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

Thomas Meitinger1, Alfonso Meinl2, Peer Bork3, Burkhard Rost4, Chris Sanders5, Martina Haussmann3 & Jan Murken1

The X-linked 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 α (TGFα). 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.

Peptideptide growth factors are involved in a range of trophic and plastic growth responses in eukaryotic cells. They include several subgroups identified on the basis of sequence similarity. A group of cystein-rich peptides with a common three-dimensional protomer structure includes nerve growth factor (NGF), placental 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, the availability of the three-dimensional (3-D) structure now greatly facilitates detailed structural and functional studies. The crystal structure of the marine NGF-dimer revealed a characteristic 3-D fold that has since been observed in other growth factors with low sequence similarity, including TGFα (refs 3,4) and PDGF-B. The common structural core of these proteins is formed by an equivalent pattern of three disulphide bridges. In this motif, a ring is formed by two disulphide bridges. A third disulphide passes through the centre of the ring (cystine knot motif1; see Fig. 1). Members of this family of proteins have 100–130 amino acid 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 which bind to specific receptors. However, the mode of dimerization is completely different for each of the three families. The most recently described member of the family of dimeric transforming growth factor-related proteins appears to be a gland cell line-derived neurotrophic factor (GDNF). We have recently described the structure of a gene which codes for a cystein-rich extracellular protein with sequence length of 131 amino acids2. Loss of function mutations in this gene cause Norrie disease, an X-linked disorder characterized by blindness, deafness and mental retardation. The gene was identified by positional cloning2 and intragenic deletions as well as point mutations have been demonstrated in patients with Norrie disease3. Clinical and morphological examination suggests the involvement of growth or angiogenic factors in the pathogenesis of the disease whose ocular manifestations are characterized by degeneration of retinal sensory cells and a highly vascular retrolental membrane4. The C-terminal cystein-rich domain (CT domain) of NDP shows homology to the C-termini of functional and structurally diverse extracellular modular proteins5,6. The closest homologues are several extracellular mucins and von Willebrand factor (vWF). The von Willebrand A4 domain, involved in development of midline glia and commissural axon pathways7, and a family of growth regulators including c-kit ligand, not and connective tissue growth factor (CTGF), 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 TGFα, 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 TGFβ and NTF. a. Schematic representation of a TGFβ-like monomer modified after rats 2 and 8. F-strands are drawn as arrows and disulfide bridges are indicated. The loop between F-strand 82 and 83 shows variable length and crystal structure in TGFβ, NTF and PDGF (V, 4). Multiple alignment of NTD and related CT-domains with TGFβ and NTF. Rat and human mRNA sequences were added to the CT-domain alignment (G - mouse, A - TGFβ, T - mouse, M - as B80-885). The structure-based alignment of TGFβ with NTF was taken from ref. 4. All secondary structures are shown in the cysteine residues numbered 1-6 and a-d. Insertion/deletions are indicated by dots. The cysteine forming an intermolecular disulfide bridge in TGFβ is marked by a star. Single contacts ( *) and hydrophobic core contacts ( headlines) in TGFβ are conserved with corresponding positions in the CT-domains. Two inserts in the NTF sequence have been deleted (GKLETVVLAENV). b. Alignment of cysteine and glycine residues in TGFβ with those in NTF. The number of amino acids (X) between conserved cysteine residues (C) is given. Cysteines involved in the cysteine knot are numbered 1-6. The additional cysteines conserved in the CT domain are numbered 1-6. A star indicates the cysteine residue involved in dimerization. c. Comparison between observed F-strands and helices in TGFβ and predicted 3D structures of TGFβ and NTF (C - crystal structure; PHD, prediction from sequence family, E - C-strand, H - helix, G - 3-strand, 3-D contact). NTF is predicted to consist predominantly of 9-11 cys, with second element roughly in agreement with those experimentally conserved in TGFβ. The central predicted F-strand (sequence NFGQSTVS in NTF) is actually a helix in TGFβ, involved in dimerization and is likely to be present in NTF; 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 significant sequence homology (Fig. 1). The highest identity and similarity values are found between NTD and the human interleukin-1 mu, MUC2, with values of 30% and 40%, respectively. Between TGFβ and related sequences with known 3-D structures, the percentage figures do not exceed 5% for identity and 25% for similarity. When comparing TGFβ with any of the NDP related sequences, similar non-significant values are obtained. However, a characteristic pattern emerges between TGFβ and NDP related sequences which only cysteines and the distances between them are compared (Fig. 2). The alignment of the NDP primary sequence to the known structure of TGFβ reveals several conserved features: (1) Cys residues at the right spacing match each of the disulfide bridges in the common structural core of the TGFβ/NTF/CT domain family. Only a few insertions and deletions have to be introduced to align TGFβ and NDP (residues in brackets). In particular, the keto- and aminoacids 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 TGFβ, is present at the equivalent position in NDP (Cys95 in NDP). (iii) A structurally important glycine with a positive phi angle, serine and conserved, is conserved in TGFβ, NDP 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 TGFβ related proteins with similar 3-D structures (Fig. 1a). 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 TGFβ can be confirmed with consensus sequence patterns derived from the alignment of CT-domains and TGFβ related sequences (data not shown). Two different consensus patterns were calculated from an alignment of all CT family members with 50 TGFβ homologs. (i) A pattern which only specifies the invariant cysteine and glycine residues conserved in the TGF family and CT-domains (Figs. 1a and 2) 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
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 TGFβ (Fig. 1A). Predicted β-stands and loop regions between the residues forming a cysteine knot of the CT domain are compatible with the observed β-strands common to TGFβ 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.

