the FEBS Journal

The tetralogy of Fallot-associated G274D mutation impairs folding of the second epidermal growth factor repeat in Jagged-1

Corrado Guarnaccia, Somdutta Dhir, Alessandro Pintar and Sándor Pongor

Protein Structure and Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy

Keywords
Alagille syndrome; disease mutation; limited proteolysis; Notch signaling; oxidative folding

Correspondence
A. Pintar or S. Pongor, AREA Science Park, Padriciano 99, I-34149 Trieste, Italy
Fax: +39 040 226555
Tel: +39 040 3757354
E-mail: pintar@icgeb.org; pongor@icgeb.org

(Received 30 June 2009, revised 25 August 2009, accepted 27 August 2009)

doi:10.1111/j.1742-4658.2009.07333.x

Notch signaling controls spatial patterning and cell-fate decisions in all metazoans. Mutations in *JAG1*, one of the five Notch ligands in man, have been associated with Alagille syndrome and with a familial form of tetralogy of Fallot. A specific G274D mutation in the second epidermal growth factor repeat of the Jagged-1 was found to correlate with tetralogy of Fallot symptoms but not with usual Alagille syndrome phenotypes. To investigate the effects of this mutation, we studied the *in vitro* oxidative folding of the wild-type and mutant peptides encompassing the second epidermal growth factor. We found that the G274D mutation strongly impairs the correct folding of the epidermal growth factor module, and folding cannot be rescued by compensative mutations. The 274 position displays very low tolerance to substitution because neither the G274S nor the G274A mutants could be refolded *in vitro*. A sequence comparison of epidermal growth factor repeats found in human proteins revealed that the pattern displayed by the second epidermal growth factor is exclusively found in Notch ligands and that G274 is absolutely conserved within this group. We carried out a systematic and comprehensive analysis of mutations found in epidermal growth factor repeats and show that specific residue requirements for folding, structural integrity and correct post-translational processing may provide a rationale for most of the disease-associated mutations.

Introduction

The Notch signaling pathway is a highly connected and tightly regulated signal transduction framework that, together with a restricted number of other signaling networks, drives developmental processes in all metazoans. Notch signaling controls cell lineage decisions in tissues derived from all three primary germ lines: endoderm, mesoderm and ectoderm, thus playing an essential role in organogenesis [1–3]. Faults in the Notch-mediated signaling network have been associated with very different pathologies, such as some cancers (T-cell acute lymphoblastic leukemia, mucoepidermoid carcinoma) [4–6]; several genetic disorders [Alagille syndrome (AGS), tetralogy of Fallot, spondylocostal dysostosis, cerebral autosomal dominant arteriopathy with subcortical infarcts] [7]; and possibly autoimmune diseases, such as multiple sclerosis [8].

Both receptors and ligands are membrane-bound proteins, and this restricts signaling to adjacent cells. Of the five Notch ligands identified in man, Jagged-1

Abbreviations
AGS, Alagille syndrome; cbEGF, calcium-binding epidermal growth factor; DSL, Delta/Serrate/LAG2; EGF, epidermal growth factor; EGF1, first epidermal growth factor; EGF2, second epidermal growth factor; Fmoc, 9-fluorenylmethyloxycarbonyl; GSH, reduced glutathione; GSSG, oxidized glutathione; MIM, Mendelian Inheritance in Man; PDB, Protein Data Bank; TFA, trifluroacetic acid; TOF, tetralogy of Fallot.

and -2 are orthologs of *Drosophila* Serrate, whereas Delta-like-1, -3 and -4 are orthologs of *Drosophila* Delta. Jagged-1 is a single pass type I membrane protein with a large extracellular region made of a poorly characterized N-terminal region, a Delta/Serrate/LAG2 (DSL) domain, a series of 16 epidermal growth factor (EGF) tandem repeats and a cysteine-rich juxtmembrane region (Fig. 1). Mutations in the *JAG1* gene have been associated with AGS [Mendelian Inheritance in Man (MIM) database: #118450], a rare genetic disorder that can affect several organs, such as the liver, heart, eye, skeleton and kidneys [9]. More than 400 mutations in the *JAG1* gene have been identified, including missense, nonsense, deletion, insertion, splice site mutations and even complete gene deletion. AGS has an autosomal dominant inheritance and haploinsufficiency has been indicated as the main mechanism for the onset of this disorder. Recent studies, however, suggest that some mutations may have a dominant-negative character, leading to truncated soluble forms of JAG1 that can compete with the membrane-bound ligand [10,11]. AGS is a complex disorder with highly variable clinical symptoms and a clear genotype–phenotype association cannot always be established, with a few exceptions. A C234Y mutation in the first epidermal growth factor (EGF1) of Jagged-1 was found in a group of patients with deafness, congenital heart defects and posterior embryotoxon [12], and a G274D mutation in the second epidermal growth factor (EGF2) repeat of Jagged-1 was found in a familial form of tetralogy of Fallot (TOF) (MIM database: #187500) [13], comprising a heart malformation involving a large ventricular septal defect, pulmonary stenosis, right ventricular hypertrophy and an overriding aorta. Although cardiac defects are frequently found in AGS patients, none of the individuals with the G274D mutation displayed any other relevant clinical feature typical of AGS. The JAG1-G274D mutant protein can actually be expressed in NIH-3T3 cells, although in two different forms [14]. A fraction of JAG1-G274D is correctly processed, presented at the cell surface, and is functional, whereas another fraction is not fully glycosylated, is retained in the intracellular compartment, and is therefore inactive. This conclusion was supported by the sensitivity of JAG1-G274D to endoglycosidase H, which removes oligomannose and hybrid N-linked oligosaccharides, but not complex carbohydrates, by the incomplete digestion of the mutant protein in cells exposed to trypsin, and by the partial activation of Notch signaling detected through a reporter gene assay. The cardiac-specific phenotype associated with this mutation was explained in terms of a high sensitivity of the developing heart to Jagged-1 levels [14].

Despite the fact that EGF repeats are widespread in extracellular proteins and that hundreds of missense mutations have been identified and associated with several genetic diseases, the structural grounds of these disorders have been investigated only in a few cases, mostly related to calcium-binding EGFs (cbEGF). Detailed biochemical studies were carried out on mutations in CRB1 [15], the human orthologue of *Drosophila* Crumbs, fibrillin-1 [16–22], low density lipoprotein receptor [23] and human factor IX [19], which are associated with Leber congenital amaurosis, Marfan syndrome, familial hypercholesterolemia and emophilia B, respectively. The overall conclusion from these studies is that, in multidomain proteins, mutations can have different effects depending on the context, and a correlation between the genotype and the phenotype is still difficult.

