Molecular modelling of the Norrie disease protein predicts a cystine knot growth factor tertiary structure

Thomas Meitinger[1], Alfons Meindl[1], Peer Bork[2,3], Burkhart Rost[2], Chris Sander[2], Martina Haasemann[4] & Jan Murken[1]

The X-lined gene for Norrie disease, which is characterized by blindness, deafness and mental retardation has been cloned recently. This gene has been thought to code for a putative extracellular factor; its predicted amino acid sequence is homologous to the C-terminal domain of diverse extracellular proteins. Sequence pattern searches and three-dimensional modelling now suggest that the Norrie disease protein (NDP) has a tertiary structure similar to that of transforming growth factor ß (TGFß). Our model identifies NDP as a member of an emerging family of growth factors containing a cystine knot motif, with direct implications for the physiological role of NDP. The model also sheds light on sequence related domains such as the C-terminal domain of mucins and of von Willebrand factor.

[1]*Abteilung für Pädiatrische Genetik, Kinderpoliklinik der Ludwig-Maximilians-Universität, Goethestrasse 29, 80336 München, Germany*
[2]*European Molecular Biology Laboratory, 69012 Heidelberg, Germany*
[3]*Max Delbrück Center of Molecular Medicine, 13122 Berlin, Germany*
[4]*Institut Jacques Monod, 75251 Paris, France*

Polypeptide growth factors are involved in a range of trophic and plastic growth responses in eukaryotic cells[1]. They include several subgroups identified on the basis of sequence homology. A group of cysteine-rich peptides with a common three dimensional protomer structure includes neurotrophins such as transforming growth factors (TGFß), nerve growth factor (NGF), platelet derived growth factor (PDGF-B), and other sequence related proteins (for review see ref. 2). While the sequence of nerve growth factor (NGF) has been known for more than 20 years[3], the availability of the three-dimensional (3-D) structure now greatly facilitates detailed structural and functional studies. The crystal structure of the murine NGF-dimer revealed a characteristic 3-D fold[4] that has since been observed in other growth factors with low sequence similarity, including TGFß2 (refs 5,6) and PDGF-B[7]. The common structural core of these proteins is formed by an equivalent pattern of three disulphide bridges[8]. In this motif, a ring is formed by two disulphide bridges. A third disulphide passes through the centre of the ring (cystine knot motif[2], see Fig. 1*a*). Monomers of this family of proteins have 100–130 amino acids residues. The six cysteine residues and their spacing are essentially the only conserved feature when the primary sequences are aligned. All of the growth factors containing the cystine knot motif form active dimers that bind to specific receptors. However, the mode of dimerization is completely different for each of the three families[2]. The most recently described member of the family of dimer-forming TGFß related proteins appears to be a glial cell line-derived neurotrophic factor (GDNF)[9].

We have recently described the structure of a gene which codes for a cysteine-rich, extracellular protein with a sequence length of 133 amino acids[10]. Loss of function mutations within this gene cause Norrie disease, an X-linked disorder characterized by blindness, deafness and mental retardation. The gene was identified by positional cloning[11,12] and intragenic deletions as well as point mutations have been demonstrated in patients with Norrie disease[10,13]. Clinical and histological examination suggests the involvement of growth or angiogenic factors in the pathogenesis of the disease whose ocular manifestations are characterized by dysorganization of retinosensory cells and a highly vascular retrolental membrane[14–16]. The C-terminal cysteine-rich domain (CT domain) of NDP shows homology to the C-termini of functionally and structurally diverse extracellular modular proteins[10,17]. The closest homologues are several extracellular mucins and von Willebrand factor (vWF[18]). The *Drosophila* slit protein, involved in development of midline glia and commisural axon pathways[19], and a family of growth regulators including cef*10*, *cyr61*, *nov* and connective tissue growth factor (CTGF)[20], also have a homologous CT domain. No function has been assigned so far to this domain.

