Research Article

Journal of PeptideScience

Received: 27 June 2008 Accepted: 6 October 2008 Published online in Wiley Interscience: 26 November 2008

(www.interscience.com) DOI 10.1002/psc.1091

Structural insights into the interaction between the Cripto CFC domain and the ALK4 receptor‡

Luisa Calvanese,[a] Angela Saporito,[b] Romina Oliva,[c] Gabriella D' Auria,[a,b] Carlo Pedone,[b] Livio Paolillo,[a,b] Menotti Ruvo,[b] Daniela Marasco[b] and Lucia Falcigno[a,b]***

The protein Cripto is the founding member of the extra-cellular EGF–CFC growth factors, which are composed of two adjacent cysteine-rich domains: the EGF-like and the CFC. Members of the EGF–CFC family play key roles in embryonic development and are also implicated in tumourigenesis. Cripto is highly over-expressed in many tumours, while it is poorly detectable in normal tissues.

Although both Cripto domains are involved in its tumourigenic activity, the CFC domain appears to play a crucial role. Indeed, through this domain, Cripto interferes with the onco-suppressive activity of Activins, either by blocking the Activin receptor ALK4 or by antagonising proteins of the TGF-β family. We have undertaken the chemical synthesis and the structural characterisation of human CFC Cripto domain. Using a combined NMR and computational approach, supported by binding studies by SPR, we have investigated the molecular basis of the interaction between h-CFC and ALK4. Binding studies indicate that the synthetic h-CFC interacts with the ALK4 receptor with a K_D in μM range, whereas it does not recognise the ActRIIB receptor. The NMR study shows that the h-CFC overall topology is determined by the presence of three disulfide bridges and that residues H120 and W124 are located between the first strand and the first loop with the side chains externally exposed. A model of the CFC–ALK4 complex has also been obtained by molecular docking and shows that all residues indicated by prior mutagenesis studies can contribute to the ALK4–CFC interaction at the protein–protein interface.

Supporting information may be found in the online version of this article

Keywords: conformational analysis; EGF–CFC family; NMR; SPR; homology modelling; docking

Introduction

The EGF–CFC family includes a group of structurally related extracellular proteins, that play essential roles in the initial stage of the embryogenesis [1,2] and in the development and progression of tumour cells [3]. The protein Cripto (188 a.a.), the founding member of this growth factors family, is found in several vertebrates such as humans, chicken, Zebra fish and frogs [4,5]. Cripto is an extracellular protein containing: (i) a *N*-terminal signal peptide; (ii) two highly conserved cysteine-rich domains of about 40 residues, known as EGF-like (a variant of the epidermal growth factor) and CFC (so called by the founder members: *C*R in the human, *F*RL-1 in Xenopus and *C*ryptic in mouse); (iii) a *C*-terminal hydrophobic region linked to the cell membrane by a glycosyl-phosphatidylinositol (GPI) moiety [5].

An important signalling pathway identified for Cripto, and supporting cell survival, involves the Activin receptor complex (comprising the Activin type II serine/threonine kinase receptor, ActRII or ActRIIB, and the Activin type I serine/threonine kinase receptor, ALK4) and a number of proteins of the TGF-β family, including Nodal [2,6–9], Vg1/Gdf1 [10], Gdf3 [11], Activin A [12], Activin B [3], and Lefty [13]. In particular, the presence of Cripto is essential for the binding of Nodal/Gdf1/Gdf3 to the Activin receptor complex (ALK4/ActRIIB).

Several studies have established that the EGF-like domain binds to Nodal while the CFC domain binds to ALK4 [3,9,14]. Possible mechanisms by which Cripto can induce carcinogenesis involve the inhibition of the onco-suppressive activity of Activins and of the TGF-β, either by direct binding [15] or by means of additional mediators [16]. Both mechanisms, however, involve the CFC domain, suggesting it for a primary role in the protein tumourigenic activity [3,5,16–18].

* *Correspondence to: Lucia Falcigno, Dipartimento di Chimica, Università degli Studi di Napoli "Federico II", via Cintia, 45, 80126, Napoli, Italy. E-mail: lucia.falcigno@unina.it*

** *Daniela Marasco, Istituto di Biostrutture e Bioimmagini del CNR, via Mezzocannone, 16, 80134, Napoli, Italy. E-mail: daniela.marasco@unina.it*

a *Dipartimento di Chimica, Università Federico II, Complesso Universitario MSA, via Cintia 45, 80126, Napoli, Italy*

b *Istituto di Biostrutture e Bioimmagini del CNR, via Mezzocannone, 16, 80134, Napoli, Italy*

c *Dipartimento di Scienze Applicate, Università degli Studi di Napoli "Parthenope" Centro Direzionale Isola C4, 80143 Napoli, Italy*

‡ *11th Naples Workshop on Bioactive Peptides.*

Recently, the crystal structures of several proteins of the TGF-β family in complex with their respective receptors have been solved, allowing a progress in the understanding of the action mechanisms within the TGF-β signalling cascade (for a complete review see Ref. 19). The structure of ALK4 alone or in complex with ligands has so far not been reported; instead, the solution structure of the mouse Cripto CFC domain has been described [20].

Here, we report the synthesis and the solution structure of the human CFC Cripto domain (sequence 112–150, Figure 1) obtained by NMR analysis in phosphate buffer at pH 7. The interaction between the synthetic h-CFC and the extra-cellular domain (ECD) of ALK4 receptor has also been investigated by surface plasmon resonance (SPR) technique, providing a quantitative evaluation of the complex formation. To further broaden our understanding of the CFC–ALK4 interaction, we have finally undertaken an *in silico* study whereby the NMR experimental structures of h-CFC have been docked to the homology model of the ALK4 ECD, based on the crystal structure of the homologue human type IA bone morphogenic protein receptor ectodomain BRIA–ECD (ALK3–ECD) [21].