To investigate the structural effects of the G274D mutation, and to attempt to correlate them with the available data obtained *in vitro* and *in vivo*, we initially synthesized a peptide corresponding to EGF2 of Jagged-1 (residues 263–295). This peptide, however,

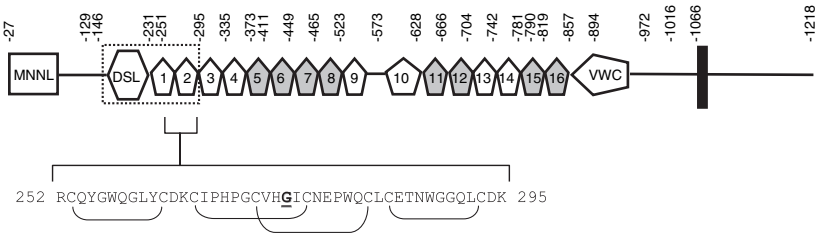

Fig. 1. Domain architecture of human Jagged-1. MNNL, N-terminal domain of Notch ligands; EGF domains are numbered progressively; potential calcium-binding EGF domains are shaded in gray; VWC, von Willebrand factor type C domain; the transmembrane segment is shown as a black bar; the receptor binding region is enclosed within a dashed rectangle. Amino acid residues corresponding to exon boundaries are shown above. The amino acid sequence of the J1ex6 peptide and the disulfide bond connectivities are also shown; the mutated glycine (G274) is shown in bold and underlined.

could not be refolded *in vitro* under the standard oxidative folding conditions used for other EGFs. Stemming from the observation that exon 6 of the *JAG1* gene encodes not only EGF2, but also part of EGF1, we speculated that exon 6 might encode an autonomously folding unit. We thus prepared a longer peptide encompassing the C-terminal part of EGF1 and the entire EGF2 (Fig. 1). This peptide, J1ex6 (residues 252–295), could be readily refolded *in vitro* and was shown to yield a structured unit with a disulfide bond topology typical of EGF repeats [24,25]. In the present study, we show that the G274D mutation associated with TOF strongly impairs the *in vitro* oxidative folding of this minimal folding and structural unit.

Results

The G274D mutation impairs folding of J1ex6

The solution structure of J1ex6 (residues 252–295) determined by NMR [Protein Data Bank (PDB) code: 2KB9] showed that the N-terminal overhang corresponding to the C-terminal part of EGF1 is not only required for folding, but is also an integral part of a structural unit that encompasses the EGF1 C-terminal region and the entire EGF2 [25]. The solution structure of this unit, including the conformation of the N-terminal overhang, is also very similar to the structure of the same region in a larger Jagged-1 construct comprising the DSL and EGF1, 2 and 3 domains, for which the crystal structure was determined recently [26] (PDB code: 2VJ2). We used this minimal folding and structural unit to address the structural grounds for the misfolding of the G274D mutant.

The results obtained with respect to the *in vitro* oxidative folding for the wild-type J1ex6 and its variants are summarized in Fig. 2. Folding of the wild-type J1ex6 lead to a largely prevalent product (> 86% by RP-HPLC area integration), with a very minor fraction of products that could be identified by LC-MS as mixed disulfides with glutathione (GSH). Because GSH is hydrophilic, adducts with GSH usually display shorter retention times in RP-HPLC compared to the native folded species. The folded species has a retention time that is only slightly shorter (~ 1 min) than that of the reduced peptide, suggesting that J1ex6 lacks a true hydrophobic core. Under the same experimental conditions, oxidative folding of the G274D mutant produced a very complex mixture. A clear separation of the different products in the mixture could not be achieved, but MS analysis revealed that most of the RP-HPLC peaks arise from adducts with one or more molecules of GSH. This suggests that, in the G274D

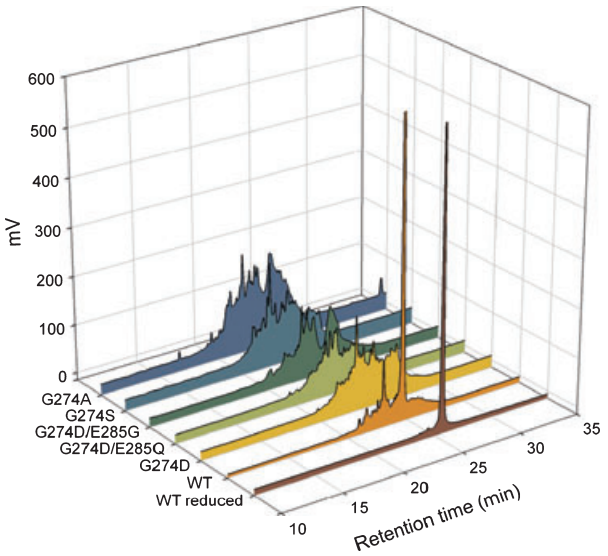

Fig. 2. Oxidative folding. RP-HPLC profiles for the *in vitro* oxidative folding of the wild-type J1ex6 and its variants in the presence of the GSH/oxidized GSH (GSSG) redox couple. The RP-HPLC profile of the purified, reduced J1ex6 peptide is also shown.

mutant, the correct folding and the complete formation of the four disulfide bonds cannot be accomplished, leaving one or more cysteines coupled to GSH and leading to shorter retention time species. By contrast, products at longer retention times may contain incorrect disulfide bonds that remain exposed to the solvent. The same results were obtained using redox buffers containing aromatic thiols, which were shown to enhance both folding rates and yields [27]: wild-type J1ex6 refolded in excellent yield, whereas the G274D mutant remained trapped in mixed disulfides forms. Figure 2 refers to a time point where equilibrium has been reached, but the same trend was observed at short refolding times. Although the RP-HPLC profile for the wild-type J1ex6 already showed a major product after 1 h, at the same time point, the profile of the G274D mutant displayed a complex pattern.

