazione con *ribbon* del complesso tra ALK4-ECD (viola) e hCFC (grigio) ottenuto dalla struttura NMR hCFC5. I residui rilevanti all'interfaccia di *binding* sono mostrati e colorati per tipo: idrofobici in grigio, polari in giallo, acidi in rosso e basici in blu

Questo complesso ha mostrato sia interazioni idrofobiche molto favorevoli sia interazioni molto critiche. Come si può osservare dalla Fig.3.19, il residuo di *binding* W<sup>12</sup> di hCFC5 si inserisce in un *patch* idrofobico formato da V<sup>46</sup>, L<sup>48</sup>, V<sup>49</sup>, F<sup>55</sup>, Y<sup>56</sup> ed esibisce interazioni

aromatiche con gli ultimi due residui. Il residuo L<sup>19</sup> di hCFC5 partecipa a questo *patch* idrofobico, mentre il residuo di *binding* H<sup>9</sup> di hCFC5 è coinvolto in un'interazione elettrostatica con il D<sup>62</sup>. Un'interazione polare-polare si osserva tra la S<sup>60</sup> di ALK4 e la T<sup>11</sup> di hCFC5. In questo complesso, però, si possono osservare anche interazioni elettrostatiche sfavorevoli tra ALK4 E<sup>47</sup>- hCFC D<sup>39</sup>, ALK4 D<sup>62</sup>- hCFC D<sup>10</sup> e ALK4 R<sup>64</sup>- hCFC H<sup>9</sup>. Questo ultimo contatto che coinvolge uno dei residui di *binding* di hCFC, ci ha convinto a scartare questo complesso.

Invece, il complesso con il più basso valore dell'energia totale tra quelli ottenuti dal *docking* della struttura NMR hCFC3, che è la struttura più rappresentativa dell'insieme NMR, ha superato tutti i nostri controlli (Fig.3.20).

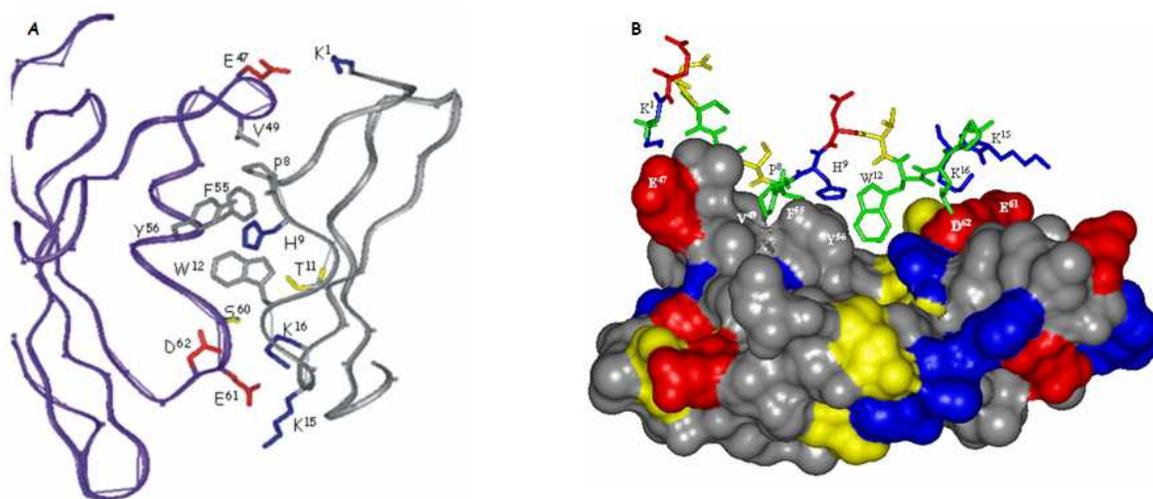


Fig.3.20: (A) Rappresentazione con *ribbon* del complesso tra ALK4-ECD (viola) e hCFC (grigio) ottenuto dalla struttura NMR hCFC3. I residui rilevanti all'interfaccia di *binding* sono mostrati e colorati per tipo: idrofobici in grigio, polari in giallo, acidi in rosso e basici in blu; (B) Rappresentazione dell'interazione all'interfaccia tra ALK4-ECD (superficie di *Connolly* con i residui colorati per tipo) e hCFC (rappresentazione con *stick*)

Il modello ha mostrato interazioni aromatiche favorevoli tra il W<sup>12</sup> di hCFC3 e F<sup>56</sup> di ALK4. Inoltre, l'H<sup>9</sup> di hCFC3 si trova tra gli anelli aromatici della Y<sup>55</sup> e F<sup>56</sup> di ALK4 potenzialmente capaci di creare interazioni aromatiche e/o  $\pi$ -cationiche.<sup>167,168</sup> Possibili interazioni elettrostatiche sono state osservate tra ALK4 E<sup>47</sup>- hCFC3 K<sup>1</sup>, ALK4 D<sup>62</sup>- hCFC3 K<sup>16</sup> e ALK4 E<sup>61</sup>- hCFC3 K<sup>15</sup>. Un'altra interazione favorevole all'interfaccia di *binding* è stata trovata tra i residui idrofobici ALK4 V<sup>49</sup>- hCFC3 P<sup>8,162</sup>

### 3.6 Studio di binding SPR tra hCFC e hEGF-like e i recettori ALK4/ActRIIB

Per stimare la capacità dei domini hCFC ed hEGF-like di riconoscere i recettori di tipo I ALK4 o di tipo II ActRIIB, sono stati effettuati esperimenti di 'Real Time Bia' con tecnica Surface Plasmon Resonance (SPR).

I saggi di *binding* sono stati effettuati attraverso l'iniezione di soluzioni dei domini hCFC ed hEGF-like a concentrazioni crescenti, nel range 0-40  $\mu\text{M}$ , e sottraendo il contributo del riferimento del canale del Fc-derivatizzato.

In Fig. 3.21A è riportato il grafico dei sensorgrammi ottenuti dalle prove di *binding* del dominio hCFC sul recettore ALK4 immobilizzato su *chip*. I sensorgrammi rivelano i profili tipici dell'affinità ligando-recettore e l'analisi cinetica dell'interazione fornisce un valore per la  $K_{on}$  di  $466 \text{ (sec}^{-1}\text{M}^{-1})$  e per la  $K_{off}$  di  $5.3 \times 10^{-3} \text{ (sec}^{-1})$ , portando a un valore per  $K_D$  di  $13 \pm 4 \mu\text{M}$ . Questo valore è consistente con quello ottenuto riportando in grafico l' $\text{RU}_{max}$  (unità di risonanza massima) per ogni singolo esperimento in funzione della concentrazione di hCFC (Fig. 3.21B) ed eseguendo un'analisi di regressione non lineare dei dati.<sup>162</sup>

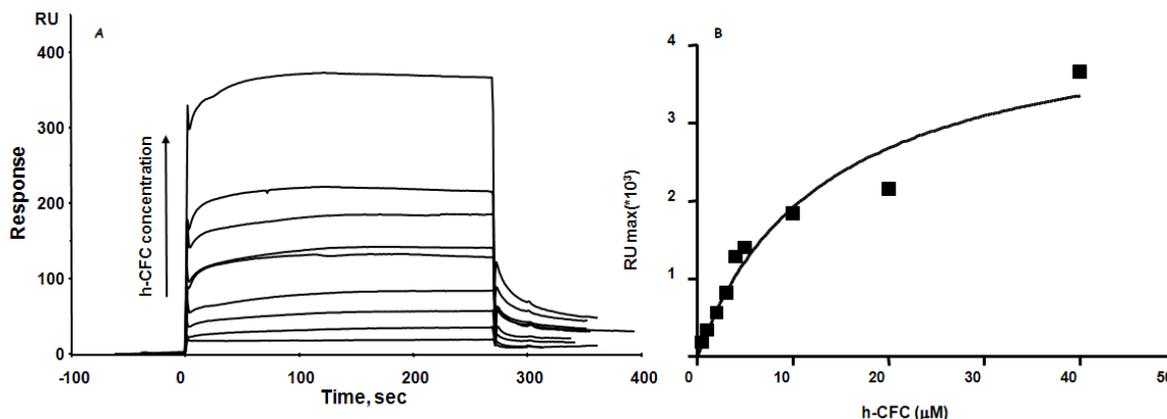


Fig.3.21: (A) Sensorgrammi SPR che mostrano l'interazione tra hCFC e la chimera ALK4-Fc immobilizzata su *chip* CM5. (B) Valore di risposta massima ( $\text{RU}_{max}$ ) di ciascun curva di *binding* riportati verso le concentrazioni di hCFC

In Fig. 3.22A sono riportati i sensorgrammi ottenuti dalle prove di *binding* del dominio hEGF-like sul recettore ALK4 immobilizzato su *chip*. Data le esigue quantità di peptide ottenuto per sintesi chimica, è stato condotto un numero minore di esperimenti dose-risposta. I sensorgrammi registrati comunque rivelano una certa affinità ligando-recettore e l'analisi cinetica dell'interazione fornisce un valore per la  $K_D$  di  $130 \pm 4 \mu\text{M}$  (Fig. 3.22B) inferiore di 10 volte all'affinità esibita dal dominio CFC nelle stesse condizioni.

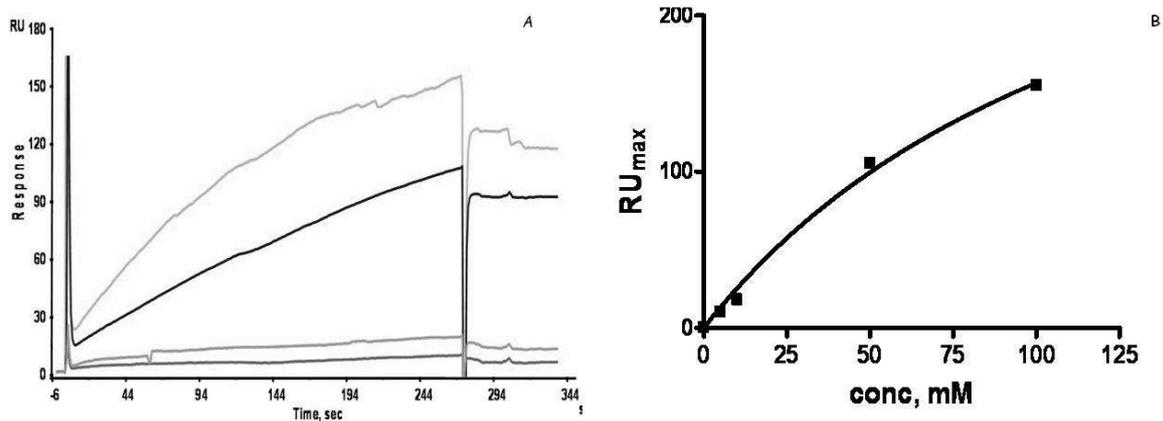


Fig.3.22: (A) Sensorgrammi SPR che mostrano l'interazione tra hEGF-like e la chimera ALK4-Fc immobilizzata su *chip* CM5. (B) Valore di risposta massima ( $RU_{max}$ ) di ciascun curva di *binding* riportati verso le concentrazioni di hEGF-like

Quindi entrambi i domini isolati della proteina Cripto mostrano affinità di *binding* verso il recettore di tipo I ALK4, però il dominio hCFC mostra un'affinità dieci volte maggiore rispetto a hEGF-like.

Al contrario i domini hCFC e hEGF-like non mostrano affinità di *binding* nei confronti del recettore di tipo II ActRIIB. Infatti, immobilizzando su *chip* la chimera ActRIIB-Fc e iniettando soluzioni a concentrazione crescente dei due domini separatamente si registrano sensorgrammi che indicano l'assenza d'interazione. A titolo di esempio si riportano in Fig.3.23 i sensorgrammi registrati per hCFC e ActRIIB.

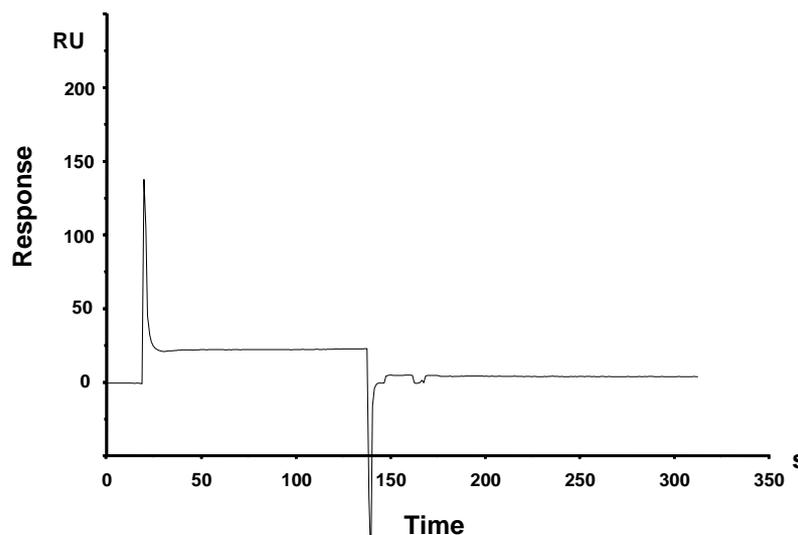


Fig.3.23: Sensorgrammi SPR che mostrano l'assenza d'interazione tra hCFC e la chimera ActRIIB-Fc immobilizzata su *chip* CM5



### 3.7.1 Modello per omologia di Nodal

Per la costruzione del modello per omologia del monomero della proteina Nodal umana abbiamo utilizzato il programma HHpred per la scelta del template. Il programma ci ha fornito cinque template possibili, quattro appartenenti alle BMP (BMP6, BMP9, BMP2 e BMP3) e uno appartenente alle GDF (GDF5). Tutti i template mostrano valori di identità di sequenza di circa il 40% (Tabella 2.1, par. 2.5).

Dei cinque template è stato effettuato un allineamento di struttura secondaria, utilizzando il programma SSM (Secondary Structure Matching)<sup>149</sup> ed è stato costruito un profilo HMM, utilizzando il programma HMMer.<sup>150</sup> A questo profilo è stata allineata la sequenza umana di Nodal (E-value  $5.7 \times 10^{-36}$ ). Per la costruzione dell'HM è stato utilizzato il programma Modeller6v2<sup>148</sup> La qualità dei modelli ottenuti è stata appurata utilizzando alcuni programmi di controllo qualità, come WhatCheck,<sup>151,152</sup> Verify3D<sup>153,154</sup> e ProQ.<sup>155,156</sup> Il modello di Nodal (Fig. 3.26) ottenuto per *homology modelling*, conserva il nodo di cisteina caratteristico della superfamiglia dei TGF- $\beta$ ; la struttura tridimensionale di Nodal è formata da due  $\beta$ -sheet antiparalleli, costituiti da due *strand* (tra i residui 11-19 e 32-40) e quattro *strand* (tra i residui 75-82, 84-90 accoppiati a 103-110 e 93-99, rispettivamente), un piccolo *turn* di elica che lega i primi due *strand* (tra i residui 20-26) e un  $\alpha$ -elica anfipatica (H3) perpendicolare ai  $\beta$ -sheet (tra i residui 55-66).

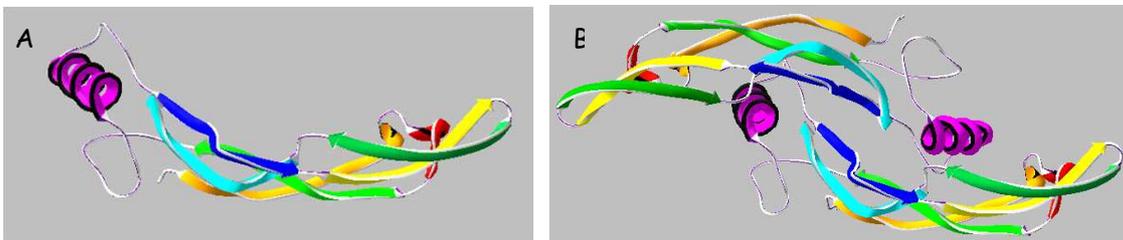


Fig. 3.26: Modello per omologia del monomero (A) e (B) dimero della proteina Nodal umana

La costruzione del modello per omologia del dimero è stata ottenuta con lo stesso procedimento descritto precedentemente per il monomero, utilizzando le strutture dimeriche dei cinque template, ottenute dal programma PDBsum.<sup>169</sup> Si può notare che il modello molecolare adotta l'arrangiamento canonico, simmetrico e allungato dei TGF- $\beta$ .

Sui modelli ottenuti si possono osservare le regioni caratteristiche del *binding* del recettore di tipo I ('*wrist epitope*') e del recettore di tipo II ('*knucle epitope*').

### 3.7.2 Analisi conformazionale del frammento Nodal53-69

Il frammento Nodal53-69 della proteina Nodal umana, la cui sequenza è riportata in Fig. 3.27,

è costituito da 17 residui amminoacidici e riproduce il tratto elicoidale H3 di Nodal. Il polipeptide è stato ottenuto mediante la sintesi in fase solida presso l'IBB del CNR di Napoli, sotto forma di derivato acetilato all'N-terminale ed amidato al C-terminale.



Fig.3.27: Sequenza amminoacidica di Nodal53-69

Per il frammento Nodal53-69 sono stati registrati spettri CD a diversi valori di pH (Fig.3.28) presso l'IBB. Gli spettri CD sono stati registrati in soluzioni acquose a concentrazione di peptide dell'ordine di  $10^{-4}$  M e a diversi tenori di TFE: in tampone fosfato a pH=6.9 (curva verde); in tampone fosfato a pH=6.9 più 10% di TFE (curva blu); in tampone fosfato a pH=6.9 più 30% di TFE (curva rossa) e in tampone fosfato a pH=6.9 più 70% di TFE (curva fucsia).

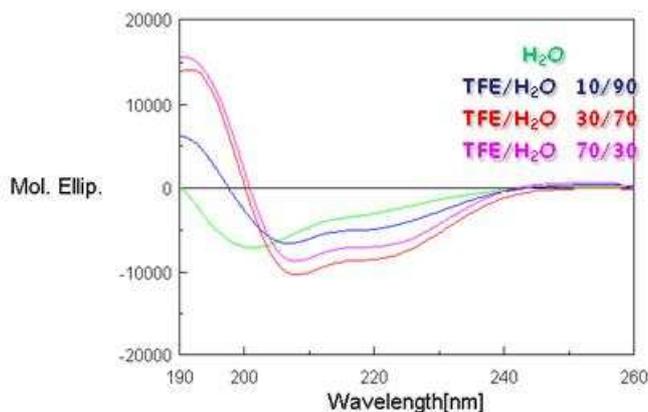


Fig.3.28: Spettri CD di Nodal53-69

Dall'osservazione degli spettri CD si è scelto di effettuare l'analisi conformazionale nella miscela TFE/H<sub>2</sub>O 30/70 (v/v) mediante NMR, perché rappresenta la soluzione con la minima quantità di TFE che ci permette di avere il frammento strutturato come il tratto corrispondente nella proteina parente. L'assegnazione sequenziale di tutte le risonanze protoniche è stata ottenuta mediante l'esame comparato degli spettri TOCSY, DQF-COSY e NOESY, secondo il metodo classico di Wüthrich.<sup>130</sup>

I *chemical shift* di tutti i sistemi di spin sono riportati in **Tabella3.3**.

**Tabella3.3:** *Chemical shift* (ppm) di Nodal53-69 in TFE/H<sub>2</sub>O 30/70 (v/v), 298 K

Residuo	NH	$\alpha$ H	$\beta$ H	$\gamma$ H	altro
Pro <sup>53</sup>		4.52	2.54, 2.10	2.08	$\delta$ H:3.45
Thr <sup>54</sup>	8.60	4.40	4.20	1.22	

<b>Asn<sup>55</sup></b>	8.48	4.78	2.92		$\gamma\text{NH}_2$ :7.54, 6.79
<b>His<sup>56</sup></b>	8.58	4.47	3.25		2H :8.40; 4H:7.23
<b>Ala<sup>57</sup></b>	8.30	4.22	1.49		
<b>Tyr<sup>58</sup></b>	8.00	4.38	3.14		2, 6H:7.12; 3, 5H:6.84
<b>Ile<sup>59</sup></b>	7.85	3.75	1.97	1.66, 1.36	$\gamma\text{CH}_3$ :0.97; $\delta\text{CH}_3$ :0.89
<b>Gln<sup>60</sup></b>	8.25	3.98	2.48, 2.16	2.37	$\epsilon\text{NH}_2$ :7.20, 6.75
<b>Ser<sup>61</sup></b>	7.86	4.22	4.04, 3.99		
<b>Leu<sup>62</sup></b>	7.77	4.13	1.88	1.60	$\delta\text{CH}_3$ :0.85
<b>Leu<sup>63</sup></b>	8.12	4.18	1.88	1.56	$\delta\text{CH}_3$ :0.88
<b>Lys<sup>64</sup></b>	7.79	4.15	1.94, 1.72	1.61, 1.52	$\delta\text{CH}_2$ :1.70; $\epsilon\text{CH}_2$ :3.00; $\epsilon\text{NH}_3$ :7.79
<b>Arg<sup>65</sup></b>	7.64	4.19	1.88, 1.78	1.48	$\delta\text{CH}_2$ :3.10; $\epsilon\text{NH}$ :7.09
<b>Tyr<sup>66</sup></b>	7.92	4.62	3.20, 2.93		2, 6H:7.20; 3, 5H:6.81
<b>Gln<sup>67</sup></b>	7.80	4.63	2.12, 2.03	2.41, 1.94	$\epsilon\text{NH}_2$ :7.44, 6.72
<b>Pro<sup>68</sup></b>		4.38	2.30, 1.88	2.02	$\delta\text{H}$ :3.68
<b>His<sup>69</sup></b>	8.42	4.72	3.34, 3.21		2H :8.60; 4H:7.35  $\text{CONH}_{2\text{ter}}$ :7.53, 7.12

Riportando in grafico le deviazioni dei *chemical shift* dei protoni  $\alpha\text{CH}$  rispetto ai loro valori di *random coil*<sup>161</sup> per ogni residuo (Fig.3.29), si osservano deviazioni negative nel tratto centrale della molecola (58-65) indicative di una struttura elicoidale .

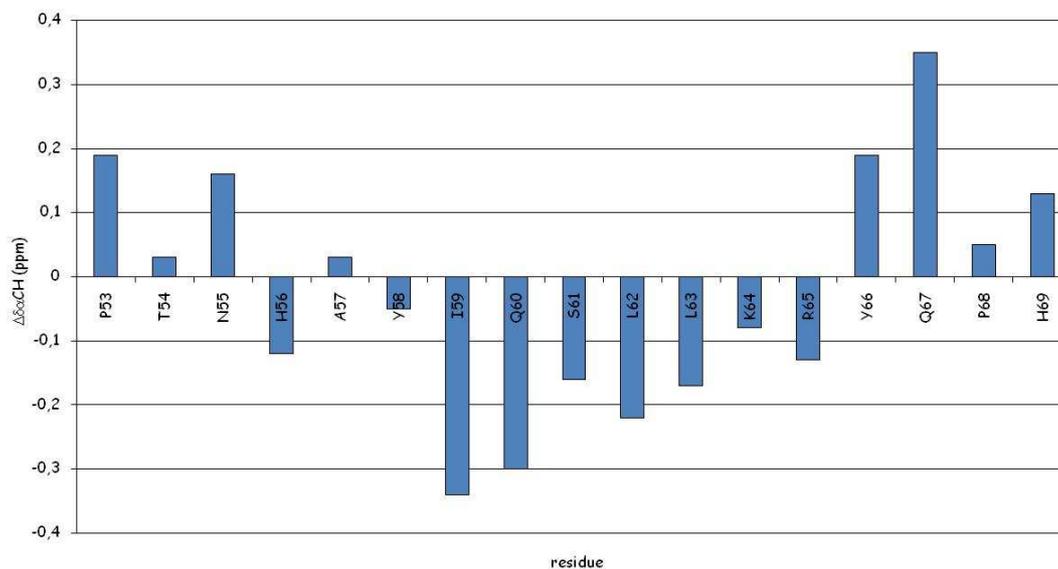


Fig.3.29: Deviazioni dei *chemical shift* (ppm) dei protoni  $\alpha\text{CH}$  di Nodal53-69 dai valori *random coil*

Questo dato è confermato dall'analisi degli spettri NOESY che, oltre all'assegnazione in sequenza dei sistemi di spin, forniscono informazioni strutturali.