Materials and Methods

Chemical Synthesis of Human CFC Domain

The human CFC domain 112–150 (h-CFC), acetylated and amidated (Figure 1), was obtained as reported for the mouse variant [14] by stepwise solid phase synthesis following standard Fmoc/tBu methodologies [22] on a 100 µmole scale (Rink amide resin, 1.1 mmol/g). A ten-fold excess of amino acids, pre-activated with 1-H-benzotriazolium, 1-[bis(dimethylamino)methylene]-hexafluorophosphate(1-),3-oxide (HBTU)/N-hydroxybenzotriazole (HOBt)/ diisopropylethylamine (DIEA) (1 : 1 : 2), was used throughout the synthesis. Coupling and deprotection times were kept at 25 and 15 min, respectively. Standard side-chain protection groups for Fmoc chemistry were used for all residues; the six cysteines, introduced as trityl (Trt) derivatives, produced after cleavage as free thiol groups.

After removal from the solid support by treatment with a TFA/H_2O/ triisopropylsilane (TIS) (94 : 3 : 3 v/v/v) mixture (RT, 4h, 1.0 ml mixture/100 mg resin) and lyophilisation, the polypeptide was reduced in 100 mM 1,4-dithiothreitol (DTT) for 30 min at room temperature, purified by reversed-phase HPLC and refolded by spontaneous oxidation in 100 mM carbonate buffer, pH 8.5, 0.5 mM EDTA as described elsewhere [14]. Refolding was followed by LC-MS analysis to assess the retention time shifts and molecular mass decrease due to disulfide bridge formation. Refolding and RP-HPLC purification conditions of the resulting products were described elsewhere [14].

Circular Dichroism Spectroscopy

CD spectra were obtained at room temperature on a Jasco J-715 dichrograph, calibrated at 290 nm with an aqueous solution of D(+)-10-camphor sulphonic acid, using 0.1 mm quartz cuvettes.

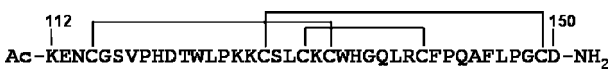

Figure 1. Amino acid sequence of human Cripto 112–150 CFC domain. The three disulfide bridges are indicated. In the text, the first residue in the sequence (K112) has been reported as K1.

Spectra of the refolded polypeptide were recorded on solutions with a Cpep $= 1.0 \times 10^{-4}$ M. Three separate 1.0×10^{-4} M solutions were prepared: (i) 10 mM phosphate buffer, pH 7.0; (ii) 10 mM phosphate buffer, pH 7.0 at 10% TFE (v/v) and (iii) 10 mM phosphate buffer, pH 7.0 at 20% TFE (v/v).

Data were collected at 0.2 nm intervals with a 20 nm min^{-1} scan speed, a 2-nm bandwidth and a 16-s response, from 260 to 190 nm. Three spectra for each sample were recorded, averaged and transformed in molar ellipticity/residue.

NMR Analysis

NMR characterisation of h-CFC was performed in phosphate buffer/D_2O (90 : 10 v/v) at 298 K. The sample was prepared by dissolving *ca* 4 mg of peptide in phosphate buffer adding D_2O for a final ratio of 90/10 (v/v). The D_2O was purchased from Sigma Aldrich (99.96% D). Final concentration was *ca* 1.2 mM and the pH value was 7.

One- and two-dimensional (2D) NMR spectra were acquired on a Varian Inova spectrometer, operating at a proton frequency of 600 MHz. The 2D experiments, such as total correlation spectroscopy (TOCSY) [23], and nuclear Overhauser effect spectroscopy (NOESY) [24], were recorded by the phase sensitive States-Haberkorn method [25]. The data file generally consisted of 512 and 4096 data points in the $\omega 1$ and $\omega 2$ dimensions, respectively. TOCSY experiment was acquired with a 70-ms mixing time while NOESY experiments were acquired with 200-ms and 300-ms mixing times. The water resonance was suppressed by the use of gradients [26].

Chemical shifts were referred to internal sodium 3-(trimethylsilyl) propionate 2,2,3,3-d_4 (TSP). Free induction decays (FIDs) were multiplied, in both dimensions, with shifted sine-bell or gaussian weighting functions and data points were zero filled to 1 K in $\omega 1$ prior to Fourier transformation. Spectra were analysed by using NMRPipe/NMRView programs [27,28]. Proton sequential assignment of amino acid spin systems (reported as Supporting Information, Table S1) was obtained according to Wüthrich [29]. NOE intensities were evaluated by the integration of cross-peaks in the 300-ms NOESY spectrum, processed by NMR Pipe, using the NMRView software [27]. They were then converted into inter-proton distances by use of the CALIBA program [30]. G37 geminal α–α' protons were chosen as reference with a distance of 2.2 Å.

SPR Binding

In order to assess the recognition by h-CFC domain towards ALK4 receptor real time binding assays to immobilised ALK4 and ActRIIB were performed.

Recombinant Human Activin RIB/ALK-4- and ActRIIB-Fc Chimeras were from R&D systems (Minneapolis, US). Reagents for protein immobilisation were from GE Healthcare. Sensograms were recorded on a BIACORE 3000 instrument (Biacore AB, Uppsala, Sweden).

The two recombinant receptors (fused to Fc) were immobilised on a CM5 dextran-coated sensor chip, on two distinct channels, by a standard amine coupling procedure, according to the manufacturer's instructions. Briefly, the carboxymethylated dextran surface was activated by a 7-min injection of a solution containing 200 mM 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and 50 mM *N*-hydroxysuccinimide (NHS), followed by the injection of ALK4 and ActRIIB (50 µg/ml solution in 10 mM sodium acetate, pH 4.0). A continuous flow of HBS (Hepes 10 mM, NaCl 150 mM, EDTA 3.4 mM)

Journal of PeptideScience

at 20 µl/min was maintained, for 3 min and capping of unreacted sites was achieved by injecting 1 M ethanolamine, pH 8.5, for 7 min. The control surface was obtained by the derivatisation of an empty channel with an Fc sample (50 µg/ml), obtained in our laboratory in previous studies.

