Molecular modelling and experimental studies of mutation and cell-adhesion sites in the fibronectin type III and whey acidic protein domains of human anosmin-1

Alexis ROBERTSON*, Gavin S. MacCOLL*, Julia A. B. NASH†, Mark K. BOEHM‡, Stephen J. PERKINS‡¹ and Pierre-Marc G. BOULOUX*

*Department of Medicine, Royal Free Campus, Royal Free and University College Medical School, University College London, Rowland Hill Street, London NW3 2PF, U.K., †MRC Centre for Developmental Neurobiology, Fourth Floor, New Hunt's House, Kings's College London, Guy's Campus, London Bridge, London SE1 1UL, U.K., and ‡Department of Biochemistry and Molecular Biology, Royal Free Campus, Royal Free and University College Medical School, University College London, Rowland Hill Street, London NW3 2PF, U.K.

Anosmin-1, the gene product of the *KAL* gene, is implicated in the pathogenesis of X-linked Kallmann's syndrome. Anosmin-1 protein expression is restricted to the basement membrane and interstitial matrix of tissues affected in this syndrome during development. The anosmin-1 sequence indicates an N-terminal cysteine-rich domain, a whey acidic protein (WAP) domain, four fibronectin type III (FnIII) domains and a C-terminal histidine-rich region, and shows similarity with cell-adhesion molecules, such as neural cell-adhesion molecule, TAG-1 and L1. We investigated the structural and functional significance of three loss-of-function missense mutations of anosmin-1 using comparative modelling of the four FnIII and the WAP domains based on known NMR and crystal structures. Three missense mutation-encoded amino acid substitutions, N267K, E514K and F517L, were mapped to structurally defined positions on the GFCC' β -sheet face of the first and third FnIII domains. Electrostatic maps demonstrated large basic surfaces containing clusters of conserved predicted heparan sulphate-binding residues adjacent to these mutation sites. To examine these modelling results anosmin-1 was expressed in insect cells. The incorporation

of the three mutations into recombinant anosmin-1 had no effect on its secretion. The removal of two dibasic motifs that may constitute potential physiological cleavage sites for anosmin-1 had no effect on cleavage. Peptides based on the anosmin-1 sequences R254–K285 and P504–K527 were then synthesized in order to assess the effect of the three mutations on cellular adhesion, using cell lines that represented potential functional targets of anosmin-1. Peptides (10 μ g/ml) incorporating the N267K and E514K substitutions promoted enhanced adhesion to 13.S.1.24 rat olfactory epithelial cells and canine MDCK1 kidney epithelial cells ($P < 0.01$) compared with the wild-type peptides. This result was attributed to the introduction of a lysine residue adjacent to the large basic surfaces. We predict that two of the three missense mutants increase the binding of anosmin-1 to an extracellular target, possibly by enhancing heparan sulphate binding, and that this critically affects the function of anosmin-1.

Key words: electrostatic maps, heparan sulphate, Kallmann's syndrome, multidomain proteins, sequence alignment.

INTRODUCTION

X-linked Kallmann's syndrome is a developmental disorder characterized by anosmia and hypogonadotrophic hypogonadism. The olfactory disturbance results from defective olfactory bulb development and hypogonadotrophic hypogonadism from failed migration of gonadotropin-releasing hormone neurons from the medial olfactory placode into the septo-preoptic area of the anterior hypothalamus. The latter results in functional gonadotropin-releasing hormone secretory deficiency [1]. Other symptoms of X-linked Kallmann's syndrome, but with variable penetrance, include bimanual synkinesis and both unilateral and bilateral renal agenesis [2–4], suggesting a fundamental role of a gene product in the organization of the corticospinal pathways and development of the kidneys respectively. Positional cloning has led to the identification of *KAL*, localized to Xp22.3 [5], as the causative gene [6,7]. It spans a 220 kb region of genomic DNA and consists of 14 exons [8]. The cDNA contains a 2.1 kb coding sequence flanked by 5'- and

3'-untranslated regions encoding a 680-residue glycoprotein, anosmin-1 (Figure 1). *KAL* transcripts have been identified in the developing olfactory bulb and metanephros, as well as in the neuroretina and spinal column [9]. Anosmin-1 is secreted and has been localized to basement membranes and interstitial matrices [10] within these tissues. Anosmin-1 is believed to be a local mediator of cell–cell or cell–matrix interactions, possibly acting as a growth/guidance factor during nephrogenesis and olfactory bulb development.

The protein sequence of anosmin-1 comprises: an N-terminal cysteine-rich domain, followed by a whey acidic protein (WAP) domain, four fibronectin type III (FnIII) domains and a C-terminal histidine-rich region [6–8,11] (Figure 1). The sequences of human, chicken (73 % similarity) and quail (74 % similarity) anosmin-1 are closely related. The WAP domain shows similarities to the structure of small protease inhibitors [12,13]. The FnIII domains shows the highest degree of similarities to those of the neural cell-adhesion molecule (N-CAM) protein family, including membrane-bound proteins such as TAG-1 and L1

Abbreviations used: FnIII, fibronectin type III; PDB, Protein Data Bank; WAP, whey acidic protein; N-CAM, neural cell-adhesion molecule; MDCK1, Madin–Darby canine kidney; N267K etc., mutation of Asn²⁶⁷ to Lys etc.

¹ To whom correspondence should be addressed (e-mail s.perkins@rfc.ucl.ac.uk).

Figure 1 The exonic and domain structure of anosmin-1
(a) Exonic structure of KAL cDNA of anosmin-1. Mutation sites are marked by lettered symbols (see Table 1). The domain structure is indicated by boxes, where SP denotes the signal peptide. The positions of the peptide sequences 32R1 and 24III are indicated by solid bars. (b) Domain structure of anosmin-1, in which the mature protein consists of A24-Y680, and each domain is represented by a circle. Residue numbering indicates the size of each domain. The location of six putative N-glycosylation sites are indicated by numbered Y symbols. Cys box, cysteine-rich region; H, histidine-rich region.

[14,15]. It is N-glycosylated at six putative sites. The anosmin-1 sequence also contains a lipid-soluble signal peptide, but no transmembrane or glycosyl phosphatidylinositol anchorage residues. Anosmin-1 has previously been expressed in mammalian cell culture systems to give a secreted cell surface-associated mature monomeric 100 kDa protein [16–18]. *In vitro* cleavage generated 55 kDa and 45 kDa surface-bound fragments [17]. The identification of disease-causing mutations frequently provides insight into protein function. Hitherto, 22 anosmin-1 mutations have been described, 18 of which are shown in Table 1 [19–25]. However, only three loss-of-function missense

Figure 2 Schematic diagram of the seven-stranded β -sheet sandwich structure of a FnIII domain
The β -strands G, F, C and C' are labelled on the upper face as shown, while the β -strands A, B and E are on the lower face.

mutations have been described, all of which result in an identical phenotype.
Difficulties in the expression of anosmin-1 have resulted in only a limited set of functional studies [16–18], and no three-dimensional structural studies have been undertaken. However, crystal structures of FnIII domains structurally homologous to those in anosmin-1 are now available. The FnIII domain is related to the constant-type immunoglobulin fold and consists of a seven-stranded β -sheet sandwich with a β -strand ABE-GFCC' topology (Figure 2). FnIII domains occur in a number of cell-adhesion proteins, such as N-CAM and neural glial cell-adhesion molecule. A total of 23 atomic structures for FnIII domains in neuroglial [26], fibronectin [27–29], tissue factor [30–32], tenascin [33], erythropoietin [34], human growth hormone [35] and granulocyte colony-stimulating factor receptor [36] includes crystal structures for two-domain and four-domain FnIII arrangements, and enables the homology modelling of the four FnIII domains within anosmin-1. The WAP domain contains a core of four disulphide bridges that is found in small inhibitory proteins, in which the positions of the eight cysteine residues are conserved. Crystal and NMR structures are available for mod-

Table 1 Position and nature of the known mutations of the KAL gene

See Figure 1 for more details. In k, l and m the deleted nucleotides are indicated in lower case. A–Q correspond to a–q in Figure 1.