To examine the possible involvement of growth factors in the pathogenesis of Norrie disease, we have tested the hypothesis that NDP adopts a 3-D structure with a cystine knot motif similar to that of TGFß, NGF and PDGF. The structural similarity is explored using sequence pattern searches, secondary structure predictions, tertiary structure modelling, and conservation mapping, that is the analysis of the placement of conserved residue properties in the 3-D model. The progression from positional cloning of the NDP gene to modelling of the 3-D structure and precise functional predictions for particular residues illustrates the power of combining

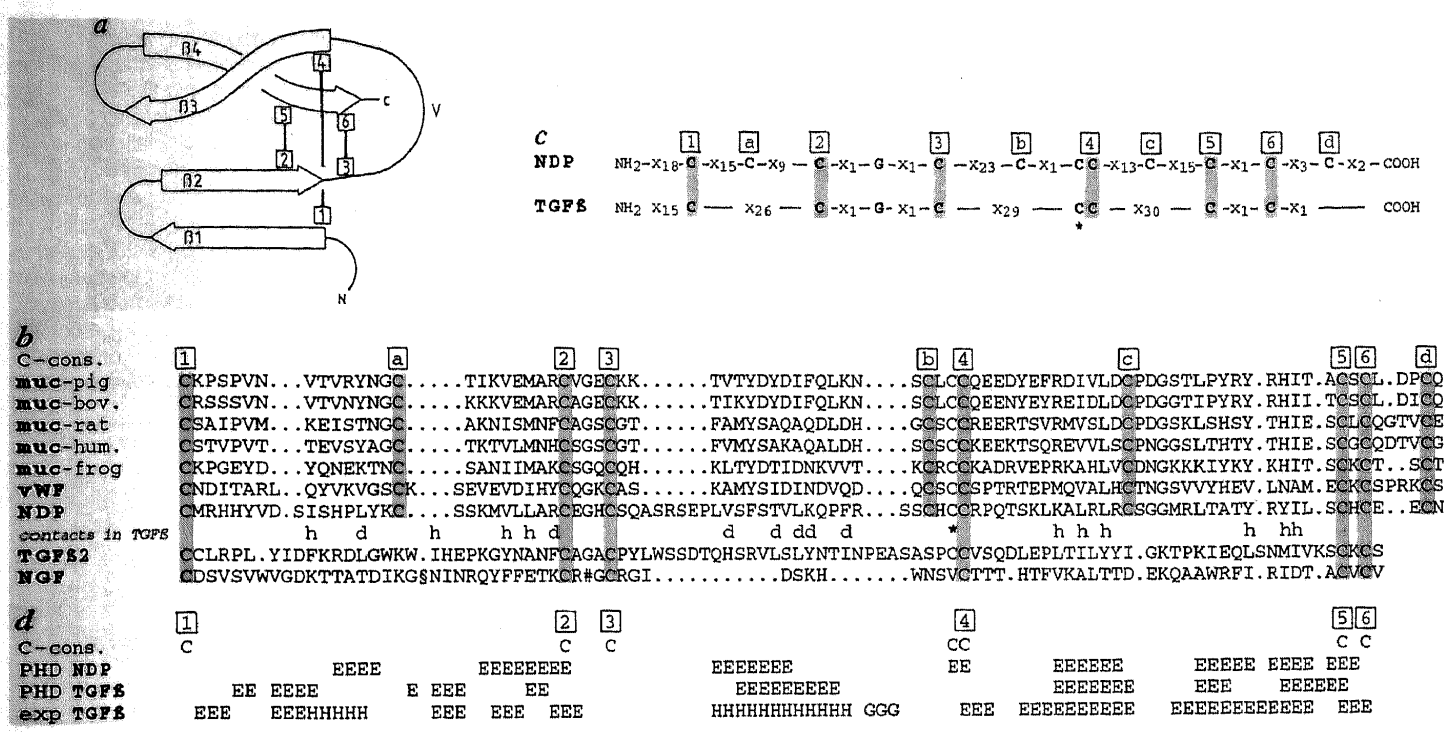

Fig. 1 Cysteine residues, sequence alignment and secondary structure prediction of TGFß and NDP. *a*, Schematic representation of a TGFß-like monomer modified after refs 2 and 8. ß-strands are drawn as arrows (ß1-ß4). The cystine knot forming residues are numbered 1–6; disulphide bridges are indicated. The loop between ß-strand ß2 and ß3 shows variable lengths and crystal structure in TGF, NGF and PDGF (V). *b*, Multiple alignment of NDP and related CT-domains with TGFß and NGF. Rat and human mucin sequences were added to the CT-domain aligment 10 (muc-rat[40], aa 732–817; muc-hum.[41], aa 880–965). The structure-based alignment of TGFß2 with NGF was taken from ref. 8. All sequences are aligned to match the cysteine residues numbered 1–6 and a-d. Insertions/deletions are indicated by dots. The cysteine forming an intermolecular disulphide bridge in TGFß is marked by a star. Dimer contacts ('d') and hydrophobic core contacts ('h') in TGFß are consistent with corresponding positions in the CT-domains. Two inserts in the NGF sequence have been deleted (§,KEVTVLAEV; #, ASNPVES). *c*, Alignment of cysteine and glycine residues in TGFß with those in NDP. The number of amino acids (X_n) between conserved cysteine residues (C) is given. Cysteines involved in the cystine knot are numbered 1–6. The additional cysteines conserved in the CT domain are numbered a-d. A star indicates the cysteine residue involved in dimerization. *d*, Comparison between observed ß-strands and helices in TGFß and predicted 2nd structures of TGFß and NDP (exp, crystal structure; PHD, prediction from sequence family; E, ß-strand; H, α-helix; G, 3_{10}-helix; space, loop). NDP is predicted to consist predominantly of ß-sheets, with strand placement roughly in agreement with those experimentally observed in TGFß. The central predicted ß-strand (sequence LVSFSTV in NDP) actually is a helix in TGFß, involved in dimerization, and is likely to be present in NDP; the incorrect prediction of this strand is probably due to the hydrophobic residues at the surface involved in the dimer contacts.