To confirm that the mixture of products obtained in the refolding of the G274D mutant is actually composed of misfolded species, we subjected it to proteolysis using proline endopeptidase, and analyzed the fragments by MALDI-TOF MS (Fig. 3). Proline endopeptidase was chosen because J1ex6 contains three prolines (P267, P269 and P279), all in the central part of the EGF2 sequence, and close to the mutated G274. Although the wild-type J1ex6 was scarcely affected after 20 h at 37 °C, the refolding mixture of the G274D mutant digested under the same conditions displayed a completely different MS profile, with an almost complete proteolysis of the G274D mutant into

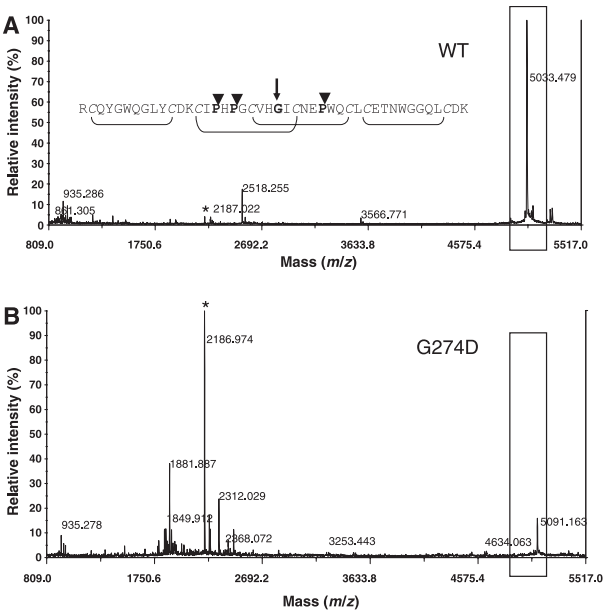

Fig. 3. Probing folding by proteolysis. MALDI-TOF analysis of the folding mixtures of (A) the wild-type (WT) and (B) the G274D mutant peptides subjected to proteolysis with proline endopeptidase for 20 h at 37 °C. Cleavage sites are indicated by triangles (▼), the mutated glycine by an arrow; the *m/z* region of the intact peptide is enclosed within a rectangle; for quantitative comparison, the intensity ratio between the *m/z* values of the intact peptide and the fragment of *m/z* = 2187 (labeled with an asterisk and corresponding to a C-terminal fragment of 16 residues) can be used.

small fragments. The much higher susceptibility of the G274D mutant to proline endopeptidase suggests the prevalence of misfolded species with a bead-like arrangement of the core disulfide bonds.

From the 3D structure of J1ex6, we speculated that misfolding of the G274D mutant could be a result of electrostatic repulsion or a steric clash with E285. If this was the case, the G274D mutant would be rescued by a compensative mutation at position 285 aimed either at neutralizing the negative charge or at reducing the steric hindrance of E285. Supported by the observation that position 285 shows a high variability (see below), we prepared the double mutants G274D/E285Q and G274D/E285G, purified them, and refolded under the same conditions used for the wild-type J1ex6. RP-HPLC profiles (Fig. 2) reveal, as for the G274D mutant, a complex pattern, suggesting that the tested putative compensative mutations cannot rescue the correct folding of the G274D mutant *in vitro*. To test the tolerance of position 274 to amino acid substitution, we prepared two additional mutants, G274S and G274A, replacing glycine with two small amino acids (i.e. serine or alanine, respectively). Also in this case, however, RP-HPLC analysis of the oxidative folding mixture (Fig. 2) showed a complex profile and the lack of a major product. In conclusion, position 274 is not tolerant to substitution, nor could putative compensative mutations rescue the folding of the G274D mutant. From a closer inspection of the J1ex6 structure, it should be noted that, to maintain the correct stereochemistry, any side chain at position 274 would point towards the interior of the domain. In other words, any amino acid other than glycine would require a substantial rearrangement of the backbone to reorient the side chain. The experiments performed in the present study demonstrated that this region of J1ex6 may be too rigid to allow for such a conformational change to occur.

The sequence pattern of EGF2 is unique

The very low tolerance of J1ex6 to amino acid substitution at position 274 lead us to investigate whether the sequence pattern associated with EGF2 is found in other proteins. A pattern search in SWISS-PROT (http://www.expasy.org/prosite/) produced 22 hits, which, surprisingly, are all related to Notch ligands in different organisms. In this dataset, G at position 274 is absolutely conserved. Extending the pattern search to trEMBL, we obtained 115 hits. A plot of Shannon entropy shows that, apart from cysteines, there are only two additional positions that display no variability at all, the first corresponds to G274 in the Jagged-1 sequence, and the second to G290 (Fig. 4). This supports the idea that, in this specially constrained type of EGF, position 274 is not tolerant to substitution.

Analysis of disease-associated mutations

Because the EGF domain is one of the most common structural building blocks in extracellular proteins [28,29], we decided to undertake a global analysis of disease-associated missense mutations found in EGF-containing proteins (Tables S1–S4). By far the most frequent disease-associated mutation found in EGF domains involves cysteine (48%) followed by arginine (11%) and glycine (10%). Although $R \rightarrow X$ and $G \rightarrow X$ mutations are also involved in polymorphism, $C \rightarrow X$ mutations are almost exclusively disease-associated. To take into account the relative abundance of certain amino acid types in EGF domains, which are notably cysteine-rich, the number of observed mutations was normalized for the amino acid content, and this mutation frequency was compared with that calculated for the reference dataset. The ratio between these two frequencies can be considered as a measure of the relative impact of a certain $AA_i \rightarrow X$ mutation in a

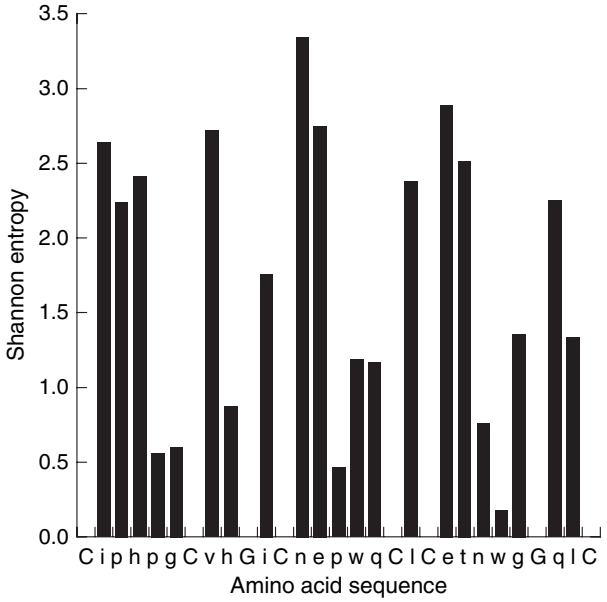

Fig. 4. Sequence variability. Sequence variability in a set of 115 EGFs matching the pattern {C-X(5)-C-X(4)-C-X(5)-C-X-C-X(8)-C} plotted as Shannon entropy versus position. Values for the Shannon entropy can vary from zero (no variability) to a maximum of 4.3. The amino acid sequence of Jagged-1 EGF2 (residues 265–293) is shown on the *x*-axis; amino acids in capital letters are totally conserved.