Key	Location	Mutation type	Nucleotide change	Amino acid change	Reference
A	Residue 267	Missense	T to A	Asn to Lys	[20]
B	Residue 517	Missense	C to G	Phe to Leu	[21]
C	Residue 514	Missense	G to A	Glu to Lys	[22]
D	Residue 423	Stop	C to T	–	[20]
E	Residue 237	Stop	G to A	–	[19]
F	Residue 257	Stop	C to T	–	[19]
	Residue 258	Stop	G to A	–	[19]
G	Residue 328	Stop	C to G	–	[21]
H	Residue 421	Stop	C to A	–	[20]
I	Exons 1–3	Deletion	–	–	[23]
J	Exons 3–5	Deletion	–	–	[22]
K	Residue 326	Deletion	CATTAcAAGGT	–	[20]
L	Residue 463	Deletion	TTTCCTgaagcgtgtgccACAA	–	[21]
M	Residue 276	Deletion	TGCCcCCAG	–	[19]
N	Nucleotide 1166	Insertion	A	–	[20]
O	Intron 5	Splice site	GT to TG	–	[24]
P	Intron 12	Splice site	–	–	[20]
Q	Nucleotides 241–251	Insertion	Duplication of CTGCGCGGCGG	–	[25]

and quail sequences have accession numbers of P33005 and A47222 respectively [6,8,11]. In order to provide a reliable three-dimensional template for the modelling of the anosmin-1 FnIII domains that is not dependent on sequence similarities in the FnIII domains, 23 FnIII sequences were aligned using 12 crystal and NMR structures from the Protein Data Bank (PDB). In the order shown in Figure 3, the 12 PDB codes were as follows: the 1CFB structure from *Drosophila* neuroglian [26], the 1FNF, 1TTG, 1TTF and 1FNA structures from human fibronectin [27–29], the 2HFT, 1BOY and 1DAN structures from human tissue factor [30–32], the 1TEN structure from tenascin [33], the 1EBP structure from erythropoietin [34], the 3HHR structure from human growth hormone [35], and the 1GCF structure from murine granulocyte colony-stimulating factor receptor [36]. In the 23 FnIII structures, the seven β -strands A, B, C, C', E, F and G in the secondary structure of each one (Figure 2) were identified using the DSSP program of Kabsch and Sander [42]. This identification of the seven β -strands in each FnIII structure was used to superimpose all 23 structures using INSIGHT II molecular modelling software (MSI, San Diego, CA, U.S.A.) with the HOMOLGY module for the comparative modelling

[illegible]

Figure 3 Comparison of the human anosmin-1 FnIII sequences with a structural analysis of 23 FnIII domains

Each FnIII structure is identified by its PDB code. The consensus position of the seven β -strands are specified as A, B, C, C', E, F and G in arrowed ranges. The alignment of the 23 FnIII sequences is based on the topological equivalence of the residues according to molecular graphics. The 23 secondary structures identified using DSSP are denoted by E (β -strand), H (α -helix), B (β -ladder), G (3_{10} -helix), T and S (loop). The 23 surface accessibility analyses of residues are denoted by 0–9 to correspond with accessibilities in the range of 0–9%, 10–19%, and so on up to 90–99%. The underlined sequences were not modelled. Asterisks indicate the 14 hydrophobic residues that were conserved to better than 65% in the alignment.

Table 2 The defined and searched loops used to construct the anosmin-1 FnIII homology models

Numbers in parentheses denote the total of residues changed.

Anosmin-1 domain	Residues	A to B	B to C	C to C'	C' to E	E to F	F to G
FnIII-1	97	1FNA (3)	1CFB (10)	1FNF (7)	Search (9)	1FNA (7)	1FNA (9)
FnIII-2	102	Search (8)	1CFB (10)	1FNF (7)	Search (9)	1FNA (7)	1FNA (9)
FnIII-3	97	Search (7)	1FNF (8)	3HHR (6)	Search (9)	1FNA (7)	1FNA (9)
FnIII-4	104	Search (6)	1EBP (11)	Search (10)	Search (10)	1FNA (7)	1FNA (9)

of protein structures starting from a known structure on INDY workstations (Silicon Graphics, Reading, Berks., U.K.). This procedure permitted the identification of topologically equivalent residues. In addition, the Lee and Richards method [43] in the COMPARER program was used to calculate the side chain solvent accessibility with a probe diameter of 0.14 nm. The use of these secondary structures, topologically equivalent residues and solvent accessibilities defined the sequence alignment of Figure 3 with the four FnIII domains from anosmin-1. The identity of each alignment of two sequences was defined as (the number of identical residues) $\times$ 100/(the total number of topologically equivalent positions in the alignment, excluding gaps in either sequence).

Construction of the anosmin-1 FnIII homology models

Homology models for the four FnIII domains of anosmin-1 (denoted FnIII-1–4) were constructed using the BIOPOLYMER, HOMOLOGY and DISCOVER modules of INSIGHT II. BIOPOLYMER was used to append additional structures to the models, whereas DISCOVER was used to perform energy minimization. The structurally conserved regions in these four domains from anosmin-1 were defined as the seven β -strands A, B, C, C', E, F and G in the crystal structure of the tenth FnIII domain of fibronectin (PDB code 1FNA). In the FnIII-1 model, structures for five of the six loops connecting these β -strands were taken from the equivalent loops from the 1FNA, 1CFB and 1FNF crystal structures, which were defined by the preservation of the loop length (Table 2). In the FnIII-2 and FnIII-3 models, four loops were similarly defined based on the 1FNA, 1CFB, 1FNF and 3HHR crystal structures. In the FnIII-4 model, two loops were defined based on the 1EBP and 1FNA crystal structures. The remaining loops were modelled using searches of a database of peptide fragments. A precalculated C α distance matrix identified the PDB loops that best fitted the corresponding C α distance matrix calculated from the structurally conserved regions of the FnIII framework that defined the starts and ends of these searched loops for a specified number of flanking and intervening residues. Side chain atoms were automatically generated for both the structurally conserved regions and structurally variable regions using the template structures and general rules for residue exchanges. The anosmin-1 homology models were refined using energy minimization, where 300 steps of steepest descent minimization were performed using DISCOVER. The energy refinements significantly improved the covalent geometry at the splice junctions with the searched loops and the bad contacts of each FnIII model as assessed by the PROCHECK program [44].

Construction of the anosmin-1 WAP homology model

The WAP domain of anosmin-1 was modelled using a similar procedure based on structural analyses of the WAP domain in the crystal structure (PDB code 1FLE) [37] and two NMR

structures (PDB codes 1REL and 2REL) [38] for elafin. The structurally conserved regions of the anosmin-1 WAP domain were defined as residues 134–140 and residues 151–172. The remaining residues 141–150 were identified as a searched loop from the PDB database of protein fragments. After replacement of the elafin side chains with those of anosmin-1, the disulphide bridge connectivity of cysteine-147–cysteine-168 was made using BIOPOLYMER, and the anosmin-1 WAP model was subjected to energy refinement and verified by an assessment with PROCHECK.

Calculation of electrostatic surfaces

The electrostatic surface of each model was calculated using the DELPHI module of INSIGHT II. A full Coulombic boundary condition was used, with the interior dielectric constant set as two and the exterior one set to 80. The solvent-accessible surface of each model was displayed using the Connolly algorithm with a solvent probe of diameter 0.14 nm. The surface was coloured red for potentials less than -5 kT (acidic), blue for potentials greater than $+5$ kT (basic) and white for neutral potentials of 0 kT. Colours between these values were produced using a linear interpolation.

Expression of wild-type and mutated anosmin-1 in an insect cell system

For expression in insect cells, a *KAL* cDNA lacking the signal peptide sequence and the terminal stop codon sequence was cloned in frame into the *Bgl*II and *Eco*RV sites of the vector pMT/BiP/V5-His (Invitrogen San Diego, CA, U.S.A.). This created a clone, pKD1, which consisted of an inducible metallothionein promoter [45], an insect signal sequence (BiP), the *KAL* cDNA and sequences coding for both the V5 and six histidine epitopes (V5-His). The K³⁴⁵KKRRK³⁵⁰ motif was excised from the clone pKD1 by site-directed mutagenesis, using the Quik-Change kit and protocol (Stratagene), to create a second clone pKDM2. Site-directed mutagenesis was also carried out on the pKD1 clone using four sets of mutagenic primers to generate the clones pKD1-R423G, pKD1-M267, pKD1-M514 and pKD1-M517 encoding anosmin-1 that incorporated the R423G, N267K, R514K and F517L amino acid substitutions respectively. All clones were verified by sequencing. The pKD1, pKDM2, pKD1-R423G, pKD1-M267, pKD1-M514 and pKD1-M517 clones and a positive control vector pMT/BiP/V5-His/GFP (Invitrogen), containing the cDNA for green fluorescent protein, were transiently transfected into Schneider S2 *Drosophila melanogaster* cells [46] by a calcium phosphate method (Invitrogen). The cultures were grown for 48 h before recombinant protein synthesis was induced by addition of CuSO₄ to a final concentration of 500 μ M. The media were harvested 24 h post induction.

