Extracellular domain of type I receptor for transforming growth factor-p: molecular modelling using protectin (CD59) as a template

T. Sakari Jokiranta[a,*], Jorma Tissari[a], Olle Teleman[b], Seppo Meri[a]

[a] Department of Bacteriology and Immunology, Huartman Institute, Haartmuninkatu 3, PO Box 21 FIN-00014, University of Helsinki, Helsinki, Finland
[b] VTT Biotechnology and Food Research, Biologinkuja I, PO Bos 1500 FIN-02044 VTT, Espoo, Finland

Received 16 August 1995; revised version received 16 October 1995

Abstract We have observed that the extracellular domain of TβRI and protectin (CD59), an inhibitor of the membrane attack complex of complement, share structural features, a distinct spacing of ten cysteines and a C-terminal 'Cys-box'. Based on these common features and the recently determined NMR-structure of protectin, a three-dimensional model for the extracellular domain of TβRI was constructed. After energy minimization and molecular dynamics simulation, a structure with four extending fingers (pes quattvordigitorum) and two clusters of charged residues was obtained. This model provides a view to the understanding of interactions between TβRI, TβRII and TGFβ during ligand recognition and signal transduction.

Keywords: Receptor; Transforming growth factor-,& Protein-serinelthreonine kinase; Protein structure, tertiary; Complement

1. Introduction

Transforming growth factors-β1–β3 (TGFβ) are a group of cytokines, which appear usually as homodimers with an M_r of about 2.5 kDa. TGFβs participate in several different local or systemic biological responses, e.g. inflammation, host defence, growth regulation, cellular differentation, development and tissue repair [1]. The TGFβ group of cytokines belongs to a larger family of structurally related proteins that include activins, inhibins, Müllerian inhibiting substance (MIS). bone morphogenetic proteins (BMPs) and osteogenic proteins (OPs) [1]. Many of these cytokines bind to cell membrane receptors that have intracellular serine/threonine kinase domains (activin receptor-like kinases; ALKs) [2–6].

The effects of TGFβs are currently known to be mediated by a series of molecular interactions. First, TβRII binds TGFβ, which is free in the fluid-phase or presented by TβRIII/endoglin [7]. Thereafter the TβRII–TGFβ complex interacts with TβRI. TβRII is a constitutively active protein kinase, but a stable complex with both TGFβ and TβRI is needed for signal propagation through phosphorylation of TβRI by TβRII in a heterodimeric or -tetrameric complex [1,5–9].

TβRI and activin receptor type IB (ActR-IB, ALK4) have nearly identical intracellular kinase domains and mediate similar responses although they interact with different type II receptors and ligands. Thus, it is likely that type I receptors determine the specificity of responses to various type II–ligand–type I complexes [7.10]. It is probable that mainly the extracellular part of TβRI is needed for the specific recognition of the TβRII–TGFβ complex [l,ll].

The extracellular parts of TβRI and ALK1-4,6 contain each of about 100 amino acids including 10 cysteines [2,3,5,6]. Their sequence homology with each other suggests a common domain structure within this family. At present no information is available about the tertiary structure of these domains, mainly because of difficulties in obtaining enough protein for NMR spectroscopy or X-ray crystallography.

We have observed that the extracellular domains of TβRI and ALK1-4,6 share structural features with another family of membrane proteins that include protectin (CD59), three domains of the receptor for urokinase-type plasminogen activator (uPAR), mouse lymphocyte protein Ly-6 and a viral membrane protein of *Herpes saimiri* (HVS-15) [12]. In addition, there is similarity to several snake venom neurotoxins [13] and a recently described human antigen E48 [14]. The basic domain in this putative superfamily ('protectin-superfamily') is characterized by the conserved distribution of 10 cysteine residues within a domain of 65–85 amino acids and a highly conserved C-terminal sequence CCXXXXCN ('cysteine box'). The index molecule protectin (CD59) is a glycosylphosphatidylinositol (GPI)-anchored membrane protein containing 77 amino acids. The tertiary structure of CD59 has been determined by two-dimensional NMR spectroscopy [15,16]. All the cysteine pairing patterns of the various family members resolved so far have been found to be analogous, although some domains lack one disulphide bridge. Thus it seems that the internal cysteine skeleton serves as a preferred and compact frame for these proteins of different function. It is likely that also the rest of the family shares the same cysteine pairing pattern. Therefore, it is possible to construct tertiary structure models for the extracellular domains of the ALK-type protein kinases by using protectin as a template in computer-aided homology modelling. In this study, we have modelled the structure of the extracellular domain of TβRI which is an important and functionally best characterized representative of the receptor-type serine/threonine protein kinases.