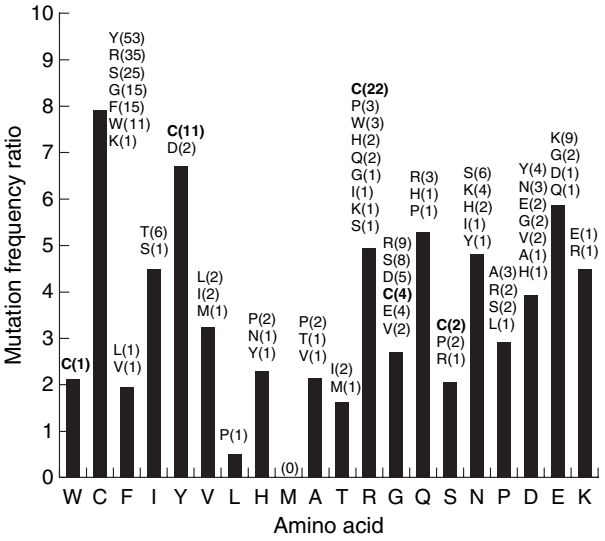

Fig. 5. Disease-associated mutations in EGF domains. The ratio between disease-associated mutation frequencies in EGF domains and the reference data set is plotted for each amino acid type. Amino acid types are shown in order of flexibility, as defined previously [41], from the least flexible (W) to the most flexible (K). The resulting amino acid and the number of occurrences for each mutation (in parenthesis) are shown above each bar. Mutations involving cysteines are shown in bold.

EGF domain (Fig. 5). Although normalization drastically reduces the weight of mutations involving cysteine, it is apparent that mutations either removing ($C \rightarrow X$) or introducing a cysteine ($X \rightarrow C$, similar to $Y \rightarrow C$ and $R \rightarrow C$) still have a great impact on EGF domains. This effect can be easily explained by the structural requirements of EGF domains, which, lacking a true hydrophobic core, are mainly stabilized by the three disulfide bridges. On the other hand, the introduction of an additional cysteine is likely to scramble the oxidative folding of EGF domains *in vivo*. Oxidation of cysteines to yield disulfide bonds is the most studied but not the only post-translational modification found in EGF domains [30]. β-hydroxylation of aspartate and asparagine, as well as different types of N- and O-glycosylation, has been reported. Although the role of β-hydroxylation remains elusive, correct O-glucosylation and O-fucosylation of serine/threonine residues has been shown to be required for correct signaling mediated by Notch receptors [31,32]. The impact of mutations involving these residue types might be related to these post-translational modifications, rather than to changes in the physico-chemical properties of a specific amino acid.

To analyze this latter aspect, we compared the disease-associated and neutral mutations in terms of the chemical distance, as measured by the Grantham score [33] (Fig. 6). Polymorphism-related mutations follow an almost bell-like distribution centered on rela-

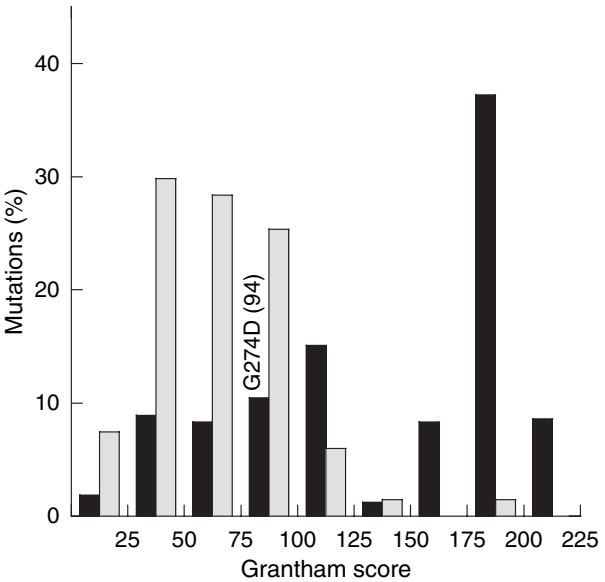

Fig. 6. Physico-chemical analysis of mutations. The percentage of disease-associated mutations (black bars) and polymorphism-related mutations (gray bars) are plotted versus their corresponding Grantham score.

tive small values of the Grantham score, whereas disease-associated mutations show an uneven distribution. Overall, it can be concluded that mutations with a high Grantham score are highly likely to be disease-associated, but the contrasting case is not true, at least for EGF domains, suggesting that the chemical distance is not the only determinant, as discussed above.

As a further step, we attempted to identify positions in the EGF scaffold that are most sensitive to mutations. This type of analysis, however, turned out to be problematic because of the very high variability in the amino acid sequence of EGF domains and in the length of the loops, which together make both sequence and structural alignments unreliable for this purpose. We thus decided to carry out this type of analysis on a coarser basis, dividing the sequence of EGFs into seven windows, w1 to w7, and partitioning mutations accordingly (Fig. 7). Polymorphism-related mutations show a relatively homogeneous distribution over the sequence, whereas disease-related mutations are mainly found in w1, w3, w4 and, to a minor extent, in w6. The relatively high frequency of disease-associated mutations in the N-terminal region most likely has no specific structural explanation, but is rather related to the strict requirement of specific amino acids (D/N) necessary for calcium coordination in calcium-binding EGF domains. On the other hand, mutations in w3 and w4 are more likely to disrupt the two-strand antiparallel β-sheet that is the main (and sometimes the only) secondary structure element in EGF domains, or to involve residues that are required for the correct formation of the interface between two consecutive EGF repeats. A separate positional analysis of cysteine mutations, which are all disease-associated, showed that they are equally distributed, with no significant prevalence of any of the six positions.