experimental and theoretical methods in the analysis of human genetic diseases.

Sequence-structure alignment

NDP and related CT-domains show a significant sequence homology (Fig. 1*b*). The highest identity and similarity values are found between NDP and the human intestinal mucin, MUC2, with values of 30% and 49%, respectively. Between TGFß and related sequences with known 3-D structures, the percentage figures do not exceed 10% for identity and 25% for similarity. When comparing TGFß with any of the NDP related sequences, similar non-significant values are obtained. However, a characteristic pattern emerges between TGFß and NDP related sequences when only cysteines and the distances between them are compared (Fig. 1*c*). The alignment of the NDP primary sequence to the known structure of TGFß reveals several conserved features: (i) Cys residues at the right spacing match each of the disulphide bridges in the common structural core of the TGFß/NGF family. Only a few insertions and deletions have to be introduced to align NDP and TGFß (6 residues in 5 gaps). In particular the knot-forming cysteines representing the core of the growth factor fold can be aligned without gaps. (ii) The Cys, which is known to be involved in the dimerization of TGFß, is present at the equivalent position in NDP (Cys95 in NDP). (iii) A structurally important glycine with a positive phi-angle, an unusual conformation, is conserved in TGFß, PDGF and the other CT-domain proteins. Differences can be observed in the length of loops which extend from cysteine residues 1 and 2 as well as from cysteine residues 4 and 5 in TGFß related proteins with similar 3-D structures (Fig. 1*a*). In contrast, the length of the loops is less variable in the CT domains.

Sequence pattern search

The significance of the similarity between NDP and TGFß can be confirmed with consensus sequence patterns[21] derived from the alignment of CT-domains and TGFß related sequences (data not shown). Two different consensus patterns were calculated from an alignment of all CT family members with 50 TGFß2 homologues: (i) a pattern which only specifies the invariant cysteine and glycine residues common to the TGFß family and CT-domains (Fig. 1*c*) and (ii) a pattern which includes additional sequence features such as conserved hydrophobic positions. With the first pattern, all members of both families are distinguished readily from all other