In Fig.3.30 sono diagrammati gli effetti NOE più rilevanti osservati per Nodal53-69; l'osservazione di contatti NOE sequenziali del tipo  $(NH_i, NH_{i+1})$  e *long-range* del tipo  $(\alpha_i, \beta_{i+3})$ ,  $(\alpha_i, NH_{i+3})$ , etc. conferma la presenza di una struttura  $\alpha$ -elicoidale.

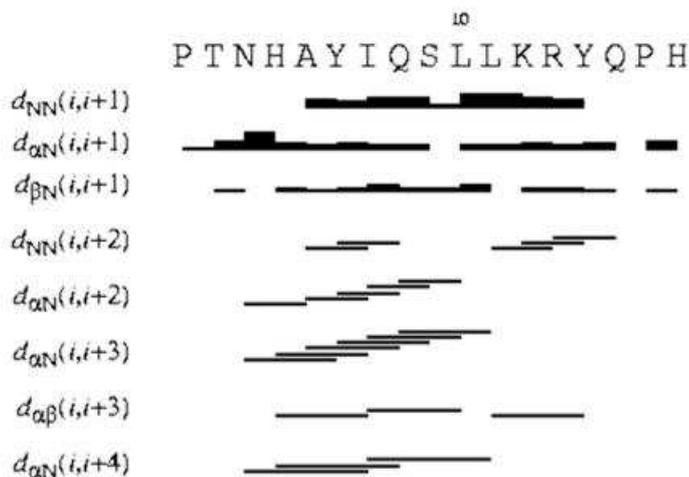


Fig.3.30:Principali effetti NOE per Nodal53-69 in TFE/H<sub>2</sub>O 30/70 (v/v). L'altezza della barra rappresenta l'intensità del NOE

Sono stati misurati un totale di 160 contatti NOE (63 intraresiduo, 60 sequenziali, 37 *long-range*), che sono stati convertiti in altrettante distanze interprotoniche dal programma CALIBA (Tabella1C, Appendice C). Queste distanze sono state utilizzate come limiti superiori per la costruzione di modelli molecolari mediante l'uso del programma CYANA di *torsion angle dynamics*.

#### Calcoli strutturali di Nodal53-69

Un primo insieme di quaranta conformeri è stato calcolato con il programma CYANA usando le distanze interprotoniche, ottenute dai NOE, come limiti superiori (upl). Per ottenere le strutture soddisfacenti i vincoli sperimentali sono state eseguite quattro corse CYANA, usando la strategia TAD, modificando opportunamente i contatti di dubbia attribuzione che maggiormente violavano i limiti superiori imposti sulle distanze.

Il valore della *target function* calcolato su 100 strutture dell'ultima corsa CYANA è di  $0.27 \pm 0.37 \text{ \AA}^2$  e quello della deviazione quadratica media calcolata (RMSD) sul *backbone* è di  $1.64 \pm 0.50 \text{ \AA}$ . Dei vincoli sperimentali inseriti nessuno risulta violato.

Delle strutture risultanti, le quaranta col più basso valore della *target function* (il valore medio della *target function* è di  $9.9 \times 10^{-2} \pm 0.7 \times 10^{-2} \text{ \AA}^2$ , il valore minimo è di  $8.8 \times 10^{-2} \text{ \AA}^2$  e l'RMSD sul *backbone* è di  $1.26 \pm 0.39 \text{ \AA}$ ) sono state sottoposte a minimizzazione energetica

usando il modulo SANDER del programma AMBER 6.0.<sup>139</sup> La statistica dell'analisi strutturale è riportata in Tabella 2C dell'Appendice C.

Le migliori dieci strutture sulle quaranta risultanti dalla minimizzazione energetica sono state selezionate in base alla loro energia e all'accordo con il dato sperimentale; in particolare, sono state scelte le strutture con energia non superiore a  $-89 \text{ kcal mol}^{-1}$ , il valore medio delle energie AMBER delle dieci strutture. Questi 10 conformeri (Fig.3.31), rappresentativi dello spazio conformazionale accessibile al peptide, sono stati analizzati con il programma di grafica molecolare MOLMOL.<sup>141</sup>

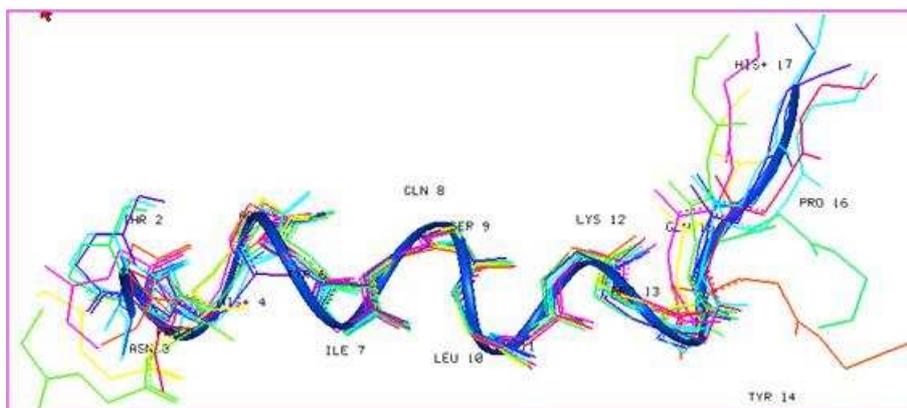


Fig.3.31: Sovrapposizione nel tratto 58-64 delle 10 strutture di Nodal53-69 selezionate dopo CYANA/AMBER

La Fig.3.31 mostra una generale somiglianza delle strutture selezionate lungo l'intero *backbone* ( $\text{RMSD } 1.67 \pm 0.61 \text{ \AA}$ ) e un'ottima sovrapposizione soprattutto nel tratto 58-64 dove è evidente la presenza di una conformazione elicoidale ( $\text{RMSD } 0.26 \pm 0.10 \text{ \AA}$ ).

E' interessante notare che il tratto rinvenuto in  $\alpha$ -elica per questo frammento di Nodal in TFE/H<sub>2</sub>O 30/70 (v/v) riproduce la stessa struttura secondaria del tratto corrispondente della proteina parente (elica H3).

In Fig.3.32 è riportata la sovrapposizione del *backbone* dei modelli molecolari ottenuti, con le catene laterali evidenziate e colorate per tipo. La struttura più vicina alla media delle dieci è evidenziata con un *ribbon* viola e le strutture sovrapposte sono orientate lungo l'asse dell'elica. Si può notare che l'elica osservata ha carattere anfipatico; infatti si ha una buona distribuzione di residui polari, basici da un lato e idrofobici dall'altro. Questa caratteristica può essere importante ai fini delle interazioni con i rispettivi recettori, come vedremo in seguito (par.3.8).

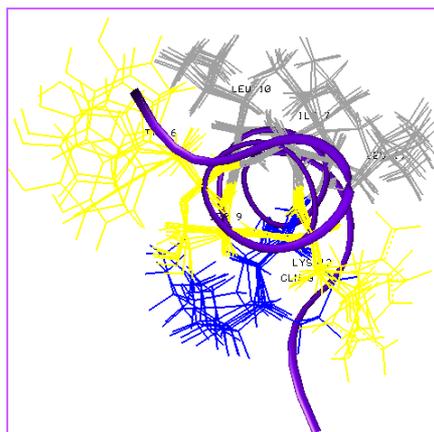


Fig.3.32: Nodal53-69: sovrapposizione delle 10 strutture CYANA/AMBER con le catene laterali colorate per tipo (grigio, idrofobico, giallo, polare, blu, basico)

### 3.7.3 Analisi conformazionale del frammento Nodal43-69

Il frammento Nodal43-69 della proteina Nodal umana, la cui sequenza è riportata in Fig.3.33, è costituito da 27 residui amminoacidici. Il polipeptide è stato ottenuto mediante la sintesi in fase solida presso l'IBB del CNR di Napoli, sotto forma di derivato acetilato all'N-terminale ed amidato al C-terminale. La cisteina è protetta sulla catena laterale con il gruppo Acm.



Fig.3.33: Sequenza amminoacidica di Nodal43-69

Rispetto al frammento precedentemente analizzato (par.3.7.2) questo peptide è allungato di dieci residui verso il tratto N-terminale della proteina Nodal nativa e contiene sempre l'elica H3 (residui 55-65). Per consentire un confronto strutturale con il frammento Nodal53-69, l'analisi conformazionale è stata effettuata nella miscela con lo stesso rapporto TFE/H<sub>2</sub>O 30/70 (v/v). Va notato che gli spettri CD in questa miscela indicano anche per questo frammento una propensione verso una struttura elicoidale.

L'assegnazione sequenziale di tutte le risonanze protoniche è stata ottenuta mediante l'esame comparato degli spettri TOCSY e NOESY, secondo il metodo classico di Wüthrich.<sup>130</sup>

Dall'analisi dello spettro DQF-COSY, è stato possibile confermare l'assegnazione dei sistemi di spin, sia delle glicine attraverso i contatti diretti NH- $\alpha$ CH sia dei protoni delle catene laterali. I *chemical shift* di tutti i sistemi di spin sono riportati in **Tabella3.4**.

**Tabella3.4:** *Chemical shift* (ppm) di Nodal43-69 in TFE/H<sub>2</sub>O 30/70 (v/v), 298 K

Residuo	NH	$\alpha$ H	$\beta$ H	$\gamma$ H	altri
<b>Cys</b> <sup>43</sup>	8.16	5.00	2.84, 2.71		
<b>Pro</b> <sup>44</sup>		4.36	2.28, 1.85	1.99	$\delta$ H :3.76, 3.64
<b>Asn</b> <sup>45</sup>	8.23	4.75	2.93		$\gamma$ NH <sub>2</sub> :7.60, 6.87
<b>Pro</b> <sup>46</sup>		4.44	2.30, 1.95	2.00	$\delta$ CH <sub>2</sub> :3.81, 3.76
<b>Val</b> <sup>47</sup>	7.98	4.07	2.11	0.96	
<b>Gly</b> <sup>48</sup>	8.22	3.97			
<b>Glu</b> <sup>49</sup>	8.07	4.31	2.11, 1.98	2.44	
<b>Glu</b> <sup>50</sup>	8.20	4.25	1.93	2.34, 2.23	
<b>Phe</b> <sup>51</sup>	7.99	4.57	3.03, 2.97		2, 6H:7.30 ; 3, 4,5H:7.22
<b>His</b> <sup>52</sup>	8.07	4.93	3.18, 3.07		2H :8.57; 4H: 7.26
<b>Pro</b> <sup>53</sup>		4.43	2.26, 1.96	2.00	$\delta$ H :3.57, 3.47
<b>Thr</b> <sup>54</sup>	8.15	4.35	4.29	1.22	
<b>Asn</b> <sup>55</sup>	8.48	4.35	2.87, 2.81		$\gamma$ NH <sub>2</sub> :7.52, 7.28
<b>His</b> <sup>56</sup>	8.68	4.46	3.29		2H:8.64; 4H:7.30
<b>Ala</b> <sup>57</sup>	8.35	4.22	1.46		
<b>Tyr</b> <sup>58</sup>	8.04	4.36	3.12		2, 6H:7.10; 3, 5H: 6.81
<b>Ile</b> <sup>59</sup>	7.87	3.73	1.93	1.62, 1.31	$\gamma$ CH <sub>3</sub> :0.92; $\delta$ CH <sub>3</sub> :0.86
<b>Gln</b> <sup>60</sup>	8.18	3.96	2.10, 1.93	2.41, 2.34	$\epsilon$ NH <sub>2</sub> :7.28, 6.80
<b>Ser</b> <sup>61</sup>	7.87	4.22	3.98		
<b>Leu</b> <sup>62</sup>	7.77	4.12	1.81, 1.54	1.54	$\delta$ CH <sub>3</sub> :0.81
<b>Leu</b> <sup>63</sup>	8.04	4.17	1.82, 1.56	1.82	$\delta$ CH <sub>3</sub> :0.86
<b>Lys</b> <sup>64</sup>	7.75	4.14	1.87	1.56, 1.47	$\delta$ CH <sub>2</sub> :1.70; $\epsilon$ CH <sub>2</sub> : 2.98; $\epsilon$ NH <sub>3</sub> :7.60
<b>Arg</b> <sup>65</sup>	7.66	4.18	1.73	1.44	$\delta$ CH <sub>2</sub> :3.09; NH:7.08
<b>Tyr</b> <sup>66</sup>	7.92	4.59	3.15, 2.91		2, 6H:7.17 ; 3, 5H:6.80
<b>Gln</b> <sup>67</sup>	7.85	4.60	2.08, 1.97	2.37, 1.71	$\epsilon$ NH <sub>2</sub> : 7.45, 6.78
<b>Pro</b> <sup>68</sup>		4.36	2.28, 1.85	1.99	$\delta$ H :3.76, 3.65
<b>His</b> <sup>69</sup>	8,47	4.70	3.32, 3.20		2H :8.61 ; 4H: 7.34  CONH <sub>2</sub> ter:7.54, 7.15

Per avere informazioni preliminari sulla struttura si considera il grafico delle deviazioni dei *chemical shift* dei protoni  $\alpha$ CH rispetto ai loro valori di *random coil*<sup>161</sup> per ogni residuo (Fig.3.34).

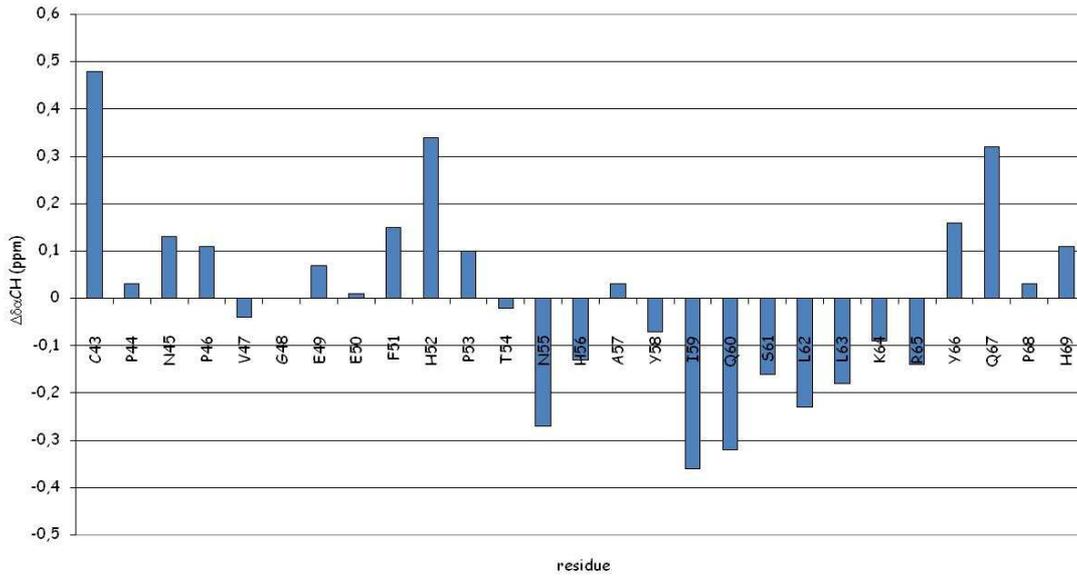


Fig.3.34: Deviazioni dei *chemical shift* (ppm) dei protoni  $\alpha$ CH di Nodal43-69 dai valori *random coil*

Si notano deviazioni positive nel tratto N-terminale e deviazioni negative nel tratto C-terminale 54-65, quest'ultime diagnostiche di una conformazione elicoidale.<sup>161</sup>

Questo dato è confermato dall'analisi degli spettri NOESY che, oltre all'assegnazione in sequenza dei sistemi di spin, forniscono informazioni strutturali.

In Fig.3.35 sono diagrammati gli effetti NOE più rilevanti osservati per Nodal43-69; l'osservazione di contatti NOE sequenziali del tipo  $(\text{NH}_i, \text{NH}_{i+1})$  e *long-range* del tipo  $(\alpha_i, \beta_{i+3})$ ,  $(\alpha_i, \text{NH}_{i+3})$ , etc. nel tratto C-terminale conferma la presenza di una struttura  $\alpha$ -elicoidale in questa regione.

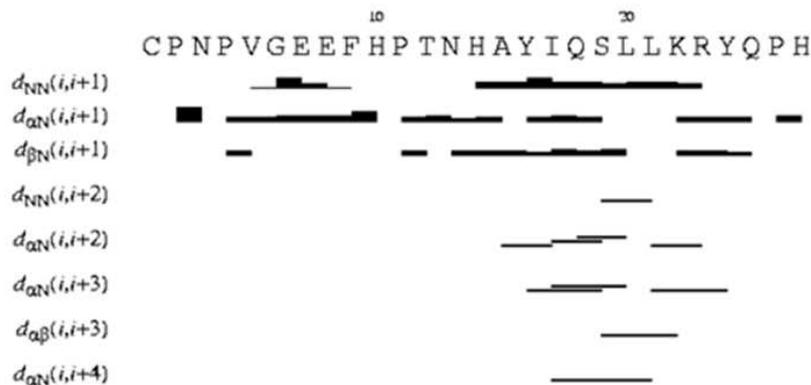


Fig.3.35: Principali effetti NOE per Nodal43-69 in TFE/H<sub>2</sub>O 30/70 (v/v). L'altezza della barra rappresenta l'intensità del NOE

Sono stati misurati un totale di 128 contatti NOE (61 intraresiduo, 24 sequenziali, 14 *long-range*), che sono stati convertiti in altrettante distanze interprotoniche dal programma

CALIBA (Tabella1D, Appendice D). Queste distanze sono state utilizzate come limiti superiori per la costruzione di modelli molecolari mediante l'uso del programma CYANA di *torsion angle dynamics*.

#### *Calcoli strutturali di Nodal43-69*

Un primo insieme di quaranta conformeri è stato calcolato con il programma CYANA usando le distanze interprotoniche, ottenute dai NOE, come limiti superiori (upl). Inoltre quattro angoli diedri  $\Phi$ , relativi a residui con una  $^3J_{\text{NH-}\alpha\text{CH}}$  minore di 6 Hz, sono stati imposti a  $-57^\circ$  con una tolleranza di  $\pm 10\%$ . Per ottenere le strutture soddisfacenti i vincoli sperimentali sono state eseguite dodici corse CYANA, usando la strategia TAD, modificando opportunamente i contatti di dubbia attribuzione che maggiormente violavano i limiti superiori imposti sulle distanze.

Il valore della *target function* calcolato su 100 strutture dell'ultima corsa CYANA è di  $6.64 \pm 1.91 \text{ \AA}^2$  e quello della deviazione quadratica media calcolata (RMSD) sul *backbone* è di  $5.37 \pm 1.06 \text{ \AA}$ . Dei vincoli sperimentali inseriti nessuno risulta violato.

Delle strutture risultanti, le quaranta col più basso valore della *target function* (il valore medio della *target function* è di  $5.53 \pm 0.20 \text{ \AA}^2$ , il valore minimo è di  $5.11 \text{ \AA}^2$  e l' RMSD sul *backbone* è di  $5.00 \pm 0.95 \text{ \AA}$ ) sono state sottoposte a minimizzazione energetica usando il modulo SANDER del programma AMBER 6.0.<sup>139</sup> La statistica dell'analisi strutturale è riportata in Tabella2D dell'Appendice D.

Le migliori dieci strutture sulle quaranta risultanti dalla minimizzazione energetica sono state selezionate in base alla loro energia e all'accordo con il dato sperimentale; in particolare, sono state scelte le strutture con energia non superiore a  $-207 \text{ kcal mol}^{-1}$ , il valore medio delle energie AMBER delle dieci strutture. Questi 10 conformeri (Fig.3.36), rappresentativi dello spazio conformazionale accessibile al peptide, sono stati analizzati con il programma di grafica molecolare MOLMOL.<sup>141</sup>

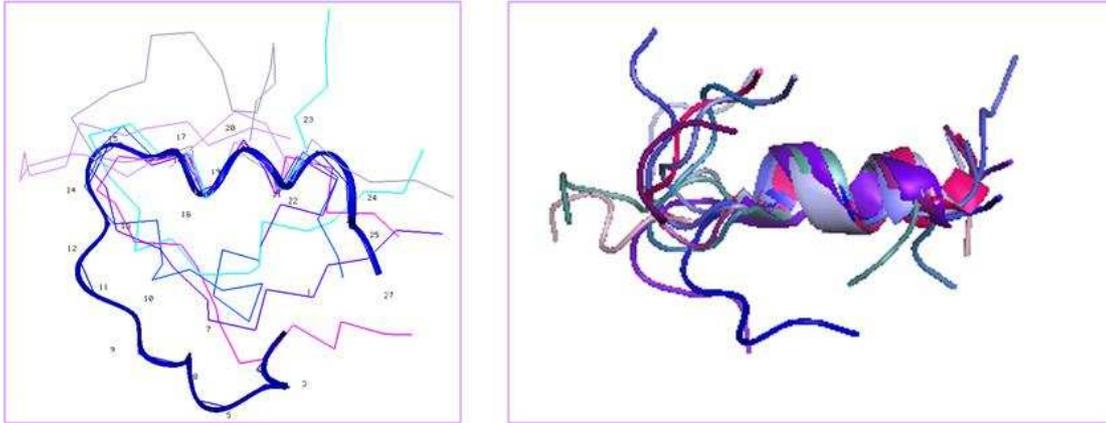


Fig.3.36: Sovrapposizione nel tratto 58-64 delle 10 strutture di Nodal43-69, selezionate dopo CYANA/AMBER

Dalla Fig.3.36 si osserva una buona sovrapposizione nel tratto 58-64 dove le strutture selezionate sono caratterizzate dalla presenza di una conformazione elicoidale con un RMSD calcolato sul *backbone*  $0.81 \pm 0.37 \text{ \AA}$ . Il resto della sequenza mostra invece una notevole flessibilità.

E' interessante notare che il tratto rinvenuto in  $\alpha$ -elica per il frammento di Nodal53-69 in TFE/H<sub>2</sub>O 30/70 (v/v) è conservato anche in questo frammento più esteso e riproduce la stessa struttura secondaria del tratto corrispondente della proteina parente. La sovrapposizione del *backbone* nel tratto 58-64 dei modelli molecolari più simili alla media, ottenuti per i frammenti 53-69 e 43-69 di Nodal, rivela un valore di RMSD pari  $0.54 \text{ \AA}$ .

#### 3.7.4 Analisi conformazionale del frammento Nodal66-90

Il frammento Nodal66-90 della proteina Nodal umana, la cui sequenza è riportata in Fig.3.37, è costituito da 25 residui amminoacidici; contiene gli ultimi quattro residui delle sequenze precedentemente analizzate e si estende nel tratto C-terminale di Nodal che costituisce il '*knuckle epitope*'. Il polipeptide è stato ottenuto mediante la sintesi in fase solida presso l'IBB del CNR di Napoli, sotto forma di derivato acetilato all'N-terminale ed amidato al C-terminale. Le cisteine sono protette sulla catena laterale con il gruppo Acm.



Fig.3.37: Sequenza amminoacidica di Nodal66-90

Anche per il frammento Nodal66-90 sono stati registrati spettri CD in diverse condizioni di solvente (Fig.3.38). Gli spettri CD sono stati registrati in soluzioni acquose a concentrazione di peptide dell'ordine di  $10^{-4}$  M e a diversi tenori di TFE: in tampone fosfato a pH=6.9 (curva rossa); in tampone fosfato a pH=6.9 più 30% di TFE (curva verde); in tampone fosfato a pH=6.9 più 50% di TFE (curva fucsia) e in tampone fosfato a pH=6.9 più 70% di TFE (curva blu).