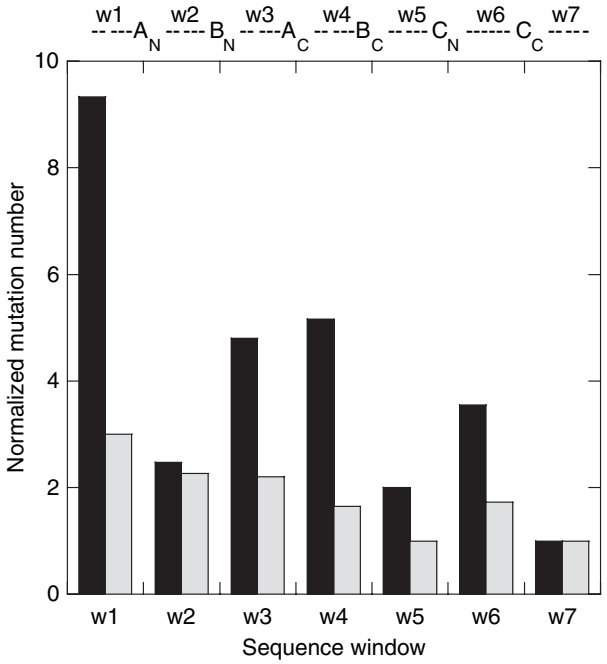

Fig. 7. Positional analysis of mutations. Disease-associated (black bars) and polymorphism-related (gray bars) mutations in EGF domains were partitioned according to their position in windows w1 to w7 and normalized for the average window size. Mutations involving cysteine were not considered. The six half-cystines are named according to the $A_NB_NA_CB_CC_NC_C$ annotation.

Discussion

The G274D mutation in EGF2 of Jagged-1 is occurring within the same window (w3 in Fig. 7) and at a position that is structurally equivalent to G1127 in fibrillin-1 and G106 in factor IX (Fig. 8). The G274D mutation, however, appears to affect folding in a more drastic way (Fig. 2) than the G1127S mutation in fibrillin-1 and the G106S mutation in factor IX [19,20]. This is likely a result of the higher constraints in the structure of this atypical EGF, as indicated by the shorter B_N-B_C loop (ten residues, compared to 13 in fibrillin-1 and 14 in factor IX) and spacing between the first and last half-cystines (the A_N–C_C distance is 27 residues in Jagged-1 EGF2, compared to 35 in fibrillin-1 and 30 in factor IX) and supported by the observation that glycine at that position is totally conserved in Notch ligands (Fig. 4). It is possible that the G274D mutation, introducing a larger charged amino acid, is more disrupting than a G → S mutation (a difference of 94 in the Grantham score, compared to 56 for a G → S mutation; Fig. 6). The misfolding of the G274S and G274A mutants (Fig. 2), however, supports the hypothesis that no amino acid other than glycine can be accommodated at that position, regardless of the substitution type. This low tolerance to substitution is consistent with the positive ϕ angle measured for G274 (Tables S5). For steric reasons, in protein structures, positive values of ϕ are observed

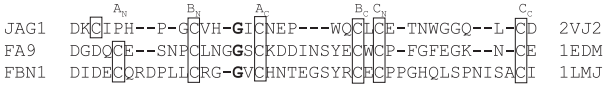

Fig. 8. Structural alignment. Multiple sequence alignment based on the structural alignment of EGF2 from Jagged-1 (JAG1; PDB code: 2VJ2), cbEGF1 from factor IX (FA9, PDB code: 1EDM) and cbEGF13 from fibrillin-1 (FBN1; PDB code: 1LMJ). Despite some discrepancy in the N-terminal region, half-cystines (boxed) and the mutated glycines (in bold) are aligned. Structure comparison was made using STAMP [42].

almost exclusively for glycine residues, and glycines that are both buried and have a positive ϕ angle tend to be highly conserved [34]. Misfolding of the G274S and G274A mutants suggests that the positive ϕ angle cannot be maintained upon introduction of any side chain, and also indicates that the backbone in this region of J1ex6 may be too rigid to allow extensive rearrangements to occur.

Additional missense mutations reported for exon 6 of *JAG1* and expected to induce an amino acid replacement include G256S in EGF1 [35], P269L [36], C271R [35], C284F [10,11,37], and W288C [10,37] in EGF2. All these six missense mutations share a common feature; they occur at residues that are either completely (positions 256, 271, 274 and 284) or very highly (positions 269 and 288) conserved in the amino acid sequence (Fig. 4). When considering all the 17 missense mutations occurring in the 16 EGF repeats of Jagged-1, ten involve either the replacement or the introduction of a cysteine, and are thus likely to be structurally disrupting (Fig. 5). Previously reported mappings of mutations over the Jagged-1 sequence [35–37] did not indicate the presence of any hot spot of critical region. Such mapping, however, was performed considering all types of possible mutations, including premature termination, and partitioning them over the 26 exons of the *JAG1* gene. Taking into account only missense mutations, which are likely to be more informative with respect to structural changes, and partitioning them over domains, rather than exons, it appears that the segment comprising the N-terminal domain, the DSL and the first two EGFs is most sensitive to missense mutations (Figure S1 and Tables S4). This is consistent with the DSL/EGF1-2 region being involved in receptor binding [26,38] and points to a key role of the N-terminal domain. From this map, it can be speculated that two additional regions, one extending over EGF12–14 and the other including the von Willebrand factor type C domain, might also play a yet unidentified structural or functional role.

