European Journal of Human Genetics (2000) 8, 280–285

www.nature.com/ejhg

ARTICLE

Small in-frame deletions and missense mutations in CADASIL: 3D models predict misfolding of Notch3 EGF-like repeat domains

Martin Dichgans[1], Harald Ludwig[1], Josef Müller-Höcker[2], Albrecht Messerschmidt[3] and Thomas Gasser[1]

[1]*Department of Neurology, Klinikum Grosshadern;* [2]*Department of Pathology; Ludwig-Maximilians-University, Munich;* [3]*Max-Planck-Institut für Biochemie, Abteilung Strukturforschung, Martinsried, Germany*

CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) is a hereditary microangiopathic condition causing stroke in young adults. The responsible gene has recently been identified as the *Notch3* gene. *Notch3* encodes a large transmembrane receptor with 34 extracellularly localised epidermal growth factor-like (EGF) repeat domains. We screened 71 unrelated CADASIL families for mutations in two exons coding for the first five EGF-like repeats and found mutations in 70% of the families (*n* = 50). Two types of mutations were identified: 48 families (96%) had missense mutations and two families (4%) had small in-frame deletions. Seven mutations occurred multiple times. All of them are C to T transitions that affect CpG dinucleotides, suggesting that their multiple occurrence is due to the hypermutability of this sequence. All mutations, including the two deletions, result in the gain or loss of a cysteine residue, thus substantiating the pivotal role of an uneven number of cysteine residues within EGF-like repeat domains of Notch3 in the pathogenesis of CADASIL. To study the potential effects of these mutations 3D homology models of the first six EGF domains were generated on the basis of NMR data from human fibrillin-1. These models predict domain misfolding for a subset of mutations. *European Journal of Human Genetics* (2000) **8**, 280–285.

Keywords: CADASIL; *Notch3*; mutation; EGF-like repeat; 3D model

Introduction

CADASIL is an increasingly recognised autosomal dominant disorder leading to cerebrovascular manifestations in early adulthood.[1] Affected individuals experience recurrent ischemic episodes with accumulating motor, sensory and cognitive deficits.[2,3] The majority of patients eventually become demented and additional manifestations have been reported including migraine (30–40%), psychiatric disturbance (30%) and epileptic seizures (10%).[2,3] Brain magnetic resonance imaging (MRI) displays diffuse white matter abnormalities and small cystic lesions suggestive of small infarcts.[4] The penetrance of MRI abnormalities in CADASIL is complete by age 35.[2] However, symptom onset varies considerably (age 14–66 years).[3] The underlying pathology is mediated by a unique non-amyloid angiopathy involving small arteries (100–400 μm) and capillaries, particularly in the brain but also in other organs.[5]

Recently, the *Notch3* gene on chromosome 19p13 was identified as the CADASIL disease gene.[6] *Notch* genes encode evolutionary conserved cell surface receptors that regulate cell fate choices in vertebrates and invertebrates during embryonic development.[7] This regulation occurs by interaction of *Notch* gene products with their transmembrane ligands on neighbouring cells.[7,8]

A striking feature of both the Notch receptors and their ligands is a large number of tandemly arranged epidermal growth factor-like (EGF-like) repeat domains which account for most of the extracellular domains of these proteins.[7] EGF-like repeats have been identified in numerous extracellular and membrane bound proteins.[9,10] Their classification is

Correspondence: Dr M Dichgans, Department of Neurology, Klinikum Grosshadern, Ludwig-Maximilians-University, Marchioninstr. 15, D-81377 Munich, Germany. Tel: +49 89 7095 3673; Fax: +49 89 7095 3677; E-mail: mdichgans@brain.nefo.med.uni-muenchen.de
Received 10 September 1999; revised 10 December 1999; accepted 21 December 1999

based mainly on the spacing and presence of six cysteine residues within approximately 40 aminoacid residues. The structures of a number of EGF-like domains have been determined, and most of them are consistent with an EGF-like disulfide bonding pattern.[10–16]