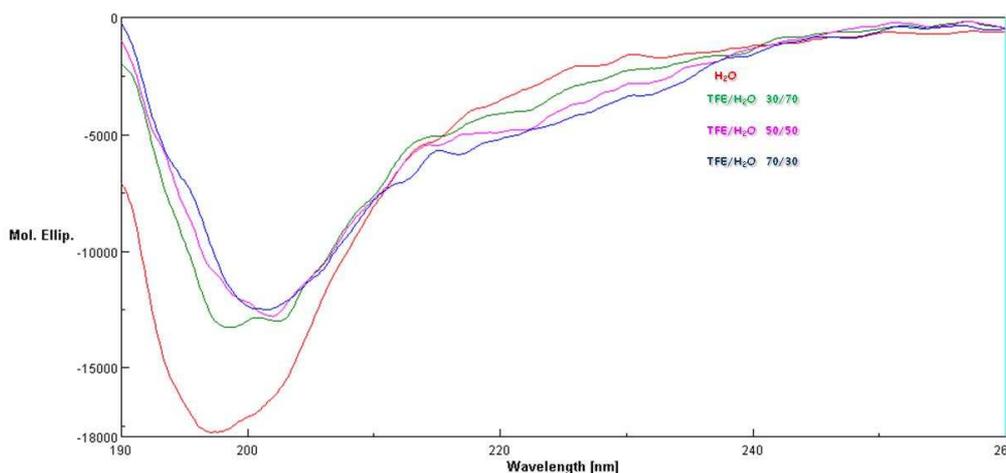


Fig.3.38: Spettri CD di Nodal66-90

L'analisi NMR è stata effettuata nella miscela TFE/H<sub>2</sub>O 30/70 (v/v) in analogia ai peptidi precedenti. Comunque in questo caso, nelle condizioni sperimentali esplorate, le curve CD rivelano la mancanza di propensione del peptide ad assumere una conformazione ordinata.

L'assegnazione sequenziale di tutte le risonanze protoniche è stata ottenuta mediante l'esame comparato degli spettri TOCSY, DQF-COSY e NOESY, secondo il metodo classico di Wüthrich.<sup>130</sup>

I *chemical shift* di tutti i sistemi di spin sono riportati in **Tabella3.5**.

**Tabella3.5:** *Chemical shift* (ppm) di Nodal66-90 in TFE/H<sub>2</sub>O 30/70 (v/v), 298 K

Residuo	NH	$\alpha$ H	$\beta$ H	$\gamma$ H	altro
<b>Tyr<sup>66</sup></b>	8.10	4.47	2.94		
<b>Gln<sup>67</sup></b>	7.95	4.58	2.29, 1.98	1.83	
<b>Pro<sup>68</sup></b>		4.28	2.26, 1.80	1.96	$\delta$ H: 3.60, 3.50
<b>His<sup>69</sup></b>	8.42	4.68	3.28, 3.18		
<b>Arg<sup>70</sup></b>	8.27	4.39	1.79, 1.70	1.59	$\delta$ CH <sub>2</sub> : 3.19; $\epsilon$ NH: 7.20, 6.68
<b>Val<sup>71</sup></b>	8.17	4.48	2.12	0.97	

<b>Pro</b> <sup>72</sup>		4.44	2.33, 1.94	2.08, 2.01	$\delta H$ : 3.88, 3.72
<b>Ser</b> <sup>73</sup>	8.31	4.51	3.91, 3.89		
<b>Thr</b> <sup>74</sup>	8.03	4.35	4.30	1.23	
<b>Cys</b> <sup>75</sup>	8.32	4.65	3.08, 2.91		
<b>Cys</b> <sup>76</sup>	8.36	4.41	3.08, 2.91		
<b>Ala</b> <sup>77</sup>	8.15	4.61	1.40		
<b>Pro</b> <sup>78</sup>		4.48	2.30, 2.00	2.04	$\delta H$ : 3.85, 3.65
<b>Val</b> <sup>79</sup>	8.04	4.12	2.07	0.96	
<b>Lys</b> <sup>80</sup>	8.35	4.40	1.87, 1.76	1.47	
<b>Thr</b> <sup>81</sup>	7.89	4.34	4.20	1.20	
<b>Lys</b> <sup>82</sup>	8.19	4.60	1.82, 1.70	1.45	$\delta CH_2$ : 1.70, $\epsilon CH_2$ : 3.00, $\epsilon NH$ : 7.59
<b>Pro</b> <sup>83</sup>		4.46	2.34	2.08, 2.00	$\delta H$ : 3.70, 3.64
<b>Leu</b> <sup>84</sup>	8.19	4.27	1.70	1.63	$\delta CH_3$ : 0.96, 0.90
<b>Ser</b> <sup>85</sup>	8.02	4.37	3.94, 3.83		
<b>Met</b> <sup>86</sup>	8.15	4.47	2.00	2.57, 2.53	
<b>Leu</b> <sup>87</sup>	7.84	4.22	1.54, 1.41	1.63	$\delta CH_3$ : 0.89, 0.83
<b>Tyr</b> <sup>88</sup>	7.83	4.57	3.12, 2.96		2, 6H: 7.11 ; 3, 5H: 6.65
<b>Val</b> <sup>89</sup>	7.72	4.06	2.10	0.94	
<b>Asp</b> <sup>90</sup>	8.20	4.68	2.92, 2.82		

In Fig.3.39 sono riportate le le deviazioni dei *chemical shift* dei protoni  $\alpha CH$  rispetto ai loro valori di *random coil*<sup>161</sup> per ogni residuo. Dal grafico si notano deviazioni positive che sono identificative della propensione del frammento ad essere beta esteso, così come il tratto corrispondente nella proteina parente.

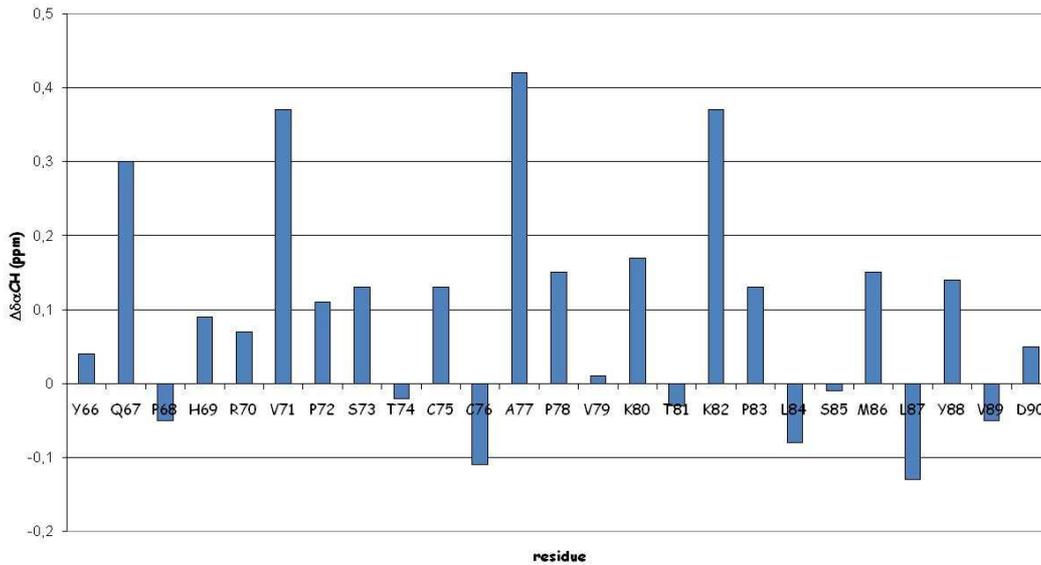


Fig.3.39: Deviazioni dei *chemical shift* (ppm) dei protoni  $\alpha$ CH di Nodal66-90 dai valori *random coil*

In Fig.3.40 è riportato il pattern di contatti NOE osservati che si mostra alquanto povero. Quindi avendo a disposizione un esiguo numero di contatti NOE (98 NOE: 46 intraresiduo, 51 sequenziali, 1 *long-range*) (Tabella1E, Appendice E) non è stato possibile calcolare un modello molecolare per questo frammento.

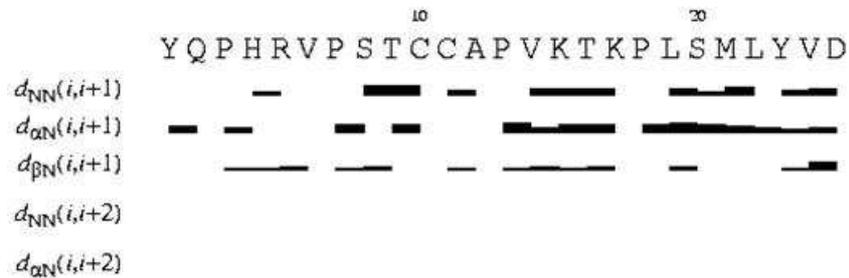


Fig.3.40: Principali effetti NOE per Nodal66-90 in TFE/H<sub>2</sub>O 30/70 (v/v). L'altezza della barra rappresenta l'intensità del NOE

*Confronto tra il modello per omologia del monomero di Nodal e i frammenti NMR*

Avendo a disposizione il modello per omologia del monomero di Nodal e i modelli molecolari ottenuti mediante NMR dei due frammenti, è stato interessante effettuare le sovrapposizioni per notare il grado di somiglianza dei frammenti dalla proteina parente. La sovrapposizione dei due frammenti con il modello per omologia (Fig.3.41) porta ad un RMSD sul *backbone* di 5.71 Å per Nodal43-69 e 1.75 Å per Nodal53-69. Il valore dell'RMSD si abbassa notevolmente se consideriamo la sovrapposizione tra il modello per omologia e i due

frammenti nel tratto elicoidale 58-64 (RMSD 1.24 Å per Nodal43-69 e 0.37 Å per Nodal53-69)



Fig.3.41: Rappresentazione con *ribbon* della sovrapposizione nel tratto elicoidale 58-64 tra il modello per omologia del monomero di Nodal (in rosa) con i frammenti NMR Nodal43-69 (in blu) e Nodal53-69 (in viola)

### 3.8 Docking tra Nodal, ALK7-ECD e ALK4-ECD

E' interessante ricordare che Nodal può legare direttamente ALK7, mentre per legare ALK4 necessita di un co-recettore come Cripto. Per poter comprendere quali sono le differenze strutturali tra i due recettori all'interfaccia di *binding* che determinano il diverso comportamento nei confronti di Nodal abbiamo costruito i modelli per omologia di ALK4-ECD e ALK7-ECD (Fig.3.42). Dal confronto delle superfici di *Connolly* dell'ECD di ALK4 e ALK7, si può notare una diversa distribuzione dei residui all'interfaccia di *binding*, infatti in ALK4 abbiamo la presenza di residui idrofobici e carichi, mentre su ALK7 sono riscontrabili aminoacidi essenzialmente idrofobici e polari.

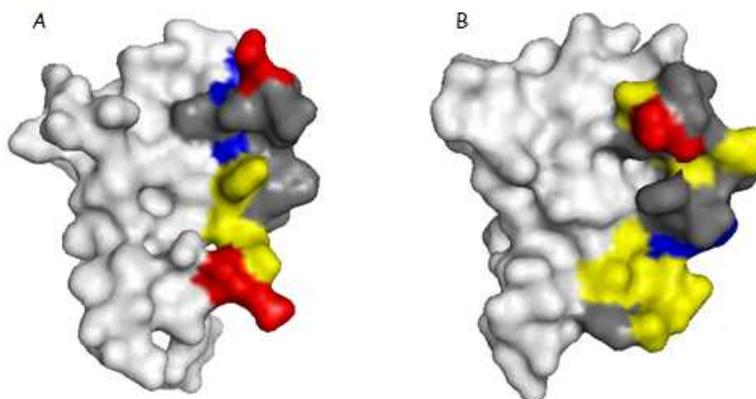


Fig.3.42: Superficie di *Connolly* dei modelli per omologia di ALK4-ECD (A) e ALK7-ECD (B).I residui all'interfaccia di *binding* sono colorati per tipo: idrofobici in grigio, polari in giallo, acidi in rosso, basici in blu

Se guardiamo la superficie di *Connolly* del 'wrist epitope' di Nodal, che comprende il tratto elicoidale 55-66, si osserva una distribuzione di residui polari-idrofobici, come visto per ALK7, questo sia nel modello per omologia sia nel modello molecolare NMR Nodal53-69 (Fig.3.43).

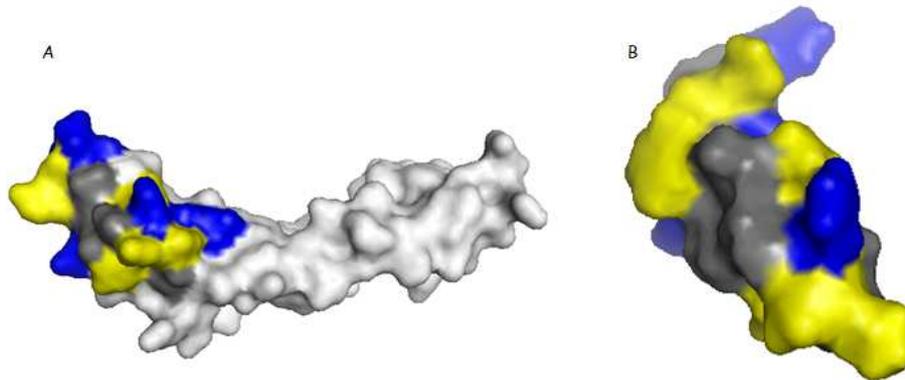


Fig.3.43: Superficie di *Connolly* del modello per omologia del monomero della proteina Nodal umana (A) e del frammento NMR Nodal53-69 (B). I residui del 'wrist epitope' sono colorati per tipo: idrofobici in grigio, polari in giallo, basici in blu

La distribuzione di residui con caratteristiche simili (polari o idrofobici) all'interfaccia di *binding* può essere uno dei motivi dell'interazione diretta tra ALK7-Nodal. Avendo a disposizione i modelli per omologia dei recettori ALK7-ECD, ALK4-ECD e del dimero della proteina Nodal, abbiamo effettuato esperimenti di *docking* tra i recettori e il ligando, per cercare di formulare una serie di ipotesi sugli elementi strutturali potenzialmente rilevanti ai fini del *binding*. Delle soluzioni del *docking* rigido, ottenute dal programma 3D-Dock,<sup>157,166</sup> sono state considerate quelle che presentavano l'interfaccia concava di *binding* dei recettori vicina al 'wrist epitope' del dimero di Nodal. Nel complesso con l'RP-Score più alto (5.46), ottenuto dal *docking* tra il recettore ALK7 e Nodal, si osservano molti contatti favorevoli: interazioni idrofobiche, che coinvolgono i due residui di Trp ( $W^{24}$  e  $W^{27}$  di un monomero) e i residui  $I^{59}$ ,  $L^{61}$ ,  $L^{62}$  del 'wrist epitope' (dell'altro monomero) di Nodal che formano un solco dove si può intascare il residuo  $F^{50}$  del recettore ALK7; interazioni tra residui polari, che coinvolgono la  $Y^{61}$  di Nodal e i residui  $S^{53}$ ,  $S^{54}$ ,  $N^{55}$  e  $N^{56}$  di ALK7 (Fig.3.44).

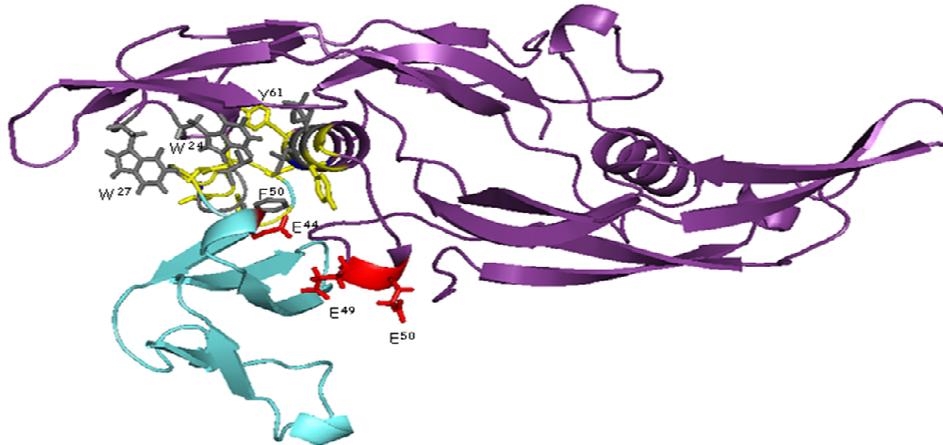


Fig.3.44: Rappresentazione con *ribbon* del complesso tra ALK7-ECD (celeste) e il dimero di Nodal (viola). I residui rilevanti all'interfaccia di *binding* sono mostrati e colorati per tipo: idrofobici in grigio, polari in giallo, acidi in rosso, basici in blu

Nel complesso che presenta l'RPSScore più alto (6.75), ottenuto dal *docking* tra il recettore ALK4 e Nodal, si osservano interazioni idrofobiche, che coinvolgono i due residui di Trp ( $W^{24}$  e  $W^{27}$  di un monomero) e i residui  $I^{59}$ ,  $L^{61}$ ,  $L^{62}$  del '*wrist epitope*' (dell'altro monomero) di Nodal e la superficie concava idrofobica, che contiene la  $F^{55}$  del recettore ALK4; interazioni tra residui polari, che coinvolgono la  $Y^{56}$  di ALK4 e il residuo  $N^{45}$  di Nodal (Fig.3.45). In questo complesso, però, si può osservare anche un'interazione elettrostatica sfavorevole tra  $E^{47}$  di ALK4 e  $E^{49}$  di Nodal.

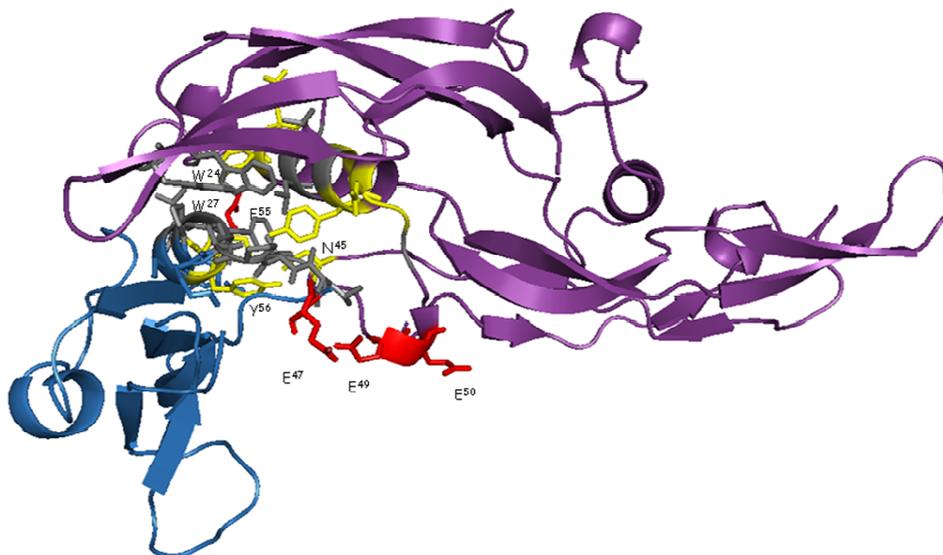


Fig.3.45: Rappresentazione con *ribbon* del complesso tra ALK4-ECD (blu) e il dimero di Nodal (viola). I residui rilevanti all'interfaccia di *binding* sono mostrati e colorati per tipo: idrofobici in grigio, polari in giallo, acidi in rosso, basici in blu

Guardando più in dettaglio il complesso tra Nodal e ALK7-ECD si può osservare che l'interazione elettrostatica sfavorevole non è presente, in quanto il residuo E<sup>44</sup> di ALK7 (che corrisponde al residuo E<sup>47</sup> di ALK4) non si trova a breve distanza dal residuo E<sup>49</sup> di Nodal (Fig.3.46).

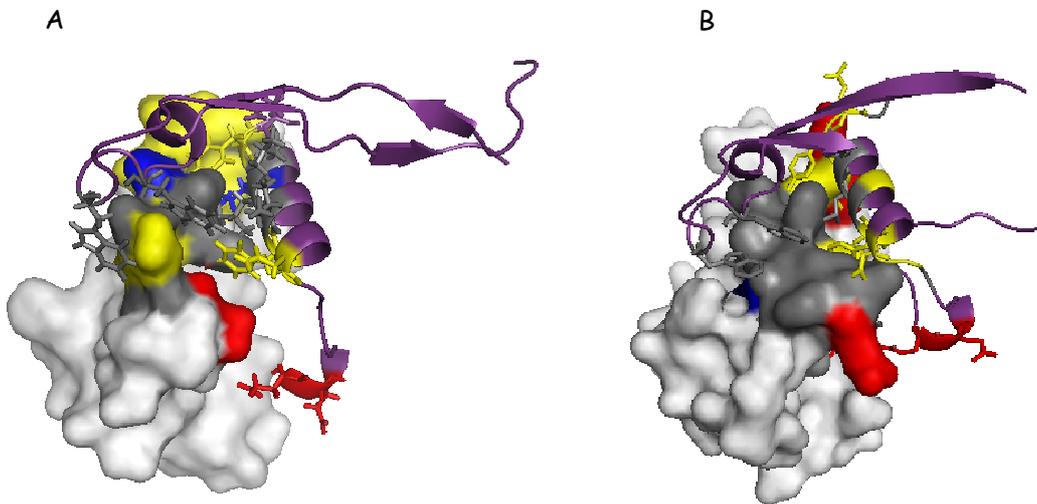


Fig.3.46: Rappresentazione dell'interfaccia di *binding* tra (A) ALK7-ECD (superficie di *Connolly* con i residui colorati per tipo) e il dimero di Nodal (rappresentato con *ribbon e stick*) e (B) ALK4-ECD (superficie di *Connolly* con i residui colorati per tipo) e il dimero di Nodal (rappresentato con *ribbon e stick*)

L'interazione elettrostatica sfavorevole, osservata nel complesso ALK4-ECD:Nodal, può essere uno dei motivi della mancata propensione delle due proteine ad interagire direttamente.