Binding assays were carried out by injecting solutions of h-CFC domain at increasing concentrations, in the range of 0–40 µM and using as reference the signal produced by non-specific binding onto the Fc-derivatised channel. The polypeptide was allowed to bind for 3 min and to dissociate for 2 min. Then a regeneration step with NaOH (50 mM, 10 µl) and Glycine pH = 2 (50 mM, 10 µl) was performed. Kinetic data were collected in triplicate for each h-CFC concentration. Data were properly manipulated to subtract the reference signal and fitted by means of BIAevaluation 4.1 software (Biacore AB, Uppsala, Sweden).

Computational Methods

Structure calculations started from 100 randomised conformers and used the standard CYANA simulated annealing schedule [31] with 20 000 torsion angle dynamics steps for conformer. Three-dimensional (3D) structures were obtained by using inter-proton distances evaluated from NOEs as upper limits, without use of stereospecific assignments. The 40 conformers with the lowest final CYANA [32] target function values (TF average value = 1.05 ± 0.10 Å^2) were subjected to restrained energy minimisation by use of the SANDER module of the AMBER 6.0 package [33]. The best ten structures in terms of fitting with experimentally derived restraints were selected from those with the lowest- residual restraint energy to represent the peptide solution structure. The molecular graphics program MOLMOL [34] was employed to perform the structural statistics analysis. Data are reported as Supporting Information (Table S2).

Homology Modelling

Selection of the template

The structure of human ALK3 receptor (pdb code: 1rew, chain C, resolution 1.86 Å) was identified as the best template for modelling ALK4 by a PSI-BLAST [35] search against the 'nr' database at the third iteration, with an *E*-value of 10^{-13} (sequence identity 27%, positives 33%). The target matched the Pfam [36] family of Activin types I and II receptor domains (PF01064) with an *E*-value of 1.7×10^{-28}, 1rew being the reference structure for the family. 1rew was also the first structure matched when searching by Hidden Markov Model (HMM)–HMM comparison [37] with the homology detection tool HHpred [38] against the PDB70_15Apr07 database, with an *E*-value of 3×10^{-27}, and when searching with a fold recognition method such as mGenThreader [39], that uses sequence profiles and predicted secondary structure, with a probability of false positive (*P*-value) of 10^{-27}.

Sequence alignment and model building

A pair-wise target-template sequence alignment was obtained by PSI-BLAST [35] and a multiple sequence alignment was obtained for the 152 best PSI-BLAST hits including the target and the template sequences by CLUSTALW [40]. Target-template alignments were also obtained by the HHpred [38] and mGenThreader [39] tools. The PSI-BLAST, HHpred and mGenThreader target-template alignments were used to build molecular models for the ALK4 ECD receptor using Modeller6v2 [41] with default parameters. The quality of all the obtained models was checked by WhatCheck [42,43], Verify3D [44,45] and ProQ [46,47].

Docking

Docking studies were performed with the 3D-Dock package from the Biomolecular Modelling Laboratory of Cancer Research, UK (http://www.bmm.icnet.uk/docking). The complexes were generated with FTDock [48], a grid-based method which performs global scans of translational and rotational space and then scores and ranks the docked solutions by means of a surface complementarity score (SCscore). The parameters for the FTDock runs included a 0.7 Å grid unit, 1.3 Å surface thickness and 12° rotation angle step size. The ALK4–ECD model was set as the static moiety while each of the ten lowest energy NMR structures of h-CFC was set as the mobile moiety. A total number of 9240 rotations were evaluated, leading to 10 000 complexes for each FTDock run. These docked solutions were then rescored and ranked using RPScore [49]. This program scores the docking solutions using an empirical residue-pair potential matrix. The matrix i90_p05_d4.5_2dp.matrix was used in this work. These solutions were then filtered by applying a distance constraint (4.5 Å) to residues known to be part of the protein–protein interface from experimental evidences. Only docking solutions that satisfied the vicinity criterion of H9 and W12 of CFC with the ALK4–ECD hydrophobic face were accepted. The remaining solutions were refined using the module MULTIDOCK that does a simple minimisation using a side-chain rotamers library [50]. Out of these, the best solution was chosen looking at the RPScore, energy, and geometry features at the interface.

Results and Discussion

Preparation of the Human CFC Domain and Circular Dichroism Spectroscopy

The CFC domain of h-Cripto was designed on the basis of the reported structure of the mouse variant [14,20], thereby it encompassed residues 112–150, with several residues on both the *N*- and *C*-termini outside the region previously assigned to this domain [51]. After purification and extensive reduction, the synthetic polypeptide was submitted to the refolding process as described in the Section on Methods. As shown in Figure 2, the polypeptide rapidly and quantitatively refolded in about 1.5 h and did not undergo any other further transformation over the next 12 h. After refolding, the domain was purified by reversed-phase HPLC obtaining a pure product (Figure 3(A) and (B)) with an experimental molecular weight (MW) in very good agreement with the calculated value ($MW_{Theor/Exper}$ = 4468.9/4469.25 amu). Also the CD spectrum, recorded in phosphate buffer at neutral pH, was similar to that reported for the mouse variant [14] (Figure 4), where no canonical conformations were observed. In order to assess the structural stability of the polypeptide, CD spectra in the presence of 10 and 20% of TFE were also recorded (Figure 4). Noticeably, the peptide structure did not significantly change upon TFE addition, suggesting the occurrence of conformations restricted by the presence of the disulfide bridges.

NMR Analysis and Molecular Modelling of h-CFC in Phosphate Buffer at pH 7

The NMR conformational analysis of h-CFC was performed in phosphate buffer at pH 7 to match physiological conditions. For clarity, in the following, the first residue in the sequence (K112) will be numbered as K1.