For Western-blot analyses, samples from the cell culture medium from each transfection were run on Nu-PAGE gels

(Invitrogen) at 200 V for 35 min. Protein transfer on to nitrocellulose was done at 30 V for 1 h using Novex transfer buffer. The blot was blocked in 5% fat-free milk powder in PBS-Tween (0.2 mM KCl, 0.15 mM KH_2PO_4 , 0.8 mM Na_2HPO_4 , 137 mM NaCl, 0.1% Tween 20 detergent, pH 7.2) for 3 h, washed twice in PBS-Tween for 5 min and once for 15 min, before an overnight incubation with 0.2 $\mu\text{g}/\text{ml}$ Qiagen tetraHis primary antibody (Qiagen) in 3% BSA. The blots were washed twice in PBS-Tween for 5 min and once for 15 min. The blot was then incubated in an anti-mouse horseradish peroxidase conjugate secondary antibody (1:20000 dilution; Sigma) in 5% milk powder and PBS-Tween for 1 h. The blots were washed twice in PBS-Tween for 5 min and once for 15 min, and were visualized with a kit protocol using ECL[®] reagents (Amersham International).

Peptide adhesion assays using neuronal and kidney cell lines

All media and fetal bovine serum were obtained from Gibco Life Technologies. Antibiotic/antimycotic (100 $\times$) and G418 were obtained from Sigma. The 13.S.1.24 cell line [47] was maintained in the following medium: Dulbecco's modified Eagle's medium/Glutamax, 10% fetal bovine serum, 1 $\times$ antibiotic/antimycotic and 0.3 mg/ml G418. The Madin-Darby canine kidney 1 (MDCK1) cell line was maintained in Dulbecco's modified Eagle's medium/Nut F-12, 10% foetal bovine serum and 1 $\times$ antibiotic/antimycotic. Peptides were synthesized at > 80% purity by Genosys, Cambridge, U.K. The 32R1-WT peptide corresponded to the anosmin-1 sequence R²⁵⁴PSRWYQFRVAAVNVHGTRGFTAPSKHFRSSK²⁸⁵. The 32R1-Mut peptide incorporated the asparagine-to-lysine substitution at residue 267 [20]. The 24III-WT peptide corresponded to the anosmin-1 sequence P⁵⁰⁴IRPKSHSKAEAVFF-TTPPCSALK⁵²⁷. The 24III-M514 and 24III-M517 peptides corresponded to the same sequence, but contained the amino acids substitutions glutamate to lysine at position 514, and phenylalanine to leucine at position 517 respectively.

The cell-adhesion assays utilized 96-well Immulon 4HBX microtitre plates (Thermo Labsystems, Helsinki, Finland). The plates were coated with the 32R1 or 24III peptides at a concentration of 10 $\mu\text{g}/\text{ml}$, where 150 μl aliquots were placed into the wells, and the plates were incubated at 37 °C overnight. The wells were washed three times with 200 μl of PBS, then 150 μl of 1% BSA in PBS was added to each well and the plate was incubated for 1 h at 37 °C, then the wells were washed for a final three times with 200 μl of PBS. Poly(L-lysine) (100 $\mu\text{g}/\text{ml}$) was used as a positive control. The peptide assays were performed using a modification of the procedures described by Soussi-Yanicostas et al. [18] and by Bauvois et al. [48]. Cells were harvested with the non-proteolytic cell-dispersal agent NoZyme (JRH Biosciences, Woodland, CA, U.S.A.) and were resuspended at a density of 3.35×10^5 cells/ml in 0.1% BSA and the appropriate serum-free medium above. Aliquots of 150 μl (5×10^4 cells) were placed into the wells and incubated at 37 °C for 1 h to allow the cells to adhere. The wells were aspirated and washed once with 200 μl of PBS to remove unbound cells. A portion of 150 μl of Rose Bengal stain (0.25% in PBS; Sigma) was added to each well and incubated for 5 min. The stain was aspirated and the wells washed twice with 200 μl of PBS. The stain was solubilized with 150 μl of 1:1 ethanol/PBS (30 min at room temperature). Quantification of the attached cells was achieved by absorbance readings at 550 nm. Assays were performed in quadruplet and repeated in three experiments. Absolute values were subjected to one-way ANOVA and a post hoc

Bonferroni test. The results for each cell type/substrate interaction were expressed as a percentage of the poly(L-lysine) value, used as a positive control, which was set as 100%.

RESULTS

Sequence alignment of known FnIII structures with the anosmin-1 FnIII domains

The modelling of the FnIII domains of anosmin-1 required a sequence alignment of these domains with known atomic structures for the FnIII domain. As 23 FnIII domain structures were known from 12 crystal and NMR structure determinations, a sequence alignment based on the topological identity of FnIII residues was derived from the superimposition of these 23 FnIII structures. This approach was more accurate than one based on identifying residue similarities using sequence-alignment algorithms [39], and has previously been used to superimpose the FnIII domains in fibronectin [29]. Accordingly, the β -sheet secondary structures in these 23 structures were quantitatively determined using the DSSP program. The centre section of Figure 3 shows that, even though the start and end of the seven core β -strands (denoted by E symbols) varied in position by up to three residues, all seven β -strands were readily identified in similar positions in the 23 structures. The solvent accessibilities of residues in these structures were calculated using the COMPARE program. The bottom section of Figure 3 shows that the positions of buried residues (denoted by 0 or 1 numbers) were consistently maintained throughout the alignment of the 23 structures, in particular in the β -strands where there is frequently an alternation of exposed and buried side chains. The resulting alignment allowed the structurally conserved region of each of the seven β -strands to be defined, and their ranges are marked as A–G above each alignment. The root-mean-square difference in position of the 49 α -carbon atoms in these β -strands was 0.16 ± 0.04 nm for 20 of the 22 co-ordinate sets relative to that of the tenth FnIII domain of fibronectin (PDB code 1FNA). The two exceptions were two structures for the first FnIII domain of growth hormone (PDB code 3HHR) in which two β -strands were displaced from the average position observed in other FnIII domains. This root-mean-square difference compared well with that reported in other modellings ([41] and references therein).

The sequence alignment in the upper section of Figure 3 shows that only one residue (tyrosine-69) was completely conserved in all 23 FnIII sequences. It was notable that the aliphatic residues were well aligned, as these have previously been noted to be important for the FnIII structure [40]. Within the length of 97 residues in the alignment, 14 hydrophobic residues were conserved to better than 65% (asterisked in Figure 3). These included the highly conserved residues tryptophan-22, tyrosine-33, leucine-63 and tyrosine-69 in the numbering of Figure 3 [29]. No hydrophilic residues were significantly conserved. The surface accessibility analysis of Figure 3 shows that these conserved residues are generally in buried locations. Despite a low level of sequence similarity in this superfamily, the 23 FnIII structures could be readily superimposed using the conserved β -strand residues, and this enabled the topological conservation of residues and their alignment to be established. In turn, this permitted the alignment of the FnIII domains in anosmin-1 with the 23 FnIII structures. The top of Figure 3 showed that this alignment was achieved using the conserved hydrophobic residues. The lengths of the four anosmin-1 FnIII sequences were close to those of the 23 FnIII structures, with minor differences being observed only in the loops connecting the seven β -strands.