2. Materials and methods

Computer-aided molecular modelling was performed with the Insight11 program package (version 2.2.1, Biosym Technologies Inc., San Diego, CA) using a Silicon Graphics Iris Indigo XZ 4000 work-station (Silicon Graphics Inc., Mountain View, CA). The Biopolymer and Homology modules of Insight11 were used to build a preliminary model of the extracellular domain of TβRI. Protectin (CD59) was found to be the optimal template for TβRI after screening of all protein structures available in the Brookhaven Protein Databank (PDB; release #68, April 1994; Brookhaven National Laboratory. Upton, NY, USA) [17]. The 10 conserved cysteines of protectin were used as a basis for homol-

*Corresponding author. Fax: (358) (0) 434-6382.
E-mail: sjokiran@cc.helsinki.fi

SSDI 0014-5793(95)01239-7

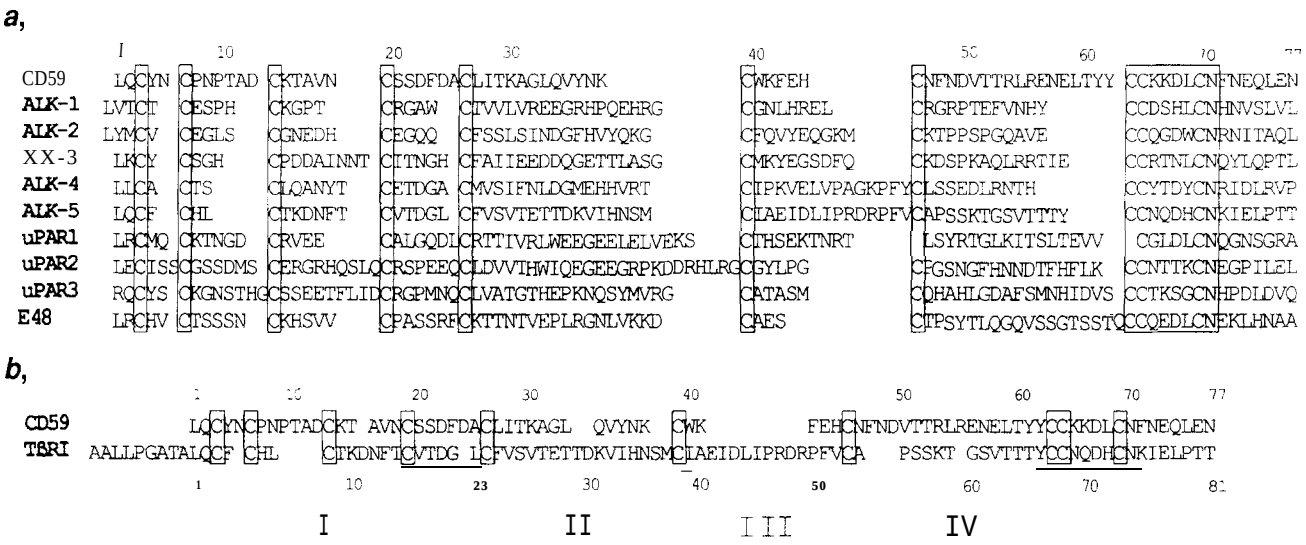

Fig. 1 Alignment of cysteine residues in protectin (CD59) and similar domains in other human membrane proteins. (a) Alignment of protectin and corresponding parts of extracellular domains of the type I TGFβ receptor family (ALK1-5), three domains of the receptor for urokinase-type plasminogen activator (uPAR1-3) and human antigen E48. The boxes indicate the alignment of cysteine residues and the C-terminal 'Cys-box'. No attempt to align other residues was made. (b) Alignment of protectin and the extracellular domain of TβRI (ALK-5). The internal 'core' residues (1–4, 15 20. 39. 52 54. 66 74) of the extracellular domain of TβRI are underlined. The sites of extending loops are numbered as I–IV. The residues of TβRI are numbered according to the frame of protectin.

ogy alignment. The most N-terminal part of TβRI was not modelled, because the sequence of the template molecule reached only two amino acids N-terminal from the first cysteine. The amino acids of the extracellular domain of TβRI were numbered by beginning from residue Leu-10 (as counted from the end of the putative hydrophobic leader sequence).