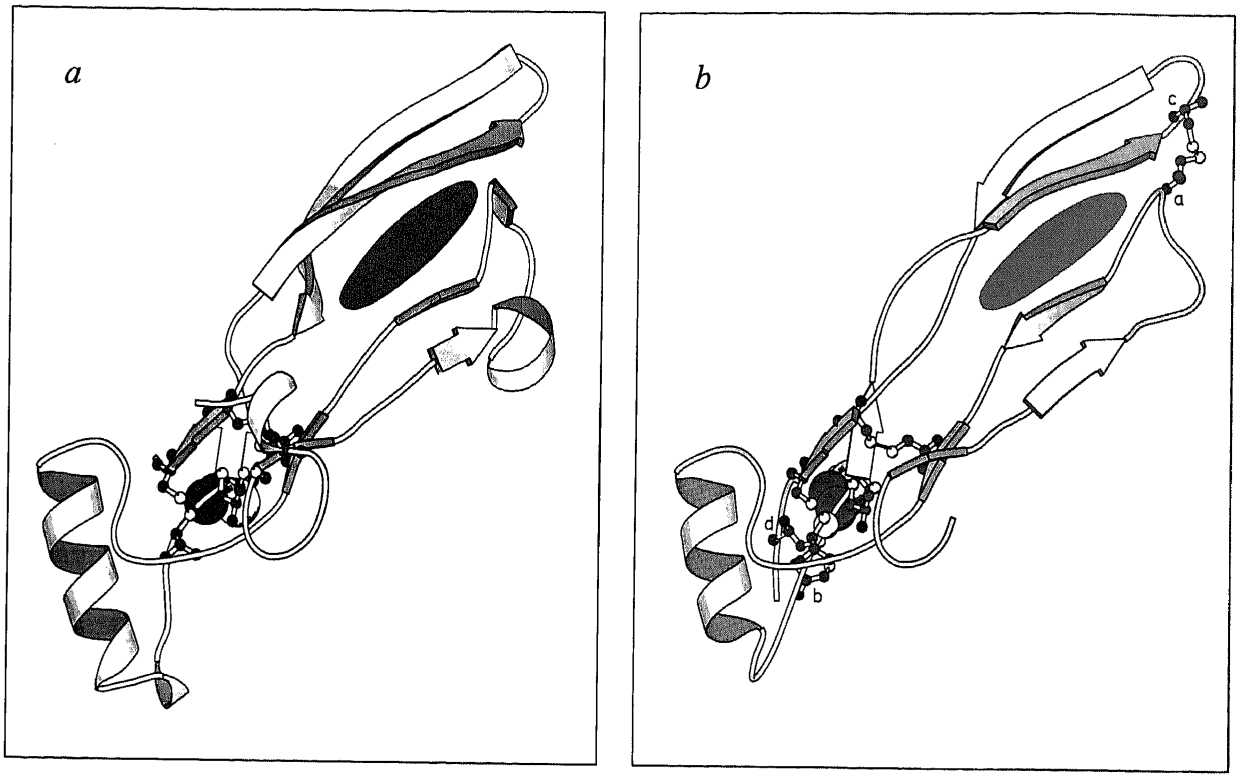

Fig. 2 Ribbon diagrams of the experimental 3-D structure of (*a*) TGFß and (*b*) the 3-D-model of NDP. The side chains of all cysteines are shown as balls and sticks. Full size atoms near the cystine knot are in the side chain of the cysteine residue involved in the dimer-forming disulphide bond. The shaded area indicates the approximate location of the hydrophobic cluster on one face of the ß-sheets. Cys residues a–d marked as in Fig. 1.

sequences in the databases. The discrimination is even clearer with the second pattern.

Secondary structure

The predicted secondary structure of NDP agrees well with that derived from the experimentally known 3-D structure of TGFß (Fig. 1*d*). Predicted ß-strands and loop regions between the residues forming a cystine knot of the CT-domain are compatible with the observed ß-strands common to TGFß and NGF. Secondary structure predictions are independent of the direct comparison of sequence patterns and become meaningful because of a recent significant increase in the performance based on multiple sequence alignments. The prediction method used—a profile based neural network method—has a sustained accuracy of about 71% in the three states helix, strand and loop, and an expected accuracy as high as 90% for residues with a very strong prediction signal[22].