Figure 4 Ribbon representations of the four homology models for the FnIII domains of anosmin-1

The seven conserved β -strands are represented as blue ribbons, with the ABE face shown on the left and the GFCC' face shown on the right in each model. The loop regions are represented as orange ribbons. The α -carbon atoms of the three missense mutations are denoted as red spheres (N267K, E514K and F517L). The α -carbon atoms of cysteine residues are denoted as yellow spheres (C368, C496-C523 and C467). The five predicted N-linked glycosylation sites are indicated as magenta spheres at the α -carbon positions of N209, N300, N470, N553 and N564. Two of the predicted heparan sulphate proteoglycan-binding sites are indicated as green spheres ($S^{278}KHFRS^{283}$ on β -strand G of FnIII-1 and the final two residues of $Q^{421}RRRPTRP^{428}$ on β -strand A of FnIII-3). The predicted cleavage site $K^{345}KKRRK^{350}$ is indicated as grey spheres on β -strand C' of FnIII-2. Another predicted cleavage site $R^{422}RR^{424}$ is located on the linker between FnIII-2 and FnIII-3, which is not modelled.

Homology modelling of the FnIII domains in anosmin-1

The sequence alignment of Figure 3 was used to construct homology models for the four anosmin-1 domains. For this, the crystal structure of the tenth FnIII domain of fibronectin at a resolution of 0.19 nm (PDB code 1FNA) [28] was used as the template to model the conserved β -strands in the FnIII domains of anosmin-1. This structure was chosen for reason of the high quality of this structure determination and the good correspondence of its sequence length with those of the four anosmin-1 FnIII domains, especially in the region of the three amino acid substitutions to be modelled (Figure 3). The sequence identity is low (6–14%), which is typical for the FnIII superfamily. The six loops joining the seven β -strands were remodelled using either the corresponding loop from other FnIII domains or database searches for suitable loops (Table 2). The resulting homology models for the four FnIII domains of anosmin-1 are shown in Figure 4. DSSP and COMPARER analyses of the models showed that their β -sheet structures and side chain accessibilities were as expected from the crystal structures (results not shown). The models were analysed using the PROCHECK program. In the four FnIII models, only five residues out of 400 were in disallowed regions of the Ramachandran plot of main chain dihedral angles, and only six proline side chains out of 400 side chains had unfavourable dihedral conformations. This outcome was comparable with that obtained by a PROCHECK analysis of the template crystal structure, which supported the validity of the four FnIII models.

The domain models were consistent with their expected structural properties. In the FnIII domain superfamily, and also in that of the structurally related nine-stranded V-type immunoglobulin fold, the GFCC' β -sheet is frequently observed to be functionally more important than the ABE β -sheet [49,50]. The former generally presents a flat surface that forms good protein–protein contacts with appropriate surfaces. There are five putative glycosylation sites in the four FnIII domains at asparagine-209, -300, -470, -553 and -564 (Figure 1b), all of which are conserved in human, chicken and quail anosmin-1. The five glycosylation sites were positioned on β -strand A and on the AB, BC and CC'

loops, none of which obscured the surface accessibility to the GFCC' β -sheet. All five glycosylation sites occur in solvent-accessible positions in the FnIII alignment of Figure 3 as expected.

Structural analysis of heparan sulphate-binding sites in anosmin-1

In anosmin-1, three conserved predicted heparan sulphate-binding-sequence motifs were identified that conform to the sequences XBBXB or XBBBXXB, where B is a basic residue and X is not basic [51,52]. These occurred at residues $S^{278}KHFRS^{283}$ on β -strand G of the FnIII-1 domain, residues $V^{325}HHYKV^{330}$ on β -strand C of the FnIII-2 domain, and residues $Q^{421}RRRPTRP^{428}$ at the beginning of the FnIII-3 domain. Two other predicted heparan sulphate-binding sites are located immediately following the four FnIII domains in anosmin-1 at residues 660–665 and residues 670–675. However, none of the four anosmin-1 FnIII sequences contains an RGD, LDV, IDAP or KLDAPT integrin-binding motif [53].

The electrostatic surfaces of the FnIII domains were examined. A large basic (blue) surface was observed on each of the FnIII-1 and FnIII-2 domains in the 0° and 180° views respectively (Figures 5a and 5b), i.e. on opposite edges of the GFCC' face as defined in Figure 2. The basic surface in the FnIII-1 domain corresponded to the residues $S^{278}KHFRS^{283}$ on β -strand G that constitutes one of the three heparan sulphate-binding motifs noted above. It also included the side chains of arginine-221, -222 and -262 in proximity to this motif, all of which were fully conserved in the human, chicken and quail sequences. The 32R1-WT peptide that includes this heparan sulphate-binding motif has strong cell-adhesive properties [18]. The basic surface in the FnIII-2 domain corresponded to the residues $V^{325}HHYKV^{330}$ and $K^{345}KKRRK^{350}$ on two adjacent β -strands C and C', together with lysine-339, arginine-384 and lysine-386, all of which were fully conserved in the human, chicken and quail sequences. In the FnIII-3 domain, two basic surfaces were observed in the upper half of this model. The larger of the two is seen in the 270° view of this domain. This is in proximity to the

Figure 5 Electrostatic surfaces of the four FnIII domains of anosmin-1

(a–d) The FnIII-1–4 domains respectively. For each domain, the four views are rotated about the vertical axis by 90° increments to the right in order to display the complete surface, in which the 0° views correspond to those shown in Figure 4. Red represents acidic regions and blue represents basic regions. Linear interpolation of the colours represents intermediate electrostatic potentials. Note the large basic (blue) regions observed in the (a) FnIII-1 (0°) and (b) FnIII-2 (180°) models.

heparan sulphate-binding motif Q⁴²¹RRRPTRP⁴²⁸ in the linker peptide joining the FnIII-2 and FnIII-3 domains. In conclusion, all the sequence motifs corresponded to large basic surfaces in the FnIII models.

Structural analysis of the interdomain arrangement in anosmin-1

In anosmin-1, the steric relationship between the basic surfaces on the separate FnIII domains will depend on the interdomain arrangement of these four domains. Information on this is available from two crystal structures. DSSP analysis (Figure 3)

showed that the two-domain crystal structure of neuroglian possessed a 12-residue linker between the ends of β -strand G in the first domain and β -strand A in the second one. The interdomain rotation angle is 175°, in which the EBA and GFCC' β -sheet faces (Figure 2) of the two domains were alternated in their orientation along the long axis of the molecule [26]. The three linkers between the four domains in the crystal structure of FnIII-7–10 in human fibronectin [21] were six or seven residues in length. In the present study, these four domains were positioned in an irregular, but extended, arrangement along the long axis of the molecule, with interdomain rotation angles

of 112°, 160° and 43°. In anosmin-1 (Figure 3), the interdomain linkers are of lengths nine, 31 and 28 residues between the FnIII-1 and FnIII-2, FnIII-2 and FnIII-3, and FnIII-3 and FnIII-4 domains respectively. It is possible that the short length of the first linker in anosmin-1 may cause the β -sheet domain surfaces of the FnIII-1 and FnIII-2 domains to alternate in their orientations. If this were so, the effect would be to create a long extended surface of basic residues, which may constitute a binding site for anionic ligands such as heparan sulphate [51,52,54,55]. Similarly, while the longer lengths of the other two linkers make it difficult to predict relative orientations for the

The formation of disulphide bridges in anosmin-1 is consistent with an extended domain arrangement in anosmin-1, as reported for many other cell-surface proteins. In the FnIII-3 domain, the α -carbon atoms of the conserved residues cysteine-496 and cysteine-523 are well-defined and separated by 0.52 nm. It was inferred that they were disulphide-linked (Figure 4c). Also in the FnIII-3 domain, cysteine-467 is located on a surface loop, and is expected to be close to cysteine-536 in the linker between the

FnIII-3 and FnIII-4 domains. As both residues are conserved in the three anosmin-1 sequences, these are likely to form another disulphide bridge. In contrast, cysteine-368 at the base of the FnIII-2 domain is distant from any other cysteine residues in the FnIII domains. It is buried (Figure 4b), and is not found in the chicken or quail sequences, so is unlikely to be disulphide-linked. The modelling of the WAP domain (see below) accounted for four further disulphide bridges. The N-terminus of anosmin-1 is at alanine-24 [16], and defines the start of the cysteine-rich Cys Box domain (Figure 1b). This contains 10 cysteine residues that are presumed to form five internal disulphide bridges. This analysis showed no evidence for interdomain disulphide bridges that might perturb the proposed extended alignment of basic surfaces in the FnIII-1, -2 and -3 domains.