On screening 50 CADASIL families for mutations along the entire coding sequence of the *Notch3* gene, Joutel *et al* found mutations in 90% of the families.[17] All mutations were located in one of the 34 EGF-like repeat domains with a strong cluster in two exons coding for the first five repeats. All mutations were missense mutations predicted to result in a loss or a gain of cysteine residue. Based on this observation the authors hypothesised that aberrant dimerisation of Notch3, due to abnormal disulfide bridging with another Notch3 molecule or with another protein, may be involved in the pathogenesis of the disorder.[17]

In the current study we performed mutational analysis in 71 unrelated CADASIL families by focusing on exons 3 and 4 which code for the first five EGF-like repeats of the *Notch3* gene product. 3D homology models were generated to study the potential effects of the observed mutations.

Patients and methods

Patients

We studied 71 index cases and their affected relatives from 71 unrelated Caucasian families (61 German, 4 Swiss, 3 Austrian, and 3 British). Probands were selected on the basis of

(1) A clinical syndrome of recurrent ischemic episodes, cognitive deterioration, migraine with aura, psychiatric disturbance, or a combination of these features;

(2) A cerebral MRI showing microangiopathic changes, and

(3) A family history consistent with autosomal dominant inheritance.

In all families the diagnosis of CADASIL had been confirmed by biopsy (n = 66) or autopsy (n = 5).[18,19]

Mutation analysis

Following informed consent, genomic DNA was isolated from peripheral blood leukocytes. Exons 3 and 4 of the *Notch3* gene were amplified using the following intronic primers:

exon 3 (F: 5'TGTGCTGCCCAACCAAGCCA;
R: 5'ACTGACCACACCCCCGACTA);
exon 4 (F: 5'TAGTCGGGGGTGTGGTCAGT;
R: 5'CCTCTGACTCTCCTGAGTAG).[17]

PCR conditions were as follows: initial denaturation 5 min at 94°C, followed by 35 cycles at 94°C for 30 s (denaturation), 65°C for 30 s (annealing) and 72°C for 30 s (extension), with a 7 min final extension at 72°C. PCR products were subjected to direct automated sequencing using dideoxy-terminator cycle sequencing (BigDye™ sequencing Ready Reaction Kit, Perkin-Elmer, Foster City, CA) and an Applied Biosystems model 377 automated sequencer (Applied Biosystems division, Perkin-Elmer Corporation, Foster City, CA). All products were sequenced on both strands. The human Notch3 cDNA and human Notch3 protein sequences were taken from GenBank (accession no. U97669) and a published paper.[17]

3D Modelling

3D modelling was done on the basis of the NMR structure of the 32nd and 33rd calcium binding (cb) EGF-like domain pair of human fibrillin-1[16] and a multiple aminoacid sequence alignment of this cbEGF-like domain pair with EGF-like domain pairs 1–2, 3–4 and 5–6 from human Notch3 (Table 1). Alignment was arranged manually with the disulfide bridge forming cysteines as anchor points. The alignment of domain pairs was chosen, since Notch3 domains 1, 3 and 5 align well on the 32nd EGF-like domain of human fibrillin-1 and Notch3 domains 2, 4 and 6 align well on the 33rd EGF-like domain of human fibrillin-1 (Table 1). Models are presented as single domains because of uncertainties regarding interdomain arrangement.

Model building was done with program 'O'.[20] Mutated side chains were put into energetically favourable conformations with option 'lego–side–chain'. Those parts of the models where insertions or deletions had to be introduced were regularised with option 'refi–zone'. Energy minimisation of the models was not carried out because it did not promise further improvement of the models.