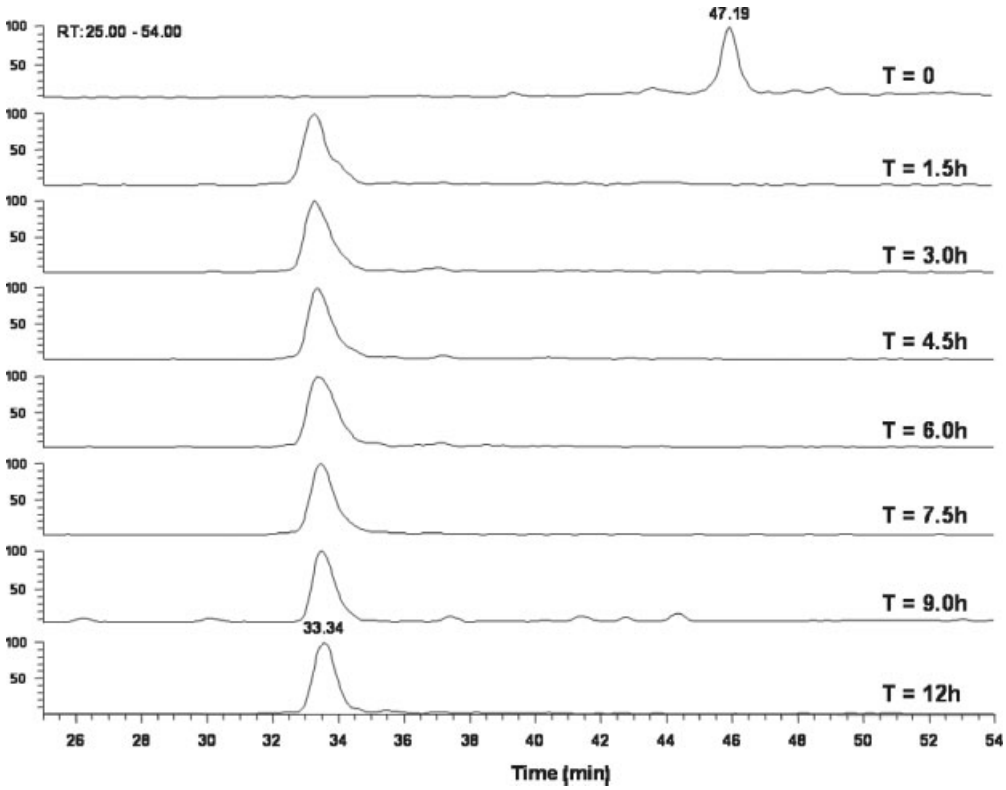

Figure 2. Refolding progression of human Cripto CFC domain. The analyses were carried out every 1.5 h, as described in Ref. 14. The polypeptide appeared refolded only after 1.5 h.

Proton assignment was carried out with the procedure suggested by Wuthrich [29] and the chemical shifts are reported in Table S1 of the Supporting Information.

A comparison of the h-CFC αCH proton chemical shifts to random coil values [52] for each residue showed that most of them presented positive deviations from random coil values, more marked in the central part of the molecule, suggesting the presence of β-sheet structures (Figure S1 of Supporting Information). This structural diagnosis was consistent with the NOE pattern. NOE effects such as (α_i, NH_{i+1}) appeared stronger than the corresponding (NH_i, NH_{i+1}) nicely confirming the occurrence of extended conformations. Some unambiguous inter-strand NOE contacts were also observed: αCH C20-αCH C29, βCH2 S6-βCH2 C20, αCH T11-βCH2 S18, βCH T11-βCH2 S18, αCH C20-NH F30, αCH C29-NH K21.

A set of 146 experimental NOE constraints for h-CFC (80 intra-residual, 40 sequential, and 26 long-range) was used for structure calculations by CYANA/AMBER programs. The best ten structures (see Section on Materials and Methods for details) were selected to represent the h-CFC solution structure.

They show a certain degree of dispersion, due to a poor number of observed NOEs, with a backbone mean global root mean square deviation (rmsd) of 4.74 $\pm$ 1.09 Å. The most ordered peptide region corresponded to the segment C20–C29, where the backbone mean global rmsd for the superposition of the final ten structures was 1.80 $\pm$ 0.56 Å. On average, the molecule was globally extended with three anti-parallel strands connected by disulfide bridges. The average molecular model of h-CFC showed an ellipsoidal compact shape with approximate dimensions of 24 Å $\times$ 15 Å $\times$ 12 Å (Figure 5).

Significantly, in all the NMR molecular models, the side chains of residues H9 and W12, reportedly involved in the binding with the receptor [51], were 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 [20,53], was found opposite to the binding site.

Binding Studies to the Immobilised Receptors by SPR Analysis

In order to evaluate the capacity of h-CFC domain to recognise the ALK4 or the ActRIIB receptors, real time binding assays were performed. Assays were carried out by injecting solutions of h-CFC domain at increasing concentrations, within the range 0–40 μM, and subtracting the reference contribute of Fc-derivatised channel. In Figure 6(A), the overlay of sensorgrams for the binding of h-CFC domain to immobilised ALK4 is reported. Sensorgrams revealed typical receptor–protein affinity profiles and kinetic analysis of the interaction afforded a k_{on} of 466 (1/M s) and k_{off} of 5.3 $\times$ 10^{-3} (1/s) values, providing a K_D of 13 $\pm$ 4 μM. This value was consistent with that obtained by plotting the RU_{max} from each single experiment against the h-CFC concentration (Figure 6(B)). Remarkably, h-CFC did not bind to ActRIIB (data not shown).

Homology Modelling of ALK4–ECD

As the 3D structure of the ALK4–ECD is unknown, we have modelled it by homology, on the basis of the crystal structure of the homologue ALK3–ECD in complex with the bone morphogenic protein 2 (BMP-2) ligand [21]. Extra-cellular domains of human ALK3 and ALK4 match the Pfam family of Activin receptors (PF01064) and are 27% sequence identical. These ECDs are