The JAG1-G274D mutant cloned into a retroviral expression vector and expressed in NIH-3T3 cells was shown to be partially retained in the intracellular compartment and partially presented at the cell surface in a functional form. The cardiac-specific phenotype associated with this mutation was explained in terms of a high sensitivity of the developing heart to Jagged-1 levels [14], in accordance with a haploinsufficiency mechanism of the disease. The severe impairment of EGF2 folding observed *in vitro* and caused by the G274D mutation may actually reflect the *in vivo* misfolding and retention in the endoplasmic reticulum of Jagged-1, and is in agreement with the prevalent intracellular localization of the mutated Jagged-1 in NIH-3T3 cells. The question arises as to whether the fraction of the mutated Jagged-1 that is presented at the cell surface is correctly folded. The results obtained in the present study suggest the opposite. There are several lines of evidence in support of this hypothesis: the dramatic impairment of the oxidative folding *in vitro* induced by the G274D mutation, the misfolding of the G274S and G274A mutants, the impossibility of rescuing the G274D mutation with compensatory mutations, the sensitivity of the G274D mutant folding mixture to proteolytic cleavage, the steric requirements at position 274, the relatively highly constrained nature of this atypical EGF, and the strict conservation of G274 emerging from sequence analysis. Thus, it is unlikely that the EGF2 containing the G274D mutation can be correctly refolded, even minimally. It is possible, however, that the structural changes induced by the G274D mutation remain confined to EGF2, and that *in vivo*, the mutated Jagged-1 can be still be correctly processed and transported to the cell surface, as observed in NIH-3T3 cells. Correct trafficking has been reported for the G1127S mutant of fibrillin-1, which is normally secreted [21], and for a C284F mutant of Jagged-1 [11]. The C284F mutant was found to be correctly processed, glycosylated and targeted to the plasma membrane, despite the fact that this mutation is expected to disrupt the C-terminal disulfide of EGF2. Of the additional missense mutations reported for exon 6 of *JAG1*, no detailed biochemical studies are available for the G256S, P269L and C271R mutants. A normal level of mRNA transcript was detected for the W288C mutant, suggesting also in this case that the protein is likely to be expressed [10]. These results suggest that large multidomain proteins such as Jagged-1 can escape degradation and undergo normal trafficking if misfolding is restricted to a small region. Depending on the type and position of the mutation, folding kinetics and post-translational modifications also may play an important role. We could not identify any straightforward correlation between missense mutations within this region of Jagged-1 and a particular phenotype. Although the G274D mutation has been reported to affect heart development almost exclusively, the other mutations are associated with more classical symptoms of Alagille syndrome (e.g. liver, heart, face, eye and skeleton defects), although with slightly different patterns.

JAG1-G274D expressed in NIH-3T3 cells was shown to activate a response in NIH-3T3 cells transfected with a reporter plasmid containing a luciferase gene downstream of a Notch-activated promoter [14].

The response varied by approximately 20–60% compared to that of the wild-type, depending on the temperature. This experiment confirmed that the JAG1-G274D expressed at the cell surface is functional, although it was not conclusive with respect to binding efficiency because the activity was normalized for the total protein content, and not for the Jagged-1 actually expressed at the surface. How can this finding be reconciled with the 'local misfolding' model proposed above? Deletion studies on mouse Jagged-1 constructs demonstrated that the DSL domain is necessary and sufficient for binding to Notch receptors, with EGF1 and 2 substantially increasing the binding [38]. Although the structural determinants of the Notch/ligands interactions are not yet known in detail, the X-ray structure of a receptor binding module comprising the DSL and EGF1-3 domains revealed the presence of a patch of highly conserved residues on the DSL domain, which were shown to be functionally important [26]. It is therefore possible that the G274D mutation, although disrupting the correct fold of EGF2, leaves the DSL and EGF1 unaffected, thereby reducing, but not abolishing, binding to the receptor. Altered flexibility in the rod-like structure of the DSL/EGF1-3 structure [26] might also affect the docking of the ligand to the receptor.

The oxidative folding *in vitro* of larger constructs comprising modules adjacent to EGF2 may provide additional clues regarding the effects of mutations on the folding, structure and flexibility of this region.

Experimental procedures

Peptide synthesis

Peptides (44 amino acid long) corresponding to residues 252–295 of human Jagged-1 and its variants were synthesized on solid phase using 9-fluorenylmethyloxycarbonyl (Fmoc)/*O*-benzotriazolyl-1,1,3,3-tetramethyluronium hexafluorophosphate chemistry on a 0.05 mmol scale. The syntheses were automatically performed on a home-built automatic synthesizer based on a Gilson Aspec XL SPE (Gilson Inc., Middleton, WI, USA). All amino acids except cysteines were introduced as double couplings using a four-fold excess of amino acid (Fmoc-AA/[(1H-6-chlorobenzotriazol-1-yl)(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide/diisopropylethylamine; 1 : 1 : 2). Cysteine residues were instead introduced by double coupling as *N*-α-Fmoc-*S*-trityl-L-cysteine pentafluorophenyl ester to avoid cysteine racemization. Peptide cleavage/deprotection was performed by treatment with trifluroacetic acid (TFA)/ethandithiol/triisopropylsilane/H_2O (90 : 5 : 2.5 : 2.5) for 3 h at room temperature. The peptides were then precipitated with diethylether, washed and freeze-dried. The crude peptides were reduced by Tris(2-carboxyethyl) phosphine hydrochloride and purified by RP-HPLC on a Zorbax 300SB-C18 9.4 × 250 mm semipreparative column (Agilent Technologies Inc., Santa Clara, CA, USA) using H_2O/0.1% TFA and MeCN/0.1% TFA as the A and B eluents, respectively. The purified peptide fractions were analyzed by LC-MS to verify purity and molecular mass. The purified reduced peptide fractions were quantified by measuring UV A_{280} using a calculated extinction coefficient of 19630 $M^{-1}\cdot cm^{-1}$ and immediately used for oxidative folding experiments.

Oxidative folding

Fractions from RP-HPLC were diluted to a final peptide concentration of 0.1 $mg\cdot mL^{-1}$ in the degassed refolding buffer (0.25 M Tris-HCl, 2 mM EDTA, 3.7 mM GSH, 3.7 mM GSSG, pH 8) and refolded for 18 h at 4 °C. The folding reactions were stopped by acid quenching (TFA addition) and analyzed by RP-HPLC using a Zorbax SB300-C18 5 μm 4.6 × 150 mm column (Agilent Technologies Inc.) connected to a Gilson analytical HPLC using UV detection at 214 nm and MS detection (Applied Biosystems API 150EX; Applied Biosystems Inc., Foster City, CA, USA). The gradient for separation was 18–38% B in 40 min with H_2O/0.1% TFA and MeCN/0.1% TFA as the A and B eluents respectively.