Questa ipotesi è avvalorata da ulteriori osservazioni. In Fig.3.47 sono riportati, per un utile confronto i due complessi ALK4-ECD:Nodal e ALK7-ECD:Nodal, ottenuti dal *docking*, e il complesso RX ALK3-ECD:BMP2.<sup>26</sup>

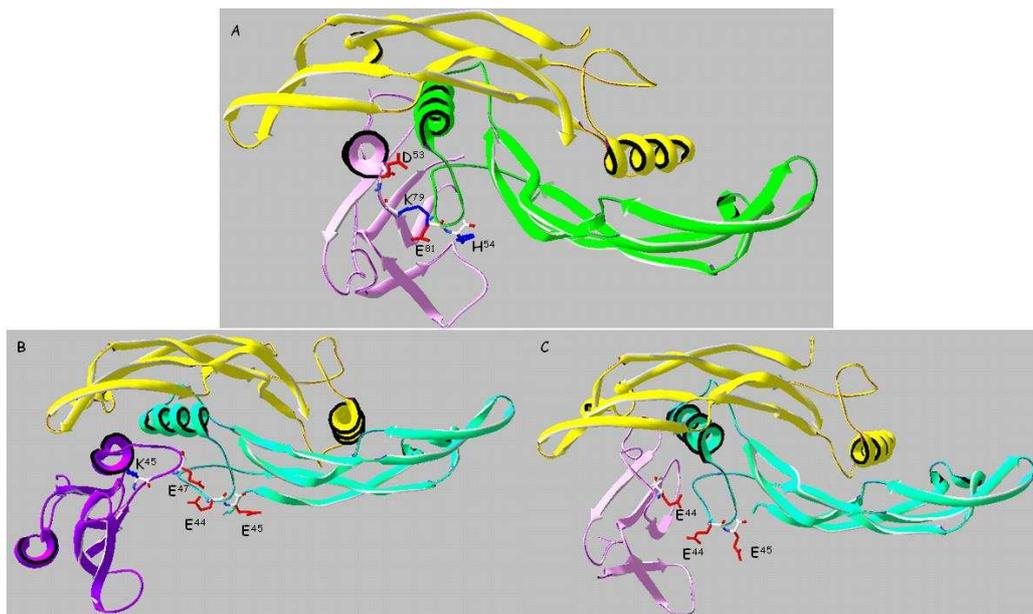


Fig.3.47: Rappresentazione con *ribbon* dei complessi (A) ALK3-ECD:BMP2 (pdbcode:1es7) (B) ALK4-ECD:Nodal e (C) ALK7-ECD:Nodal. I monomeri dei ligandi sono colorati in giallo e verde; i recettori ALK in violetto. Sono riportati in *stick* alcuni residui del *loop* pre-elica dei ligandi e del *loop* 23 dei recettori

Nelle figure 3.47 sono evidenziate le catene laterali di alcuni residui appartenenti al *loop* pre-elica dei ligandi TGF- $\beta$  e al *loop* 23 dei recettori di tipo I (par.1.2). Va ricordato che il *loop* pre-elica, cioè il *loop* flessibile che precede l'elica H3 dei TGF- $\beta$ , è un motivo di riconoscimento molto conservato in tutti i recettori di tipo I.<sup>69</sup> Si è scelto di riportare il confronto con il complesso ALK3-ECD:BMP2, perché la BMP2 ha un *loop* pre-elica lungo quanto quello di Nodal. Altri ligandi TGF- $\beta$  (ad esempio TGF- $\beta$ 3) hanno un *loop* più corto e interagiscono in maniera diversa con i recettori di tipo I. I modelli per omologia di ALK4-ECD e ALK7-ECD rivelano una diversa lunghezza del *loop* 23, che si affaccia sulla superficie di *binding* e può causare ingombro sterico.

Dal complesso RX ALK3-ECD:BMP2 si nota che il residuo D<sup>53</sup> del *loop* pre-elica di BMP2 può stabilire interazioni elettrostatiche favorevoli con il residuo K<sup>79</sup> del *loop* 23 del recettore I. La posizione corrispondente a D<sup>53</sup> di BMP2 è occupata da un analogo residuo acido (E<sup>49</sup>) nella proteina Nodal. Andando a confrontare il *loop* 23 dei recettori ALK4 e ALK7 con ALK3, si osserva che in ALK4 questo *loop* è più lungo rispetto agli altri due (Fig.3.48). Guardando in dettaglio i residui che protrudono dal *loop* 23 di ALK4 c'è un residuo di Lys (K<sup>45</sup>) corrispondente alla K<sup>79</sup> di ALK3, ma anche un Glu (E<sup>47</sup>) corrispondente a E<sup>81</sup> di ALK3. Data la maggiore estensione del *loop* 23 di ALK4, questo si allunga all'interfaccia di *binding* verso Nodal esponendo la catena laterale di E<sup>47</sup>, anziché quella di K<sup>45</sup> come invece accade con ALK3.

Quindi nell'interazione ALK4-ECD:Nodal si stabiliscono contatti elettrostatici sfavorevoli tra residui analoghi: E<sup>47</sup> del *loop* 23 di ALK4 ed E<sup>49</sup> presente sul *loop* pre-elica di Nodal.

Andando ad osservare il *loop* 23 nel recettore ALK7 non si trova più un residuo di Lys corrispondente alla K<sup>79</sup> di ALK3, mentre è sempre presente un residuo acido (E<sup>44</sup>), corrispondente ad E<sup>47</sup> di ALK4 e ad E<sup>81</sup> di ALK3. Tuttavia, in questo caso, poiché il *loop* 23 è più corto rispetto a quello di ALK4, non c'è più lo stesso avvicinamento sfavorevole tra il Glu di ALK7 e quello presente (E<sup>49</sup>) sul *loop* pre-elica di Nodal.

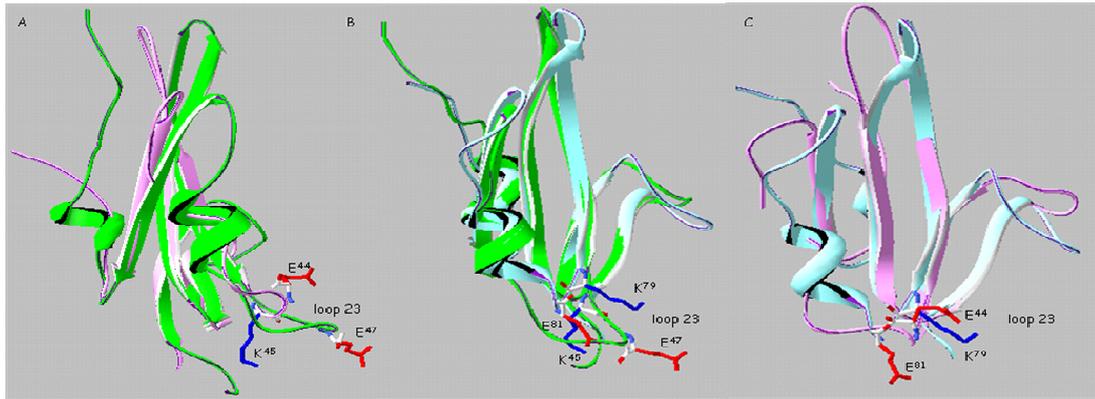


Fig.3.48: Rappresentazione con *ribbon* della sovrapposizione dei recettori di tipo I (A) ALK7 in rosa e ALK4 in verde (B) ALK4 in verde e ALK3 in blu (C) ALK7 in rosa e ALK3 in blu. Sono riportati in *stick* alcuni residui del *loop* 23 dei recettori

## CAPITOLO 4

### Conclusioni

La ricerca scientifica, svolta durante il triennio di dottorato, ha riguardato lo studio conformazionale in soluzione di domini proteici di fattori di crescita funzionalmente correlati. In particolare, sono stati studiati dal punto di vista strutturale domini delle proteine Cripto, Nodal ed ALK4, coinvolte nella formazione di un complesso recettoriale che media la trasduzione del segnale di molti fattori della famiglia dei TGF- $\beta$ .

Nodal è una proteina della famiglia dei TGF- $\beta$  e ha un ruolo fondamentale durante l'embriogenesi in cui controlla la formazione del mesoderma e l'asimmetria destra-sinistra nello sviluppo dei vertebrati. L'azione di Nodal si propaga attraverso il legame con i recettori serina/treonina chinasi ALK4 e ActRIIB e con il co-recettore Cripto. In individui adulti, Nodal e Cripto, sono espressi in molti tipi di tumori, incluso il melanoma, il tumore della mammella, del colon e dello stomaco.

Le funzioni sia di Cripto che di Nodal sono mediate dalla loro reciproca interazione e dall'interazione con il recettore ALK4. In particolare, è stato dimostrato che Cripto lega Nodal con il dominio EGF ed il recettore ALK4 con il dominio CFC. L'interruzione di una qualunque di queste interazioni provoca un *signaling* alterato dei due fattori di crescita. Pertanto, essi sono considerati dei *target* di enorme rilevanza per lo sviluppo di farmaci antitumorali altamente efficaci e specifici. Lo sviluppo di antagonisti, tuttavia, prescinde da una conoscenza accurata delle strutture molecolari coinvolte nel riconoscimento e delle affinità relative tra i singoli componenti dei complessi recettoriali o di loro sottodomini.

A tale scopo, in questo lavoro di ricerca, è stata intrapresa l'indagine strutturale del dominio CFC di Cripto e di sottodomini della proteina Nodal. Lo studio previsto del dominio EGF-like di Cripto è stato impedito dalla mancanza di adeguate quantità di polipeptide ottenuto per sintesi chimica. I sottodomini di Nodal esaminati includono i residui 43-90 e corrispondono a regioni della proteina parente appartenenti al '*wrist epitope*' (elica H3) o al '*knuckle epitope*'. L'interesse verso Nodal è determinato dalle molteplici interazioni che questa proteina stabilisce all'interno del complesso recettoriale e dalla sua incapacità, insieme ai fattori di

crescita GDF1 e GDF3, di trasdurre segnali biochimici verso i mediatori Smad2/3 in assenza di Cripto. Questa peculiarità può dipendere da differenze di struttura primaria (o superiori) con le proteine omologhe Attivina A e B o con le meglio caratterizzate dal punto di vista strutturale, le BMP, che invece attivano il *signaling* intracellulare indipendentemente dalla presenza di Cripto. In letteratura è noto che Nodal lega il dominio EGF-like di Cripto nel *signaling* attraverso il recettore ALK4, ma che è anche in grado di trasmettere il segnale attraverso il legame diretto al recettore ALK7.

In una prima fase è stata condotta un'analisi conformazionale mediante NMR del dominio CFC, derivante dalla proteina da topo (mCFC) e umana (hCFC), determinandone la struttura tridimensionale a pH fisiologico. I modelli molecolari ottenuti esibiscono una forma complessivamente ellissoidale ed il *fold*ing è globalmente esteso con tre *strand* collegati dai tre ponti disolfuro. È importante notare che i residui ritenuti responsabili del *binding* con il recettore ALK4, (H<sup>9</sup> e W<sup>12</sup>) non sono localizzati nel *core* proteico, ma si ritrovano posizionati tra la fine del primo *strand* e l'inizio del primo *loop*, con le catene laterali esposte esternamente. Opposto al sito di *binding*, è localizzato un cluster di residui idrofobici il cui ruolo non è ben chiarito, ma si può ipotizzare che svolga una funzione rilevante per la formazione d'interazioni sia intramolecolari, ad esempio per stabilizzare la struttura dell'intera proteina legandosi al dominio EGF-like, sia intermolecolari. Inoltre, al di sopra e al di sotto del sito di *binding*, sono localizzate delle zone superficiali ad elevata concentrazione di residui basici, che potrebbero costituire dei potenziali siti d'attacco da parte del recettore.

I modelli molecolari ottenuti in soluzione per il dominio CFC di Cripto costituiscono il primo esempio di strutture tridimensionali sperimentali per questa classe di fattori di crescita.

Per approfondire lo studio strutturale è stato costruito un modello per omologia del dominio extracellulare (ECD) del recettore umano di tipo I dell'Attivina, ALK4-ECD, che interagisce con il dominio CFC della proteina Cripto durante il *signaling* di Nodal e un modello per il dominio extracellulare del recettore umano di tipo I dell'Attivina, ALK7-ECD, che invece interagisce direttamente con la proteina Nodal senza bisogno di Cripto. Dal confronto delle superfici di *Connolly* dell'ECD di ALK4 e ALK7, è stato possibile osservare una diversa distribuzione dei residui all'interfaccia di *binding*: in ALK4 sono presenti residui idrofobici e carichi, mentre in ALK7 si riscontrano essenzialmente aminoacidi idrofobici e polari (Fig.4.1).

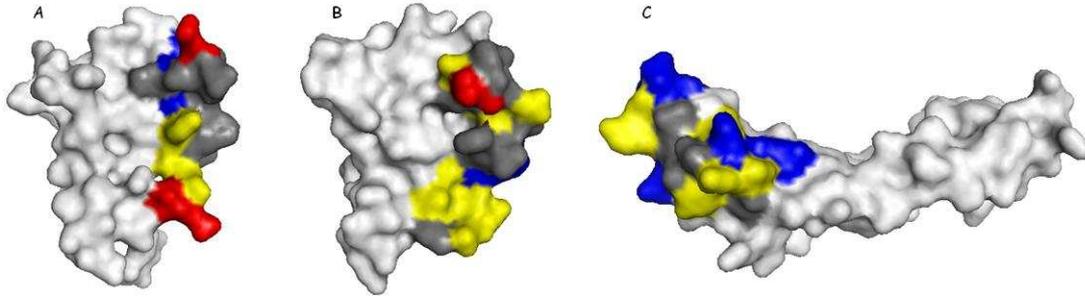


Fig.4.1: Superfici di *Connolly* dei modelli per omologia di: (A) ALK4-ECD; (B) ALK7-ECD; (C) monomero di Nodal. I residui all'interfaccia di *binding* sono colorati per tipo

Queste osservazioni hanno una grande rilevanza nella definizione delle superfici d'interazione con i possibili *partner* proteici.

Per investigare, quindi, l'interazione tra il dominio CFC di Cripto e il recettore ALK4, abbiamo effettuato il *docking* delle migliori dieci strutture NMR di hCFC con il modello per omologia del dominio extracellulare di ALK4, utilizzando il programma di *docking* rigido 3D-Dock. Le soluzioni del *docking* sono state selezionate considerando i risultati degli studi di mutagenesi che indicano i residui H<sup>9</sup> e W<sup>12</sup> di hCFC cruciali per il *binding* con il recettore ALK4. In base alle analisi condotte è stato scelto un modello rappresentativo del complesso hCFC/ALK4-ECD che presenta buona complementarità di superficie, contatti elettrostatici e idrofobici favorevoli, basso valore dell'energia totale (Fig.4.2). Ad esempio, sono state osservate interazioni aromatiche favorevoli tra il W<sup>12</sup> di hCFC e F<sup>56</sup> di ALK4. Inoltre, l'H<sup>9</sup> di hCFC si trova tra gli anelli aromatici della Y<sup>55</sup> e F<sup>56</sup> di ALK4 potenzialmente capaci di creare interazioni aromatiche e/o  $\pi$ -cationiche. Possibili interazioni elettrostatiche sono state osservate tra ALK4 E<sup>47</sup>- hCFC K<sup>1</sup>, ALK4 D<sup>62</sup>- hCFC K<sup>16</sup> e ALK4 E<sup>61</sup>- hCFC K<sup>15</sup>. Un'altra interazione favorevole all'interfaccia di *binding* è stata trovata tra i residui idrofobici ALK4 V<sup>49</sup>- hCFC P<sup>8</sup>.

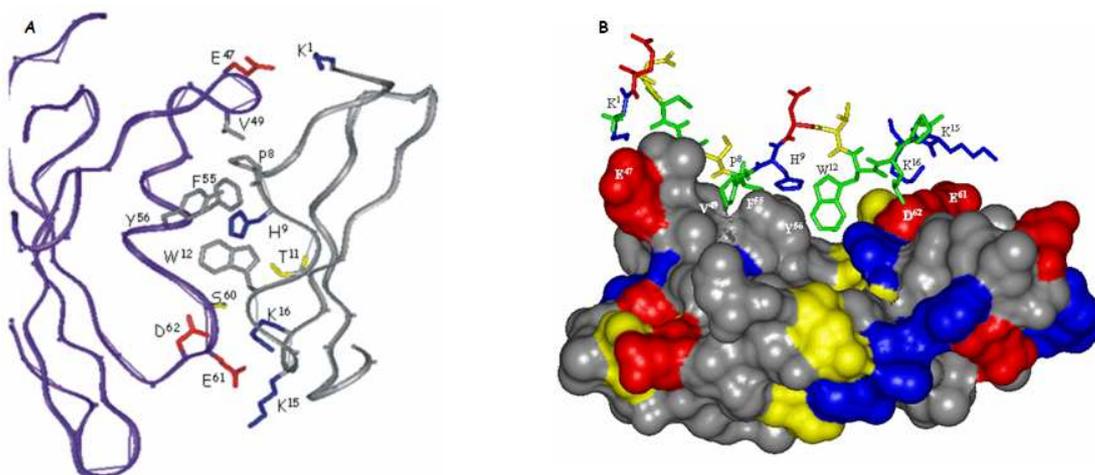


Fig.4.2: (A) Rappresentazione con *ribbon* del complesso ottenuto dal *docking* di ALK4-ECD (viola) e hCFC (grigio). I residui rilevanti all'interfaccia di *binding* sono mostrati e colorati per tipo: idrofobici in grigio, polari in giallo, acidi in rosso e basici in blu; (B) Rappresentazione dell'interazione all'interfaccia tra ALK4-ECD (superficie di *Connolly* con i residui colorati per tipo) e hCFC (rappresentazione con *stick*)

Questo modello, consistente con i precedenti dati di mutagenesi, fornisce una base strutturale per la progettazione di analoghi modificati, che potranno confermare le ipotesi di interazione molecolare tra CFC ed ALK4 ed eventualmente funzionare come antagonisti.

In una fase successiva è stata studiata la proteina Nodal, di cui ad oggi non si conosce la struttura tridimensionale né allo stato solido né in soluzione.

Nodal è un omodimero di 220 amminoacidi con 14 cisteine, che formano 6 ponti disolfuro intra-catena e uno inter-catena che lega i due monomeri. Non disponendo di informazioni strutturali, è stato costruito un modello per omologia sia del monomero sia del dimero della variante umana. Per la costruzione dei modelli per omologia sono stati utilizzati cinque template con identità di sequenza del 40%: BMP2, BMP3, BMP7, BMP9 e GDF5. Il modello del monomero di Nodal ottenuto per *homology modelling*, conserva il nodo di cisteina caratteristico della superfamiglia dei TGF- $\beta$  e la struttura tridimensionale è caratterizzata da due  $\beta$ -sheet antiparalleli e un  $\alpha$ -elica anfipatica perpendicolare ai  $\beta$ -sheet. I modelli ottenuti conservano le regioni caratteristiche di *binding* del recettore di tipo I ('*wrist epitope*') e del recettore di tipo II ('*knuckle epitope*').

Avendo a disposizione i modelli per omologia dei recettori ALK7-ECD, ALK4-ECD e del dimero della proteina Nodal, abbiamo effettuato esperimenti di *docking* tra i recettori e il ligando, per cercare di formulare una serie di ipotesi sugli elementi strutturali potenzialmente rilevanti ai fini del *binding*. Delle soluzioni del *docking* rigido, ottenute dal programma 3D-

Dock, sono state considerate quelle che presentavano l'interfaccia concava di *binding* dei recettori vicina al '*wrist epitope*' del dimero di Nodal. Sono stati scelti come modelli rappresentativi del complesso Nodal:ALK4-ECD e Nodal:ALK7-ECD, quelli che presentavano una buona complementarità di superficie, e il più alto valore di RPScore (Fig.4.3).

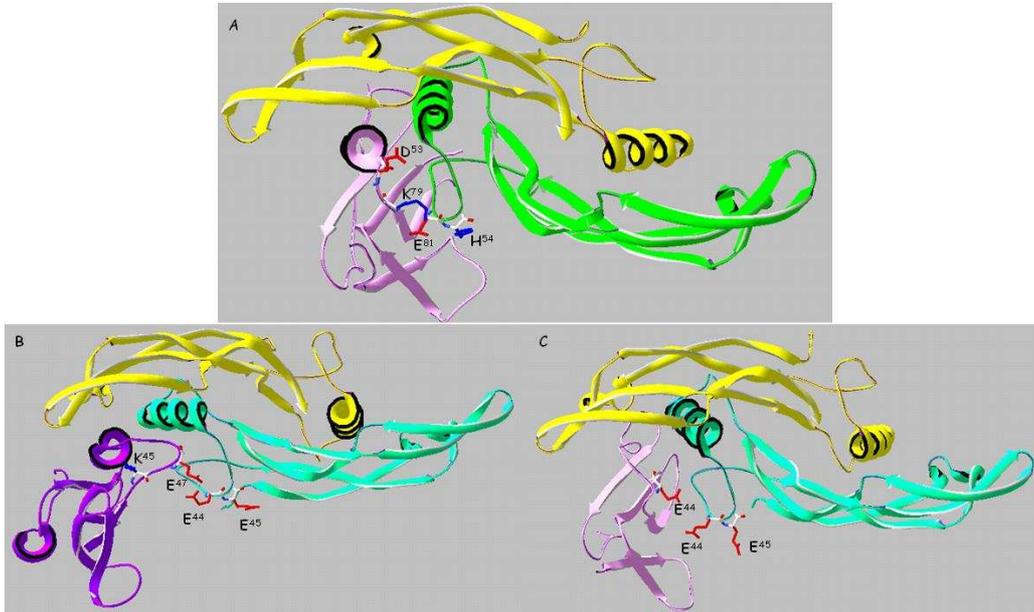


Fig.4.3: Rappresentazione con *ribbon* dei complessi (A) ALK3-ECD:BMP2 (pdbcode:1es7) (B) ALK4-ECD:Nodal e (C) ALK7-ECD:Nodal. I monomeri dei ligandi sono colorati in giallo e verde; i recettori ALK in violetto. Sono riportati in *stick* alcuni residui del *loop* pre-elica dei ligandi e del *loop* 23 dei recettori

Un'attenta analisi dei modelli per omologia dei recettori ALK7 e ALK4 ci ha permesso di evidenziare elementi strutturali differenti, oltre alla diversa distribuzione di residui amminoacidici precedentemente indicata, all'interfaccia di *binding*. Infatti, i modelli per omologia di ALK4-ECD e ALK7-ECD rivelano una diversa lunghezza del *loop* 23, che si affaccia sulla superficie di *binding* e può causare ingombro sterico. ALK7-ECD presenta un *loop* 23 più corto, di lunghezza paragonabile allo stesso nella struttura RX di ALK3-ECD; in ALK4-ECD, invece, il *loop* 23 è più lungo. Questa maggiore lunghezza del *loop* in ALK4 comporta un avvicinamento critico del residuo acido (E<sup>47</sup>) presente in questo tratto, che comporta interazioni elettrostatiche sfavorevoli con analoghi residui acidi presenti sul *loop* pre-elica di Nodal. Diversamente, il *loop* 23 più corto in ALK7-ECD, sebbene contenga conservato lo stesso residuo di acido glutammico (E<sup>44</sup>), non arriva a stabilire interazioni sfavorevoli con il *loop* pre-elica di Nodal. Considerando poi l'analoga distribuzione di residui polari/idrofobici all'interfaccia di *binding* di ALK7-ECD e Nodal, si può supporre che anche questo possa essere un elemento che favorisce la loro diretta interazione. Parallelamente all'analisi *in silico*, per

avere informazioni strutturali in soluzione della proteina umana di Nodal, sono stati sintetizzati tre suoi frammenti, che riproducono tratti di sequenza presumibilmente coinvolti nel legame con i suoi *partner* proteici. Di questi frammenti è stata effettuata un'analisi conformazionale in soluzione mediante NMR, che ha evidenziato, nelle condizioni sperimentali adoperate, una fedele riproduzione strutturale dei relativi tratti nella proteina parente.

Per poter validare le ipotesi d'interazione, avanzate sulla base dei dati strutturali raccolti, sono stati avviati esperimenti di *binding* 'Real Time Bia' con tecnica SPR.

Questi esperimenti hanno confermato l'affinità di *binding* tra il recettore di tipo I ALK4 e il dominio CFC, come già riportato in letteratura. Inoltre, è stato evidenziato che ALK4 è in grado di legare anche il dominio EGF-like isolato, benché con un'affinità notevolmente ridotta: la  $k_D$  è pari a circa 100  $\mu\text{M}$ , mentre quella di CFC è di circa 10  $\mu\text{M}$ . Entrambi i domini EGF-like e CFC, invece, non sono in grado di legare il recettore di tipo II ActRIIB.

I saggi di *binding* dei peptidi di Nodal con i recettori di tipo I e di tipo II e con Cripto sono in corso e potranno chiarire il ruolo delle regioni riprodotte nel riconoscimento delle varie "interfacce" presenti nel complesso recettoriale. Questo dato è di particolare rilevanza poiché i domini studiati, oltre ad essere degli ottimi modelli per lo studio delle proprietà conformazionali delle proteine di origine e dei *tool* ottimali per definirne le regioni di interazioni, possono essere impiegati direttamente come antagonisti solubili, diventando quindi degli *hit* con importanti proprietà biologiche oppure degli *scaffold* utilizzabili per il *design* di inibitori specifici ed efficaci.