3-D similarity between NDP and TGFß

From the alignment of NDP with TGFß, a full 3-D model of NDP protein was built by replacing the side chains in the TGFß structure with those of NDP. After optimization, the final model (Fig. 2) reveals other features that support the physiological relevance of this model structure. First, two additional disulphides can be formed by four of the remaining cysteines of NDP. The side chains of two cysteines each end up within a few Å of each other in the TGFß-based NDP model. This is a non-trivial consequence of mapping the NDP residues onto the 3-D structure of TGFß; the probability for such a constellation to occur by chance is very low. Second, a spatially compact network of hydrophobic residues at the hairpin end of TGFß (top right in Fig. 2*b*) links the four ß-strands. Almost all of these residues (9 out of 10, marked "h" in Fig. 1*b*) retain their hydrophobic cluster in the 3-D model. Third, the dimer interface in TGFß is formed primarily by one face of the ß-sheet (the palm of a slightly curled hand) interacting with the helix. Residues in this interface (marked "d" in Fig. 1*b*) remain predominantly hydrophobic in the NDP and CT sequences (7 out of 8, where the stem of Lys and Arg is considered hydrophobic). Finally, the single remaining unpaired cysteine of NDP occurs in exactly the same spatial position as the unpaired cysteine residue known to form an intermolecular disulphide link in the TGFß dimer.

Discussion

The structural similarity between NDP and TGFß is supported by comparison of disulphide patterns, sequence pattern searches, all-atom-model building after alignment, and conservation mapping. Our 3-D model of NDP is based on the known TGFß crystral structure. Although the model is in agreement with a cystine knot like structure, it has to be considered as a first approximation and is not a substitute for an experimentally derived tertiary structure. However, it provides the basis for precise predictions of functional details which can be tested experimentally.

A detailed prediction of the dimerization mode follows from the fact that NDP is recognizably more similar to TGFß than to NGF and to PDGF. Not only is the intermolecular disulphide bridge linking the two monomers predicted, but also the head-to-tail orientation that is characteristic of TGFß, and different from that of NGF. Such an orientation is supported by the presence of a sequence insertion common only to NDP and TGFß (and not present in NGF or PDGF) which corresponds to an a-helix involved in dimer interface contacts. The dimer

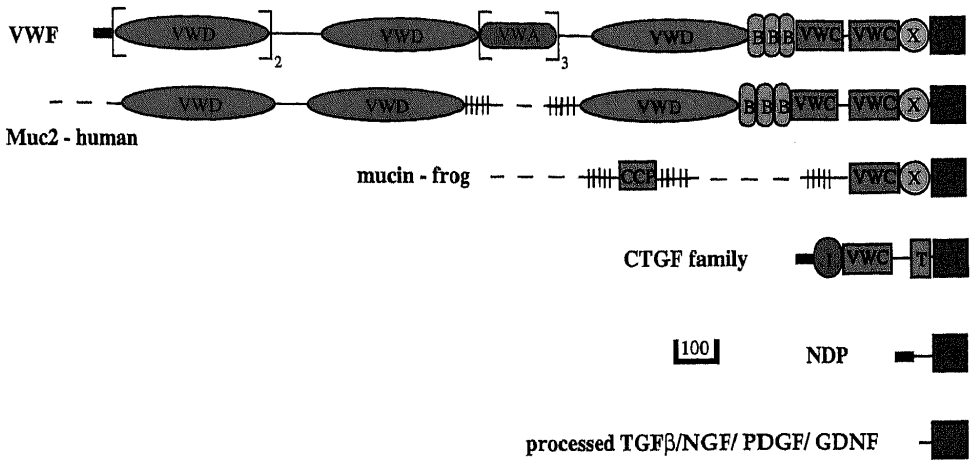

Fig. 3 Modular architecture of proteins containing the CT-domain. VWA, VWC, VWD, B and X are domains in von Willebrand factor also present in other extracellular modular proteins[42]; T, module first found in thrombospondin; I, module common to insulin-like growth factor binding proteins; CCP, domains found in complement control proteins; CT, C-terminal domain; thick bar, signal peptide; 100, approximate length of 100 amino acids. CTGF, connective tissue growth factor.

contacts in TGFß are made in part by large hydrophobic residues on the surface of the ß-sheet and the α-helix, partly conserved in the sequence of NDP. The details of dimerization may, of course, turn out to be different.

By sequence similarity, NDP belongs to a family of domains which also occur at the C-termini of a class of extracelluar modular proteins that includes vWF, several mucins, *slit* and a family of growth regulators related to connective tissue growth factor. The 3-D structure prediction of the NDP monomer can be extended to all members of this family and provides a first indication as to the function of this domain (Fig. 3). Also, the proposed mode of dimerization is likely to be valid at least for the closer relatives of NDP such as mucins and vWF. Remarkably, dimerization of vWF has been shown experimentally to occur via disulphide linkage within the C-terminal domain[23], consistent with our hypothesis. The prediction of the crucial role of Cys2773 in vWF (or equivalent residues in the other CT-modules) in dimerization provides a precise and testable hypothesis: mutants of NDP or vWF in these cysteines should be unable to dimerize.