Resilience to proteolysis was evaluated as follows. Equal amounts (~ 3 mg) of purified J1ex6 and of the G274D mutant were refolded as described above. Equal aliquots (~ 100 μg in 1 mL) of each folding mixture was quenched by addition of 20 μL of TFA and quickly desalted by RP-HPLC on a C18 analytical column (Zorbax SB300-C18, 4.6 × 150 mm; Agilent Technologies Inc.). The full range of peptides (including folding intermediates and mixed disulfides) was collected for both peptides and, in the case of J1ex6, the purified peak alone corresponding to the native folded form was also collected for comparison. Amounts of approximately 20 μg of each peptide were subjected to proteolysis with proline endopeptidase (peptide/protease; 20 : 1) for 20 h at 37 °C in 20 μL of ammonium acetate buffer (0.1 M, pH 5.8) containing 2.5 mM $CaCl_2$. Aliquots (1 μL) of the digestion mixtures were mixed with 9 μL of HCCA matrix (10 $mg\cdot mL^{-1}$) and analyzed by MALDI-TOF (Applied Biosystems 4800 Proteomics Analyzer) in reflector positive ion mode.

Sequence analysis

Sequence retrieval, filtering and analysis were carried out using in-house written PERL scripts. Multiple sequence alignment was performed using CLUSTALW (http://www.ebi.ac.uk/clustalw/).

Pattern searches (http://www.expasy.org/prosite/) in SWISS-PROT (release 55.1) or TREMBL (release 38.1) databases were carried out using either the {C-X(8)-C-X(1,2)-C-X(5)-C-X(4)-C-X(5)-C-X-C-X(8)-C} eight-cysteine motif that includes the EGF2 signature and the preceding disulfide bond loop or the {C-X(5)-C-X(4)-C-X(5)-C-X-C-X(8)-C} six-cysteine motif that characterizes EGF2. Sequence variability was estimated from the Shannon entropy calculated using the Sequence Variability Server (http://immunax.dfci.harvard.edu/Tools/svs.html).

Sequences of EGF domains containing annotated disease-associated mutations were retrieved from SWISS-PROT (release 55.6). Only EGF domains with three-disulfide bonds were considered for the present study, thus discarding the laminin and integrin-like EGF domains, which have one additional disulfide bond. Domain boundaries were considered as annotated in SWISS-PROT. Disease-associated mutations and neutral mutations (polymorphism) were collected separately. A total of 325 disease-associated mutations from 105 EGF domains in 21 different proteins were obtained (Tables S1 and S2). The neutral mutation dataset consisted of a total of 67 mutations from 64 EGF domains in 38 proteins (Tables S3). As a reference dataset, we used a collection of all disease-associated mutations described in the MIM database [39] and annotated in SWISS-PROT. This dataset comprises a total of 4236 mutations from 436 genes, regardless of protein function, cellular localization and domain type [40]. To compare the frequency of each disease-associated mutation type observed in EGF domains with that in the reference dataset, all disease mutations of the type $AA_i \rightarrow X$, where X is any amino acid, were collected, summed up for each amino acid type AA_i, and divided by the number of occurrences of AA_i, to obtain a normalized mutation frequency F_i for the EGF domain dataset and f_i for the reference dataset. The ratio F_i/f_i between these two frequencies was plotted for each amino acid type. To account for the very different size of the two datasets, the number of observed mutations in the reference dataset was first downscaled to the size of the EGF dataset.

Disease-associated and neutral mutations in EGF domains were also analyzed in terms of the Grantham score [33] associated with every mutation type. The Grantham score is a composite measure of the chemical distance between two amino acid types, and takes into account the molecular volume, polarity and side-chain composition of amino acid pairs. Grantham scores are in the range 5–215, with a higher number reflecting less conservative changes.

Mapping of disease-associated and neutral mutations onto the sequence of EGF domains was achieved by dividing the EGF sequence into seven windows, w1 to w7 (with w1 comprising the N-terminal residues, w2 to w6 comprising the residues delimited by disulfide bonds half-cystines, and w7 the C-terminal linker residues), counting the mutations occurring in each window, and dividing these values by the average number of residues in the window.

Acknowledgements

We thank Franco Pagani (ICGEB) for critically reading the manuscript.

References

1 Hurlbut GD, Kankel MW, Lake RJ & Artavanis-Tsakonas S (2007) Crossing paths with Notch in the hyper-network. *Curr Opin Cell Biol* **19**, 166–175.
2 Ehebauer M, Hayward P & Arias AM (2006) Notch, a universal arbiter of cell fate decisions. *Science* **314**, 1414–1415.
3 Bray SJ (2006) Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol* **7**, 678–689.
4 Roy M, Pear WS & Aster JC (2007) The multifaceted role of Notch in cancer. *Curr Opin Genet Dev* **17**, 52–59.
5 Koch U & Radtke F (2007) Notch and cancer: a double-edged sword. *Cell Mol Life Sci* **64**, 2746–2762.
6 Miele L, Golde T & Osborne B (2006) Notch signaling in cancer. *Curr Mol Med* **6**, 905–918.
7 Gridley T (2003) Notch signaling and inherited disease syndromes. *Hum Mol Genet* **12**, R9–R13.
8 John GR, Shankar SL, Shafit-Zagardo B, Massimi A, Lee SC, Raine CS & Brosnan CF (2002) Multiple sclerosis: re-expression of a developmental pathway that restricts oligodendrocyte maturation. *Nat Med* **8**, 1115–1121.
9 Spinner NB, Colliton RP, Crosnier C, Krantz ID, Hadchouel M & Meunier-Rotival M (2001) Jagged1 mutations in alagille syndrome. *Hum Mutat* **17**, 18–33.
10 Boyer J, Crosnier C, Driancourt C, Raynaud N, Gonzales M, Hadchouel M & Meunier-Rotival M (2005) Expression of mutant JAGGED1 alleles in patients with Alagille syndrome. *Hum Genet* **116**, 445–453.
11 Boyer-Di Ponio J, Wright-Crosnier C, Groyer-Picard MT, Driancourt C, Beau I, Hadchouel M & Meunier-Rotival M (2007) Biological function of mutant forms of JAGGED1 proteins in Alagille syndrome: inhibitory effect on Notch signaling. *Hum Mol Genet* **16**, 2683–2692.
12 Le Caignec C, Lefevre M, Schott JJ, Chaventre A, Gayet M, Calais C & Moisan JP (2002) Familial deafness, congenital heart defects, and posterior embryotoxon caused by cysteine substitution in the first epidermal-growth-factor-like domain of jagged 1. *Am J Hum Genet* **71**, 180–186.
13 Eldadah ZA, Hamosh A, Biery NJ, Montgomery RA, Duke M, Elkins R & Dietz HC (2001) Familial Tetralogy of Fallot caused by mutation in the jagged1 gene. *Hum Mol Genet* **10**, 163–169.
14 Lu F, Morrissette JJ & Spinner NB (2003) Conditional JAG1 mutation shows the developing heart is more sensitive than developing liver to JAG1 dosage. *Am J Hum Genet* **72**, 1065–1070.