In conclusione, questo lavoro di ricerca ha consentito di ottenere informazioni strutturali, finora inedite, su una serie di fattori proteici che intervengono nel processo di *signaling* Nodal/Cripto. È stata, infatti, determinata la struttura 3D del dominio CFC, ottenuto per sintesi chimica, della proteina Cripto umana e murina. Questo è stato il primo dato strutturale sperimentale presentato per i fattori EGF-CFC. Analogamente sono state determinate le strutture in soluzione di alcuni frammenti sintetici di Nodal, corrispondenti a delle putative regioni di legame con i recettori di membrana e con Cripto. Le ipotesi di legame sono state suggerite da studi di *docking* molecolare eseguiti utilizzando sia le strutture sperimentali ottenute sia i modelli molecolari dei recettori ALK4 e ALK7 e di Nodal ottenuti mediante *homology modelling*. Sulla base dei complessi teorici selezionati e dei risultati sperimentali di legame diretto è possibile formulare nuove ipotesi sugli elementi strutturali o sui singoli residui che intervengono nel riconoscimento e che sono potenzialmente rilevanti ai fini delle

mutue interazioni. Queste informazioni possono risultare utili per la messa a punto di esperimenti di mutagenesi che possono confermare le regioni di legame e nella progettazione di molecole antagoniste.

## Bibliografia

1. Kingsley DM. The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. *Genes Dev* 1994;8(2):133-146.
2. Hogan BL. Bone morphogenetic proteins: multifunctional regulators of vertebrate development. *Genes Dev* 1996;10(13):1580-1594.
3. Massague J, Chen YG. Controlling TGF-beta signaling. *Genes Dev* 2000;14(6):627-644.
4. Attisano L, Wrana JL. Signal transduction by the TGF-beta superfamily. *Science* 2002;296(5573):1646-1647.
5. Chang H, Brown CW, Matzuk MM. Genetic analysis of the mammalian transforming growth factor-beta superfamily. *Endocr Rev* 2002;23(6):787-823.
6. Roberts AB, Flanders KC, Heine UI, Jakowlew S, Kondaiah P, Kim SJ, Sporn MB. Transforming growth factor-beta: multifunctional regulator of differentiation and development. *Philos Trans R Soc Lond B Biol Sci* 1990;327(1239):145-154.
7. Roberts AB, Kim SJ, Sporn MB. Is there a common pathway mediating growth inhibition by TGF-beta and the retinoblastoma gene product? *Cancer Cells* 1991;3(1):19-21.
8. Roberts AB, McCune BK, Sporn MB. TGF-beta: regulation of extracellular matrix. *Kidney Int* 1992;41(3):557-559.
9. Moses HL, Yang EY, Pietenpol JA. Regulation of epithelial proliferation by TGF-beta. *Ciba Found Symp* 1991;157:66-74; discussion 75-80.
10. Sovak MA, Arsura M, Zanieski G, Kavanagh KT, Sonenshein GE. The inhibitory effects of transforming growth factor beta1 on breast cancer cell proliferation are mediated through regulation of aberrant nuclear factor-kappaB/Rel expression. *Cell Growth Differ* 1999;10(8):537-544.
11. Villanueva A, Garcia C, Paules AB, Vicente M, Megias M, Reyes G, de Villalonga P, Agell N, Lluís F, Bachs O, Capella G. Disruption of the antiproliferative TGF-beta signaling pathways in human pancreatic cancer cells. *Oncogene* 1998;17(15):1969-1978.
12. Engel ME, Datta PK, Moses HL. Signal transduction by transforming growth factor-beta: a cooperative paradigm with extensive negative regulation. *J Cell Biochem Suppl* 1998;30-31:111-122.
13. Howell GM, Sun L, Ziober BL, Wu SP, Brattain MG. The role of growth regulatory aberrations in progression of human colon carcinoma. *Cancer Metastasis Rev* 1993;12(3-4):275-286.
14. Kim YS, Yi Y, Choi SG, Kim SJ. Development of TGF-beta resistance during malignant progression. *Arch Pharm Res* 1999;22(1):1-8.
15. Miyaki M, Iijima T, Konishi M, Sakai K, Ishii A, Yasuno M, Hishima T, Koike M, Shitara N, Iwama T, Utsunomiya J, Kuroki T, Mori T. Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene* 1999;18(20):3098-3103.
16. Dennler S, Goumans MJ, ten Dijke P. Transforming growth factor beta signal transduction. *J Leukoc Biol* 2002;71(5):731-740.
17. Shi Y, Massague J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* 2003;113(6):685-700.
18. de Caestecker M. The transforming growth factor-beta superfamily of receptors. *Cytokine Growth Factor Rev* 2004;15(1):1-11.
19. Daopin S, Piez KA, Ogawa Y, Davies DR. Crystal structure of transforming growth factor-beta 2: an unusual fold for the superfamily. *Science* 1992;257(5068):369-373.
20. Schlunegger MP, Grutter MG. An unusual feature revealed by the crystal structure at 2.2 Å resolution of human transforming growth factor-beta 2. *Nature* 1992;358(6385):430-434.

21. Knaus P, Sebald W. Cooperativity of binding epitopes and receptor chains in the BMP/TGFbeta superfamily. *Biol Chem* 2001;382(8):1189-1195.
22. Lin SJ, Lerch TF, Cook RW, Jardetzky TS, Woodruff TK. The structural basis of TGF-beta, bone morphogenetic protein, and activin ligand binding. *Reproduction* 2006;132(2):179-190.
23. Mittl PR, Priestle JP, Cox DA, McMaster G, Cerletti N, Grutter MG. The crystal structure of TGF-beta 3 and comparison to TGF-beta 2: implications for receptor binding. *Protein Sci* 1996;5(7):1261-1271.
24. Hart PJ, Deep S, Taylor AB, Shu Z, Hinck CS, Hinck AP. Crystal structure of the human TbetaR2 ectodomain--TGF-beta3 complex. *Nat Struct Biol* 2002;9(3):203-208.
25. Scheufler C, Sebald W, Hulsmeyer M. Crystal structure of human bone morphogenetic protein-2 at 2.7 Å resolution. *J Mol Biol* 1999;287(1):103-115.
26. Kirsch T, Sebald W, Dreyer MK. Crystal structure of the BMP-2-BRIA ectodomain complex. *Nat Struct Biol* 2000;7(6):492-496.
27. Allendorph GP, Vale WW, Choe S. Structure of the ternary signaling complex of a TGF-beta superfamily member. *Proc Natl Acad Sci U S A* 2006;103(20):7643-7648.
28. Griffith DL, Keck PC, Sampath TK, Rueger DC, Carlson WD. Three-dimensional structure of recombinant human osteogenic protein 1: structural paradigm for the transforming growth factor beta superfamily. *Proc Natl Acad Sci U S A* 1996;93(2):878-883.
29. Greenwald J, Groppe J, Gray P, Wiater E, Kwiatkowski W, Vale W, Choe S. The BMP7/ActRII extracellular domain complex provides new insights into the cooperative nature of receptor assembly. *Mol Cell* 2003;11(3):605-617.
30. Brown MA, Zhao Q, Baker KA, Naik C, Chen C, Pukac L, Singh M, Tsareva T, Parice Y, Mahoney A, Roschke V, Sanyal I, Choe S. Crystal structure of BMP-9 and functional interactions with pro-region and receptors. *J Biol Chem* 2005;280(26):25111-25118.
31. Nickel J, Kotsch A, Sebald W, Mueller TD. A single residue of GDF-5 defines binding specificity to BMP receptor IB. *J Mol Biol* 2005;349(5):933-947.
32. Schreuder H, Liesum A, Pohl J, Kruse M, Koyama M. Crystal structure of recombinant human growth and differentiation factor 5: evidence for interaction of the type I and type II receptor-binding sites. *Biochem Biophys Res Commun* 2005;329(3):1076-1086.
33. Thompson TB, Woodruff TK, Jardetzky TS. Structures of an ActRIIB:activin A complex reveal a novel binding mode for TGF-beta ligand:receptor interactions. *Embo J* 2003;22(7):1555-1566.
34. Greenwald J, Vega ME, Allendorph GP, Fischer WH, Vale W, Choe S. A flexible activin explains the membrane-dependent cooperative assembly of TGF-beta family receptors. *Mol Cell* 2004;15(3):485-489.
35. Hinck AP, Archer SJ, Qian SW, Roberts AB, Sporn MB, Weatherbee JA, Tsang ML, Lucas R, Zhang BL, Wenker J, Torchia DA. Transforming growth factor beta 1: three-dimensional structure in solution and comparison with the X-ray structure of transforming growth factor beta 2. *Biochemistry* 1996;35(26):8517-8534.
36. Galloway SM, McNatty KP, Cambridge LM, Laitinen MP, Juengel JL, Jokiranta TS, McLaren RJ, Luiro K, Dodds KG, Montgomery GW, Beattie AE, Davis GH, Ritvos O. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nat Genet* 2000;25(3):279-283.
37. Shimasaki S, Moore RK, Otsuka F, Erickson GF. The bone morphogenetic protein system in mammalian reproduction. *Endocr Rev* 2004;25(1):72-101.
38. Nishimatsu S, Thomsen GH. Ventral mesoderm induction and patterning by bone morphogenetic protein heterodimers in *Xenopus* embryos. *Mech Dev* 1998;74(1-2):75-88.

39. Aono A, Hazama M, Notoya K, Taketomi S, Yamasaki H, Tsukuda R, Sasaki S, Fujisawa Y. Potent ectopic bone-inducing activity of bone morphogenetic protein-4/7 heterodimer. *Biochem Biophys Res Commun* 1995;210(3):670-677.
40. Suzuki A, Kaneko E, Maeda J, Ueno N. Mesoderm induction by BMP-4 and -7 heterodimers. *Biochem Biophys Res Commun* 1997;232(1):153-156.
41. Israel DI, Nove J, Kerns KM, Kaufman RJ, Rosen V, Cox KA, Wozney JM. Heterodimeric bone morphogenetic proteins show enhanced activity in vitro and in vivo. *Growth Factors* 1996;13(3-4):291-300.
42. Shimmi O, Umulis D, Othmer H, O'Connor MB. Facilitated transport of a Dpp/Scw heterodimer by Sog/Tsg leads to robust patterning of the *Drosophila* blastoderm embryo. *Cell* 2005;120(6):873-886.
43. Innis CA, Shi J, Blundell TL. Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. *Protein Eng* 2000;13(12):839-847.
44. Thompson TB, Lerch TF, Cook RW, Woodruff TK, Jardetzky TS. The structure of the follistatin:activin complex reveals antagonism of both type I and type II receptor binding. *Dev Cell* 2005;9(4):535-543.
45. Derynck R. TGF-beta-receptor-mediated signaling. *Trends Biochem Sci* 1994;19(12):548-553.
46. Andersson O. Genetic analysis of ligand-receptor interactions in the TGF-b superfamily during early embryonic development. Stockholm: Karolinska Institutet; 2006.
47. Mathews LS, Vale WW. Expression cloning of an activin receptor, a predicted transmembrane serine kinase. *Cell* 1991;65(6):973-982.
48. Attisano L, Wrana JL, Cheifetz S, Massague J. Novel activin receptors: distinct genes and alternative mRNA splicing generate a repertoire of serine/threonine kinase receptors. *Cell* 1992;68(1):97-108.
49. Baarends WM, van Helmond MJ, Post M, van der Schoot PJ, Hoogerbrugge JW, de Winter JP, Uilenbroek JT, Karels B, Wilming LG, Meijers JH, et al. A novel member of the transmembrane serine/threonine kinase receptor family is specifically expressed in the gonads and in mesenchymal cells adjacent to the mullerian duct. *Development* 1994;120(1):189-197.
50. di Clemente N, Wilson C, Faure E, Boussin L, Carmillo P, Tizard R, Picard JY, Vigier B, Josso N, Cate R. Cloning, expression, and alternative splicing of the receptor for anti-Mullerian hormone. *Mol Endocrinol* 1994;8(8):1006-1020.
51. Lin HY, Wang XF, Ng-Eaton E, Weinberg RA, Lodish HF. Expression cloning of the TGF-beta type II receptor, a functional transmembrane serine/threonine kinase. *Cell* 1992;68(4):775-785.
52. Kawabata M, Chytil A, Moses HL. Cloning of a novel type II serine/threonine kinase receptor through interaction with the type I transforming growth factor-beta receptor. *J Biol Chem* 1995;270(10):5625-5630.
53. Liu F, Ventura F, Doody J, Massague J. Human type II receptor for bone morphogenic proteins (BMPs): extension of the two-kinase receptor model to the BMPs. *Mol Cell Biol* 1995;15(7):3479-3486.
54. Nohno T, Ishikawa T, Saito T, Hosokawa K, Noji S, Wolsing DH, Rosenbaum JS. Identification of a human type II receptor for bone morphogenetic protein-4 that forms differential heteromeric complexes with bone morphogenetic protein type I receptors. *J Biol Chem* 1995;270(38):22522-22526.
55. Rosenzweig BL, Imamura T, Okadome T, Cox GN, Yamashita H, ten Dijke P, Heldin CH, Miyazono K. Cloning and characterization of a human type II receptor for bone morphogenetic proteins. *Proc Natl Acad Sci U S A* 1995;92(17):7632-7636.

56. Franzen P, ten Dijke P, Ichijo H, Yamashita H, Schulz P, Heldin CH, Miyazono K. Cloning of a TGF beta type I receptor that forms a heteromeric complex with the TGF beta type II receptor. *Cell* 1993;75(4):681-692.
57. ten Dijke P, Ichijo H, Franzen P, Schulz P, Saras J, Toyoshima H, Heldin CH, Miyazono K. Activin receptor-like kinases: a novel subclass of cell-surface receptors with predicted serine/threonine kinase activity. *Oncogene* 1993;8(10):2879-2887.
58. ten Dijke P, Yamashita H, Ichijo H, Franzen P, Laiho M, Miyazono K, Heldin CH. Characterization of type I receptors for transforming growth factor-beta and activin. *Science* 1994;264(5155):101-104.
59. Tsuchida K, Mathews LS, Vale WW. Cloning and characterization of a transmembrane serine kinase that acts as an activin type I receptor. *Proc Natl Acad Sci U S A* 1993;90(23):11242-11246.
60. Tsuchida K, Sawchenko PE, Nishikawa S, Vale WW. Molecular cloning of a novel type I receptor serine/threonine kinase for the TGF beta superfamily from rat brain. *Mol Cell Neurosci* 1996;7(6):467-478.
61. Yamaji N, Celeste AJ, Thies RS, Song JJ, Bernier SM, Goltzman D, Lyons KM, Nove J, Rosen V, Wozney JM. A mammalian serine/threonine kinase receptor specifically binds BMP-2 and BMP-4. *Biochem Biophys Res Commun* 1994;205(3):1944-1951.
62. Greenwald J, Fischer WH, Vale WW, Choe S. Three-finger toxin fold for the extracellular ligand-binding domain of the type II activin receptor serine kinase. *Nat Struct Biol* 1999;6(1):18-22.
63. Groppe J, Hinck CS, Samavarchi-Tehrani P, Zubieta C, Schuermann JP, Taylor AB, Schwarz PM, Wrana JL, Hinck AP. Cooperative assembly of TGF-beta superfamily signaling complexes is mediated by two disparate mechanisms and distinct modes of receptor binding. *Mol Cell* 2008;29(2):157-168.
64. Wells RG, Gilboa L, Sun Y, Liu X, Henis YI, Lodish HF. Transforming growth factor-beta induces formation of a dithiothreitol-resistant type I/Type II receptor complex in live cells. *J Biol Chem* 1999;274(9):5716-5722.
65. Gilboa L, Nohe A, Geissendorfer T, Sebald W, Henis YI, Knaus P. Bone morphogenetic protein receptor complexes on the surface of live cells: a new oligomerization mode for serine/threonine kinase receptors. *Mol Biol Cell* 2000;11(3):1023-1035.
66. Groppe J, Greenwald J, Wiater E, Rodriguez-Leon J, Economides AN, Kwiatkowski W, Affolter M, Vale WW, Belmonte JC, Choe S. Structural basis of BMP signalling inhibition by the cystine knot protein Noggin. *Nature* 2002;420(6916):636-642.
67. Weber D, Kotzsch A, Nickel J, Harth S, Seher A, Mueller U, Sebald W, Mueller TD. A silent H-bond can be mutationally activated for high-affinity interaction of BMP-2 and activin type IIB receptor. *BMC Struct Biol* 2007;7:6.
68. Massague J. TGF-beta signal transduction. *Annu Rev Biochem* 1998;67:753-791.
69. Keller S, Nickel J, Zhang JL, Sebald W, Mueller TD. Molecular recognition of BMP-2 and BMP receptor IA. *Nat Struct Mol Biol* 2004;11(5):481-488.
70. Cook RW, Thompson TB, Kurup SP, Jardetzky TS, Woodruff TK. Structural basis for a functional antagonist in the transforming growth factor beta superfamily. *J Biol Chem* 2005;280(48):40177-40186.
71. Muenster U, Harrison CA, Donaldson C, Vale W, Fischer WH. An activin-A/C chimera exhibits activin and myostatin antagonistic properties. *J Biol Chem* 2005;280(44):36626-36632.
72. Huse M, Muir TW, Xu L, Chen YG, Kuriyan J, Massague J. The TGF beta receptor activation process: an inhibitor- to substrate-binding switch. *Mol Cell* 2001;8(3):671-682.

73. Heldin CH, Miyazono K, ten Dijke P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* 1997;390(6659):465-471.
74. Padgett RW, Das P, Krishna S. TGF-beta signaling, Smads, and tumor suppressors. *Bioessays* 1998;20(5):382-390.
75. Raftery LA, Sutherland DJ. TGF-beta family signal transduction in Drosophila development: from Mad to Smads. *Dev Biol* 1999;210(2):251-268.
76. Attisano L, Wrana JL. Smads and Smads in TGF beta signalling. *Curr Opin Cell Biol* 1998;10(2):188-194.
77. Tsukazaki T, Chiang TA, Davison AF, Attisano L, Wrana JL. SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor. *Cell* 1998;95(6):779-791.
78. Licona-Limon P, Soldevila G. The role of TGF-beta superfamily during T cell development: new insights. *Immunol Lett* 2007;109(1):1-12.
79. Imamura T, Takase M, Nishihara A, Oeda E, Hanai J, Kawabata M, Miyazono K. Smad6 inhibits signalling by the TGF-beta superfamily. *Nature* 1997;389(6651):622-626.
80. Nakao A, Roijer E, Imamura T, Souchelnytskyi S, Stenman G, Heldin CH, ten Dijke P. Identification of Smad2, a human Mad-related protein in the transforming growth factor beta signaling pathway. *J Biol Chem* 1997;272(5):2896-2900.
81. Hata A, Lagna G, Massague J, Hemmati-Brivanlou A. Smad6 inhibits BMP/Smad1 signaling by specifically competing with the Smad4 tumor suppressor. *Genes Dev* 1998;12(2):186-197.
82. Yamada M, Szendro PI, Prokscha A, Schwartz RJ, Eichele G. Evidence for a role of Smad6 in chick cardiac development. *Dev Biol* 1999;215(1):48-61.
83. Zhao J, Crowe DL, Castillo C, Wuenschell C, Chai Y, Warburton D. Smad7 is a TGF-beta-inducible attenuator of Smad2/3-mediated inhibition of embryonic lung morphogenesis. *Mech Dev* 2000;93(1-2):71-81.
84. Zhao J, Shi W, Chen H, Warburton D. Smad7 and Smad6 differentially modulate transforming growth factor beta -induced inhibition of embryonic lung morphogenesis. *J Biol Chem* 2000;275(31):23992-23997.
85. Brunet LJ, McMahon JA, McMahon AP, Harland RM. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science* 1998;280(5368):1455-1457.
86. Gong Y, Krakow D, Marcelino J, Wilkin D, Chitayat D, Babul-Hirji R, Hudgins L, Cremers CW, Cremers FP, Brunner HG, Reinker K, Rimoin DL, Cohn DH, Goodman FR, Reardon W, Patton M, Francomano CA, Warman ML. Heterozygous mutations in the gene encoding noggin affect human joint morphogenesis. *Nat Genet* 1999;21(3):302-304.
87. Brown CB, Boyer AS, Runyan RB, Barnett JV. Requirement of type III TGF-beta receptor for endocardial cell transformation in the heart. *Science* 1999;283(5410):2080-2082.
88. Lewis KA, Gray PC, Blount AL, MacConell LA, Wiater E, Bilezikjian LM, Vale W. Betaglycan binds inhibin and can mediate functional antagonism of activin signalling. *Nature* 2000;404(6776):411-414.
89. Marchuk DA. Genetic abnormalities in hereditary hemorrhagic telangiectasia. *Curr Opin Hematol* 1998;5(5):332-338.
90. Onichtchouk D, Chen YG, Dosch R, Gawantka V, Delius H, Massague J, Niehrs C. Silencing of TGF-beta signalling by the pseudoreceptor BAMBI. *Nature* 1999;401(6752):480-485.
91. Cheng SK, Olale F, Bennett JT, Brivanlou AH, Schier AF. EGF-CFC proteins are essential coreceptors for the TGF-beta signals Vg1 and GDF1. *Genes Dev* 2003;17(1):31-36.
92. Rosa FM. Cripto, a multifunctional partner in signaling: molecular forms and activities. *Sci STKE* 2002;2002(158):PE47.
93. Schier AF, Shen MM. Nodal signalling in vertebrate development. *Nature* 2000;403(6768):385-389.

94. Chen C, Ware SM, Sato A, Houston-Hawkins DE, Habas R, Matzuk MM, Shen MM, Brown CW. The Vg1-related protein Gdf3 acts in a Nodal signaling pathway in the pre-gastrulation mouse embryo. *Development* 2006;133(2):319-329.
95. Shen MM. Decrypting the role of Cripto in tumorigenesis. *J Clin Invest* 2003;112(4):500-502.
96. Ciccocicola A, Dono R, Obici S, Simeone A, Zollo M, Persico MG. Molecular characterization of a gene of the 'EGF family' expressed in undifferentiated human NTERA2 teratocarcinoma cells. *Embo J* 1989;8(7):1987-1991.
97. Saloman DS, Bianco C, Ebert AD, Khan NI, De Santis M, Normanno N, Wechselberger C, Seno M, Williams K, Sanicola M, Foley S, Gullick WJ, Persico G. The EGF-CFC family: novel epidermal growth factor-related proteins in development and cancer. *Endocr Relat Cancer* 2000;7(4):199-226.
98. Schiffer SG, Foley S, Kaffashan A, Hronowski X, Zichittella AE, Yeo CY, Miatkowski K, Adkins HB, Damon B, Whitman M, Salomon D, Sanicola M, Williams KP. Fucosylation of Cripto is required for its ability to facilitate nodal signaling. *J Biol Chem* 2001;276(41):37769-37778.
99. Yan YT, Liu JJ, Luo Y, E C, Haltiwanger RS, Abate-Shen C, Shen MM. Dual roles of Cripto as a ligand and coreceptor in the nodal signaling pathway. *Mol Cell Biol* 2002;22(13):4439-4449.
100. Parisi S, D'Andrea D, Lago CT, Adamson ED, Persico MG, Minchiotti G. Nodal-dependent Cripto signaling promotes cardiomyogenesis and redirects the neural fate of embryonic stem cells. *J Cell Biol* 2003;163(2):303-314.
101. Sakuma R, Ohnishi Yi Y, Meno C, Fujii H, Juan H, Takeuchi J, Ogura T, Li E, Miyazono K, Hamada H. Inhibition of Nodal signalling by Lefty mediated through interaction with common receptors and efficient diffusion. *Genes Cells* 2002;7(4):401-412.
102. Schier AF. Nodal signaling in vertebrate development. *Annu Rev Cell Dev Biol* 2003;19:589-621.
103. Yeo C, Whitman M. Nodal signals to Smads through Cripto-dependent and Cripto-independent mechanisms. *Mol Cell* 2001;7(5):949-957.
104. Gray PC, Harrison CA, Vale W. Cripto forms a complex with activin and type II activin receptors and can block activin signaling. *Proc Natl Acad Sci U S A* 2003;100(9):5193-5198.
105. Adkins HB, Bianco C, Schiffer SG, Rayhorn P, Zafari M, Cheung AE, Orozco O, Olson D, De Luca A, Chen LL, Miatkowski K, Benjamin C, Normanno N, Williams KP, Jarpe M, LePage D, Salomon D, Sanicola M. Antibody blockade of the Cripto CFC domain suppresses tumor cell growth in vivo. *J Clin Invest* 2003;112(4):575-587.
106. Macias-Silva M, Abdollah S, Hoodless PA, Pirone R, Attisano L, Wrana JL. MADR2 is a substrate of the TGFbeta receptor and its phosphorylation is required for nuclear accumulation and signaling. *Cell* 1996;87(7):1215-1224.
107. Kretzschmar M, Liu F, Hata A, Doody J, Massague J. The TGF-beta family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. *Genes Dev* 1997;11(8):984-995.
108. Adamson ED, Minchiotti G, Salomon DS. Cripto: a tumor growth factor and more. *J Cell Physiol* 2002;190(3):267-278.
109. Woodruff TK, Lyon RJ, Hansen SE, Rice GC, Mather JP. Inhibin and activin locally regulate rat ovarian folliculogenesis. *Endocrinology* 1990;127(6):3196-3205.
110. van Schaik RH, Wierikx CD, Timmerman MA, Oomen MH, van Weerden WM, van der Kwast TH, van Steenbrugge GJ, de Jong FH. Variations in activin receptor, inhibin/activin subunit and follistatin mRNAs in human prostate tumour tissues. *Br J Cancer* 2000;82(1):112-117.
111. Su GH, Bansal R, Murphy KM, Montgomery E, Yeo CJ, Hruban RH, Kern SE. ACVR1B (ALK4, activin receptor type 1B) gene mutations in pancreatic carcinoma. *Proc Natl Acad Sci U S A* 2001;98(6):3254-3257.