Results

Mutation detection

Seventy-one unrelated CADASIL families were analysed for mutations in the cluster region (exons 3 and 4) of the *Notch3*

Table 1 Multiple aminoacid sequence alignment between the 32nd and 33rd cbEGF-like domains from human fibrillin-1 and EGF-like domain pairs 1–2, 3–4 and 5–6 from Notch3

```
HBFIB1 31–32 SAVDMDECKEPDVCKHG–QCINTDG–SYRCECPFGYILAGNECV–DTDECSVGNPCG–NGTCKNVIGG———FECTCEEGFEPGPMMTCE–
NOTCH3 01–02 ——AAPPCLDGSPCANGGRCTQLPSREAACLCPPGWV——GERCQL–EDPCHSG–PCAGRGVCQSSVVAGTARFSCRCPRGFR–GPD——CSL
NOTCH3 03–04 ———PDPCLS–SPCAHGARCSVGPDGRFLCSCPPGYQ——GRSCRSDVDECRVGEPCRHGGTCLNTP–GS———FRCQCPAGYT–GPL——CEN
NOTCH3 05–06 ——PAVPCAP–SPCRNGGTCRQSGDLTYDCACLPGFE——GQNCEVNVDDCP–GHRCLNGGTCVDGVN–T———YNCQCPPEWT–GQF——CTE
                    *     *  *  *         *  *  *    *  *    *  *  *  *  *  *           *  *      *    *
```

The alignment was done manually with the pattern of cysteines (black boxed) as anchor points. *Consensus residues.

Table 2 *Notch3* mutations in 50 CADASIL families

No. of families	Notch3 *nucleotide*	Notch3 *mutation*	*Amino acid exchange*	*Exon*	*Domain*
6	346	CGT → TGT	Arg90 → Cys	N3	EGF 2
1	356	TGC → TTC	Cys93 → Phe*	N3	EGF 2
3	406	CGT → TGT	Arg110 → Cys	N3	EGF 2
1	428	TGC → TTC	Cys117 → Phe	N4	EGF 2
1	445	TGC → TTC	Cys123 → Phe*	N4	EGF 3
10	475	CGC → TGC	Arg133 → Cys	N4	EGF 3
5	499	CGC → TGC	Arg141 → Cys	N4	EGF 3
1	508	TGC → TCC	Cys144 → Ser*	N4	EGF 3
1	508	TGC → TAC	Cys144 → Tyr*	N4	EGF 3
1	525	TAC → TGC	Tyr150 → Cys*	N4	EGF 3
2	535	CGC → TGC	Arg153 → Cys	N4	EGF 3
5	583	CGC → TGC	Arg169 → Cys	N4	EGF 4
1	599	TGC → TAC	Cys174 → Tyr	N4	EGF 4
6	622	CGC → TGC	Arg182 → Cys	N4	EGF 4
1	625	TGC → CGC	Cys183 → Arg	N4	EGF 4
1	625	TGC → AGC	Cys183 → Ser*	N4	EGF 4
1	631	TGT → CGT	Cys185 → Arg	N4	EGF 4
1	658	TGT → TTT	Cys194 → Phe*	N4	EGF 4
1	317–331	deletion	D80–S84del*	N4	EGF 2
1	537–545	deletion	R153–C155del*	N4	EGF 3

All mutations are predicted to involve cysteine residues. In 38 families the mutation is predicted to create a cysteine, whereas in 12 families the mutation is predicted to delete a cysteine residue. Nucleotide numbers are given based on the human *Notch3* cDNA sequence taken from GenBank (accession no. U97669); *novel mutation.

gene by direct sequencing of genomic DNA. Mutations were found in 50 (70%) of the families. Of the 20 different mutations identified 18 were missense mutations (90%) and two were deletions (10%). Nine mutations are novel (Table 2).

Missense mutations All the missense mutations are predicted to involve cysteine residues, ie to replace the wild-type aminoacid with a cysteine residue (eight mutations) or to replace one of the cysteine residues with another aminoacid (ten mutations). Seven mutations were found more than once, ie 2–10 times. All of them are C to T transitions leading to the substitution of an arginine by a cysteine residue. Eleven mutations were found only once. Ten of them are predicted to replace a cysteine residue with another aminoacid.