Structural analysis of three missense substitutions in anosmin-1

All three missense-encoded residue substitutions (Table 1) were located to the GFCC' β -sheet of the FnIII-1 and FnIII-3 domains on the β -strands F and G or the loop between them (Figures 4a and 4c). These regions of the protein structures were well-defined in the models (Table 2) because no sequence insertions or deletions were required. All three affected residues were conserved in the human, chicken and quail sequences, suggesting that they were important for protein folding or function. The FnIII models suggested the following explanations for the effect of each of the three substitutions.

First, the N267K substitution at the C-terminus of β -strand F of FnIII-1 is clearly adjacent to the predicted heparan sulphate proteoglycan-binding site in this domain (see above). The asparagine-267 side chain is predicted to be buried (Figure 3), where it may form hydrogen bonds with main chain groups in the FG loop to stabilize this. Therefore the substitution can either disrupt the protein folding at the FG loop to prevent the secretion of anosmin-1, or alter the function of anosmin-1 as the result of the introduction of an additional basic residue near a large basic surface.

Secondly, the E514K substitution is found at one end of the FG loop of FnIII-3, and is clearly adjacent to the largest basic surface in the FnIII-3 domain and the linker residues glutamine-421–proline-428. The glutamate-514 side chain is predicted to be buried, where it may form a salt-bridge interaction with either the side chain of histidine-510 (or arginine-510 in the chicken or quail sequences) or arginine-457 (or histidine-457 in the chicken or quail sequences). Again the substitution can affect either the folding and secretion of anosmin-1, or it can disrupt the function of anosmin-1 through the introduction of a basic charge.

Thirdly, the F517L substitution is found in proximity to the E514K substitution on the FG loop of FnIII-3. While this involves the replacement of a large bulky hydrophobic residue with a smaller one, it is predicted to be on the surface of the FnIII-3 domain. While an effect on the secretion of anosmin-1 cannot be ruled out (although see below), the proximity of this substitution to the large basic surface in the FnIII-3 domain suggests that this may be important for a direct or indirect hydrophobic interaction between anosmin-1 and a potential ligand.

Homology model for the WAP domain of anosmin-1

The WAP domain structure in anosmin-1 is common to serine protease inhibitors such as elafin, human seminal plasma inhibitor and the Na⁺,K⁺-ATPase inhibitor. It possesses a conserved core of four disulphide bridges [37,38]. The analysis of Figure 6 showed that α -helix and β -sheet structures were largely absent, and that only five side chains were buried out of 38. The

Figure 7 Western blot of media samples from pKD1, pKDM2 and pKD1-R423G-transfected S2 cells

Lanes 1 and 2 (pKD1 and pKDM2 respectively) show the presence of identical bands representing the entire mature anosmin-1 protein (100 kDa + 5 kDa V5-His epitope) and the C-terminal fragment (45 kDa + 5 kDa V5-His). A V5-His-tagged green fluorescent protein was run in lane 3 as a positive control. Lane 4 represents a negative control using insect cells alone. Lane a (pKD1-R423G) shows the presence of bands representing the entire mature anosmin-1 protein (100 kDa + 5 kDa V5-His epitope) and the C-terminal fragment (45 kDa + 5 kDa V5-His).

homology modelling of this domain in anosmin-1 was based on one crystal structure and two NMR structures for elafin. The sequence alignment was readily completed on the basis of seven conserved cysteine residues and six other residues (Figure 6a), giving a high sequence identity of 37%. The sequence alignment initially suggested a structural change of cysteine-147 in anosmin-1 in relation to cysteine-83 of elafin. The homology modelling explained this by showing that this cysteine residue was relocated to the opposite side of an unchanged protein surface loop. This rearrangement enabled the construction of the expected cysteine-147–cysteine-168 disulphide bridge with a correct separation of 0.52 nm between the cysteine α -carbon atoms (Figure 6b). This agreed with the α -carbon separations of the four bridges in elafin, which were 0.56–0.59 nm. The WAP model was verified using PROCHECK, where all 46 residues occurred in the allowed regions of a Ramachandran plot, with only the marginal exception of alanine-144, and the side chain dihedral angles of all eight cysteine residues were in favourable conformations.

Unlike the FnIII domains in anosmin-1, the electrostatic surface of the anosmin-1 WAP model exhibited an overall acidic negative charge (Figure 6c). It is therefore unlikely to participate in heparan sulphate binding. The scissile peptide bond in elafin at alanine-91–methionine-92 is equivalent in position to glycine-142–phenylalanine-143 in anosmin-1, and is at the bottom of the view in Figure 6(b). This is identified as a potential WAP-binding site for proteases or receptors [1]. Of interest is that this peptide bond lies on the opposite side of the WAP domain to the location of the N- and C-termini (Figure 6b). This means that the attachment of the WAP domain to the cysteine-rich Cys Box and FnIII-1 domains renders the scissile peptide bond in the WAP domain accessible for interactions with proteases or receptors.

Expression studies of the N267K, E514K and F517L substitutions

To determine whether the three missense mutations influenced the cellular processing, folding and secretion of anosmin-1, *in vitro* expression studies were undertaken. Anosmin-1 was successfully expressed in an insect cell system. Figure 7 (lane 1) showed that recombinant wild-type anosmin-1 tagged with a C-terminal histidine tag was secreted into the cell medium, where it was detected by Western blotting, both in its integral (100 kDa plus the 5 kDa V5-His epitope) as well as in its C-terminus cleaved forms (45 kDa plus 5 kDa V5-His). Even though insect cells and not mammalian cells were used, no differences in size or

cleavage were identified between the insect-derived and human proteins [16,17]. The three anosmin-1 mutants incorporating N267K, E514K and F517L were generated by site-directed mutagenesis, and were also expressed using this transfection system. Western blotting showed that all three translated products containing the N267K, E514K and F517L substitutions were secreted into the cell medium from the transfected cells, ruling out a secretory defect as the consequence of the substitutions with lysine at positions 267 and 514, and with leucine at position 517.

Expression studies of the KKKRRK and RRR motifs and anosmin-1 cleavage

Anosmin-1 is cleaved in cell culture to a 55 kDa membrane-associated and potential 45 kDa diffusible fragments [17]. Three other N-CAM proteins, neuron-related cell-adhesion molecule, neurofascin and neural glial cell-adhesion molecule, are all formed as heterodimers that consist of α - and β -subunits derived from a single precursor protein by an *in vivo* proteolytic cleavage within FnIII domains at a common paired dibasic sequence motif [56–58]. Similar motifs in several pre-proteins have been shown to be recognition and cleavage sites for endoproteases [59]. The sequencing of the N-terminus of the 45 kDa fragment of anosmin-1, in order to identify the cleavage site, was not successful for reason of a blocked N-terminus. Accordingly, to investigate whether dibasic cleavage motifs may be responsible for the cleavage of anosmin-1, the sequences and models of anosmin-1 were analysed. Only two of the paired dibasic motifs in the anosmin-1 sequence would allow cleavage that would result in two anosmin-1 fragments with the appropriate sizes of approx. 55 kDa and 45 kDa observed in these *in vitro* experiments. These are the sequences K³⁴⁵KKRRK³⁵⁰ and R⁴²²RR⁴²⁴, located in the FnIII-2 domain and FnIII-2/-3 linker region respectively. Both motifs have sequence similarity to a cleavage site found in LAR [17,60]. However the anosmin-1 model showed that the K³⁴⁵KKRRK³⁵⁰ motif is located in the FnIII-2 domain at the edge of the GFCC' β -sheet on β -strand C'. This coincides with part of the large basic surface attributed to a predicted heparan sulphate-binding site. As it is known that the 55 kDa fragment of anosmin-1 with the FnIII-1 and FnIII-2 domains adheres to cell surfaces, a cleavage at this position is therefore unlikely.