112. Shi S, Ge C, Luo Y, Hou X, Haltiwanger RS, Stanley P. The threonine that carries fucose, but not fucose, is required for cripto to facilitate nodal signaling. *J Biol Chem* 2007.
113. Bianco C, Adkins HB, Wechselberger C, Seno M, Normanno N, De Luca A, Sun Y, Khan N, Kenney N, Ebert A, Williams KP, Sanicola M, Salomon DS. Cripto-1 activates nodal- and ALK4-dependent and -independent signaling pathways in mammary epithelial Cells. *Mol Cell Biol* 2002;22(8):2586-2597.
114. Marasco D, Saporito A, Ponticelli S, Chambery A, De Falco S, Pedone C, Minchiotti G, Ruvo M. Chemical synthesis of mouse cripto CFC variants. *Proteins* 2006;64(3):779-788.
115. de la Cruz JM, Bamford RN, Burdine RD, Roessler E, Barkovich AJ, Donnai D, Schier AF, Muenke M. A loss-of-function mutation in the CFC domain of TDGF1 is associated with human forebrain defects. *Hum Genet* 2002;110(5):422-428.
116. Calvanese L, Saporito A, Marasco D, D'Auria G, Minchiotti G, Pedone C, Paolillo L, Falcigno L, Ruvo M. Solution structure of mouse Cripto CFC domain and its inactive variant Trp107Ala. *J Med Chem* 2006;49(24):7054-7062.
117. Lohmeyer M, Harrison PM, Kannan S, DeSantis M, O'Reilly NJ, Sternberg MJ, Salomon DS, Gullick WJ. Chemical synthesis, structural modeling, and biological activity of the epidermal growth factor-like domain of human cripto. *Biochemistry* 1997;36(13):3837-3845.
118. Foley SF, van Vlijmen HW, Boynton RE, Adkins HB, Cheung AE, Singh J, Sanicola M, Young CN, Wen D. The CRIPTO/FRL-1/CRYPTIC (CFC) domain of human Cripto. Functional and structural insights through disulfide structure analysis. *Eur J Biochem* 2003;270(17):3610-3618.
119. Thornton JM. Disulphide bridges in globular proteins. *J Mol Biol* 1981;151(2):261-287.
120. Baron M, Norman DG, Harvey TS, Handford PA, Mayhew M, Tse AG, Brownlee GG, Campbell ID. The three-dimensional structure of the first EGF-like module of human factor IX: comparison with EGF and TGF-alpha. *Protein Sci* 1992;1(1):81-90.
121. Mer G, Hietter H, Kellenberger C, Renatus M, Luu B, Lefevre JF. Solution structure of PMP-C: a new fold in the group of small serine proteinase inhibitors. *J Mol Biol* 1996;258(1):158-171.
122. Merrifield B. Solid phase synthesis. *Science* 1986;232(4748):341-347.
123. Bax A, D. G. Davis MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J Magn Res* 1985;65:355-360.
124. Bax A. Practical aspects of two-dimensional transverse NOE spectroscopy. *J Magn Res* 1985;63:207-213.
125. Piantini U, Sorensen O, Ernst R. Multiple quantum filters for elucidating NMR coupling networks. *J Am Chem Soc* 1982;104(24):6800-6801.
126. States D, Haberhorn R, Ruben D. A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J Magn Res* 1982;48:286-292.
127. Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 1992;2(6):661-665.
128. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6(3):277-293.
129. Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 2004;278:313-352.
130. Wuthrich K. *NMR of Proteins and Nucleic Acids*. New York: Wiley; 1986.
131. Guntert P, Braun W, Wuthrich K. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol* 1991;217(3):517-530.
132. Guntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* 2004;278:353-378.

133. Bystrov VF. Spin-spin coupling and the conformational states of peptide systems. *Progress in Nuclear Magnetic Resonance Spectroscopy* 1976;10(2):41-82.
134. Karplus M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J Am Chem Soc* 1963;85(18):2870-2871.
135. Guntert P, Wuthrich K. Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J Biomol NMR* 1991;1(4):447-456.
136. Mc Cammon J, Harvey S. *Dynamics of Protein and Nucleic Acids*. Cambridge University Press, London; 1987.
137. Brooks III C, Karplus M, Pettitt B. *Proteins: A theoretical Perspective of Dynamics, Structure and Thermodynamics Advances in Chemical Physics* 1988;71.
138. van Gunsteren WF, Berendsen H. *Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry*. *Angew Chem Int Ed Engl* 1990;29:992-1023.
139. Pearlman DA, CD, Caldwell JW, Ross WS, Cheatham TE, III, DeBolt S, FD, Seibel G, Kollman P. AMBER, a package of computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun* 1995;91:1-41.
140. Weiner S, Kollmann P, Case D, Singh U, Ghio C, Alagona G, Profeta S, Weiner P. An all-atom force field for simulations of proteins and nucleic acids. *J Computat Chem* 1986;7:230-252.
141. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 1996;14(1):51-55, 29-32.
142. Rost B. Twilight zone of protein sequence alignments. *Protein Eng* 1999;12(2):85-94.
143. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* 1986;5(4):823-826.
144. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25(17):3389-3402.
145. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* 1999;287(4):797-815.
146. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005;33(Web Server issue):W244-248.
147. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 2005;21(7):951-960.
148. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234(3):779-815.
149. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 2004;60(Pt 12 Pt 1):2256-2268.
150. Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998;14(9):755-763.
151. Hoofst RW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature* 1996;381(6580):272.
152. Vriend G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 1990;8(1):52-56, 29.
153. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 1991;253(5016):164-170.
154. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature* 1992;356(6364):83-85.

155. Wallner B, Elofsson A. Can correct protein models be identified? *Protein Sci* 2003;12(5):1073-1086.
156. Wallner B, Elofsson A. Identification of correct regions in protein models using structural, alignment, and consensus information. *Protein Sci* 2006;15(4):900-913.
157. Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 1997;272(1):106-120.
158. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A* 1992;89(6):2195-2199.
159. Moont G, Gabb HA, Sternberg MJ. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins* 1999;35(3):364-373.
160. Jackson RM, Gabb HA, Sternberg MJ. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J Mol Biol* 1998;276(1):265-285.
161. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 1991;222(2):311-333.
162. Calvanese L, Saporito A, Oliva R, D'Auria G, Pedone C, Paolillo L, Ruvo M, Marasco D, Falcigno L. Structural insight into the interaction between the Cripto CFC domain and the ALK4 receptor. *J Pept Sci* 2008;in press.
163. Reissmann E, Jornvall H, Blokzijl A, Andersson O, Chang C, Minchiotti G, Persico MG, Ibanez CF, Brivanlou AH. The orphan receptor ALK7 and the Activin receptor ALK4 mediate signaling by Nodal proteins during vertebrate development. *Genes Dev* 2001;15(15):2010-2022.
164. Harrison CA, Gray PC, Koerber SC, Fischer W, Vale W. Identification of a functional binding site for activin on the type I receptor ALK4. *J Biol Chem* 2003;278(23):21129-21135.
165. Kelber JA, Shani G, Booker EC, Vale WW, Gray PC. Cripto is a noncompetitive activin antagonist that forms analogous signaling complexes with activin and nodal. *J Biol Chem* 2008;283(8):4490-4500.
166. Sternberg MJ, Gabb HA, Jackson RM, Moont G. Protein-protein docking. Generation and filtering of complexes. *Methods Mol Biol* 2000;143:399-415.
167. Fernandez-Recio J, Romero A, Sancho J. Energetics of a hydrogen bond (charged and neutral) and of a cation-pi interaction in apoflavodoxin. *J Mol Biol* 1999;290(1):319-330.
168. Meurisse R, Brasseur R, Thomas A. Aromatic side-chain interactions in proteins. Near- and far-sequence His-X pairs. *Biochim Biophys Acta* 2003;1649(1):85-96.
169. Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res* 2001;29(1):221-222.

## Appendice A

**TabellaA1:** Distanze (Å) ottenute dai NOE di hCFC in tampone fosfato/D<sub>2</sub>O 90/10 (v/v) pH 7

1	NLYS+	HN	1	NLYS+	QG	6.88	#peak	66
1	NLYS+	HN	1	NLYS+	HA	3.42	#peak	64
1	NLYS+	HN	1	NLYS+	QB	4.18	#peak	65
1	NLYS+	HA	1	NLYS+	QB	3.57	#peak	168
1	NLYS+	HA	2	GLU-	QB	3.72	#peak	168
1	NLYS+	QB	2	GLU-	HA	3.57	#peak	166
1	NLYS+	QB	2	GLU-	HN	4.98	#peak	52
1	NLYS+	HA	2	GLU-	HN	3.19	#peak	51
1	NLYS+	HA	23	TRP	QB	3.66	#peak	167
2	GLU-	HN	2	GLU-	QB	4.24	#peak	52
2	GLU-	HA	2	GLU-	QB	3.07	#peak	164
2	GLU-	HA	3	ASN	HN	3.91	#peak	81
3	ASN	HN	3	ASN	QB	4.79	#peak	82
3	ASN	QB	3	ASN	HD21	5.10	#peak	86
3	ASN	HD22	3	ASN	HD21	2.35	#peak	6
3	ASN	QB	3	ASN	HD22	5.32	#peak	122
3	ASN	QB	4	CYSS	HN	4.45	#peak	32
3	ASN	HA	4	CYSS	HN	2.96	#peak	33
4	CYSS	HN	4	CYSS	HA	2.96	#peak	33
4	CYSS	HN	4	CYSS	QB	5.04	#peak	34
4	CYSS	SG	22	CYSS	SG	2.10	#peak	0
4	CYSS	SG	22	CYSS	CB	3.10	#peak	0
4	CYSS	CB	22	CYSS	SG	3.10	#peak	0
6	SER	HA	8	PRO	QD	3.89	#peak	161
6	SER	QB	20	CYSS	QB	4.11	#peak	175
7	VAL	HN	7	VAL	QG2	3.90	#peak	85
7	VAL	HN	7	VAL	HA	2.69	#peak	83
7	VAL	HN	7	VAL	HB	2.66	#peak	84
7	VAL	HA	7	VAL	QQG	4.43	#peak	153
7	VAL	QQG	8	PRO	HA	7.22	#peak	134
8	PRO	HA	9	HIS	HN	2.88	#peak	67
10	ASP-	HN	10	ASP-	HA	2.92	#peak	74
11	THR	HN	11	THR	QG2	4.59	#peak	117
11	THR	HA	11	THR	QG2	3.67	#peak	148
11	THR	QG2	11	THR	HB	3.37	#peak	170
11	THR	HN	11	THR	HA	2.69	#peak	114
11	THR	HA	12	TRP	HN	2.77	#peak	53
11	THR	HB	12	TRP	HN	3.07	#peak	54
11	THR	HA	12	TRP	HD1	3.72	#peak	102
11	THR	QG2	12	TRP	HN	4.24	#peak	56
11	THR	HA	18	SER	QB	3.73	#peak	141
11	THR	HB	18	SER	QB	6.58	#peak	169
12	TRP	HN	12	TRP	QB	3.14	#peak	55
12	TRP	HA	12	TRP	QB	3.33	#peak	128

12 TRP	QB	12 TRP	HD1	3.33	#peak	105
12 TRP	HN	12 TRP	HD1	3.30	#peak	5
12 TRP	HA	12 TRP	HE3	3.23	#peak	88
12 TRP	HE3	16 LYS+	QB	5.97	#peak	92
12 TRP	HE3	16 LYS+	QG	4.45	#peak	93
12 TRP	QB	16 LYS+	QD	4.69	#peak	189
12 TRP	HA	13 LEU	HN	4.44	#peak	96
14 PRO	QG	15 LYS+	HA	5.29	#peak	145
15 LYS+	HA	15 LYS+	QG	3.41	#peak	146
14 PRO	HA	15 LYS+	HN	2.67	#peak	60
15 LYS+	HN	15 LYS+	QB	6.16	#peak	61
15 LYS+	HA	16 LYS+	HN	2.43	#peak	68
17 CYSS	SG	38 CYSS	SG	2.10	#peak	0
17 CYSS	SG	38 CYSS	CB	3.10	#peak	0
17 CYSS	CB	38 CYSS	SG	3.10	#peak	0
18 SER	HA	18 SER	QB	3.60	#peak	133
18 SER	QB	19 LEU	HN	5.06	#peak	48
18 SER	HA	19 LEU	HN	2.54	#peak	46
19 LEU	HN	19 LEU	HA	2.54	#peak	46
19 LEU	HN	19 LEU	QB	6.24	#peak	49
19 LEU	HN	19 LEU	HG	4.29	#peak	50
19 LEU	HA	19 LEU	HG	3.57	#peak	136
19 LEU	HA	19 LEU	QQD	5.82	#peak	137
19 LEU	HA	20 CYSS	HN	2.35	#peak	40
19 LEU	HG	20 CYSS	HB2	2.96	#peak	192
20 CYSS	HN	20 CYSS	QB	4.89	#peak	42
20 CYSS	HA	20 CYSS	QB	3.71	#peak	126
20 CYSS	HA	21 LYS+	HN	2.50	#peak	22
20 CYSS	QB	21 LYS+	HN	4.72	#peak	24
20 CYSS	HA	29 CYSS	HA	2.54	#peak	124
20 CYSS	SG	29 CYSS	SG	2.10	#peak	0
20 CYSS	SG	29 CYSS	CB	3.10	#peak	0
20 CYSS	CB	29 CYSS	SG	3.10	#peak	0
20 CYSS	HA	30 PHE	HN	3.19	#peak	27
21 LYS+	HN	21 LYS+	QB	5.28	#peak	26
21 LYS+	HN	29 CYSS	HA	3.38	#peak	23
22 CYSS	HN	22 CYSS	HA	4.40	#peak	57
22 CYSS	HN	22 CYSS	QB	3.64	#peak	58
23 TRP	QB	23 TRP	HD1	4.48	#peak	113
22 CYSS	HA	23 TRP	HN	2.66	#peak	20
22 CYSS	QB	28 ARG+	QD	4.96	#peak	190
23 TRP	HN	23 TRP	QB	4.09	#peak	21
23 TRP	HA	23 TRP	QB	3.64	#peak	129
23 TRP	HA	23 TRP	HD1	6.00	#peak	109
23 TRP	QB	23 TRP	HD1	4.04	#peak	110
23 TRP	QB	25 GLY	QA	3.62	#peak	163

23	TRP	HN	26	GLN	HN	4.06	#peak	3
25	GLY	HN	25	GLY	QA	3.29	#peak	62
26	GLN	QG	27	LEU	HN	4.98	#peak	43
27	LEU	HN	27	LEU	HG	2.35	#peak	44
27	LEU	HA	27	LEU	HG	2.92	#peak	138
27	LEU	HN	27	LEU	QQD	5.66	#peak	45
27	LEU	HA	27	LEU	QQD	5.06	#peak	139
27	LEU	HN	27	LEU	QB	3.58	#peak	44
27	LEU	HA	28	ARG+	HN	2.31	#peak	35
27	LEU	QQD	28	ARG+	HN	7.03	#peak	39
27	LEU	QB	28	ARG+	HN	4.37	#peak	38
27	LEU	QB	28	ARG+	QD	4.53	#peak	191
28	ARG+	HN	28	ARG+	HA	3.42	#peak	36
28	ARG+	HN	28	ARG+	QB	3.71	#peak	38
28	ARG+	QH2	28	ARG+	HE	3.67	#peak	19
28	ARG+	HN	28	ARG+	QG	6.73	#peak	37
28	ARG+	HA	28	ARG+	HE	2.88	#peak	118
28	ARG+	QB	28	ARG+	HE	3.43	#peak	120
28	ARG+	QB	28	ARG+	QD	3.80	#peak	191
28	ARG+	QD	28	ARG+	HE	4.83	#peak	119
29	CYSS	HA	29	CYSS	QB	3.45	#peak	127
29	CYSS	HA	30	PHE	HN	2.50	#peak	28
29	CYSS	QB	31	PRO	QD	3.92	#peak	172
30	PHE	HN	30	PHE	HA	3.19	#peak	29
30	PHE	HN	30	PHE	QB	4.93	#peak	30
30	PHE	HA	30	PHE	QB	3.64	#peak	131
30	PHE	HN	30	PHE	QE	6.73	#peak	4
30	PHE	HA	30	PHE	QE	5.97	#peak	97
30	PHE	QB	30	PHE	QE	6.41	#peak	99
30	PHE	QB	30	PHE	QD	6.30	#peak	111
30	PHE	QE	31	PRO	QD	8.52	#peak	98
30	PHE	HA	31	PRO	QD	3.62	#peak	130
30	PHE	QB	31	PRO	QD	4.34	#peak	176
31	PRO	HA	32	GLN	HN	2.88	#peak	69
32	GLN	QB	34	PHE	HA	4.04	#peak	144
32	GLN	QB	34	PHE	QE	9.13	#peak	107
33	ALA	HN	33	ALA	HA	3.87	#peak	76
33	ALA	HN	33	ALA	QB	4.02	#peak	75
33	ALA	HA	33	ALA	QB	3.29	#peak	157
33	ALA	QB	34	PHE	HN	6.22	#peak	94
33	ALA	HA	34	PHE	HN	3.26	#peak	90
34	PHE	HN	34	PHE	HA	3.11	#peak	89
34	PHE	HN	34	PHE	QB	3.80	#peak	91
34	PHE	HA	34	PHE	QB	3.99	#peak	142
34	PHE	HN	34	PHE	HZ	4.94	#peak	8
34	PHE	QB	34	PHE	QD	7.18	#peak	115

Appendice A

34	PHE	HA	34	PHE	QD	5.51	#peak	114
34	PHE	HN	35	LEU	QQD	6.01	#peak	95
34	PHE	HA	35	LEU	QQD	4.60	#peak	147
35	LEU	HN	35	LEU	HA	2.66	#peak	70
35	LEU	HG	36	PRO	QD	4.34	#peak	181
35	LEU	QQD	36	PRO	QD	7.92	#peak	185
36	PRO	QG	37	GLY	QA	6.37	#peak	156
36	PRO	QD	37	GLY	QA	4.00	#peak	158
36	PRO	QB	37	GLY	QA	3.96	#peak	159
37	GLY	QA	38	CYSS	HN	4.72	#peak	78
38	CYSS	HN	38	CYSS	HA	3.80	#peak	77
38	CYSS	HN	38	CYSS	QB	4.00	#peak	79
39	CASP-	HA	39	CASP-	QB	3.69	#peak	143

---

**TabellaA2:** Statistica strutturale di hCFC in tampone fosfato/D<sub>2</sub>O 90/10 (v/v) pH 7**CYANA structural statistics**

	NMR restraints
distance restraints	146
intraresidue	80
sequential ( $ i - j  = 1$ )	40
medium-range ( $1 <  i - j  \leq 4$ )	26
	Violation statistics (100 structures)
CYANA target function ( $\text{\AA}^2$ )	$1.54 \pm 0.70 \text{\AA}^2$
mean global backbone rmsd	$4.13 \pm 0.95 \text{\AA}$
mean global heavy rmsd	$5.09 \pm 0.92 \text{\AA}$
	Violation statistics (40 structures)
CYANA target function ( $\text{\AA}^2$ )	$1.05 \pm 0.10 \text{\AA}^2$
mean global backbone rmsd	$4.21 \pm 0.90 \text{\AA}$
mean global heavy rmsd	$5.10 \pm 0.90 \text{\AA}$
Constraint violations ( $\text{\AA}$ ):	
<b>Number residual distance &gt; 0.2 <math>\text{\AA}</math></b>	<b>0</b>
AMBER structural statistics	
	Violation statistics (10 structures)
0.2 < d $\leq$ 0.3	$5.2 \pm 2.0$
0.3 < d $\leq$ 0.4	$1.9 \pm 1.5$
0.4 < d $\leq$ 0.5	$1.3 \pm 1.2$
maximum violation	$0.48 \pm 0.18$
	AMBER Energies (kcal mol <sup>-1</sup> )
distance constraint	$18.1 \pm 5.6$
van der Waals	$-169.6 \pm 12.4$
total	$-157.2 \pm 43.3$
	RMSD to the Averaged Coordinates( $\text{\AA}$ )
Backbone (1-39)	$4.74 \pm 1.09$
All heavy (1-39)	$6.02 \pm 1.23$
Backbone (4-12)	$2.77 \pm 0.78$
All heavy (4-12)	$4.06 \pm 0.81$
Backbone (4-20)	$3.25 \pm 0.85$
All heavy (4-20)	$4.47 \pm 0.95$
Backbone (15-20)	$1.40 \pm 0.37$
All heavy (15-20)	$2.64 \pm 0.62$
Backbone (20-29)	$1.80 \pm 0.56$
All heavy (20-29)	$3.74 \pm 0.99$
Backbone (15-29)	$2.79 \pm 0.85$

All heavy (15-29)	4.44 ± 1.03
Backbone (29-38)	2.73 ± 0.56
All heavy (29-38)	4.23 ± 0.74