Journal of PeptideScience

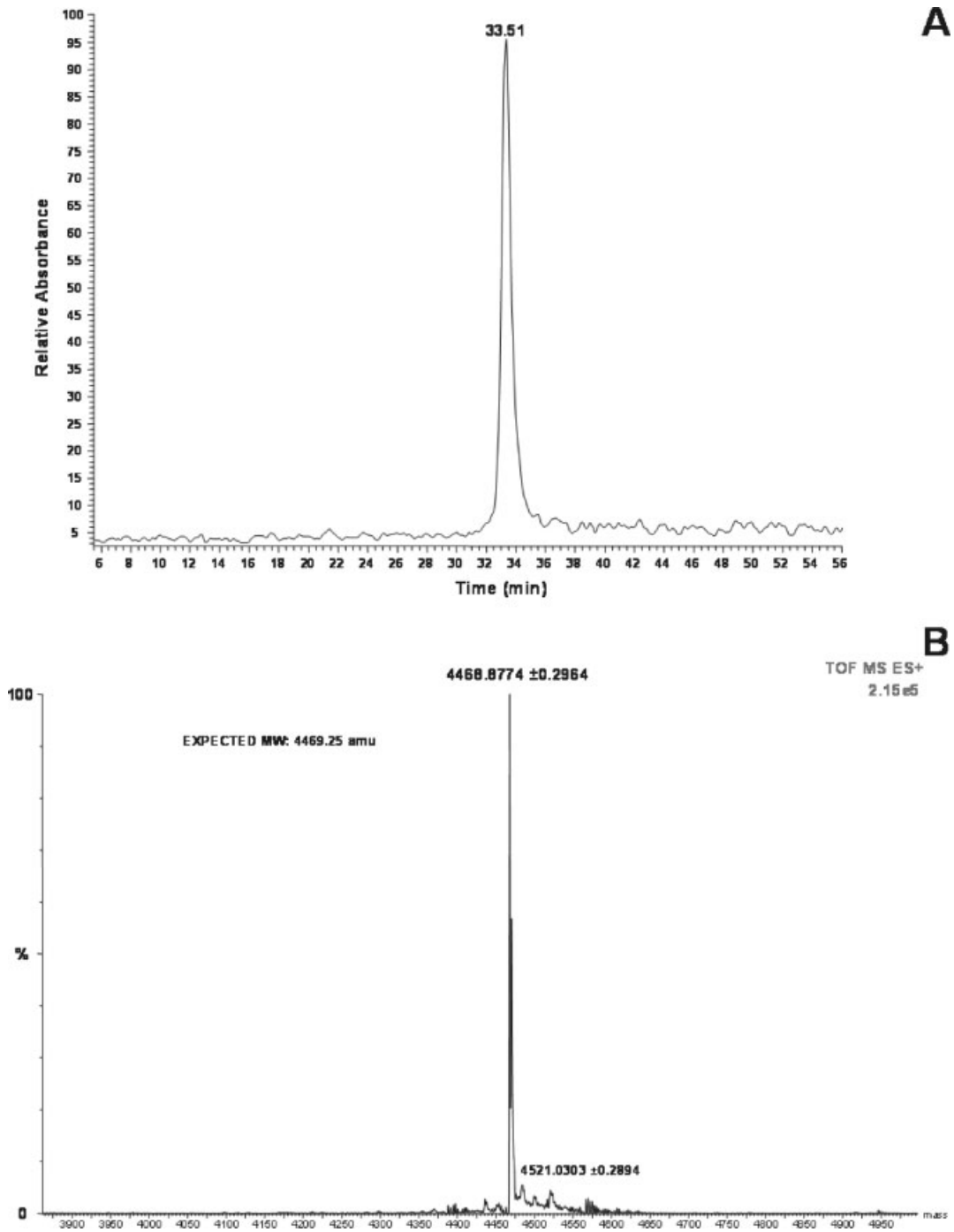

Figure 3. The refolded and purified human CFC was characterised by HPLC (A) and mass spectrometry analyses (B) observing a single, very clean peak at about 33 min. The experimental MW was in agreement with the experimental value.

hydrophilic cysteine-rich ligand-binding domains characterised by the consensus cysteine box: $CCX_{4-5}CN$ [54].

PSI-BLAST [35], HHpred [38] and mGenThreader [39] searches all identified the structure of human ALK3 receptor (pdb code: 1rew, resolution of 1.86 Å) as the best template for the homology modelling of human ALK4 receptor ectodomain (*E*-values below 10^{-13}). Moreover, the three tools gave target-template alignments with all ten cysteine residues involved in disulfide bridges correctly aligned and coincident in 72 out of the 91 aligned positions (Figure 7(A)), allowing to trustfully identify an extended protein core. The three alignments from PSI-BLAST, HHpred and mGenThreader were all used to build molecular models for the ALK4–ECD receptor. The best model according to the predicted quality scores, resulted to be that built on the basis of the HHpred alignment. This model has been selected for further analyses and is shown in Figure 7(B).

The ALK4–ECD model, exhibiting the same global three-finger fold of ALK3–ECD domain, shows two β-sheets composed by two (residues 5–9 and 17–21) and three (residues 24–32, 35–43 and 63–69) strands, respectively; a short α-helix segment between

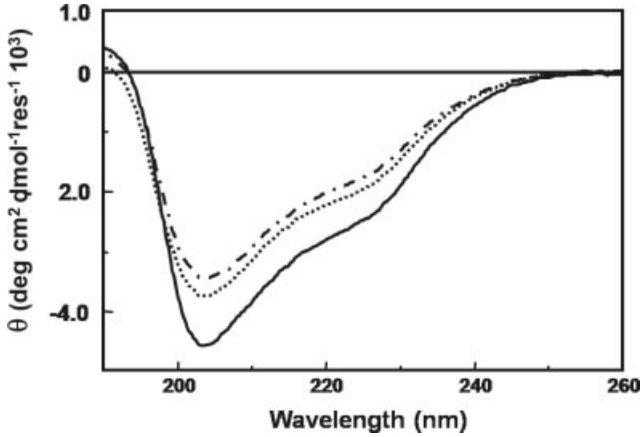

Figure 4. CD spectra of the refolded h-CFC within the range of 190–260 nm. The solid line refers to the sample in 10 mM phosphate buffer, pH 7.0; the dotted line refers to the sample in 10 mM phosphate buffer, pH 7.0 at 10% of TFE (v/v) and the dashed line refers to the sample in 10 mM phosphate buffer, pH 7.0 at 20% of TFE (v/v).