3-D similarity between NDP and TGFβ
From the alignment of NDP with TGFβ, a full 3-D model of NDP protein was built by replacing the side chains in the TGFβ structure with those of NDP. After optimization, the final model (Fig. 2B) reveals other features that support the physiological relevance of the model structure. First, two additional disulfides 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 TGFβ-based NDP model. This is an trivial consequence of mapping the NDP residues onto the 3-D structure of TGFβ, the probability for such a constellation to occur by chance is very low. Second, a spatially compact network of hydrophobic residues at the hinge end of TGFβ (top right in Fig. 2B) links the four β-strands. Almost all of these residues (9 out of 10, marked “h” in Fig. 1B) retain their hydrophobic cluster in the 3-D model. Third, the dimer interface of TGFβ is formed primarily by one face of the β-sheet (the palm of a slightly curved hand) interacting with the helix. Residues in this interface (marked “d” in Fig. 1B) remain predominantly hydrophobic in the NDP and CT sequences (7 out of 8), whereas the stem of β-sheets 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 disulfide link in the TGFβ dimer.

Discussion
The structural similarity between NDP and TGFβ is supported by comparing disulfide patterns, sequence pattern searches, full atom model building for alignment, and conservation mapping. Our 3-D model of NDP is based on the known TGFβ crystal structure. Although the model is in agreement with a cysteine knotlike structure, it has to be considered as a first approximation and 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 recognizable more similar to TGFβ than to NGF and to PDGF. Not only is the intermolecular disulfide bridge linking the two monomers predicted, but also the head-to-tail orientation that is characteristic of TGFβ, and different from that of NGF. Such an orientation is supported by the presence of a sequence insertion common only to NDP and TGFβ (and not present in NGF or PDGF) which corresponds to a helix involved in dimer interface contacts. The dimer...
contains in TGFβ are made in part by large hydrophobic residues on the surface of the A-sheet and the B-helix, partly conserved in the sequence of NDP. The details of dimerisation may, of course, turn out to be different.

By sequence similarity, NDP belongs to a family of domains which also occur at the C-terminals of a class of extracellular modular proteins that includes VWF, several mucins, SM and a family of growth factors related to congenital tissue growth factor. The V-D3 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 dimerisation is likely to be valid at least for the closer relatives of NDP such as mucin and VWF. Remarkably, dimerisation of VWF has been shown experimentally to occur via disulphide linkage within the C-terminal domain, consistent with our hypothesis. The production of the crucial role of Cys276 in VWF (or equivalent residues in the other CT-modules) in dimerisation provides a precise and testable hypothesis: mutants of NDP or VWF in these cysteines should be unable to dimerise.

Further examples of peptides with eight membered rings formed by two disulfide connected backbone include endotoxins and neurotoxins. All these peptides have in common that they affect their target cells through surface receptors and that they proceed from the extracellular domains of the receptors. The suggestion has been made that the receptor specificity has been associated with sequence alterations in the B-helix between cysteine mutant residues 3 and 5 (Fig. 4). In contrast, the putative binding sites for NDP are located in the hairpin loop located in the antiparallel B-strand region. Possible evidence for the mode of receptor binding in NDP comes from mutagenesis, observed in Norrie disease patients. Eight out of 10 point mutations observed so far (data not shown) result in translational stop, four mutations alter cysteine involved disulfide bridges. The other three mutations, V64Q, L61F and R89P are either located within the predicted B-helix region (R89P) or at the centre of the predicted antiparallel B-sheet 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 cancer and autoimmune disease. Growth factor gene mutations have also been implicated in the pathogenesis of diseases like protein syndrome and multiple

inducing neoplasia. A monogenic disease model has been established in mice, disruption of the TGFβ gene by hemizygous recombination causes hair follicle and eye abnormalities similar to the waved mutation. With the identification of NDP as a putative peptide growth factor, a monogenic disease model becomes available for the family of cystoerect growth factors. Norrie disease is a neurodegenerative condition affecting the eye and the brain with the phenotype mimicked by the expression pattern of the disease gene. The most prominent histological finding in this disease is degeneration of retinal ganglion cells and sensory cells of the retina. The highly vascular remodelling of the retina is similar to the secondary effect of psychiatric disorders may develop later in life although marked intrasinular differences have been observed. Bilateral neuronal loss can occur and has been described in the absence of mental disturbances. Nerve cells depend on their target cells for development and survival. 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 CT-domains of extracellular proteins, and new members to the expanding family of growth factors containing a triple helix motif. Molecular details can be read from the model below the protein, a secondary effect on the retina and its shape investigated experimentally. Animal models in combination with immunohistochemical staining are to be used to investigate the cells expressing NDP as well as their corresponding target cells. Expression and/or identification of the inhibitor 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 whether other treatment of the retina will also be open to investigation. For all these areas of future research, the 3-D NDP structure model will serve as a valuable guide.
are stored in terms of the number of properties that define each epitope on a particular protein (antigen). In the future, such decisions will also be automated. The program PROACT is a service to database search with a number of pattern matching methods. The range of logic and semantic scores is used to identify matching patterns. A threshold is then selected to separate the relevant from the non-relevant sequences. In many instances, all the data produced are false positives, and if two unrelated sequences are identical or nearly identical, the close match score would fail to separate the two different sequences from the unrelated database sequences.

Secondary structure prediction. The second model method used is a profile based neural network method. As the CLUST family represents a rather large set of proteins, a alignment of all sequences previously

Received 14 June; accepted 24 August 1989.