Further examples of peptides with eight membered rings formed by two disulphide connected backbones include endothelins and neurotoxins[24]. All these peptides have in common that they affect their target cell through cell-surface receptors and that loops protruding from the core structure determine receptor specificity. In TGFß, receptor specificity has been associated with sequence alterations in the α-helix between cystine knot residues 3 and 5 (Fig. 1*a*)[25]. In contrast, the putative binding sites for NGF are located in the hairpins loops located in the antiparallel ß-strand regions[2]. Possible evidence for the mode of receptor binding in NDP comes from mutations observed in Norrie disease patients. Eight out of 15 point mutations observed so far (data not shown) result in translational stops; four mutations affect cysteines involved in disulphide bridges. The other three mutations, V60Q, L61F and R90P are either located within the predicted α-helix region (R90P) or at the centre of the predicted anti-parallel ß-sheets and probably act by disrupting the protein fold or participating directly in NDP-receptor interactions.

Growth factors are involved in the pathogenesis of diseases with multifactorial origins such as cancers and autoimmune disease[26,27]. Growth factor gene mutations have also been implicated in the pathogenesis of phacomatoses like Proteus syndrome and multiple endocrine neoplasia[28]. A monogenic disease model has been established in mice: disruption of the TGFα gene by homologous recombination causes hair follicle and eye abnormalities similar to the waved-1 mutation[29]. With the identification of NDP as a putative peptide growth factor, a monogenic disease model becomes available for the family of cystine knot growth factors. Norrie disease is a neurodegenerative condition affecting the eyes and the brain with the phenotype mirrored by the expression pattern of the disease gene. The most prominent histological finding in this disease is a dysgenesis of ganglion cells and sensory cells of the retina[14]. The highly vascular retrolental membrane which is a consistent finding in Norrie disease is likely to be a secondary effect. Psychiatric disorders may develop later in life although marked intrafamiliar differences have been observed. Bilateral sensineuronal hearing loss can also occur and has been described in the absence of mental disturbances[30]. Nerve cells depend on their target cells for development and survival[31]. Loss of function of a factor involved in the organisation and maintenance of neurons in the retina as well as in the central nervous system provides a plausible explanation for the observed symptoms of the disease.

The proposed 3-D structures of NDP and of related C-terminal domains of extracellular proteins add new members to the expanding family of growth factors containing a cystine knot motif. Molecular details can be read from the model before the protein is expressed and its shape investigated experimentally. Animal models in combination with immunohistochemical studies will be used to investigate the cells excreting NDP as well as their corresponding target cells. Expression cloning will facilitate the identification of the NDP-receptor and recombinant proteins will be used to study NDP-receptor interactions. Whether NDP can function as a rescue factor in Norrie disease or in otherwise damaged retina will also be open to investigation[32]. For all these areas of future research, the 3-D NDP structure model will serve as a valuable guide.

Methodology

Sequence homology. Sequences were taken from SWISSPROT, PIR and EMBL sequence databases. Homology searches were done using the database PIR International (relase 36, Martinsried Institute for Protein sequences) and the programs FASTA[33] and ISSC[34]. The sequences were aligned to give maximum identity with a minimal number of gaps. From the alignment a consensus pattern based on 11 steric and physicochemical amino acid properties is derived[21]. The pattern reflects combinations of properties required at each position. Deviations from the pattern, given a candidate sequence,

Acknowledgements
We thank J.G. Sgouros for advice on sequence analysis, G. Vriend, L. Holm, R.C. Wade and B. Lorenz for help in model building and E. Auerswald for comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (T.M., A.M. & P.B.).

are scored in terms of the number of properties that deviate at a particular position (mismatches); insertions and deletions are also penalized[21]. The program PROPATS performs a sensitive database search with consensus patterns. Sequences with a low mismatch score are identified as containing the pattern, provided their scores can be distinguished from the random background of unrelated database sequences (many mismatches). With this procedure false positives are unlikely: if two unrelated families were artificially aligned, the subsequent database search would fail to separate the two input families from the unrelated database sequences.