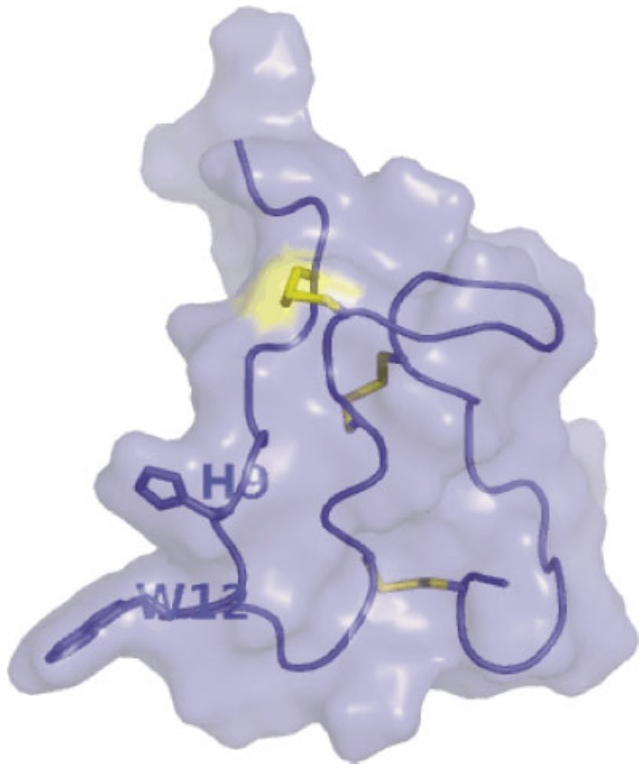

Figure 5. Connolly surface and ribbon representation of CYANA/AMBER average molecular model of h-CFC at pH 7. The side chains of binding site residues (H9 and W12) and the disulfide bridges are shown in stick. Figure made with Pymol (DeLano (2002) http://www.pymol.org).

residues 54–58, and an additional α-helix segment between residues 75–78. A hydrophobic face, involving residues 48–58, followed by a polar patch (S28, S59, S60, N65, T66, S83) is also observed. The opposite side of the molecule is highly hydrophilic, with a patch succession of acidic (E20, D22, E47), polar (T10, S11, Q14, N16, T18) and basic (H37, H38, R40, R76) residues.

Mutagenesis studies, carried out by Harrison *et al*. [55] on ALK4, indicated five hydrophobic amino acid residues (here L13, I43, V46, L48, P50) as those relevant for the ALK4 binding to Activin and, together with contributions from other residues, liable to provide an interface for ligand binding. It has also been proposed that TGF-β ligands show a high-binding affinity for this hydrophobic concave surface on type I receptors, i.e. ALKs [19]. Residues V46, L48 and P50, however, are in the region of ALK4 that has low sequence similarity with ALK3. The ALK4 largely hydrophobic region, above mentioned, precedes a short α-helix on residues 54–58. Residues V46 and I43, but not L13, are in close proximity of this hydrophobic patch that can play a role in the ligand binding. A more detailed look at this patch shows that F55 and Y56, which when partially mutated to affect Activin signalling [55], are well oriented to possibly bind CFC binding site (H9 and W12). Although it has not yet been demonstrated whether Activin and CFC share a common binding site on ALK4 receptor, the possible interactions of residues F55 and Y56, along with the whole hydrophobic patch, towards CFC were considered in analysing docking complexes. This choice is indeed supported by considering the crucial relevance of aromatic rings for receptor binding [56].

Docking of h-CFC NMR Structures onto ALK4–ECD Homology Model

Once built a model for the ALK4 ectodomain, we docked the ten selected NMR structures of h-CFC into this homology model by the rigid body docking program 3D-Dock [48,57]. Docking solutions were filtered by taking advantage of mutagenesis studies indicating that the H9 and W12 residues of h-CFC have a crucial role in the binding to ALK4 [9,14,51]. As a consequence of this filtering procedure, the 10 000 initial docking solutions were reduced to about ten for each NMR structure. By visual inspection of the filtered structures, only those complexes exhibiting H9 and W12 of h-CFC well oriented towards the binding concave hydrophobic interface of ALK4 were retained. These docking solutions, ranked by their RPScore values, were energy minimised by MULTIDOCK refinement.

The model of the complex with the lowest total energy is reported in Figure 8, and it shows favourable aromatic interactions between h-CFC W12 and ALK4 F56. Furthermore, h-CFC H9 lies between ALK4 Y55 and F56 aromatic rings potentially able to make aromatic and/or π-cationic interactions [58,59]. Possible electrostatic interactions are observed between ALK4 E47–h-CFC K1, ALK4 D62–h-CFC K16 and ALK4 E61–h-CFC K15. Another favourable interaction at the binding interface is found between the hydrophobic residues ALK4 V49–h-CFC P8.

Conclusions

Cripto, a growth factor involved in morphogenesis and in tumour progression, interacts with ALK4 and is necessary for both Nodal recruitment to the ALK4/ActRIIB receptor complex and for Nodal signalling. Structural mapping of functional sites of ALK4 could provide insights into the mechanisms underlying the Cripto-dependent tumour formation and progression and could pave the way to the design of antagonists.

To elucidate the interaction of the Cripto CFC domain with the ALK4 receptor, we undertook the synthesis and conformational analysis of human CFC domain by NMR, investigated the kinetic and thermodynamic parameters of this interaction by SPR, and modelled the ALK4–ECD by homology and then built a model for the h-CFC/ALK-ECD complex.