Using the insect S2 cell expression system, two mutants were created in which the K³⁴⁵KKRRK³⁵⁰ motif was deleted (clone pKDM2) or the R⁴²²RR⁴²⁴ motif was mutated into R⁴²⁴GR⁴²⁴ (clone pKD1-R423G). Using Western-blot analyses (Figure 7), both mutant proteins were shown to be secreted into the cell medium. Lanes 1 and 2 (pKD1- and pKDM2-transfected S2 media fractions respectively) show the presence of identical bands representing the mature protein (100 kDa plus 5 kDa V5-His) and the cleaved C-terminal fragment (45 kDa plus 5 kDa V5-His). Lane a (pKD1-R423G-transfected S2 media fraction) also showed the same two bands as in lanes 1 and 2. These experiments rule out the involvement of either of these two motifs in the cleavage of the 100 kDa mature protein, and favour the explanation that the primary role of these basic surfaces in the models is heparan sulphate binding.

Peptide-adhesion studies of the three missense mutations

Peptide-adhesion experiments were used to investigate the functional consequence of the three missense mutations, where short peptide sequences incorporating each amino acid substitution were synthesized for cell adhesion assays. Previously, a synthetic oligopeptide sequence 32R1-WT from the first FnIII domain of

Figure 8 Cell-adhesion-assay experiments using the 32R1 and 24III peptides with the MDCK1 and 13.S.1.24 cell lines

The cell adhesion in each case is represented as a percentage of the cell adhesion to poly(L-lysine) (%PLL), and the probability values (p) are shown below. The error bars represent S.E.M.s for each data set.

anosmin-1 was shown to have cell adhesive properties similar to that of the full recombinant protein [18]. To confirm and extend these results, two 32-residue peptides that corresponded to arginine-254–lysine-285 in the FnIII-1 domain and labelled 32R1-WT and 32R1-Mut respectively (Figure 1a) were synthesized to evaluate the effect of the asparagine-to-lysine substitution on cell adhesion. Canine MDCK1 kidney epithelial cells and rat 13.S.1.24 olfactory epithelial cells were used for the assays (Figure 8), as these cell types originate from tissues representing potential functional targets for anosmin-1. For both cell lines, cell adherence to Immulon plates coated with 32R1-Mut with the lysine substitution was significantly higher ($P < 0.005$) than those coated with wild-type 32R1-WT peptide. As a negative control, Chinese-hamster ovary cells was observed to adhere to plates coated with the mutant peptide with a similar affinity as that of the wild-type peptide (results not shown). Differences in the adhesion results between the cell lines are attributable to different levels of expression of heparan sulphate on cell surfaces by different cell lines [18,61].

The E514K and F517L substitutions were also investigated using peptide adhesion studies with both the MDCK1 and 13.S.1.24 cell lines. These used three 24-residue peptides derived from the FnIII-3 domain of anosmin-1 (proline-504–lysine-527). One corresponded to the native sequence (peptide 24III-WT), one contained the E514K substitution (peptide 24III-M514) and one contained the F517L substitution (peptide 24III-M517) (Figure 1a). Figure 8 shows that the 24III and 24III-M517 peptides had a low adhesive affinity for the MDCK1 and 13.S.1.24 cell types, which was similar to that of the negative control using BSA-coated wells. However, the 24III-M514 peptide was observed to effect a significantly greater cell adhesion ($P < 0.001$) than the 24III and 24III-M517 peptides. This effect was similar to that of the positive poly(L-lysine) control, suggesting a sensitivity of anosmin-1 function to charge alteration.

DISCUSSION

Our data provide for the first time a three-dimensional model of five of the six domains in anosmin-1 using a procedure that was validated by conformational checks. Several predicted features of this model were then tested by experiment. Homology modelling is not a substitute for an experimental atomic structure

determinations, but has provided sufficient information to indicate spatial relationships between amino acid residues in the anosmin-1 sequence. It is known that anosmin-1 interacts with heparan sulphate, and its binding is affected by heparitinase [16,18]. A large basic charged patch was identified on each of the FnIII-1, -2 and -3 domains of anosmin-1, each corresponding to a predicted heparan sulphate-binding site, but these were not present on the FnIII-4 or WAP domains. The three-dimensional models showed that the basic patches on the FnIII-1 and FnIII-3 domains were proximate to the disease-causing N267K, E514K and F517L amino acid substitutions. This relationship could not be predicted from the sequence alone.

The low yields observed when using mammalian cell cultures to generate recombinant anosmin-1 [23] prompted a search for a better method of protein production. Insect cell culture systems have been successfully used to express biologically active, secreted, glycosylated human proteins [62]. In the present study it has allowed us to generate sufficient quantities of wild-type and mutant anosmin-1 for analysis by Western blot. The further development of this expression system should provide larger quantities for structural studies by scattering, ultracentrifugation and crystallography. Expression studies showed that all three mutant proteins were secreted at similar levels to that of the wild-type protein, and all these underwent cleavage, indicating that a folding or secretion defect was not the consequence of the mutations. The alternative possibility that the basic patch on the FnIII-2 domain might represent a dibasic protease cleavage site [17] was ruled out by expression studies that removed either of the two candidate dibasic motifs. This outcome favoured the involvement of this basic patch as a site for heparan sulphate-binding and adhesive interactions.

The peptide-adhesion experiments confirmed the previous report [18] in which the 32R1-WT peptide from the first FnIII domain demonstrated cell adhesive properties similar to intact anosmin-1 in cell-adhesion assays. The three-dimensional model showed that most of the nine basic residues in the 32R1-WT peptide (residues 64–95 in the FnIII-1 sequence of Figure 3) occurred in the longest β -strand in the anosmin-1 model and the following large loop. This extended structure may resemble an unstructured, but flexible, conformation expected for the free 32R1-WT peptide, and would explain why this sequence showed similar biological activity to that of anosmin-1. Our *in vitro* cell adhesion data extended these earlier reports by showing that the introduction of a basic residue corresponding to N267K augmented the adhesive properties of this 32R1 anosmin-1 peptide. The 24III-WT peptide (residues 75–98 in the FnIII-3 sequence: Figure 3) occupies the same secondary structure in the FnIII-3 model as that of the 32R1 peptide in the FnIII-1 model. Even though it has not been shown that cell binding of the 24III-WT peptide competes with that of native anosmin-1, it is interesting that a similar substitution (E514 K) in the third FnIII domain conferred a significant adhesive property to the 24III-M514 peptide. The specificity of these effects was suggested by the fact that both the 32R1-WT and 24III-WT peptides showed biological effects with olfactory epithelial cells. We and others [9,10] have shown that anosmin-1 is expressed and secreted from the outer layer of the developing olfactory bulb, and the above result is consistent with anosmin-1 acting as a local guidance cue for olfactory neuroepithelial cells, which induce bulb histogenesis on contact with the developing forebrain [63]. Consistent with this hypothesis, anosmin-1 also promotes cell adhesion and axon clustering *in vitro* [18]. Furthermore, none of the peptide sequences described above stimulated olfactory bulb neurogenesis using a rodent bioassay, which measured olfactory bulb neurite outgrowth (results not shown). This suggests that

anosmin-1 does not affect olfactory bulb development directly, and that changes in the adhesive properties of anosmin-1 may account for the disease-causing nature of the amino acid substitutions modelled in this study. The substitutions to a basic residue at positions 267 and 514 suggested that this increased effect observed in the cell adhesion studies may be due to the increase in basic charge in these regions of the protein structure. As the F517L substitution did not change the adhesive properties of this region of anosmin-1, its effect is more likely to be the consequence of its proximity to the basic surface identified by the modelling.

The use of homology modelling and experiments that tested rational predictions made from these models has clarified the role of functionally important residues in anosmin-1. They have provided initial views of the nature of three disease-causing amino acid substitutions. The development of the kidney and olfactory system depends on appropriate signals that are thought to involve components of the extracellular matrix including heparan sulphate proteoglycans [61]. The involvement of anosmin-1 in these processes may explain the kidney and olfactory abnormalities seen in X-linked Kallmann's patients. The present study has opened the way for further investigations into the functional and structural properties of wild-type and mutated anosmin-1, which will clarify the understanding of its actions during renal and olfactory development.

We thank Dr Gilles Sicard of the Université Claude Bernard in Lyon for the 13.S.1.24 cell line and Dr Colin Hughes of the National Institute for Medical Research, London for the MDCK1 cells. A.R. and G.S.M. are supported by Pharmacia and Upjohn, and S.J.P. is supported by the Wellcome Trust.