Deletions We found one 15-bp and one 9-bp deletion predicted to result in the loss of five and three aminoacids, respectively. In both cases the deleted aminoacids include one cysteine residue. Both deletions occur at sites containing DNA repeat sequences. The 317–331del sequence is flanked by a 7bp-CCCCTGT repeat while the 537–545del sequence is flanked by a 4bp-GCCG repeat. In both cases one of the repeats is deleted by the mutation which is consistent with misalignment during DNA replication.[21]

Polymorphisms In addition to disease-causing mutations there were several silent nucleotide polymorphisms (Table 3).

Discussion

Our study extends the spectrum of disease-causing mutations in CADASIL by adding small in-frame deletions. Both deletions are predicted to result in the loss of one cysteine residue. In conjunction with the missense mutations reported here and previously[17,22,23] these deletions substantiate the pivotal role of an odd number of cysteine residues within EGF-like repeat domains of *Notch3* in the pathogenesis of this disorder.

The mechanisms by which mutations in CADASIL become pathogenic are currently unknown. However, at least two possibilities have to be considered. The first is domain misfolding. As illustrated by the modelled 3D structures of the first six EGF-like repeat domains (Figure 1) each domain is predicted to contain an N-terminal two-stranded β-sheet followed by a second shorter sheet, or double hairpin at the C-terminus. The six cysteines form three disulfide bonds which stabilise the two β-strand substructures. Some of the mutations are predicted to be incompatible with this folding pattern. This is obvious for the two deletions which destroy sequences involved into β-strand formation and domain stabilisation (Figure 1). Domain misfolding may further be predicted for a group of missense mutations where large side chains of the replacing residues are predicted to interfere with adjacent parts of the native structure (Cys to Arg; Cys to

Table 3 Silent nucleotide polymorphisms in exons 3 and 4 of the *Notch3* gene

Notch3 *nucleotide*	*Nucleotide change*	*Amino acid exchange*	*Exon*	*Domain*
360	CAG → CAA	none	N3	EGF 2
381	ACC → ACT	none	N3	EGF 2
423	CCT → CCC	none	N4	EGF 2
546	CGA → CGG	none	N4	EGF 3
684	GCG → GCA	none	N4	EGF 5