Journal of PeptideScience

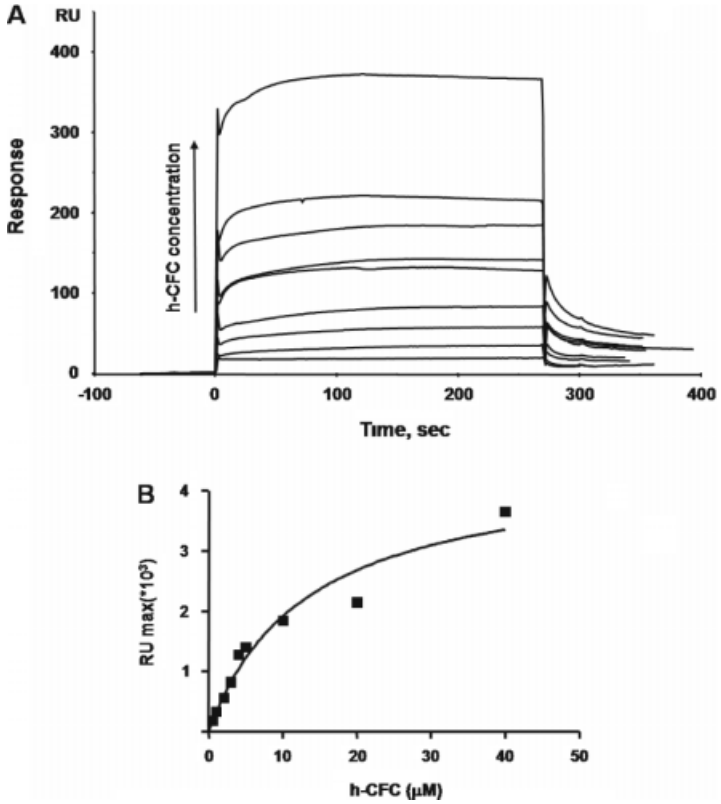

Figure 6. (A) Overlay plot of SPR sensorgrams showing the interaction between h-CFC and ALK4-Fc chimera immobilized on a CM5 dextran-coated sensor chip. Peptide was injected at concentrations ranging from 0–40 μM and a flow rate of 30 μl/min. (B) Response maximum values of each binding curve are reported against h-CFC concentrations.

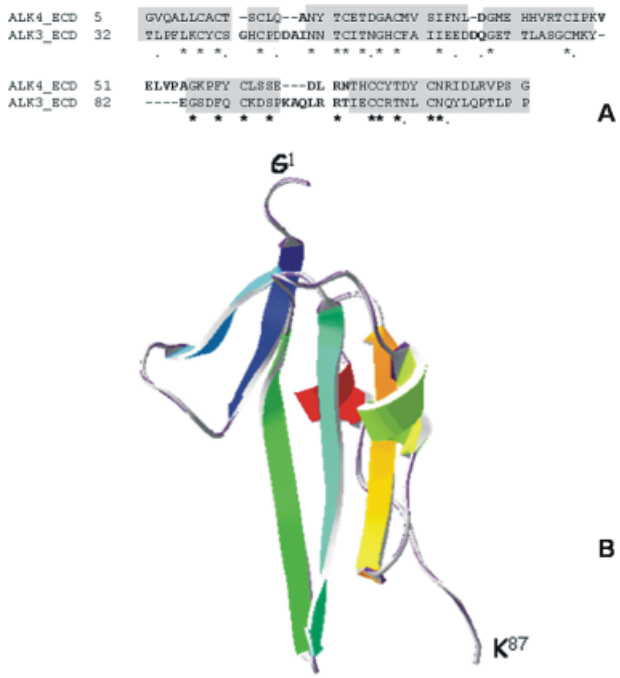

Figure 7. (A) Target-template alignment from HHpred, based on HMM–HMM comparison. Grey boxes outline regions that are conserved as compared to the target-template alignments obtained by PSI-BLAST and mGenThreader. (B) Ribbon representation of ALK4–ECD model, built by Modeller6v2 starting from the ALK3 structure (pdb code: 1rew) coordinates on the basis of the HHpred alignment. For the residues shown, see text. The colour code is based on the sequence of secondary structure elements.

Molecular models of the h-CFC domain were obtained by NMR data. The h-CFC overall topology is determined by the presence of three disulfide bridges (C4–C22, C20–C29, C17–C38), and shows an ellipsoidal shape with some side chains pointing externally. In particular, residues H9 and W12, indicated as the binding site, are not buried in the protein core, but are located between the first strand and the first loop with the side chains externally exposed and readily accessible to ALK4. As expected, the h-CFC molecular models are very similar to the mouse CFC NMR models that we previously reported [20]. The ability of h-CFC domain to recognise ALK4 receptor was confirmed by SPR experiments providing a K_D value in the micromolar range.

We built a model for ALK4–ECD by homology using the crystal structure of the homologue ALK3–ECD in complex with the BMP-2 ligand as a template. The ALK4–ECD model shows a large hydrophobic region where residues I43, V46, L48, P50, indicated as crucial for the binding to Activins, are found. The two aromatic residues, F55 and Y56 that protrude from this patch appear properly oriented for inter-molecular interactions. Although these residues are not considered crucial for the ALK4/Activins binding [55], they instead could be crucial in determining the ALK4/CFC binding specificity. Finally we docked the h-CFC selected NMR structures onto the ALK4–ECD homology model. The models number severely decreased upon applying the constraint that H9 and W12 of h-CFC are essential for ALK4 binding. These models were further assessed against experimental biological data and against an array of *in silico* analysis techniques; only one complex structure was, then, selected for further analyses. This model presents all residues indicated by mutagenesis studies to contribute to the ALK4–CFC interaction at the protein–protein interface (Figure 8).

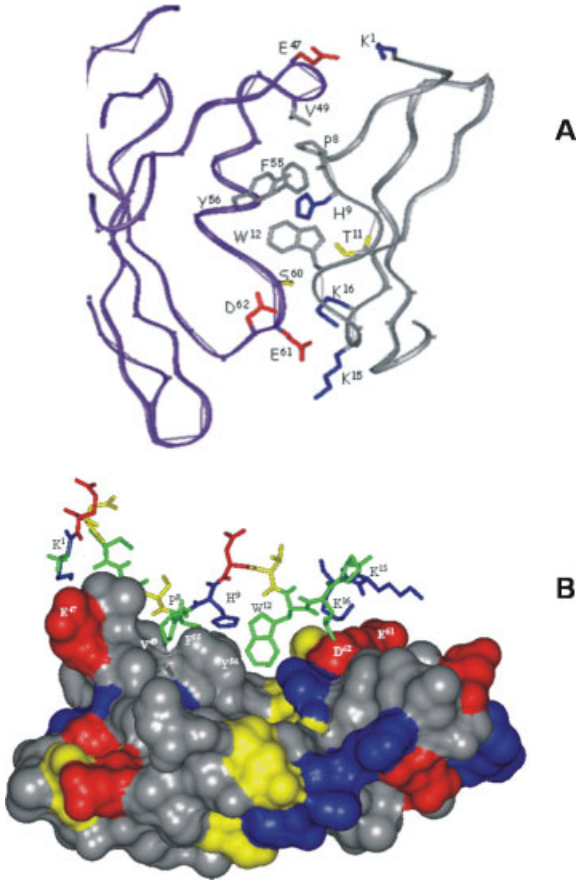

Figure 8. (A) Ribbon representation of the proposed model for the interaction between ALK4–ECD (purple) and h-CFC (grey). The relevant residues at the binding interface are highlighted and labelled. (B) Representation of the interaction at the interface of ALK4–ECD (Connolly surface with residues coloured by type) and h-CFC (stick model).