The CD59 template structure was kindly provided by M. Fletcher and D. Neuhaus (Laboratory for Molecular Biology, MRC, Cambridge) (PDB accession code 1cdq, structure #1). It represents the one of the 50 structures calculated from NMR-data that has the lowest distance constraint violation energy [15]. Four loops (I–IV) were needed to complete the preliminary model structure (Fig. 1). The best available loop structures were searched from **all** protein structures deposited in PDB using anchors of 3–5 residues. In choosing the loop structures the following criteria were used: amino acid sequence, distance between the first C_α-atoms of the anchors, rms-deviation, secondary structure (α-helix, β-sheet) prediction based on the sequence of TβRI and estimation of suitability to the overall tertiary structure. Model loops I (residues 5–14 of TβRI) and II (residues 21–38) were obtained from the snake venom toxin family members neurotoxin-B (PDB accession code 1nxb) and neurotoxin-II (1nor), respectively. Loops III (residues 40–51) and IV (residues 55–65) were parts of human immunoglobulin light chains 2rhe and 2mcg, respectively.

The preliminary model structure of TβRI was soaked in a waterbox of 1392 water molecules (dimensions of the box were 44 x 52 × 31 Å) to achieve about 6 Å of water surrounding the whole protein. Using the Discover module (version 2.Y.5) the water molecules and the protein were allowed to relax by diminishing, gradually the fixations of the protein between successive energy minimizations. The energy minimizations were performed using both the steepest descents and conjugate gradient algorithms. Molecular dynamics simulations were performed by gradually increasing the temperature of the system. The target was simulated 1.0 ps at 100 K, 20.0 ps at 300 K and 196.0 ps at 600 K using a 1.0 fs time step throughout. The calculations* were made under full periodic boundary conditions using a 12–13 Å cut-off distance and CVFF as forcefield. To adjust the waterbox to an altered conformation of the protein, the dimensions of the waterbox and the amount of water molecules were changed three times (amount of H_2O molecules was 176X. 152 I and 1490, respectively). The average pressure in the waterbox during the simulation at 600 K was 4800 bars.

After a total of 217 ps molecular dynamics simulation, 10 structures were selected from the last 196 ps (at 600 K) using a local potential