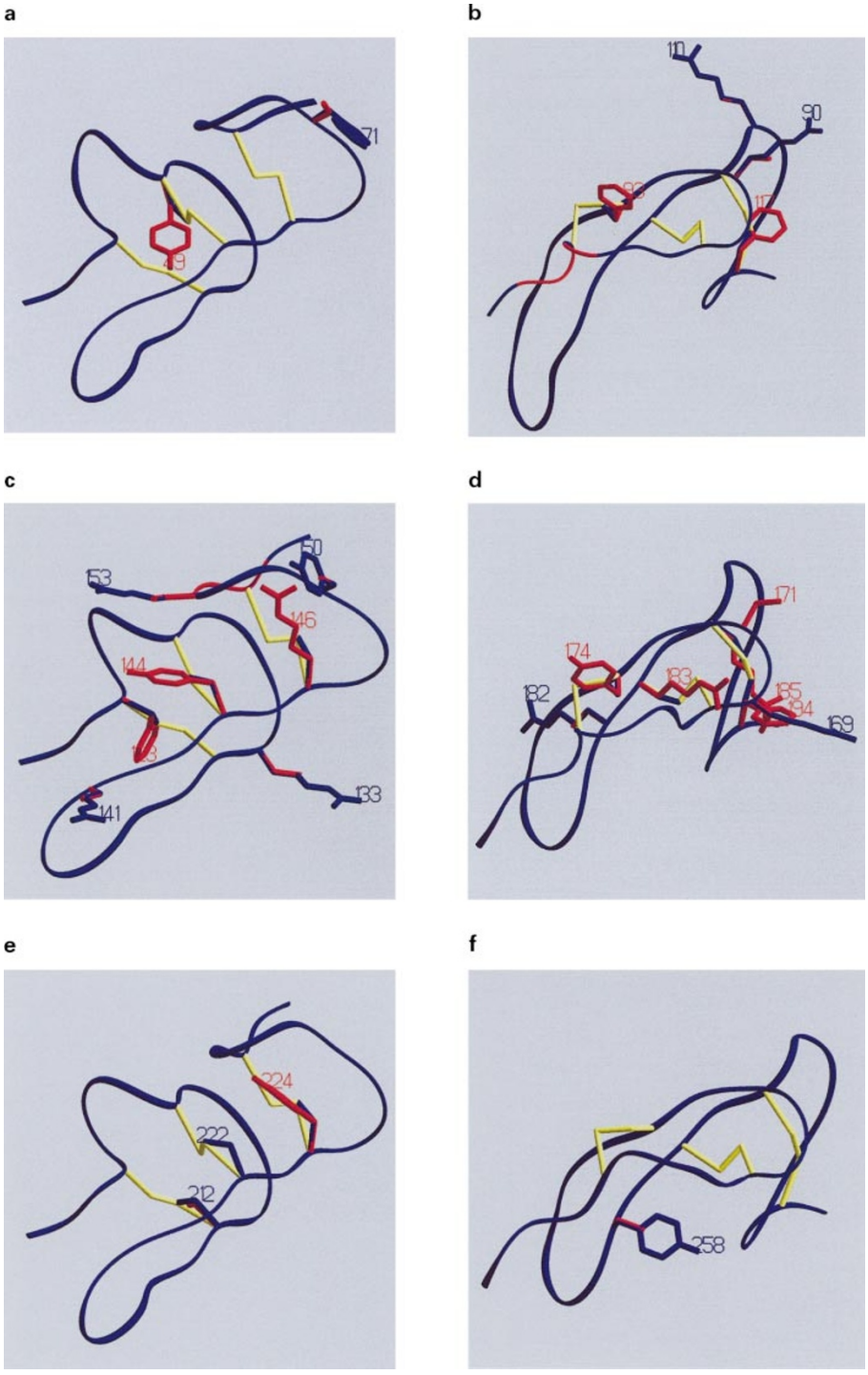

Figure 1 3D models of the predicted main chain traces of the first six Notch3 EGF-like repeat domains (**a**–**f**). The main chain traces are shown in blue (N-termini at the lower left). Each domain contains a two-stranded β-sheet followed by a second shorter sheet, or double hairpin. The six cysteines form three disulfide bonds (yellow). The two-stranded β-sheet is stabilised by two crossing disulfide bonds (1st to 3rd and 2nd to 4th cysteine residue), and the double hairpin is fixed by the 5th to 6th – disulfide bond. The 4th and 5th cysteines are only one amino acid apart, so that the two β-strand substructures are closely connected. Wild-type side chains (blue) are shown at positions affected by missense mutations (mutated side chains in red). The two deletions are indicated as a red section within the main chain trace. For the purpose of completedness mutations identified in this study were combined with previously reported mutations: Cys49Tyr; Trp71Cys; Cys146Arg; Gly171Cys; Cys212Ser; Cys222Gly; Cys224Tyr; Tyr258Cys.[17] This Figure was produced using SETOR.[26]

Phe; Cys to Tyr; Figure 1, data not shown). In fact, it seems likely that all mutations substituting cysteine residues cause protein misfolding since these mutations disrupt disulfide bonds likely to be necessary for a correct fold of the respective domain (Figure 1). The consequences of mutations generating new cysteine residues (Arg to Cys; Tyr to Cys; Trp to Cys; Gly to Cys) are difficult to predict. However, with an odd number of cysteines, new permutations of the bridging scheme might be favoured. This would also cause domain misfolding in these mutations.

Another possibility consists in the formation of intermolecular cross-links with other Notch3 molecules (homodimers) or other cysteine-containing proteins (heterodimers). It is conceivable from Figure 1, that mutations replacing an amino acid with a cysteine residue could provide the basis for such interactions in particular, since these mutations affect residues predicted to be situated on the surface of the molecule. Mutations that replace cysteine residues could have a similar effect by leaving an unpaired and thus reactive cysteine residue.