Secondary structure prediction. The prediction method used[35] is a profile based neural network method. As the CT-family represents a rather divergent set of proteins, an alignment calculated previously (Fig. 1*b*) was used as input profile (for details see ref. 22). The program is available on e-mail (Internet: PredictProtein@Embl-Heidelberg.De, "help"). The experimentally derived TGFß structure was taken from the Brookhaven protein databank[36] (pre-release; 1TGI)

3-D model building. An initial 3-D model of NDP was built by replacing the side chains of the TGFß structure with those of NDP and optimizing the model by side chain rotamer searches using the program WHAT IF[37]. In a subsequent procedure energy minimization was carried out by the program SYBYL (version 6.0, TRIPOS associates Inc. St. Louis, MO, USA) using the AMBER forcefield[38]. MOLSCRIPT[39] was used for the ribbon drawing.

Received 14 June; accepted 24 August 1993.

1. Sporn, M.B. & Roberts, A.B. (eds) In *Peptide growth factors and their receptors* (Springer, Berlin 1990).
2. McDonald, N.Q. & Hendrickson, W.A. A structural superfamily of growth factors containing a cystine knot motif. *Cell* **73**, 421–424 (1993).
3. Angeletti, R.H. & Bradshaw, R.A. Nerve growth factor from mouse submaxillary gland: amino acid sequence. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2417–2420 (1971).
4. McDonald, N.Q. *et al.* New protein fold revealed by a 2,3-Å resolution crystal structure of nerve growth factor. *Nature* **354**, 411–414 (1991).
5. Schlunegger, M.P. & Grutter, M.G. An unusual feature revealed by the crystal structure at 2,2Å resolution of human transforming growth factor-ß2. *Nature* **358**, 430–434 (1992).
6. Daopin, S., Piez, K.A., Ogawa, Y. & Davies, D.R. Crystal structure of transforming growth factor ß2: An unusual fold for the superfamily. *Science* **257**, 369–373 (1992).
7. Oefner, C., D'Arcy, A., Winkler, F.K., Eggiman, B. & Hosang, M. Crystal structure of human platelet-derived growth factor BB. *EMBO* J. **11**, 3921–3926 (1992).
8. Swindells, M.B. Daopin, S., Cohen, G.H. & Davies, D. Structural similarity between transforming growth factor. ß2 and nerve growth factor *Science* **258**, 1160–1162 (1992).
9. Lin, L-F.H. *et al.* GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. *Science* **260**, 1130–1132 (1993).
10. Meindl, A. *et al.* Norrie disease is caused by mutations in an extracellular protein resembling C-terminal globular domain of mucins. *Nature Genet.* **2**, 139–143 (1992).
11. Berger, W. *et al.* Isolation of a candidate gene for Norrie disease by positional cloning. *Nature Genet.* **1**, 199–203 (1992).
12. Chen, Z-Y. *et al.* Isolation and characterisation of a candidate gene for Norrie disease. *Nature Genet.* **1**, 204–208 (1992).
13. Berger, W. *et al.* Mutations in the candidate gene for Norrie disease. *Hum. molec. Genet.* **1**, 461–465, (1992).
14. Warburg, M., Norrie's disease: A congenital progressive oculo-acousticocerebral degeneration. *Acta Ophthalmol.* **89**, 1–147 (1966).
15. Apple, D.J., Fishman, G.A. & Goldberg, M.F. Ocular histopathology of Norrie's disease. *Am. J. Ophthalmol.* **78**, 196–203 (1974).
16. Parsons, M.A., Curtis D., Blank C.E., Hughes H.N. & McCatney A.C.E. The ocular pathology of Norrie disease in a fetus of 11 weeks' gestational age. *Graefe's Arch. clin. exp. Ophthalmol.* **230**, 248–251 (1992).
17. Rothberg, J.M. & Artavanis-Tsakonas, S. Modularity of the slit protein: characterisation of a conserved carboxy-terminal sequence in secreted proteins and a motif implicated in extracellular protein interactions. *J. molec. Biol.* **227**, 367–370 (1992).
18. Mancuso, D.J. *et al.* Structure and gene for human von Willebrand factor. *J. biol. Chem.* **264**, 19514–19527 (1989).