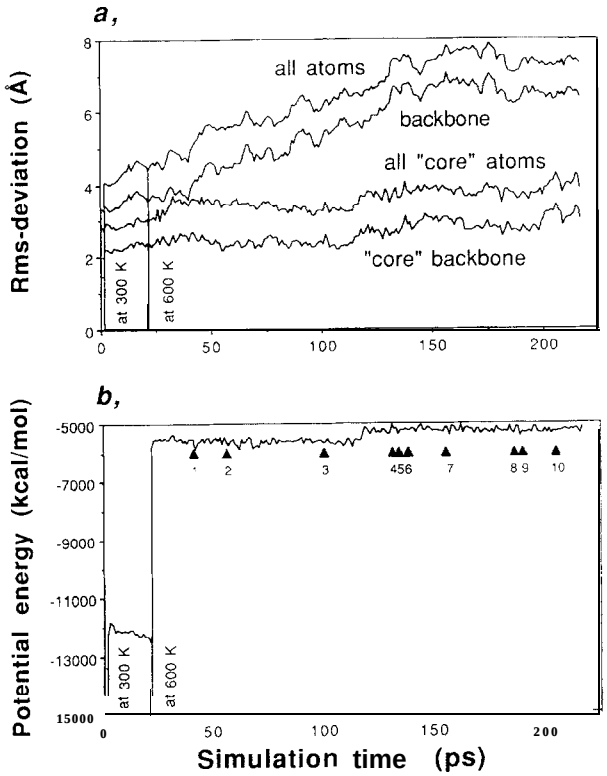

Fig. 7. Rms-deviation (rmsd) and potential energy of the model of the TβRI extracellular domain as a function of molecular dynamics simulation time. (a) Rmsd of all atoms or backbone atoms (C, C_α, N, O) of either all residues or 'core' residues of the model. Rmsd values were calculated by comparing the coordinates of the model during dynamics simulation to the coordinates of the preliminary model. (b) Potential energy of the model as a function of dynamics simulation time. The ten arrowheads indicate time points for selecting structures at local potential energy minima. These ten model structures were subjected to energy-minimization (structures #4–10 are shown in Fig. 3). The size of the waterbox was changed three times: at 1 ps, 21 ps and 118 ps.

Fig. 3. Tertiary model structures of the extracellular domain of TβRI. Backbone traces of seven energy-minimized model structures after superimposition of the 'core' residues. Each of the seven structures were energy-minimized from a local potential energy minimum found during dynamics simulation at 600 K. The spectrum between yellow and red indicates the order of structures as a function of simulation time in the order shown in Fig. 2b (yellow = #4, red = #10). Disulphide bridges are shown in grey.

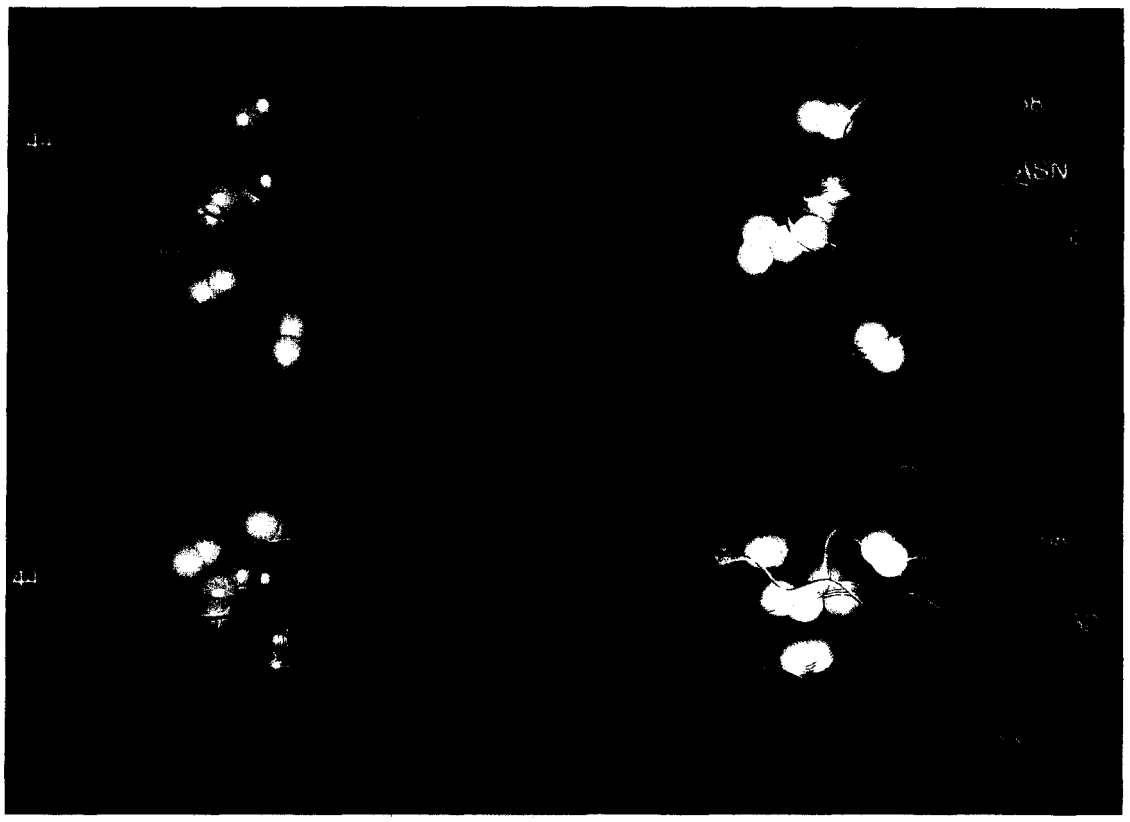

Fig. 4. A comparison between the template structure protectin (right) and the extracellular domain of TβRI (left). The TβRI structure is the one with the lowest potential energy of the ten energy-minimized structures (structure #10; see Fig. 2b). Both structures are shown as top- and side-projections. Sulphur atoms of cysteines are shown in yellow