Seven mutations occurred multiple times. As previously shown, the multiple occurrence of mutations in CADASIL is not explained by founder effects.[17] Interestingly, all the multiple occuring mutations found in this study and in the study by Joutel[17] are C to T transitions affecting CpG dinucleotides. CpG dinucleotides have been shown to undergo germ-line transitions to TG and CA at frequencies six to seven times the base mutation rate.[24] Thus we assume, that the multiple occurrence of these mutations may be attributed to the hypermutability of CpG dinucleotides, which involves methylation-mediated deamination of cytosines.[24]

The reasons for the strong clustering of CADASIL mutations within exons coding for N-terminal EGF repeats are still unknown. A comparison of exons 3 and 4 to more downstream coding sequences of *Notch3* failed to reveal a significant increase in GC content or in the frequency of CpG dinucleotides (results not shown). Thus a simple DNA-based explanation for the clustering of mutations is lacking. It has already been pointed out that the N-terminal repeats are less conserved in between Notch3 and otherwise highly homologous Notch proteins.[17] In fact, sequence alignment with other Notch proteins revealed that Notch3 lacks a region corresponding to parts of EGF-repeats 2 and 3 in the homologous proteins[25] (own results, data not shown). The loss of this region is equivalent to a net loss of one EGF-repeat. Interestingly, the second Notch3 EGF-repeat domain differs from most other Notch3 EGF domains by an additional three to four aminoacid residues that are located in between the third and fourth cysteine residue. As indicated in Figure 1b this is predicted to result in a longer two-stranded β-sheet, elongated in the direction of the N-terminus. It is conceivable, that this could affect the interdomain arrangement in this region. However, whether structural features relate to the clustering of mutations within N-terminal repeats of the Notch3 receptor awaits further studies.

The broad clinical spectrum of CADASIL[2,3] raises the question of genotype–phenotype correlations. The clinical and MRI phenotype in the two families with the deletions did not diverge from the usual spectrum. Both index cases had a history of stroke. One of them had developed cognitive deficits, the other suffered from migraine with aura. Also, age at onset for single manifestations in these individuals was within the expected range. This adds to previous clinical and neuroimaging evidence indicating, that the *Notch3* genotype has no major influence on the phenotype.[3,22]

In conclusion we have found a number of new mutations including two small in-frame deletions. These deletions are relevant in that they substantiate the pivotal role of an odd number of cysteine residues within EGF-like repeat domains in the pathogenesis of CADASIL. The 3-D models suggest domain misfolding at least for a subset of mutations including the two deletions.

Acknowledgements

We gratefully acknowledge the collaboration of all CADASIL patients and their families, and the numerous physicians who referred cases. We are further indebted to Dr A Joutel and Dr E Tournier-Lasserve (Paris) who provided us with the primer sequences prior to their publication. This work was supported by the Deutsche Forschungsgemeinschaft (Di 722/I-1). Part of this work are part of the thesis of H Ludwig (in preparation).

Accession numbers and URLs for data in this article are as follows: GenBank, http://www.ncbi.nlm.nih.gov/Entrez/nucleotide.htlm (Notch3 cDNA [U97669]). Online Mendelian inheritance in man (OMIM), http://www3.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?125310 (for CADASIL [MIM 125310]).

References

1 Tournier-Lasserve E, Joutel A *et al*: Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy maps on chromosome 19q12. *Nat Genet* 1993; **3**: 256–259.
2 Chabriat H, Vahedi K, Iba Zizen MT *et al*: Clinical spectrum of CADASIL: a study of 7 families. *Lancet* 1995; **346**: 934–939.
3 Dichgans M, Mayer M, Uttner I *et al*: The phenotypic spectrum of CADASIL: clinical findings in 102 cases. *Ann Neurol* 1998; **44**: 731–739.
4 Chabriat H, Levy C, Taillia H *et al*: Patterns of MRI lesions in CADASIL. *Neurology* 1998; **51**: 452–457.
5 Ruchoux MM, Maurage CA: CADASIL: cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *J Neuropathol Exp Neurol* 1997; **56**: 947–964.
6 Joutel A, Corpechot C, Ducros A *et al*: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature* 1996; **383**: 707–710.
7 Artavanis-Tsakonas S, Matsuno K, Fortini ME: Notch signaling. *Science* 1995; **268**: 225–232.
8 Weinmaster G: The ins and outs of Notch signaling. *Mol Cell Neurosci* 1997; **9**: 91–102.
9 Doolittle RF, Feng DF, Johnson MS: Computer-based characterization of epidermal growth factor precursor. *Nature* 1984; **307**: 558–560.
10 Campbell ID, Bork P: Epidermal growth factor-like modules. *Curr Opin Struct Biol* 1993; **3**: 385–392.