15 Davis JA, Handford PA & Redfield C (2007) The N1317H substitution associated with Leber congenital amaurosis results in impaired interdomain packing in human CRB1 epidermal growth factor-like (EGF) domains. *J Biol Chem* **282**, 28807–28814.
16 Downing AK, Knott V, Werner JM, Cardy CM, Campbell ID & Handford PA (1996) Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* **85**, 597–605.
17 Suk JY, Jensen S, McGettrick A, Willis AC, Whiteman P, Redfield C & Handford PA (2004) Structural consequences of cysteine substitutions C1977Y and C1977R in calcium-binding epidermal growth factor-like domain 30 of human fibrillin-1. *J Biol Chem* **279**, 51258–51265.
18 Yuan X, Werner JM, Lack J, Knott V, Handford PA, Campbell ID & Downing AK (2002) Effects of the N2144S mutation on backbone dynamics of a TB-cbEGF domain pair from human fibrillin-1. *J Mol Biol* **316**, 113–125.
19 Whiteman P, Downing AK, Smallridge R, Winship PR & Handford PA (1998) A Gly –> Ser change causes defective folding in vitro of calcium-binding epidermal growth factor-like domains from factor IX and fibrillin-1. *J Biol Chem* **273**, 7807–7813.
20 Whiteman P, Smallridge RS, Knott V, Cordle JJ, Downing AK & Handford PA (2001) A G1127S change in calcium-binding epidermal growth factor-like domain 13 of human fibrillin-1 causes short range conformational effects. *J Biol Chem* **276**, 17156–17162.
21 Whiteman P & Handford PA (2003) Defective secretion of recombinant fragments of fibrillin-1: implications of protein misfolding for the pathogenesis of Marfan syndrome and related disorders. *Hum Mol Genet* **12**, 727–737.
22 Whiteman P, Willis AC, Warner A, Brown J, Redfield C & Handford PA (2007) Cellular and molecular studies of Marfan syndrome mutations identify co-operative protein folding in the cbEGF12-13 region of fibrillin-1. *Hum Mol Genet* **16**, 907–918.
23 Saha S, Boyd J, Werner JM, Knott V, Handford PA, Campbell ID & Downing AK (2001) Solution structure of the LDL receptor EGF-AB pair: a paradigm for the assembly of tandem calcium binding EGF domains. *Structure* **9**, 451–456.
24 Guarnaccia C, Pintar A & Pongor S (2004) Exon 6 of human Jagged-1 encodes an autonomously folding unit. *FEBS Lett* **574**, 156–160.
25 Pintar A, Guarnaccia C, Dhir S & Pongor S (2009) Exon 6 of human JAG1 encodes a conserved structural unit. *BMC Struct Biol* **9**, 43.
26 Cordle J, Johnson S, Tay JZ, Roversi P, Wilkin MB, de Madrid BH, Shimizu H, Jensen S, Whiteman P, Jin B *et al.* (2008) A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nat Struct Mol Biol* **15**, 849–857.
27 Gurbhele-Tupkar MC, Perez LR, Silva Y & Lees WJ (2008) Rate enhancement of the oxidative folding of lysozyme by the use of aromatic thiol containing redox buffers. *Bioorg Med Chem* **16**, 2579–2590.
28 Campbell ID & Bork P (1993) Epidermal growth factor-like modules. *Curr Opin Struct Biol* **3**, 385–392.
29 Wouters MA, Rigoutsos I, Chu CK, Feng LL, Sparrow DB & Dunwoodie SL (2005) Evolution of distinct EGF domains with specific functions. *Protein Sci* **14**, 1091–1103.
30 Harris RJ & Spellman MW (1993) O-linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **3**, 219–224.
31 Haines N & Irvine KD (2003) Glycosylation regulates Notch signalling. *Nat Rev Mol Cell Biol* **4**, 786–797.
32 Stanley P (2007) Regulation of Notch signaling by glycosylation. *Curr Opin Struct Biol* **17**, 530–535.
33 Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864.
34 Overington J, Donnelly D, Johnson MS, Sali A & Blundell TL (1992) Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. *Protein Sci* **1**, 216–226.
35 Warthen DM, Moore EC, Kamath BM, Morrissette JJ, Sanchez P, Piccoli DA, Krantz ID & Spinner NB (2006) Jagged1 (JAG1) mutations in Alagille syndrome: increasing the mutation detection rate. *Hum Mutat* **27**, 436–443.
36 Ropke A, Kujat A, Graber M, Giannakudis J & Hansmann I (2003) Identification of 36 novel Jagged1 (JAG1) mutations in patients with Alagille syndrome. *Hum Mutat* **21**, 100.
37 Crosnier C, Driancourt C, Raynaud N, Dhorne-Pollet S, Pollet N, Bernard O, Hadchouel M & Meunier-Rotival M (1999) Mutations in JAGGED1 gene are predominantly sporadic in Alagille syndrome. *Gastroenterology* **116**, 1141–1148.
38 Shimizu K, Chiba S, Kumano K, Hosoya N, Takahashi T, Kanda Y, Hamada Y, Yazaki Y & Hirai H (1999) Mouse jagged1 physically interacts with notch2 and other notch receptors. Assessment by quantitative methods. *J Biol Chem* **274**, 32961–32969.
39 Hamosh A, Scott AF, Amberger JS, Bocchini CA & McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* **33**, D514–D517.
40 Vitkup D, Sander C & Church GM (2003) The amino-acid mutational spectrum of human genetic disease. *Genome Biol* **4**, R72.
41 Vihinen M, Torkkila E & Riikonen P (1994) Accuracy of protein flexibility predictions. *Proteins* **19**, 141–149.
42 Russell RB & Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* **14**, 309–323.

Supporting information

The following supplementary material is available:

Fig. S1. Plot of missense mutations found in *JAG*1 and associated with Alagille syndrome.

Table S1. List of EGF-containing proteins and the names of the mutation-associated diseases.

Table S2. List of disease-associated missense mutations found in EGF repeats.

Table S3. List of neutral mutations found in EGF repeats.

Table S4. List of missense mutations found in *JAG*1 and associated with Alagille syndrome.

Table S5. Solvent accessible surface, ϕ angles and conservation of glycines in J1ex6.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.