energy minimum of the entire system (water molecules and the protein) as a criterion. The 10 structures were energy minimized in their waterboxes using the conjugate gradient algorithm until the maximum derivative was below 0.001 kcal·Å^{-1}. Coordinates of one of the ten structures (#10) have been deposited in the Brookhaven Protein Databank (accession code 1TBI).

3. Results

3. I. *Sequence similarity between TβRI and protectin*

Sequence homology between protectin and TβRI/ALK1–4 is limited to the 10 cysteines, amino acids surrounding them and to the CCXXXXCN 'Cys-box' (Fig. la). As seen in Fig. lb the number of amino acids between the neighbouring cysteines is different in TβRI and protectin. Thus a total of four loops were constructed to the model from coordinates in the Brookhaven PDB to generate a preliminary model structure.

3.2. Course of *the modelling*

The preliminary model structure of the extracellular domain of TβRI was soaked in a waterbox and subjected to energy minimization and molecular dynamics simulation. The course of molecular dynamics simulation was analyzed by calculating root mean square deviations (rmsd) of individual structures at various time points compared to the preliminary tertiary structure (Fig. 2a). Rmsd was calculated for four different sets of atoms: (1) all atoms in all residues; (2) backbone atoms (C,, C, 0, N) in all residues; (3) all atoms in 'core' residues, corresponding to internal residues adopted from protectin coordinates: 1–4, 15-20, 39, 52-54, 66-74; and (4) backbone atoms in 'core' residues.

The preliminary model structure underwent a series of conformational changes during the energy minimization procedure, as the rmsd for all residues after the minimization was somewhat high (3.30 Å for all atoms, 2.77 Å for backbone atoms). During dynamics simulation the rmsd of all residues (all or backbone atoms) increased in a fairly linear fashion to reach a plateau after an about 130 ps simulation (Fig. 2a). The rmsd-plateau level was about 7 Å for all atoms of the molecule and 6 Å for all backbone atoms.

During minimization the rms of the 'core' residues changed 1.96 Å for all atoms and 1.38 Å for the backbone. These changes constitute about 50% of total rmsd after a 217 ps dynamics simulation (1.96 vs. 3.85 for all 'core' atoms and 1.38 vs. 3.0 for backbone atoms of the 'core'). As seen in Fig. 2a, the rmsd of 'core' residues varied within a relatively small range during the 196 ps simulation at 600 K (3–4 Å for all atoms and 2-3 Å for the backbone). Potential energy of the model structure varied only a little during dynamics simulation except during changes in temperature or in the size of the waterbox (see section 2). No radical global potential energy minimum was reached during the simulation procedure (Fig. 2b).

3.3. *Analysis of the model*

Ten structures each representing a local potential energy minimum of the waterbox-system were subjected to energy-minimization. After minimization the potential energies of the ten model structures were 750–1150 kcal/mol while potential energy of the similarly energy minimized template structure was 665 kcal/mol. The rmsd values against the preliminary model before any minimization or dynamics simulation were 4.8-7.5 Å for all atoms of the molecule (data not shown).

Seven of the ten energy-minimized structures are shown in Fig. 3. These structures cover the last 90 ps of the dynamics simulation, when the rmsd of the whole molecule had reached the plateau described above. After superimposing the 'core' residues of these structures it can be seen that the core appears relatively stable whereas the loops, especially their distal parts, shift somewhat out of the average plane of the corresponding loops. This is also observed numerically as the rmsd between the core residues of each possible pair of the seven energy-minimized structures was below 2.5 Å for all core atoms and below 1.8 Å for backbone of the core.