---

## Appendice B

**TabellaB1:** Distanze (Å) ottenute dai NOE di mCFC in tampone fosfato/D<sub>2</sub>O 90/10 (v/v) pH 7

1	NLYS	HN	1	NLYS	HA	2.85	#	peak	108
1	NLYS	HN	1	NLYS	QB	4.41	#	peak	110
1	NLYS	HN	1	NLYS	QG	4.33	#	peak	109
1	NLYS	HA	1	NLYS	QB	3.57	#	peak	257
1	NLYS	HA	1	NLYS	QG	3.57	#	peak	256
1	NLYS	HA	2	GLU-	HN	3.53	#	peak	71
1	NLYS	QB	2	GLU-	HA	4.45	#	peak	267
1	NLYS	QG	3	HIS+	HD2	7.43	#	peak	181
2	GLU-	HN	2	GLU-	HA	3.34	#	peak	72
2	GLU-	HN	2	GLU-	QB	3.67	#	peak	73
2	GLU-	HN	2	GLU-	QG	6.08	#	peak	74
2	GLU-	HN	3	HIS+	HN	4.25	#	peak	10
2	GLU-	HA	2	GLU-	QB	3.71	#	peak	264,266
2	GLU-	HA	2	GLU-	QG	4.12	#	peak	265
2	GLU-	HA	3	HIS+	HN	2.31	#	peak	159
2	GLU-	QB	3	HIS+	HN	3.96	#	peak	163
2	GLU-	QG	3	HIS+	HN	5.24	#	peak	164
3	HIS+	HN	3	HIS+	HA	2.66	#	peak	158
3	HIS+	HN	3	HIS+	QB	3.18	#	peak	162,161
3	HIS+	HN	3	HIS+	HE1	6.00	#	peak	13
3	HIS+	HA	3	HIS+	QB	2.91	#	peak	211,212
3	HIS+	QB	3	HIS+	HD2	3.16	#	peak	179,180
3	HIS+	HA	4	CYSS	HN	2.50	#	peak	65
3	HIS+	QB	4	CYSS	HN	3.73	#	peak	69
3	HIS+	HA	24	HIS+	HA	5.00	#	peak	210
3	HIS+	QB	24	HIS+	HA	5.00	#	peak	246
4	CYSS	HN	4	CYSS	HA	4.44	#	peak	66
4	CYSS	HN	4	CYSS	QB	3.45	#	peak	68
4	CYSS	HA	4	CYSS	QB	3.70	#	peak	226
4	CYSS	HN	24	HIS+	HA	5.00	#	peak	67
4	CYSS	HA	22	CYSS	QB	4.10	#	peak	227
4	CYSS	SG	22	CYSS	SG	2.10	#	peak	
4	CYSS	SG	22	CYSS	CB	3.10	#	peak	
4	CYSS	CB	22	CYSS	SG	3.10	#	peak	
6	SER	HA	6	SER	QB	3.03	#	peak	250,251
6	SER	HA	7	ILE	HN	3.04	#	peak	143
7	ILE	HN	7	ILE	HA	3.71	#	peak	144
7	ILE	HN	7	ILE	HB	3.31	#	peak	146
7	ILE	HN	7	ILE	QG1	4.14	#	peak	148
7	ILE	HA	7	ILE	HB	3.53	#	peak	255
7	ILE	HA	7	ILE	QD1	2.75	#	peak	259
7	ILE	HA	7	ILE	QG1	4.48	#	peak	258
7	ILE	HB	7	ILE	QD1	3.46	#	peak	302

7 ILE	HA	8 LEU	HN	3.82	#	peak	120
7 ILE	QG1	8 LEU	HN	4.98	#	peak	127
8 LEU	HN	8 LEU	HA	2.78	#	peak	119
8 LEU	HN	8 LEU	QB	3.85	#	peak	126
8 LEU	HN	8 LEU	QQD	4.16	#	peak	129
8 LEU	HA	8 LEU	QB	3.46	#	peak	247
8 LEU	HA	8 LEU	QQD	4.35	#	peak	249
8 LEU	HN	9 HIS+	HD2	6.00	#	peak	15
8 LEU	HA	9 HIS+	HD2	3.64	#	peak	188
8 LEU	QQD	9 HIS+	QB	8.20	#	peak	293,292
8 LEU	HN	19 LEU	QB	6.04	#	peak	125
9 HIS+	QB	9 HIS+	HD2	3.37	#	peak	190,191
9 HIS+	HD2	19 LEU	HG	5.00	#	peak	193
11 THR	HN	11 THR	HA	3.19	#	peak	115
11 THR	HN	11 THR	HB	4.21	#	peak	116
11 THR	HN	11 THR	QG2	4.22	#	peak	117
11 THR	HA	11 THR	HB	2.31	#	peak	239
11 THR	HB	11 THR	QG2	3.49	#	peak	280
11 THR	HA	12 TRP	HE3	5.00	#	peak	177
11 THR	HB	12 TRP	QB	5.25	#	peak	279
12 TRP	HA	12 TRP	QB	4.40	#	peak	203
12 TRP	QB	12 TRP	HD1	4.32	#	peak	172,173
12 TRP	HA	13 LEU	HN	2.54	#	peak	38
12 TRP	QB	13 LEU	HN	4.61	#	peak	42
12 TRP	QB	13 LEU	HN	4.11	#	peak	41
12 TRP	HE3	13 LEU	QB	7.43	#	peak	181
13 LEU	HN	13 LEU	HA	3.23	#	peak	39
13 LEU	HN	13 LEU	QB	4.02	#	peak	43
13 LEU	HN	13 LEU	HG	3.34	#	peak	44
13 LEU	HN	13 LEU	QQD	6.02	#	peak	45
13 LEU	HA	13 LEU	QB	3.79	#	peak	217,218
13 LEU	HA	13 LEU	HG	2.69	#	peak	219
13 LEU	HA	13 LEU	QQD	4.46	#	peak	222
13 LEU	HN	14 PRO	QD	7.00	#	peak	40
13 LEU	HA	14 PRO	QD	3.20	#	peak	213,214
13 LEU	QQD	14 PRO	QD	5.50	#	peak	270
14 PRO	HA	15 LYS+	HN	2.74	#	peak	95
14 PRO	QG	15 LYS+	HN	4.57	#	peak	97
15 LYS+	HN	15 LYS+	HA	2.85	#	peak	96
15 LYS+	HN	15 LYS+	QB	4.02	#	peak	98
15 LYS+	HN	15 LYS+	QG	6.15	#	peak	100
15 LYS+	HA	15 LYS+	QG	3.34	#	peak	238
15 LYS+	HA	15 LYS+	QB	3.39	#	peak	237no
16 LYS+	HA	16 LYS+	QB	4.05	#	peak	278,277
16 LYS+	HA	16 LYS+	QG	4.03	#	peak	268
17 CYSS	HA	17 CYSS	QB	4.65	#	peak	231

## Appendice B

17	CYSS	SG	38	CYSS	SG	2.10	#	peak	
17	CYSS	SG	38	CYSS	CB	3.10	#	peak	
17	CYSS	CB	38	CYSS	SG	3.10	#	peak	
18	SER	HA	18	SER	QB	3.41	#	peak	225,224
18	SER	HA	19	LEU	HN	2.69	#	peak	52
18	SER	QB	19	LEU	HN	4.26	#	peak	54,53
18	SER	HA	29	CYSS	HA	5.31	#	peak	201
18	SER	QB	30	LEU	HN	7.00	#	peak	59
19	LEU	HN	19	LEU	HG	4.25	#	peak	55
19	LEU	HA	19	LEU	QB	6.12	#	peak	233
19	LEU	HA	19	LEU	HG	2.77	#	peak	234
19	LEU	HN	30	LEU	HN	5.63	#	peak	8
19	LEU	HN	32	GLN	HN	4.97	#	peak	9
20	CYSS	HN	20	CYSS	QB	4.43	#	peak	93
20	CYSS	HA	20	CYSS	QB	3.56	#	peak	199
20	CYSS	HA	21	ARG+	HN	2.66	#	peak	46
20	CYSS	QB	21	ARG+	HN	4.42	#	peak	49
20	CYSS	HN	22	CYSS	QB	5.21	#	peak	94
20	CYSS	HA	29	CYSS	HA	2.43	#	peak	198
20	CYSS	QB	29	CYSS	HA	4.61	#	peak	205
20	CYSS	SG	29	CYSS	SG	2.10	#	peak	
20	CYSS	SG	29	CYSS	CB	3.10	#	peak	
20	CYSS	CB	29	CYSS	SG	3.10	#	peak	
20	CYSS	QB	30	LEU	HN	7.00	#	peak	62
20	CYSS	HA	30	LEU	HN	5.00	#	peak	56
21	ARG+	HN	21	ARG+	HA	3.19	#	peak	48
21	ARG+	HN	22	CYSS	QB	3.96	#	peak	50
21	ARG+	HA	22	CYSS	HN	2.54	#	peak	103
21	ARG+	QB	22	CYSS	QB	5.22	#	peak	300,296
21	ARG+	HN	27	LEU	QQD	7.64	#	peak	51
21	ARG+	HN	29	CYSS	HA	2.90	#	peak	47
21	ARG+	QG	22	CYSS	HN	6.27	#	peak	106
22	CYSS	HN	22	CYSS	HA	3.42	#	peak	102
22	CYSS	HN	22	CYSS	QB	3.29	#	peak	104,105
22	CYSS	HA	22	CYSS	QB	3.45	#	peak	215
22	CYSS	HA	23	TRP	HN	2.24	#	peak	31
22	CYSS	QB	23	TRP	HN	5.37	#	peak	34
22	CYSS	HN	29	CYSS	HA	5.00	#	peak	101
23	TRP	HN	23	TRP	HA	3.30	#	peak	30
23	TRP	HN	23	TRP	QB	3.83	#	peak	33
23	TRP	HN	23	TRP	HD1	6.00	#	peak	5
23	TRP	HA	23	TRP	QB	4.13	#	peak	209
23	TRP	HA	23	TRP	HE3	3.34	#	peak	168
23	TRP	QB	23	TRP	HD1	3.75	#	peak	167
23	TRP	QB	23	TRP	HE3	5.38	#	peak	171
23	TRP	HA	24	HIS+	HA	5.00	#	peak	206

23	TRP	HA	24	HIS+	QB	5.18	#	peak	208
23	TRP	QB	24	HIS+	HN	4.15	#	peak	84
23	TRP	HN	26	GLN	HN	2.77	#	peak	4
23	TRP	HN	26	GLN	QE2	7.00	#	peak	6
23	TRP	HN	27	LEU	QB	6.53	#	peak	36
23	TRP	HD1	28	HIS+	QB	5.00	#	peak	166
23	TRP	HD1	28	HIS+	HD2	3.99	#	peak	29
23	TRP	HE1	28	HIS+	HD2	4.18	#	peak	3
24	HIS+	HN	24	HIS+	HA	2.69	#	peak	80
24	HIS+	HN	24	HIS+	QB	4.16	#	peak	82
24	HIS+	HA	24	HIS+	QB	3.53	#	peak	244
24	HIS+	HA	24	HIS+	HD2	3.04	#	peak	186
24	HIS+	QB	24	HIS+	HD2	4.75	#	peak	187
24	HIS+	HA	28	HIS+	HD2	5.60	#	peak	194
25	GLY	HN	25	GLY	QA	3.56	#	peak	157,156
25	GLY	HN	26	GLN	HN	2.69	#	peak	23
25	GLY	QA	26	GLN	HN	3.62	#	peak	152
26	GLN	HN	26	GLN	HA	2.67	#	peak	151
26	GLN	HN	26	GLN	QB	3.22	#	peak	153,154
26	GLN	HN	26	GLN	QG	4.63	#	peak	155
26	GLN	QB	26	GLN	QE2	4.85	#	peak	133,134
26	GLN	QG	26	GLN	QE2	5.24	#	peak	135
26	GLN	HN	27	LEU	HN	5.16	#	peak	18
26	GLN	HA	27	LEU	HN	3.04	#	peak	87
26	GLN	QB	27	LEU	HN	6.16	#	peak	88
26	GLN	QG	27	LEU	HN	6.31	#	peak	89
27	LEU	HN	27	LEU	QB	4.29	#	peak	90
27	LEU	HN	27	LEU	HG	2.66	#	peak	91
27	LEU	HN	27	LEU	QQD	7.14	#	peak	92
27	LEU	HG	28	HIS+	QB	3.54	#	peak	291
28	HIS+	HA	28	HIS+	QB	3.48	#	peak	261,260
28	HIS+	HA	28	HIS+	HD2	4.06	#	peak	195
28	HIS+	QB	28	HIS+	HD2	3.64	#	peak	196,197
29	CYSS	HA	29	CYSS	QB	3.90	#	peak	204
29	CYSS	HA	30	LEU	HN	2.54	#	peak	57
29	CYSS	QB	30	LEU	HN	6.09	#	peak	60,61
30	LEU	HN	30	LEU	HA	3.68	#	peak	58
30	LEU	HN	30	LEU	QB	3.10	#	peak	63
30	LEU	HN	30	LEU	QQD	6.08	#	peak	64
30	LEU	HA	30	LEU	QQD	5.62	#	peak	220
30	LEU	HA	32	GLN	HN	3.83	#	peak	70
32	GLN	HN	32	GLN	HA	2.28	#	peak	75
32	GLN	HN	32	GLN	QB	4.18	#	peak	76
32	GLN	HN	32	GLN	QG	6.88	#	peak	77
33	THR	HN	33	THR	HA	3.08	#	peak	112
33	THR	HN	33	THR	HB	3.44	#	peak	113

Appendice B

33	THR	HN	33	THR	QG2	4.23	#	peak	114
33	THR	HA	33	THR	HB	2.73	#	peak	242
33	THR	HB	33	THR	QG2	3.23	#	peak	252
33	THR	HA	34	PHE	HN	3.08	#	peak	112
33	THR	HB	34	PHE	HN	3.44	#	peak	113
33	THR	QG2	34	PHE	HN	4.23	#	peak	114
34	PHE	HN	39	ASP-	HN	5.00	#	peak	11
36	PRO	HA	37	GLY	HN	3.16	#	peak	138
36	PRO	QG	37	GLY	HN	4.19	#	peak	142
36	PRO	QD	37	GLY	HN	4.43	#	peak	140,141
37	GLY	HN	37	GLY	QA	3.56	#	peak	139
37	GLY	HN	38	CYSS	HN	5.20	#	peak	19
37	GLY	QA	38	CYSS	HN	4.21	#	peak	136,137
38	CYSS	HN	38	CYSS	HA	2.50	#	peak	130
38	CYSS	HN	38	CYSS	QB	3.52	#	peak	131
38	CYSS	HA	38	CYSS	QB	4.65	#	peak	231
39	CASP-	HN	39	CASP-	HA	2.43	#	peak	85
39	CASP-	HN	39	CASP-	QB	4.48	#	peak	86
39	CASP-	HA	39	CASP-	QB	3.30	#	peak	236

---

**TabellaB2:** Statistica strutturale di mCFC in tampone fosfato/D<sub>2</sub>O 90/10 (v/v) pH7**CYANA structural statistics**

	NMR restraints
distance restraints	200
intraresidue	109
sequential ( $ i - j  = 1$ )	54
medium-range ( $1 <  i - j  \leq 4$ )	37
	Violation statistics (100 structures)
CYANA target function ( $\text{\AA}^2$ )	$2.34 \pm 1.87 \text{\AA}^2$
mean global backbone rmsd	$3.86 \pm 0.75 \text{\AA}$
mean global heavy rmsd	$4.88 \pm 0.79 \text{\AA}$

	Violation statistics (40 structures)
CYANA target function ( $\text{\AA}^2$ )	$1.03 \pm 0.12 \text{\AA}^2$
mean global backbone rmsd	$3.24 \pm 0.58 \text{\AA}$
mean global heavy rmsd	$4.13 \pm 0.57 \text{\AA}$

Constraint violations ( $\text{\AA}$ ):**Number residual distance > 0.2  $\text{\AA}$**  0

## AMBER structural statistics

	Violation statistics (10 structures)
$0.2 < d \leq 0.3$	$4.4 \pm 2.2$
$0.3 < d \leq 0.4$	$1.7 \pm 0.9$
$0.4 < d \leq 0.5$	$0.6 \pm 0.7$
maximum violation	$0.42 \pm 0.09$

AMBER Energies ( $\text{kcal mol}^{-1}$ )

distance constraint	$23.2 \pm 6.5$
van der Waals	$-192.7 \pm 8.9$
total	$-73.6 \pm 33.6$

RMSD to the Averaged Coordinates( $\text{\AA}$ )

Backbone (1-39)	$4.40 \pm 1.19$
All heavy (1-39)	$5.55 \pm 1.31$
Backbone (4-12)	$2.32 \pm 0.53$
All heavy (4-12)	$3.69 \pm 0.67$
Backbone (4-20)	$1.60 \pm 0.36$
All heavy (4-20)	$3.22 \pm 0.61$
Backbone (15-20)	$3.29 \pm 0.74$
All heavy (15-20)	$4.73 \pm 0.85$
Backbone (20-29)	$1.89 \pm 0.47$
All heavy (20-29)	$3.21 \pm 0.68$
Backbone (15-29)	$0.88 \pm 0.26$

All heavy (15-29)	$1.88 \pm 0.47$
Backbone (29-38)	$2.77 \pm 0.74$
All heavy (29-38)	$4.13 \pm 0.85$

---

## Appendice C

**TabellaC1:** Distanze (Å) ottenute dai NOE di Nodal53-69 in TFE/H<sub>2</sub>O 30:70 (v:v)

1	NPRO	HA	2	THR	HN	4.71	#peak	15
1	NPRO	QD	3	ASN	HA	5.35	#peak	149
2	THR	HA	2	THR	QG2	5.60	#peak	169
2	THR	HA	3	ASN	HN	3.04	#peak	18
2	THR	HB	3	ASN	HN	4.63	#peak	19
2	THR	QG2	3	ASN	HN	6.53	#peak	21
3	ASN	HN	3	ASN	HA	2.53	#peak	17
3	ASN	HN	3	ASN	QB	3.69	#peak	20
3	ASN	HA	3	ASN	QB	3.66	#peak	151
3	ASN	HA	4	HIS+	QB	4.85	#peak	150
3	ASN	HA	4	HIS+	HN	2.20	#peak	30
3	ASN	HA	5	ALA	HN	2.93	#peak	31
3	ASN	HA	6	TYR	HN	3.55	#peak	60
3	ASN	QB	6	TYR	HN	5.08	#peak	64
3	ASN	HA	7	ILE	HN	4.35	#peak	78
4	HIS+	HN	4	HIS+	HA	4.71	#peak	15
4	HIS+	HN	4	HIS+	QB	4.58	#peak	16
4	HIS+	HA	5	ALA	HN	3.22	#peak	32
4	HIS+	HA	4	HIS+	QB	3.10	#peak	162
4	HIS+	QB	5	ALA	HN	4.03	#peak	34
4	HIS+	QB	5	ALA	HA	5.41	#peak	171
4	HIS+	HA	7	ILE	HN	4.09	#peak	79
4	HIS+	HA	7	ILE	HB	3.33	#peak	163
4	HIS+	HA	7	ILE	QG2	6.16	#peak	164
4	HIS+	HA	8	GLN	HN	4.38	#peak	38
5	ALA	HN	5	ALA	HA	2.93	#peak	33
5	ALA	HN	5	ALA	QB	3.49	#peak	37
5	ALA	QB	5	ALA	HA	3.43	#peak	177
5	ALA	HN	6	TYR	HN	2.93	#peak	1
5	ALA	HN	6	TYR	QB	5.31	#peak	35
5	ALA	HA	6	TYR	HN	3.65	#peak	62
5	ALA	QB	6	TYR	HN	4.71	#peak	66
5	ALA	QB	7	ILE	HN	6.53	#peak	84
5	ALA	HN	7	ILE	HN	3.72	#peak	3
5	ALA	HA	7	ILE	HN	5.00	#peak	74
5	ALA	HN	8	GLN	QB	5.94	#peak	36
5	ALA	HA	8	GLN	HN	3.98	#peak	40
5	ALA	QB	8	GLN	HN	6.53	#peak	50
6	TYR	HN	6	TYR	HA	2.83	#peak	61
6	TYR	HN	6	TYR	QB	3.26	#peak	63
6	TYR	HN	7	ILE	HN	3.19	#peak	14
6	TYR	HN	7	ILE	HB	5.07	#peak	65

Appendice C

6 TYR	HA	7 ILE	HN	3.26	#peak	73
6 TYR	HA	6 TYR	QB	2.87	#peak	165
6 TYR	QB	7 ILE	HN	3.86	#peak	81
6 TYR	QB	7 ILE	HA	6.50	#peak	198
6 TYR	QB	8 GLN	HN	6.20	#peak	43
6 TYR	HN	8 GLN	HN	4.64	#peak	2
6 TYR	HA	8 GLN	HN	5.50	#peak	39
6 TYR	HA	9 SER	HN	5.00	#peak	73
7 ILE	HN	7 ILE	HA	2.99	#peak	80
7 ILE	HN	7 ILE	QG1	4.15	#peak	85
7 ILE	HA	7 ILE	HB	3.26	#peak	199
7 ILE	HA	7 ILE	QG1	4.29	#peak	201
7 ILE	HN	7 ILE	QG2	5.04	#peak	86
7 ILE	HN	7 ILE	QD1	6.10	#peak	87
7 ILE	HA	7 ILE	QG2	3.85	#peak	204
7 ILE	HN	8 GLN	HN	2.63	#peak	4
7 ILE	QG1	8 GLN	HN	6.50	#peak	49
7 ILE	HA	8 GLN	HN	3.65	#peak	42
7 ILE	HB	8 GLN	HN	3.06	#peak	47
7 ILE	QG2	8 GLN	HN	4.78	#peak	51
7 ILE	QD1	8 GLN	HN	5.87	#peak	52
7 ILE	HA	9 SER	HN	5.00	#peak	80
7 ILE	HA	10 LEU	HN	3.49	#peak	104
7 ILE	HA	10 LEU	QB	3.99	#peak	200
7 ILE	HA	10 LEU	HG	3.51	#peak	202
7 ILE	HA	10 LEU	QQD	5.42	#peak	205
7 ILE	HA	11 LEU	HN	4.52	#peak	56
8 GLN	HN	8 GLN	HA	2.70	#peak	41
8 GLN	HN	8 GLN	QB	3.79	#peak	44
8 GLN	HN	8 GLN	QG	4.85	#peak	45
8 GLN	QB	8 GLN	QE2	6.10	#peak	126
8 GLN	HA	9 SER	HN	3.37	#peak	76
8 GLN	QB	9 SER	HN	4.29	#peak	77
8 GLN	QB	9 SER	HA	5.38	#peak	173
8 GLN	HN	9 SER	HN	2.63	#peak	4
8 GLN	QG	9 SER	HA	6.50	#peak	174
8 GLN	HN	10 LEU	QB	5.25	#peak	48
8 GLN	HA	10 LEU	HN	4.06	#peak	103
8 GLN	HA	11 LEU	HN	3.62	#peak	55
9 SER	HN	9 SER	HA	2.86	#peak	74
9 SER	HN	9 SER	QB	3.56	#peak	75
9 SER	HA	9 SER	QB	3.06	#peak	170
9 SER	HN	10 LEU	HN	3.98	#peak	12
9 SER	QB	10 LEU	HN	4.45	#peak	102
9 SER	QB	10 LEU	HG	6.50	#peak	196