19. Rothberg, J.M., Jacobs, J.R., Goodman, C.S. & Atavanis-Tsakonas, S. Slit: an extracellular protein necessary for development of midline glia and commisural axon pathways contains both EGF and LRR domains. *Genes Dev.* **4**, 2169–2187 (1990).
20. Bradham, D.M., Igarashi, A., Potter, R.L. & Grotendorst G.R. Connecting tissue growth factor: a cysteine-rich mitogen secreted by human vascular endothelial cells is related to the SRC-induced immediate early gene product CEF-10. *J. cell Biol.* **114**, 1285–1294 (1986).
21. Rohde, K. & Bork, P. A fast, sensitive pattern matching approach for protein sequences. *Comp. Appl. Biosci.* **6**, 183–189 (1993).
22. Rost, B. & Sander, C. Prediction of protein structure at better than 70% accuracy. *J. molec. Biol.* **232**, 584–599 (1993).
23. Voorberg, J. *et al.* Assembly and routing of von Willebrand factor variants: the requirements for disulphide-linked dimerization reside within the carboxy-terminal 151 amino acids. *J. cell Biol.* **113**, 195–205 (1991).
24. Tamaoki, H. *et al.* Solution conformation of endothelin determined by means of 1H-NMR spectroscopy and distance geometry calculations. *Protein Eng.* **4**, 509–518 (1991).
25. Qian, S.W. *et al.* Identification of a structural domain that distinguishes the actions of the type 1 and 2 isoforms of transforming growth factos ß on endothelial cells. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6290–6294 (1992).
26. Steel, C.M. Peptide regulatory factors and malignancy in *Peptide regulatory factors* p.121 (Edward Arnold, London, 1989).
27. Arend W.P. & Dayer, J.M. Cytokines and cytokine inhibitors or antagonists in rheumatoid arthritis. *Arthritis Rheum.* **33**, 305 (1990).
28. Kousseff, B.G. Sipple syndrome with lichen amyloidosis as a paracrinopathy. *Am. J. med. Genet.* **42**, 751–753 (1992).
29. Luetteke, N.C. *et al.* TGFa deficiency results in hair follicle and eye abnormalities in targeted and waved-1 mice. *Cell* **73**, 263–278 (1993).
30. Skevas, A., Kastanioudakis, I., Daniilidis, B. & Exarchakos, G. Norrie-Warburg Syndrom. *Laryngo-Rhino-Otol.* **71**, 534–536 (1992).
31. Mudge, A.W. Motor neurons find their factors. *Nature* **363**, 213 (1993).
32. LaVail M. *et al.* Multiple growth factors, cytokines and neurotrophins rescue photoreceptors form the damaging effects of constant light. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11249–11253 (1992).
33. Pearson, W.R., & Lipman, D.J. Improved tools for biological sequence comparison. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444–2448 (1988).
34. Argos, P. A sensitive procedure to compare amino acid sequences. *J. molec. Biol.* **193**, 385–391 (1988).
35. Rost, B., Schneider, R. & Sander, C. Progress in protein structure prediction? *Trends Biochem. Sci.* **18**, 120–123 (1993).
36. Bernstein, S.C. *et al.* The protein databank: computer based archival file for macromolecular studies. *J. molec. Biol.* **112**, 535–542 (1978).
37. Vriend, G. WHAT IF: a molecular modelling and drug design program *J. molec. Graphics* **8**, 52–56 (1990).
38. Weiner, S.J. *et al.* A new force field for molecular simulation of nucleic acids and proteins. *J. Am. chem. Soc.* **106**, 765–784 (1984).
39. Kraulis, P.J. MOLSCRIPT: a program to produce both detailed a schematic plots of protein structures. *J. appl. Crystallogr.* **24**, 946–950 (1991).
40. Gum, J.R. *et al.* The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. *J. biol. Chem.* **267**, 21375–21383 (1992).
41. Xu, G. *et al.* cDNA for the carboxyl-terminal region of a rat intestinal mucin-like peptide. *J. biol. Chem.* **267**, 5401–5407 (1992).
42. Bork, P. Mobile modules and motifs. *Curr. Op. struc. Biol.* **2**, 413–421 (1992).