As compared with the NMR structure of the template molecule protectin, the overall model structure of the extracellular domain of TβRI was quite similar (Fig. 4). The folding pattern in the model of the extracellular domain of TβRI appears like a pes quattvordigitorum with three fingers pointing forward and one backwards. As indicated by visual and Ramachandran plot analysis (data not shown) the model structure of the extracellular domain of TβRI contains 1.5 turns of a-helix (residues 43–49). Both the template molecule protectin and the model structure contain five long β-sheet structures. In addition, the model has short stretches of B-sheet.

3.4. *Properties of the model structure*

As charged amino acids and surface hydrophobic patches are often involved in intermolecular protein-protein interactions, we analyzed their distribution in the 3D-model of TβRI. The charged amino acids seem to be located in two clusters on the surface of the extracellular domain of TβRI (Fig. 5). One cluster is located distally between the first and second loops and contains eight charged residues: three lysines (10, 3 1,75), three aspartic (11, 30, 71) and two glutamic acids (27, 77). In the middle of this cluster there are also three histidines (6, 34, 72). The other cluster is on the 'back-finger', loop III. It contains six charged residues: two arginines (47, 49), three aspartic acids (18, 43, 48) and one glutamic acid (41). The uncertainty regarding local conformation in loop III does not affect the existence of this cluster, since the obtained alternative conformations all preserve the cluster. Outside these two clusters there is only one charged residue in the modelled structure (Lys-58). On the surface of the model there is one major hydrophobic cluster (residues 22, 39, 42, 44, 45, 51) located within and near the second charge cluster.

4. Discussion

Modelling of the structure of the extracellular domain of TβRI was based on the NMR-structure of protectin (CD59) [15,16]. Protectin has ten cysteine residues, three of which are located in a C-terminal cysteine box. This cysteine pattern is conserved in the extracellular domain of TβRI as well as in the other members of the TβRI/ALK-family and uPAR (with the exception of uPAR domain #1 with eight cysteine residues) (Fig. 1). The pairing of the cysteines also appears to be conserved as all structures of the superfamily members determined so far have shown an analogous pairing pattern (protectin, uPAR domain #1 and twelve snake venom neurotoxins) [12,181. On this basis we made the fundamental assumption that the cysteine pairs are formed similarly in the extracellular domain of TβRI. Although it is likely that TβRI has this cysteine pairing pattern it remains to be experimentally verified. In the