In particular, a close fitting between the h-CFC H9 and W12 with the ALK4 F55 and Y56 side chains is found. The groove is completed by favourable electrostatic and hydrophobic interactions. These indications could be helpful in designing mutagenesis experiments. Although we consciously know that any model generated *in silico* for a protein–protein complex cannot be expected to be completely accurate, we find that our model is nonetheless consistent with prior mutagenesis data and can therefore be quite useful in designing modified analogues to help in a more through understanding of the CFC/ALK4 molecular interaction.

Supporting information

Supporting information may be found in the online version of this article.

Acknowledgements

This project was supported by projects FIRB2003, *N° RBNE03PX83_005* and FIRB n° *RBRN07BMCT*. The technical assistance of Mr Leopoldo Zona and Dr Giuseppe Perretta is also gratefully acknowledged.

References

1. 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.
2. Yan YT, Liu JJ, Luo YEC, 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.
3. 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.
4. 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.
5. Shen MM. Decrypting the role of Cripto in tumorigenesis. *J. Clin. Invest.* 2003; **112**(4): 500–502.
6. 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.
7. 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.
8. Schier AF. Nodal signaling in vertebrate development. *Annu. Rev. Cell Dev. Biol.* 2003; **19**: 589–621.
9. Yeo C, Whitman M. Nodal signals to Smads through Cripto-dependent and Cripto-independent mechanisms. *Mol. Cells* 2001; **7**(5): 949–957.
10. 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.
11. 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.
12. 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.
13. Cheng SK, Olale F, Brivanlou AH, Schier AF. Lefty blocks a subset of TGFbeta signals by antagonizing EGF-CFC coreceptors. *PLoS Biol.* 2004; **2**(2): E30.
14. 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.
15. Risbridger GP, Schmitt JF, Robertson DM. Activins and inhibins in endocrine and other tumors. *Endocr. Rev.* 2001; **22**(6): 836–858.
16. Shani G, Fischer WH, Justice NJ, Kelber JA, Vale W, Gray PC. GRP78 and Cripto form a complex at the cell surface and collaborate to inhibit transforming growth factor beta signalling and enhance cell growth. *Mol. Cell Biol.* 2008; **28**(2): 666–677.
17. Bianco C, Normanno N, Salomon DS, Ciardiello F. Role of the cripto (EGF-CFC) family in embryogenesis and cancer. *Growth Factors* 2004; **22**(3): 133–139.
18. Gray PC, Shani G, Aung K, Kelber J, Vale W. Cripto binds transforming growth factor beta (TGF-beta) and inhibits TGF-beta signalling. *Mol. Cell Biol.* 2006; **26**(24): 9268–9278.
19. 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.
20. 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.
21. Kirsch T, Sebald W, Dreyer MK. Crystal structure of the BMP-2-BRIA ectodomain complex. *Nat. Struct. Biol.* 2000; **7**(6): 492–496.
22. Fields GB, Noble RL. Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.* 1990; **35**(3): 161–214.
23. Bax A, Davis DG. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* 1985; **65**: 355–360.

24. Bax A. Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson.* 1985; **63**: 207–213.
25. States DJ, Haberhorn RA, Ruben DJ. A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.* 1982; **48**: 286–292.
26. 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.
27. 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.
28. Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* 2004; **278**: 313–352.
29. Wuthrich K. *NMR of Proteins and Nucleic Acids*. Wiley: New York, 1986.
30. 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.
31. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* 1997; **273**: 283–298.
32. Guntert P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* 2004; **278**: 353–378.
33. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, III, DeBolt S, Ferguson D, 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.
34. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* 1996; **14**(1): 51–55, 29–32.
35. 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.
36. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. Pfam: clans, web tools and services. *Nucleic Acids Res.* 2006; **34**: (Database issue): D247–D251.
37. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 2005; **21**(7): 951–960.
38. 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–W248.
39. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* 1999; **287**(4): 797–815.
40. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; **22**(22): 4673–4680.
41. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 1993; **234**(3): 779–815.
42. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature* 1996; **381**(6580): 272.
43. Vriend G. WHAT IF: a molecular modelling and drug design program. *J. Mol. Graphics* 1990; **8**(1): 52–56, 29.
44. 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.
45. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature* 1992; **356**(6364): 83–85.
46. Wallner B, Elofsson A. Can correct protein models be identified? *Protein Sci.* 2003; **12**(5): 1073–1086.
47. Wallner B, Elofsson A. Identification of correct regions in protein models using structural, alignment, and consensus information. *Protein Sci.* 2006; **15**(4): 900–913.
48. Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* 1997; **272**(1): 106–120.
49. Moont G, Gabb HA, Sternberg MJ. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins* 1999; **35**(3): 364–373.
50. 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.
51. Minchiotti G, Manco G, Parisi S, Lago CT, Rosa F, Persico MG. Structure-function analysis of the EGF-CFC family member Cripto identifies residues essential for nodal signalling. *Development* 2001; **128**(22): 4501–4510.
52. 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.
53. 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.
54. 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.
55. 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.
56. 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.
57. Sternberg MJ, Gabb HA, Jackson RM, Moont G. Protein-protein docking. Generation and filtering of complexes. *Methods Mol. Biol.* 2000; **143**: 399–415.
58. 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.
59. 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.