REFERENCES

- Lutz, B., Rugarli, E. I., Eichele, G. and Ballabio, A. (1993) X-linked Kallmann syndrome. A neuronal targeting defect in the olfactory system? *FEBS Lett.* **325**, 128–134
- Quinton, R., Duke, V. M., De Zoysa, P. A. and Bouloux, P.-M. G. (1996) The neurobiology of Kallmann's syndrome. *Hum. Reprod.* **2**, 121–127
- Kirk, J. M. W., Grant, D. B., Besser, G. M., Shalet, S., Quinton, R., Smith, C. S., White, M. C., Edwards, O. M. and Bouloux, P.-M. G. (1994) Unilateral renal aplasia in X-linked Kallmann's syndrome. *Clin. Genet.* **46**, 260–262
- Duke, V., Quinton, R., Gordon, I., Bouloux, P.-M. G. and Woolf, A. S. (1998) Proteinuria, hypertension and chronic renal failure in X-linked Kallmann's syndrome, a defined genetic cause of solitary functioning kidney. *Nephrol. Dial. Transplant.* **13**, 1998–2003
- Meitinger, T., Heye, B., Petit, C., Leveilliers, J., Golla, A., Moraine, C., Dalla Piccola, B., Sippell, W. G., Murken, J. and Ballabio, A. (1990) Definitive localization of X-linked Kallmann syndrome (hypogonadotropic hypogonadism and anosmia) to Xp22.3: Close linkage to the hypervariable repeat sequence CRI-S232. *Am. J. Hum. Genet.* **47**, 664–669
- Franco, B., Guioli, S., Pragliola, A., Incerti, B., Bardoni, B., Tonlorenzi, R., Carrozzo, R., Maestrini, E., Pieretti, M., Taillon-Miller, P. et al. (1991) A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. *Nature (London)* **353**, 529–534
- Legouis, R., Hardelin, J.-P., Leveilliers, J., Clavierie, J.-M., Compain, S., Wunderle, V., Millasseau, P., Le Paslier, D., Cohen, D. and Caterina, D. (1991) The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell (Cambridge, Mass.)* **67**, 423–435
- del Castillo, I., Cohen-Salmon, M., Blanchard, S., Lutfalla, G. and Petit, C. (1992) Structure of the X-linked Kallmann syndrome gene and its homologous pseudogene on the Y chromosome. *Nat. Genet.* **2**, 305–310
- Duke, V. M., Winyard, P. J. D., Thorogood, P., Soothill, P., Bouloux, P.-M. G. and Woolf, A. S. (1995) *KAL*, a gene mutated in Kallmann's syndrome, is expressed in the first trimester of human development. *Mol. Cell. Endocrinol.* **110**, 73–79
- Hardelin, J.-P., Julliard, A. K., Moniot, B., Soussi-Yanicostas, N., Verney, C., Schwanzel-Fukuda, M., Lievre, C. A. and Petit, C. (1999) Anosmin-1 is a regionally restricted component of basement membranes and interstitial matrices during organogenesis: Implications for the developmental anomalies of X chromosome-linked Kallmann syndrome. *Dev. Dyn.* **215**, 26–44