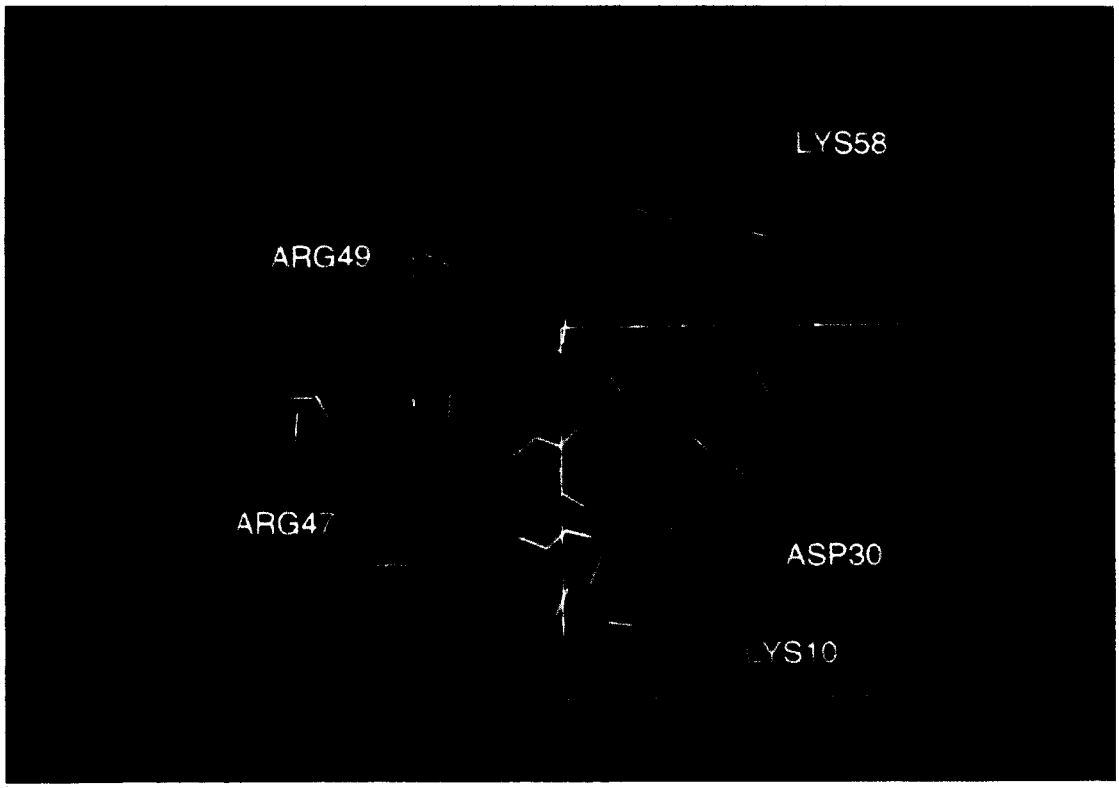

Fig. 5. Distribution of charged residues in the model of the extracellular domain of TβRI. Side chains of charged residues as seen from above and clustered in two boxes. Side chains of histidines are shown in yellow. Side chains of the positively and negatively charged amino acids are shown in blue and red, respectively.

less likely case that the cysteine pairing is significantly different in TβRI and the 'protectin superfamily' the model is not valid.

As seen in Fig. 2a the conformation of the core structures of the model, adopted directly from the coordinates of protectin, changed only moderately during the modelling procedure (rmsd for the backbone < 3 A). The conformation of the whole model changed much more, probably because of variation in the extending loops. However, a plateau in the rmsd for all atoms was reached at about 130 ps. This rmsd-plateau may reflect attainment of a compact or semi-compact model structure, as indicated by a rather small variation in the rmsd values during this time (Fig. 2a) and small rmsd of the core of the seven energy-minimized structures chosen from the course of the last 90 ps of dynamics simulation (rmsd for the backbone of the core < 1.8 A). As a rather stable core structure was reached during dynamics simulation in water at 600 K, the tertiary structure of the modelled domain seems to be firmly stabilized by the five disulphide bridges.

The overall rmsd level of about 7 Å for all atoms of the molecule and 6 Å for the whole backbone suggests that some parts of the preliminary model structure were quite different from the final model conformation. This is explained mainly by changes in the conformation of the loops. However, considering the large amount of variation in the loops, model structures of the outer halves of loops II and IV (residues 26-33 and 57-64) must be considered undetermined. There is also some uncertainty regarding the placement of the short a-helix in loop III. On the other hand, the structure of loop I appears well determined. After minimization, the potential energies of the ten model structures were somewhat higher than that of the template structure. The likeliest cause for this is the computational procedure, i.e. energy minimization from a conformation taken from a high temperature simulation. The energy minimization will produce a conformation which is still closer to a high temperature state than to a room temperature state, and it will therefore have a higher potential energy (compare the increase in potential energy on raising the temperature, Fig. 2b). Ideally, the proper way to compare the potential energies would be to anneal the ten conformations, i.e. lower the temperature slowly from 600 K to 300 K during simulation. This amounts to a considerable amount of computer time and was deemed to be outside the scope of this work.

The five cysteine pairs are probably a major factor in stabilizing the tertiary structure of the members of the protectin-superfamily. Thus it is not surprising that the model structure of TβRI resembles protectin in containing three 'fingers' protruding forward in a plane (Fig. 4). Both molecules are rich in P-sheet structure and contain one a-helix, although the latter is found at a different position. The two molecules show differences in the surface distribution of charged and hydrophobic residues. These differences probably reflect different biological properties of the molecules.

The model structure of TβRI may provide some insight into the physiological activities of TβRI. It is known that the extracellular part of TβRI is needed to specifically recognize the TβRII–TGFβ-complex [7,10]. However, the exact nature of the binding of the TβRII–TGFβ complex to TβRI is not known. Further, as the specificity of the biological response to TGFβ related cytokines (activins, inhibins, BMPs, MIS and OPs) in different cell types seems to be defined by the particular type