Appendice C

9	SER	QB	10	LEU	QQD	7.93	#peak	197
9	SER	QB	11	LEU	QB	6.50	#peak	195
9	SER	QB	13	ARG+	HN	5.81	#peak	109
10	LEU	HN	10	LEU	HA	2.78	#peak	96
10	LEU	HN	10	LEU	QB	3.20	#peak	105
10	LEU	HA	10	LEU	QB	4.05	#peak	192
10	LEU	HN	10	LEU	HG	2.56	#peak	106
10	LEU	HN	10	LEU	QQD	5.21	#peak	107
10	LEU	HA	10	LEU	HG	2.70	#peak	193
10	LEU	HA	10	LEU	QQD	4.46	#peak	194
10	LEU	HA	11	LEU	HN	3.49	#peak	54
10	LEU	HN	11	LEU	HN	2.40	#peak	5
10	LEU	QB	11	LEU	HN	3.27	#peak	57
11	LEU	HN	11	LEU	HA	2.96	#peak	53
11	LEU	HN	11	LEU	QB	2.87	#peak	57
11	LEU	HN	11	LEU	HG	3.19	#peak	58
11	LEU	HN	11	LEU	QQD	5.21	#peak	59
11	LEU	HA	11	LEU	HG	2.53	#peak	187
11	LEU	HA	11	LEU	QQD	4.59	#peak	189
11	LEU	HN	12	LYS+	HN	2.40	#peak	5
11	LEU	HA	12	LYS+	HN	3.59	#peak	95
11	LEU	QQD	12	LYS+	HN	6.20	#peak	101
11	LEU	HA	13	ARG+	QD	5.41	#peak	183
11	LEU	HN	13	ARG+	HN	4.25	#peak	6
11	LEU	HA	14	TYR	QB	5.05	#peak	182
12	LYS+	HN	12	LYS+	HA	2.78	#peak	96
12	LYS+	HN	12	LYS+	QB	4.12	#peak	98
12	LYS+	HN	12	LYS+	QG	5.87	#peak	100
12	LYS+	HN	13	ARG+	HN	2.81	#peak	11
12	LYS+	HA	13	ARG+	HN	3.25	#peak	108
12	LYS+	QB	13	ARG+	HN	4.19	#peak	111
12	LYS+	QB	13	ARG+	HA	4.22	#peak	176
12	LYS+	QG	13	ARG+	HN	6.40	#peak	113
12	LYS+	QG	13	ARG+	HA	5.41	#peak	179
12	LYS+	HN	14	TYR	HN	4.73	#peak	13
12	LYS+	QZ	13	ARG+	HN	4.43	#peak	11
12	LYS+	HA	15	GLN	QG	5.21	#peak	190
13	ARG+	HN	13	ARG+	QB	3.20	#peak	112
13	ARG+	HN	13	ARG+	QD	6.10	#peak	110
13	ARG+	HN	13	ARG+	QG	3.82	#peak	114
13	ARG+	HA	13	ARG+	QB	4.22	#peak	178
13	ARG+	HA	13	ARG+	QD	5.21	#peak	172
13	ARG+	HA	13	ARG+	QG	3.30	#peak	177
13	ARG+	QB	13	ARG+	QD	4.02	#peak	209
13	ARG+	QG	13	ARG+	QD	3.00	#peak	210

## Appendice C

13	ARG+	HN	15	GLN	HN	5.00	#peak	11
13	ARG+	HN	14	TYR	HN	2.96	#peak	8
13	ARG+	HA	14	TYR	HN	3.53	#peak	68
13	ARG+	QB	14	TYR	HN	4.09	#peak	71
13	ARG+	QG	14	TYR	HN	5.84	#peak	72
14	TYR	HN	14	TYR	HA	2.99	#peak	67
14	TYR	HN	14	TYR	QB	3.99	#peak	69
14	TYR	HA	14	TYR	QB	4.52	#peak	155
14	TYR	HA	15	GLN	HN	3.35	#peak	88
14	TYR	QB	15	GLN	HN	5.05	#peak	90
14	TYR	QB	15	GLN	HA	4.92	#peak	155
14	TYR	HA	16	PRO	QD	3.33	#peak	154
15	GLN	HN	15	GLN	QB	3.92	#peak	93
15	GLN	HN	15	GLN	HA	2.89	#peak	88
15	GLN	HA	15	GLN	QB	3.79	#peak	158
15	GLN	HN	15	GLN	QG	4.81	#peak	92
15	GLN	QG	15	GLN	QE2	6.10	#peak	117
15	GLN	HA	15	GLN	QG	4.58	#peak	157
15	GLN	HN	16	PRO	QD	5.54	#peak	89
15	GLN	HA	16	PRO	QD	3.66	#peak	154
15	GLN	QB	16	PRO	QD	4.19	#peak	207
16	PRO	HA	17	CHIS+	HN	2.93	#peak	23
16	PRO	QB	17	CHIS+	HN	5.08	#peak	27
16	PRO	QD	17	CHIS+	HN	6.50	#peak	24
16	PRO	QG	17	CHIS+	HN	6.14	#peak	28
17	CHIS+	HN	17	CHIS+	HA	3.22	#peak	22
17	CHIS+	HN	17	CHIS+	QB	4.42	#peak	25
17	CHIS+	HA	17	CHIS+	QB	6.10	#peak	152

---

**TabellaC2:** Statistica strutturale di Nodal53-69 in TFE/H<sub>2</sub>O 30:70 (v:v)**CYANA structural statistics**

	NMR restraints
distance restraints	160
intraresidue	63
sequential ( $ i - j  = 1$ )	60
medium-range ( $1 <  i - j  \leq 4$ )	37
torsion angle restraints	0
	Violation statistics (100 structures)
CYANA target function ( $\text{\AA}^2$ )	$0.27 \pm 0.37 \text{\AA}^2$
mean global backbone rmsd	$1.64 \pm 0.50 \text{\AA}$
mean global heavy rmsd	$2.59 \pm 0.57 \text{\AA}$

	Violation statistics (40 structures)
CYANA target function ( $\text{\AA}^2$ )	$9.9 \text{ E-}02 \pm 7.4 \text{ E-}03$
mean global backbone rmsd	$1.26 \pm 0.39 \text{\AA}$
mean global heavy rmsd	$2.15 \pm 0.50 \text{\AA}$
Residual distance constraint violations ( $\text{\AA}$ ):	
<b>Number &gt; 0.2 <math>\text{\AA}</math></b>	0

**AMBER structural statistics**

	Violation statistics (10 structures)
<b>0.2 &lt; d &lt; 0.3</b>	$2.50 \pm 1.57$
<b>0.3 <math>\leq</math> d &lt; 0.4</b>	$3.60 \pm 0.92$
<b>d <math>\geq</math> 0.4</b>	$3.70 \pm 1.27$
Maximum violation	$0.57 \pm 0.13$
AMBER energies (kcal mol <sup>-1</sup> )	
Distance constraint	$26.03 \pm 4.00$
Van der Waals	$-83.18 \pm 3.10$
Total	$-89.44 \pm 11.71$
rmsd to the averaged coordinates ( $\text{\AA}$ )	
Backbone (53-69)	$1.67 \pm 0.61 \text{\AA}$
All heavy (53-69)	$2.75 \pm 0.81 \text{\AA}$
Backbone (58-64)	$0.26 \pm 0.10 \text{\AA}$
All heavy (58-64)	$1.29 \pm 0.39 \text{\AA}$

## Appendice D

TabellaD1: Distanze (Å) ottenute dai NOE di Nodal43-69 in TFE/H<sub>2</sub>O 30:70 (v:v)

1	NCYA	HA	1	NCYA	QB	3,46	#peak	134
1	NCYA	HA	2	PRO	QD	3,84	#peak	133
1	NCYA	HN	1	NCYA	QB	4,32	#peak	52
2	PRO	HA	3	ASN	HN	2,64	#peak	31
3	ASN	HN	3	ASN	QB	3,36	#peak	38
4	PRO	QB	5	VAL	HN	4,20	#peak	76
4	PRO	HA	5	VAL	HN	4,08	#peak	70
4	PRO	QD	5	VAL	HN	4,34	#peak	73
5	VAL	HN	5	VAL	HB	3,12	#peak	75
5	VAL	HA	5	VAL	HB	2,88	#peak	172
5	VAL	HN	5	VAL	QQG	4,72	#peak	77
5	VAL	HN	6	GLY	HN	7,20	#peak	8
5	VAL	HA	6	GLY	HN	4,20	#peak	36
5	VAL	QQG	6	GLY	HN	6,05	#peak	41
6	GLY	HN	6	GLY	QA	3,36	#peak	37
6	GLY	HN	7	GLU	HN	3,10	#peak	7
6	GLY	QA	7	GLU	HN	3,84	#peak	58
7	GLU	HN	7	GLU	QB	3,60	#peak	62
7	GLU	HN	7	GLU	QG	5,04	#peak	60
7	GLU	HN	8	GLU	HN	4,30	#peak	7
7	GLU	HA	8	GLU	HN	3,84	#peak	39
8	GLU	HA	8	GLU	QB	3,36	#peak	158
8	GLU	HN	9	PHE	HN	7,20	#peak	8
8	GLU	HA	9	PHE	HN	3,84	#peak	71
9	PHE	HN	9	PHE	QB	3,36	#peak	74
9	PHE	HA	10	HIS+	HN	3,14	#peak	56
10	HIS+	HA	10	HIS+	QB	3,68	#peak	140
10	HIS+	HA	11	PRO	QD	4,20	#peak	139
10	HIS+	QB	11	PRO	QD	4,88	#peak	183
10	HIS+	HN	11	PRO	QD	5,39	#peak	57
11	PRO	HA	12	THR	HN	4,20	#peak	49
11	PRO	QB	12	THR	HN	4,16	#peak	53
12	THR	HN	12	THR	HB	3,36	#peak	50
12	THR	HN	12	THR	QG2	4,01	#peak	55
12	THR	HA	13	ASN	HN	3,84	#peak	25
13	ASN	HN	13	ASN	HA	3,36	#peak	25
13	ASN	HA	14	HIS+	HN	4,70	#peak	22
13	ASN	QB	14	HIS+	HN	4,20	#peak	24
13	ASN	HN	13	ASN	QD2	4,98	#peak	10
14	HIS+	HA	14	HIS+	QB	3,36	#peak	148
14	HIS+	HN	14	HIS+	HA	3,08	#peak	21
14	HIS+	HN	14	HIS+	QB	3,60	#peak	23
14	HIS+	HA	15	ALA	HN	4,39	#peak	29

## Appendice D

14	HIS+	HN	15	ALA	HN	3,84	#peak	1
14	HIS+	QB	15	ALA	HN	4,56	#peak	32
15	ALA	HN	15	ALA	HA	2,64	#peak	30
15	ALA	HN	15	ALA	QB	4,24	#peak	35
15	ALA	QB	15	ALA	HA	3,88	#peak	162
15	ALA	HN	16	TYR	HN	3,84	#peak	4
15	ALA	QB	16	TYR	HN	4,03	#peak	65
15	ALA	HN	16	TYR	QB	4,30	#peak	33
15	ALA	QB	17	ILE	HN	5,93	#peak	91
15	ALA	HA	17	ILE	HN	6,00	#peak	0
16	TYR	HN	16	TYR	QB	3,60	#peak	59
16	TYR	HA	16	TYR	QB	3,36	#peak	151
16	TYR	HN	16	TYR	HA	3,07	#peak	67
16	TYR	HN	17	ILE	HN	2,96	#peak	11
16	TYR	QB	17	ILE	HN	5,04	#peak	87
16	TYR	HA	17	ILE	HN	4,30	#peak	83
16	TYR	HA	19	SER	HN	6,00	#peak	0
17	ILE	HN	17	ILE	HB	2,64	#peak	90
17	ILE	HA	17	ILE	QG1	3,65	#peak	179
17	ILE	HA	17	ILE	QD1	3,88	#peak	180
17	ILE	HA	17	ILE	QG2	5,08	#peak	180
17	ILE	HN	17	ILE	HA	3,36	#peak	86
17	ILE	HN	17	ILE	QG1	3,36	#peak	93
17	ILE	HN	18	GLN	HN	3,84	#peak	9
17	ILE	HA	18	GLN	HN	4,02	#peak	48
17	ILE	HB	18	GLN	HN	3,84	#peak	46
17	ILE	QG2	18	GLN	HN	4,24	#peak	47
17	ILE	HA	19	SER	HN	6,00	#peak	0
17	ILE	HA	20	LEU	HN	6,00	#peak	100
17	ILE	HA	21	LEU	HN	6,60	#peak	69
18	GLN	HA	18	GLN	QB	3,36	#peak	175
18	GLN	HA	18	GLN	QG	3,60	#peak	174
18	GLN	HN	18	GLN	HA	3,36	#peak	42
18	GLN	HN	18	GLN	QB	3,12	#peak	45
18	GLN	HN	18	GLN	QG	4,12	#peak	43
18	GLN	QB	19	SER	HN	4,20	#peak	89
18	GLN	HN	19	SER	HN	3,84	#peak	9
18	GLN	HA	19	SER	HN	4,20	#peak	85
18	GLN	QB	19	SER	HN	4,44	#peak	90
18	GLN	HA	20	LEU	HN	6,00	#peak	0
19	SER	HN	19	SER	HA	3,36	#peak	84
19	SER	HN	19	SER	QB	3,36	#peak	85
19	SER	HN	20	LEU	HN	4,32	#peak	17
19	SER	QB	20	LEU	HN	3,84	#peak	99
19	SER	HN	21	LEU	HN	6,00	#peak	0
19	SER	HA	22	LYS+	QB	6,60	#peak	160

Appendice D

19	SER	HA	22	LYS+	QD	6,60	#peak	161
20	LEU	HN	20	LEU	HA	3,36	#peak	98
20	LEU	HN	20	LEU	QB	3,60	#peak	103
20	LEU	HN	20	LEU	HG	3,96	#peak	103
20	LEU	HN	20	LEU	QQD	4,24	#peak	104
20	LEU	HN	21	LEU	HN	3,84	#peak	12
21	LEU	HN	21	LEU	QB	3,60	#peak	64
21	LEU	HN	21	LEU	HG	3,84	#peak	63
21	LEU	HN	21	LEU	HA	3,36	#peak	68
21	LEU	HN	21	LEU	QQD	5,08	#peak	66
21	LEU	HN	22	LYS+	HN	3,84	#peak	12
21	LEU	HA	23	ARG+	HN	6,00	#peak	0
21	LEU	HA	24	TYR	HN	6,60	#peak	0
22	LYS+	HN	22	LYS+	HA	3,36	#peak	98
22	LYS+	HN	22	LYS+	QB	3,60	#peak	101
22	LYS+	HN	22	LYS+	QD	4,56	#peak	102
22	LYS+	HN	22	LYS+	QG	4,20	#peak	103
22	LYS+	HA	22	LYS+	QB	3,68	#peak	168
22	LYS+	HA	22	LYS+	QG	3,60	#peak	170
22	LYS+	HA	23	ARG+	HN	4,20	#peak	109
22	LYS+	QG	23	ARG+	HN	6,17	#peak	107
22	LYS+	HN	23	ARG+	HN	4,30	#peak	19
22	LYS+	QB	23	ARG+	HN	4,20	#peak	105
22	LYS+	QB	24	TYR	HN	6,00	#peak	81
23	ARG+	HA	23	ARG+	QB	3,36	#peak	164
23	ARG+	HA	23	ARG+	QG	3,60	#peak	166
23	ARG+	HN	23	ARG+	HA	3,00	#peak	109
23	ARG+	HN	23	ARG+	QB	3,60	#peak	106
23	ARG+	HA	24	TYR	HN	4,08	#peak	78
23	ARG+	QB	24	TYR	HN	4,56	#peak	82
24	TYR	HN	24	TYR	QB	3,36	#peak	80
24	TYR	HA	25	GLN	HN	4,52	#peak	88
24	TYR	QB	25	GLN	HN	4,88	#peak	96
25	GLN	HN	25	GLN	HA	3,32	#peak	88
25	GLN	HN	25	GLN	QG	4,20	#peak	97
25	GLN	HN	26	PRO	QD	4,57	#peak	95
26	PRO	HA	27	CHIS+	HN	3,84	#peak	25
27	CHIS+	HA	27	CHIS+	QB	4,32	#peak	145
27	CHIS+	HN	27	CHIS+	QB	4,50	#peak	26

**TabellaD2:** Statistica strutturale di Nodal43-69 in TFE/H<sub>2</sub>O 30:70 (v:v)**CYANA structural statistics**

	<b>NMR restraints</b>
distance restraints	128
intraresidue	61
sequential ( $ i - j  = 1$ )	24
medium-range ( $1 <  i - j  \leq 4$ )	14
torsion angle restraints	4
	<b>Violation statistics (100 structures)</b>
CYANA target function ( $\text{\AA}^2$ )	$6.64 \pm 1.91 \text{\AA}^2$
mean global backbone rmsd	$5.37 \pm 1.06 \text{\AA}$
mean global heavy rmsd	$6.60 \pm 0.95 \text{\AA}$
	<b>Violation statistics (40 structures)</b>
CYANA target function ( $\text{\AA}^2$ )	$5.53 \pm 0.20 \text{\AA}^2$
mean global backbone rmsd	$5.00 \pm 0.95 \text{\AA}$
mean global heavy rmsd	$6.28 \pm 0.87 \text{\AA}$
	<b>Residual Distance Constraint Violations (<math>\text{\AA}</math>)</b>
number $> 0.2 \text{\AA}$	0
	<b>Angle Constraint Violations (<math>^\circ</math>)</b>
number $> 5.0^\circ$	0
<b>AMBER structural statistics</b>	
	<b>Violation statistics (10 structures)</b>
<b>0.2 &lt; d &lt; 0.3</b>	$0.60 \pm 0.80$
<b>0.3 ≤ d &lt; 0.4</b>	$0.20 \pm 0.40$
<b>d ≥ 0.4</b>	$0.10 \pm 0.30$
<b>Maximum violation</b>	$0.28 \pm 0.08$
<b>AMBER energies (kcal mol<sup>-1</sup>)</b>	
Distance constraint	$2.23 \pm 0.79$
Van der Waals	$-118.47 \pm 13.32$
Total	$-206.62 \pm 18.10$
<b>rmsd to the averaged coordinates (<math>\text{\AA}</math>)</b>	
Backbone (43-69)	$6.26 \pm 2.33 \text{\AA}$
All heavy (43-69)	$7.69 \pm 2.31 \text{\AA}$
Backbone (58-64)	$0.81 \pm 0.37 \text{\AA}$
All heavy (58-64)	$2.06 \pm 0.55 \text{\AA}$

## Appendice E

**TabellaE1:** Distanze (Å) ottenute dai NOE di Nodal66-90 in TFE/H<sub>2</sub>O 30:70 (v:v)

1	NTYR	HN	1	NTYR	HA	3.98	#peak	65
1	NTYR	HN	1	NTYR	QB	4.43	#peak	66
1	NTYR	HN	2	GLN	QB	4.06	#peak	67
1	NTYR	HA	2	GLN	HN	3.06	#peak	81
2	GLN	HN	2	GLN	HB2	3.19	#peak	82
3	PRO	HA	4	HIS+	HN	3.45	#peak	13
3	PRO	QB	4	HIS+	HN	5.41	#peak	16
4	HIS+	HN	4	HIS+	QB	4.38	#peak	14
4	HIS+	HN	5	ARG+	HN	4.28	#peak	1
4	HIS+	QB	5	ARG+	HN	4.98	#peak	36
5	ARG+	HN	5	ARG+	HA	3.45	#peak	35
5	ARG+	HN	5	ARG+	QB	3.69	#peak	38
5	ARG+	HN	5	ARG+	QG	5.85	#peak	39
5	ARG+	HN	5	ARG+	QD	6.38	#peak	36
5	ARG+	QB	6	VAL	HN	4.36	#peak	53
6	VAL	HA	6	VAL	QQG	6.04	#peak	101
6	VAL	HN	6	VAL	HB	3.32	#peak	62
6	VAL	HN	6	VAL	QQG	5.91	#peak	55
7	PRO	HA	8	SER	HN	2.93	#peak	29
7	PRO	QB	8	SER	HN	5.18	#peak	32
7	PRO	QG	8	SER	HN	6.50	#peak	33
8	SER	HN	8	SER	QB	4.02	#peak	30
8	SER	HN	9	THR	HN	2.70	#peak	5
8	SER	QB	9	THR	HN	4.32	#peak	73
9	THR	HN	9	THR	QG2	4.42	#peak	79
9	THR	HN	9	THR	HA	2.89	#peak	71
9	THR	HN	9	THR	HB	3.00	#peak	72
9	THR	HN	10	CYA	HN	2.70	#peak	5
9	THR	HA	10	CYA	HN	2.96	#peak	25
9	THR	QG2	10	CYA	HN	6.26	#peak	26
10	CYA	HN	10	CYA	QB	3.53	#peak	27
11	CYA	QB	12	ALA	HN	5.25	#peak	60
11	CYA	HN	12	ALA	HN	3.39	#peak	2
12	ALA	HN	12	ALA	HA	4.74	#peak	56
12	ALA	HN	12	ALA	QB	4.12	#peak	64
13	PRO	HA	14	VAL	HN	2.63	#peak	68
13	PRO	QB	14	VAL	HN	5.11	#peak	75
13	PRO	QG	14	VAL	HN	5.25	#peak	76
14	VAL	HA	14	VAL	QQG	5.44	#peak	106
14	VAL	HN	14	VAL	HA	3.39	#peak	69
14	VAL	HN	14	VAL	HB	3.06	#peak	70
14	VAL	HN	14	VAL	QQG	5.77	#peak	80
14	VAL	HA	15	LYS+	HN	2.60	#peak	19

Appendice E

14	VAL	HB	15	LYS+	HN	3.39	#peak	20
14	VAL	HN	15	LYS+	HN	3.16	#peak	3
14	VAL	QQG	15	LYS+	HN	6.37	#peak	24
15	LYS+	HN	15	LYS+	HA	3.12	#peak	18
15	LYS+	HN	15	LYS+	QB	4.02	#peak	21
15	LYS+	HN	15	LYS+	QG	6.38	#peak	23
15	LYS+	QB	16	THR	HN	4.98	#peak	86
15	LYS+	HN	16	THR	HN	3.26	#peak	4
15	LYS+	HA	16	THR	HN	3.00	#peak	83
16	THR	HN	16	THR	HB	3.69	#peak	84
16	THR	HN	16	THR	HA	2.79	#peak	83
16	THR	HN	16	THR	QG2	4.88	#peak	87
16	THR	QG2	16	THR	HB	3.49	#peak	105
16	THR	HN	17	LYS+	HN	3.09	#peak	7
16	THR	HA	17	LYS+	HN	3.03	#peak	41
16	THR	HB	17	LYS+	HN	3.88	#peak	42
16	THR	QG2	17	LYS+	HN	6.23	#peak	49
17	LYS+	HN	17	LYS+	QB	4.22	#peak	46
17	LYS+	HN	17	LYS+	QG	6.01	#peak	48
18	PRO	HA	19	LEU	HN	2.86	#peak	40
19	LEU	HN	19	LEU	QB	2.87	#peak	47
19	LEU	HN	19	LEU	HA	3.88	#peak	42
19	LEU	HN	19	LEU	QQD	5.81	#peak	50
19	LEU	HA	20	SER	HN	2.70	#peak	72
19	LEU	HN	20	SER	HN	3.29	#peak	6
19	LEU	QB	20	SER	HN	3.96	#peak	78
19	LEU	QQD	20	SER	HN	5.09	#peak	80
20	SER	HN	20	SER	HA	2.89	#peak	71
20	SER	HA	21	MET	HN	3.03	#peak	58
20	SER	HN	21	MET	HN	3.92	#peak	9
20	SER	HN	20	SER	QB	3.92	#peak	73
20	SER	HA	22	LEU	HN	4.64	#peak	89
21	MET	HN	21	MET	HA	3.52	#peak	57
21	MET	HN	21	MET	QB	4.10	#peak	63
21	MET	HN	21	MET	QG	6.50	#peak	61
21	MET	HA	22	LEU	HN	3.22	#peak	88
21	MET	HN	22	LEU	HN	2.83	#peak	10
22	LEU	HA	22	LEU	QB	3.63	#peak	102
22	LEU	HA	22	LEU	QQD	6.19	#peak	104
22	LEU	HN	22	LEU	QB	3.59	#peak	91
22	LEU	HN	22	LEU	HA	2.63	#peak	92
22	LEU	HA	23	TYR	HN	3.63	#peak	92
23	TYR	HA	24	VAL	HN	4.35	#peak	95
23	TYR	QB	24	VAL	HN	5.35	#peak	97
23	TYR	HN	23	TYR	QB	3.59	#peak	93
23	TYR	HN	23	TYR	QD	6.88	#peak	12

Appendice E

23	TYR	HN	24	VAL	HN	3.55	#peak	11
24	VAL	HA	25	ASP	HN	2.86	#peak	43
24	VAL	HB	25	ASP	HN	2.83	#peak	45
24	VAL	HN	24	VAL	HA	3.29	#peak	96
24	VAL	HN	24	VAL	HB	3.12	#peak	98
24	VAL	HN	24	VAL	QQG	6.37	#peak	99
24	VAL	HN	25	CASP	HN	3.22	#peak	8
25	ASP	HN	25	ASP	QB	3.92	#peak	44
24	VAL	QQG	25	ASP	HN	6.63	#peak	50

---