- 11 Legouis, R., Cohen-Salmon, M., del Castillo, I., Levilliers, J., Capy, L., Mornon, J.-P. and Petit, C. (1993) Characterization of the chicken and quail homologues of the human gene responsible for the X-linked Kallmann syndrome. *Genomics* **17**, 516–518.
- 12 Dandekar, A. M., Robinson, E. A., Appella, E. and Qasba, P. K. (1982) Complete sequence analysis of cDNA clones encoding rat whey phosphoprotein: Homology to a protease inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3987–3991.
- 13 Hennighausen, L. G. and Sippel, A. E. (1982) Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. *Nucleic Acids Res.* **10**, 2677–2684.
- 14 Lander, A. D. (1989) Understanding the molecules of neural cell contacts—emerging patterns of structure and function. *Trends Neurosci.* **12**, 189–195.
- 15 Edelman, G. M. and Crossin, K. L. (1991) Cell-adhesion molecules: Implications for a molecular histology. *Annu. Rev. Biochem.* **60**, 155–190.
- 16 Soussi-Yanicostas, N., Hardelin, J.-P., del Mar Arroyo-Jimenez, M., Ardouin, O., Legouis, R., Levilliers, J., Traincard, F., Betton, J.-M., Cabanie, L. and Petit, C. (1996) Initial characterization of anosmin-1, a putative extracellular matrix protein synthesized by definite neuronal cell populations in the central nervous system. *J. Cell Sci.* **109**, 1749–1757.
- 17 Rugarli, E. I., Ghezzi, C., Valsecchi, V. and Ballabio, A. (1996) The Kallmann syndrome gene product expressed in COS cells is cleaved on the cell surface to yield a diffusible component. *Hum. Mol. Genet.* **5**, 1109–1115.
- 18 Soussi-Yanicostas, N., Faivre-Sarrailh, C., Hardelin, J.-P., Levilliers, J., Rougon, G. and Petit, C. (1998) Anosmin-1 underlying the X-chromosome-linked Kallmann syndrome is an adhesion molecule that can modulate neurite growth in a cell-type specific manner. *J. Cell Sci.* **111**, 2953–2965.
- 19 Hardelin, J.-P., Levilliers, J., del Castillo, I., Cohen-Salmon, M., Legouis, R., Blanchard, S., Compain, S., Bouloux, P.-M. G., Kirk, J., Moraine, C. et al. (1992) X-chromosome-linked Kallmann syndrome: stop mutations validate the candidate gene. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8190–8194.
- 20 Hardelin, J.-P., Levilliers, J., Blanchard, S., Carel, J.-C., Leutenegger, M., Pinard-Bertellette, J. P., Bouloux, P.-M. G. and Petit, C. (1993) Heterogeneity in the mutations responsible for X-chromosome-linked Kallmann syndrome. *Hum. Mol. Genet.* **2**, 373–377.
- 21 Georgopoulos, N. A., Pralong, F. P., Seidman, C. E., Seidman, J. G., Crowley, W. F. and Vallejo, M. (1997) Genetic heterogeneity evidenced by low incidence of *KAL-1* gene mutations in sporadic cases of gonadotropin-releasing hormone deficiency. *J. Clin. Endocrinol. Metab.* **82**, 213–217.
- 22 Maya-Nunez, G., Zenteno, J. C., Ulloa-Aguirre, A., Kofman-Alfaro, S. and Mendez, J. P. (1998) A recurrent missense mutation in the *KAL* gene in patients with X-linked Kallmann's syndrome. *J. Clin. Endocrinol. Metab.* **83**, 1650–1653.
- 23 Maya-Nunez, G., Cuevas-Covarrubias, S., Zenteno, J. C., Ulloa-Aguirre, A., Kofman-Alfaro, S. and Mendez, J. P. (1998) Contiguous gene syndrome due to deletion of the first three exons of the Kallmann gene and complete deletion of the steroid sulphatase gene. *Clin. Endocrinol.* **48**, 713–718.
- 24 O'Neill, M. J., Tridjaja, B., Smith, M. J., Bell, K. M., Warne, G. L. and Sinclair, A. H. (1998) Familial Kallmann syndrome: A novel splice acceptor mutation in the *KAL* gene. *Hum. Mutat.* **11**, 340–342.
- 25 Gu, W. X., Colquhoun-Kerr, J. S., Kopp, P., Bode, H. H. and Jameson, J. L. (1998) A novel amino terminal mutation in the *KAL-1* gene in a large pedigree with X-linked Kallmann syndrome. *Mol. Genet. Metab.* **65**, 59–61.
- 26 Huber, A. H., Wang, Y. E., Bieber, A. J. and Bjorkman, P. J. (1994) Crystal structure of tandem Type III fibronectin domains from *Drosophila* neuroglian at 2.0 Å. *Neuron* **12**, 717–731.
- 27 Main, A. L., Harvey, T. S., Baron, M., Boyd, J. and Campbell, I. D. (1992) The three-dimensional structure of the tenth Type III module of fibronectin: An insight into RGD-mediated interactions. *Cell (Cambridge, Mass.)* **71**, 671–678.
- 28 Dickinson, C. D., Veerapandian, B., Dai, X. P., Hamlin, R. C., Xuong, N. H., Ruoslahti, E. and Ely, K. R. (1994) Crystal structure of the tenth Type III cell adhesion module of human fibronectin. *J. Mol. Biol.* **236**, 1079–1092.
- 29 Leahy, D. J., Aukhil, I. and Erickson, H. P. (1996) 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell (Cambridge, Mass.)* **84**, 155–164.
- 30 Harlos, K., Martin, D. M. A., O'Brien, D. P., Jones, E. Y., Stuart, D. I., Polikarpov, I., Miller, A., Tuddenham, E. G. D. and Boys, C. W. G. (1994) Crystal structure of the extracellular region of human tissue factor. *Nature (London)* **370**, 662–666.
- 31 Banner, D. W., D'Arcy, C., Chene, C., Winkler, F. K., Guha, A., Konigsberg, W. H., Nemerson, Y. and Kirchhofer, D. (1996) The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature (London)* **380**, 41–46.
- 32 Muller, Y. A., Ultsch, M. H. and de Vos, A. M. (1996) The crystal structure of the extracellular domain of human tissue factor refined to 1.7 Å resolution. *J. Mol. Biol.* **256**, 144–159.
- 33 Leahy, D. J., Hendrickson, W. A., Aukhil, I. and Erickson, H. P. (1992) Structure of a fibronectin Type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science (Washington, D.C.)* **258**, 987–991.
- 34 Livnah, O., Stura, E. A., Johnson, D. L., Middleton, S. A., Mulcahy, L. S., Wrighton, N. C., Dower, W. J., Jolliffe, L. K. and Wilson, I. A. (1996) Functional mimicry of a protein hormone by a peptide agonist—the EPO receptor complex at 2.8 Å. *Science (Washington, D.C.)* **273**, 464–471.
- 35 de Vos, A. M., Ultsch, M. and Kossiakoff, A. A. (1992) Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. *Science (Washington, D.C.)* **255**, 306–312.
- 36 Yamasaki, K., Naito, S., Anaguchi, H., Ohkubo, T. and Ota, Y. (1997) Solution structure of an extracellular domain containing the WSxWS motif of the granulocyte colony-stimulating factor receptor and its interaction with ligand. *Nat. Struct. Biol.* **4**, 498–504.
- 37 Tsunemi, M., Matsuura, Y., Sakakibara, S. and Katsube, Y. (1996) Crystal structure of an elastase-specific inhibitor elafin complexed with porcine pancreatic elastase determined at 1.9 Å resolution. *Biochemistry* **35**, 11570–11576.
- 38 Francart, C., Dauchez, M., Alix, A. J. P. and Lippens, G. (1997) Solution structure of R-elafin, a specific inhibitor of elastase. *J. Mol. Biol.* **268**, 666–677.
- 39 Sternberg, M. J. E. (1996) *Protein Structure Prediction: Practical Approach*, Oxford University Press, Oxford.
- 40 Bateman, A., Jouet, M., MacFarlane, J., Du, J. S., Kenwick, S. and Chothia, C. (1996) Outline structure of the human L1 cell adhesion molecule and the sites where mutations cause neurological disorders. *EMBO J.* **15**, 6050–6059.
- 41 Jenkins, P. V., Pasi, K. J. and Perkins, S. J. (1998) Molecular modelling of ligand and mutation sites of the Type A domains of human von Willebrand factor and their relevance to von Willebrand's disease. *Blood* **91**, 2032–2044.
- 42 Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
- 43 Lee, B. and Richards, F. M. (1971) An interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
- 44 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
- 45 Bunch, T. A., Grinblat, Y. and Goldstein, L. B. (1988) Characterization and use of the *Drosophila* metallothionein promoter in cultured *Drosophila melanogaster* cells. *Nucleic Acids Res.* **16**, 1043–1061.
- 46 Schneider, I. (1972) Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* **27**, 353–365.
- 47 Coronas, V., Feron, F., Hen, R., Sicard, G., Jourdan, F. and Moyse, E. (1997) *In vitro* induction of apoptosis or differentiation by dopamine in an immortalized olfactory neuronal cell line. *J. Neurochem.* **69**, 1870–1881.
- 48 Bauvois, B., Rouillard, D., Sanceau, J. and Wietzerbin, J. (1992) IFN-gamma and transforming growth factor-beta-1 differently regulate fibronectin and laminin receptors of human differentiating monocytic cells. *J. Immunol.* **148**, 3912–3919.
- 49 Boehm, M. K., Mayans, M. O., Thornton, J. D., Begent, R. H. J., Keep, P. A. and Perkins, S. J. (1996) Extended glycoprotein structure of the seven domains in human carcinoembryonic antigen by X-ray and neutron solution scattering and an automated curve fitting procedure: implications for cellular adhesion. *J. Mol. Biol.* **259**, 718–736.
- 50 Chothia, C. and Jones, E. Y. (1997) The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* **66**, 823–862.
- 51 Margalit, H., Fischer, N. and Ben-Sasson, S. A. (1993) Comparative analysis of structurally defined heparin binding sequences reveals a distinct spatial distribution of basic residues. *J. Biol. Chem.* **268**, 19228–19231.
- 52 Hileman, R. E., Fromm, J. R., Weiler, J. M. and Linhardt, R. J. (1998) Glycosaminoglycan–protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *Bioessays* **20**, 156–167.
- 53 Yokosaki, Y., Matsuura, N., Higashiyama, S., Murakami, I., Obara, M., Yamakido, M., Shigeto, N., Chen, J. and Sheppard, D. (1998) Identification of the ligand binding site for the integrin $\alpha_5\beta_1$ in the third fibronectin Type III repeat of tenascin-C. *J. Biol. Chem.* **273**, 11423–11428.
- 54 Schlessinger, J., Lax, I. and Lemmon, M. (1995) Regulation of growth-factor activation by proteoglycans: What is the role of the low-affinity receptors? *Cell (Cambridge, Mass.)* **83**, 357–360.
- 55 Busby, T. F., Argraves, W. S., Brew, S. A., Pechik, I., Gilliland, G. L. and Ingham, K. C. (1995) Heparin binding by fibronectin module III-13 involves 6 discontinuous basic residues brought together to form a cationic cradle. *J. Biol. Chem.* **270**, 18558–18562.
- 56 Burgoon, M. P., Grumet, M., Mauro, V., Edelman, G. M. and Cunningham, B. A. (1991) Structure of the chicken neuron-glia cell-adhesion molecule, Ng-CAM—origin of the polypeptides and relation to the Ig superfamily. *J. Cell Biol.* **112**, 1017–1029.
- 57 Kayyem, J. F., Roman, J. M., de la Rosa, E. J., Schwarz, U. and Dreyer, W. J. (1992) Bravo/Nr-CAM is closely related to the cell-adhesion molecules L1 and Ng-CAM and has a similar heterodimer structure. *J. Cell Biol.* **118**, 1259–1270.

-
- 58 Volkmer, H., Hassel, B., Wolff, J. M., Frank, R. and Rathjen, F. G. (1992) Structure of the axonal surface recognition molecule neurofascin and its relationship to a neural subgroup of the immunoglobulin superfamily. *J. Cell Biol.* **118**, 149–161
- 59 Barr, P. J. (1991) Mammalian subtilisins: The long-sought dibasic processing endoproteases. *Cell (Cambridge, Mass.)* **66**, 1–3
- 60 Streuli, M., Krueger, N. X., Ariniello, P. D., Tang, M., Munro, J. M., Blattler, W. A., Adler, D. A., Disteche, C. M. and Saito, H. (1992) Expression of the receptor-linked protein tyrosine phosphatase LAR: Proteolytic cleavage and shedding of the CAM-like extracellular region. *EMBO J.* **11**, 897–907
- 61 Bernfield, M., Götte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J. and Zako, M. (1999) Functions of cell surface heparan sulfate proteoglycans. *Annu. Rev. Biochem.* **68**, 729–777
- 62 Farnell, P. J., Behie, L. A. and Iastro, K. (1999) Transformed Lepidopteran insect cells: New sources of recombinant human tissue plasminogen activator. *Biotech. Bioeng.* **64**, 426–433
- 63 Gong, Q. and Shipley, M. T. (1995) Evidence that pioneer olfactory axons regulate telencephalon cell cycle kinetics to induce the formation of the olfactory bulb. *Neuron* **14**, 91–101
-

Received 20 February 2001/23 April 2001; accepted 21 May 2001