I receptors. the model could help understand the mechanism of this specificity as well as the mechanisms behind the multifunctional nature of these cytokines. Practically, the most important application of the current model is that it provides a basis for site-directed mutagenesis studies of the type I family of TGFβ and activin receptors. The surface of the modelled domain of TβRI bears two charge clusters (residues 10, 11, 27, 30, 31, 71, 75, 77 and 18, 39–54), the latter located next to the hydrophobic patch. The former charge cluster is located on TβRI analogously to a recently described functional site of erabutoxin a [19], a snake toxin homologous to protectin and the extracellular domain of TβRI. It is possible that the two clusters are involved in the biological functions of TβRI and thus provide a fascinating target for mutagenesis studies. The present model and the current biochemical data suggest that these clusters could constitute two distinct binding sites in the extracellular domain of TβRI, perhaps one for TGFβ and the other for TβRII.

Acknowledgements: We thank Drs. M. Fletcher and D. Neuhaus, MRC, Cambridge, UK, for kindly providing us with the coordinates of protectin (CD59) prior to publication. We thank Dr. L. Kuutti for kind advice on modelling and Dr. J. Keski-Oja for careful reading of the manuscript. This study was supported by the Academy of Finland, the Sigrid Juselius Foundation, and the University of Helsinki, Finland.

References

[1] Massaque, J., Attisano, L. and Wrana, J.L. (1994) Trends Cell Biol. 4, 172–178.
[2] ten-Dijke, P., Ichijo, H., Franzén, P., Schulz, P., Saras, J., Toshima, H., Heldin, C.-H. and Miyazono, K. (1993) Oncogene 8, X79-2887.
[3] ten-Dijke, P., Yamashita, H., Sampath, T.K., Reddi, A.H., Estevez, M., Riddle, D.L., Ichijo, H., Heldin, C.H. and Miyazono, K. (1994) J. Biol. Chem. 269, 16985–16988.
[4] ten-Dijke, P., Yamashita, H., Ichijo, H., Franzen, P., Laiho, M., Miyazono, K. and Heldin, C.H. (1994) Science 264, 101–104.
[5] Attisano, L., Carcamo, J., Ventura, F., Weis, F.M., Massague, J. and Wrana, J.L. (1993) Cell 75, 671-680.
[6] Franzen, P., ten-Dijke, P., Ichijo, H., Yamashita, H., Schulz, P., Heldin, C.H. and Miyazono, K. (1993) Cell 75, 681-692.
[7] Wrana, J.L., Attisano, L., Wieser, R., Ventura, F. and Massaque, J. (1994) Nature 370, 341–347.
[8] Ventura, F., Doody, J., Liu, F., Wrana, J.L. and Massaque, J. (1994) EMBO J. 13, 5581-5589.
[9] Yamashita, H., ten Dijke, P., Franzén, P., Miyazono, K. and Heldin, C.-H. (1994) J. Biol. Chem. 269, 20172–20178.
[10] Carcamo, J., Weis, F.M., Ventura, F., Wieser, R., Wrana, J.L., Attisano, L. and Massague, J. (1994) Mol. Cell Biol. 14, 3810–3821.
[11] Derynck, R. (1994) Trends Biol. Sci. 19, 548-553.
[12] Meri, S. (1994) Immunologist 2, 149–155.
[13] Endo, T. and Tamiya, N. (1987) Pharmacol. Ther. 34, 403–451.
[14] Brakenhoff, R.H., Gerretsen, M., Knippels, E.M.C., van Dijk, M., van Essen, H., Weghuis, D.O., Sinke, R.J., Snow, G.B. and van Dongen, G.A.M.S. (1995) J. Cell Biol. 129, 1677–1689.
[15] Fletcher, C.M., Harrison, R.A., Lachmann, P.J. and Neuhaus, D. (1994) Structure 2, 185–199.
[16] Kieffer, B., Driscoll, P.C., Campbell, I.D., Willis, A.C., van der Merwe, P.A. and Davis, S.J. (1994) Biochemistry 33, 4471–4482.
[17] Bernstein, F.C., Koezle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 5355542.
[18] Ploug, M. and Ellis, V. (1994) FEBS Lett. 349, 163–168.
[19] Trémeau, O., Lemaire, C., Drevet, P., Pinkasfeld, S., Ducancel, F., Boulain, J.-C. and Ménez, A. (1995) J. Biol. Chem. 270, 9362–9369.