11 Moy FJ, Li YC, Rauenbuehler P, Winkler ME, Scheraga HA, Montelione GT: Solution structure of human type-alpha transforming growth factor determined by heteronuclear NMR spectroscopy and refined by energy minimization with restraints. *Biochemistry* 1993; **32**: 7334–7353.
12 Nagata K, Kohda D, Hatanaka H *et al*: Solution structure of the epidermal growth factor-like domain of heregulin-alpha, a ligand for P180erbB-4. *EMBO J* 1994; **13**: 3517–3523.
13 Graves BJ, Crowther RL, Chandran C, Rumberger JM, Li S, Huang KS, Presky DH, Familletti PC, Wolitzky BA, Burns DK: Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature* 1994; **367**: 532–538.
14 Hansen AP, Petros AM, Meadows RP, Fesik SW: Backbone dynamics of a two-domain protein: 15N relaxation studies of the amino-terminal fragment of urokinase-type plasminogen activator. *Biochemistry* 1994; **33**: 15418–15424.
15 Smith BO, Downing AK, Driscoll PC, Dudgeon TJ, Campbell ID: The solution structure and backbone dynamics of the fibronectin type I and epidermal growth factor-like pair of molecules of tissue plasminogen activator. *Structure* 1995; **3**: 823–833.
16 Downing AK, Knott V, Werner JM, Cardy CM, Campbell ID, Handford PA: Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* 1996; **85**: 597–605.
17 Joutel A, Vahedi K, Corpechot C *et al*: Strong clustering and stereotyped nature of Notch3 mutations in CADASIL patients. *Lancet* 1997; **350**: 1511–1515.
18 Baudrimont M, Dubas F, Joutel A, Tournier-Lasserve E, Bousser MG: Autosomal dominant leukoencephalopathy and subcortical ischemic stroke. A clinicopathological study. *Stroke* 1993; **24**: 122–125.
19 Ruchoux MM, Chabriat H, Bousser MG, Baudrimont M, Tournier-Lasserve E: Presence of ultrastructural arterial lesions in muscle and skin vessels of patients with CADASIL. *Stroke* 1994; **25**: 2291–2922.
20 Jones TA, Zou YJ, Kjeelgard M: Improved methods for building protein models in electron-density maps and location of errors in these models. *Acta Crystallogr* 1991; A **47**: 110–119.
21 Krawczak M, Cooper DN: Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum Genet* 1991; **86**: 425–441.
22 Dichgans M, Filippi M, Brüning R *et al*: Quantitative MRI in CADASIL: correlation with disability and cognitive performance. *Neurology* 1999; **52**: 1361–1367.
23 Oberstein SA, Ferrari MD, Bakker E *et al*: Diagnostic Notch3 sequence analysis in CADASIL: three new mutations in Dutch patients. Dutch CADASIL Research Group. *Neurology* 1999; **52**: 1913–1915.
24 Cooper DN, Antonarakis SE, Krawczak M: The nature and mechanisms of human gene mutation. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The Metabolic and Molecular Bases of Inherited Disease*. 7th edn. McGraw-Hill: New York, 1995, pp 259–291.
25 Lardelli M, Dahlstrand J, Lendahl U: The novel Notch homologue mouse *Notch 3* lacks specific epidermal growth factor-repeats and is expressed in proliferating neuroepithelium. *Mech Dev* 1994; **46**: 123–136.
26 Evans SV: SETOR: hardware lighted three dimensional solid model representations of macromolecules. *J Mol Graph* 1990; **11**: 134–138.