The Crystal Structure of Fibroblast Growth Factor (FGF) 19 Reveals Novel Features of the FGF Family and Offers a Structural Basis for Its Unusual Receptor Affinity[†,‡]

Nicholas J. Harmer,* Luca Pellegrini, Dima Chirgadze, Juan Fernandez-Recio, and Tom L. Blundell

Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom

Received July 24, 2003; Revised Manuscript Received October 29, 2003

ABSTRACT: The 22 members of the FGF family have been implicated in cell proliferation, differentiation, survival, and migration. They are required for both development and maintenance of vertebrates, demonstrating an exquisite pattern of affinities for both protein and proteoglycan receptors. FGF19, one of the most divergent human FGFs, is unique in binding solely to one receptor, FGFR4. We have used molecular replacement to solve the crystal structure of FGF19 at 1.3 Å resolution using five superimposed FGF structures as the search model. The structure shows that two novel disulfide bonds found in FGF19, one of which appears to be conserved among several of the other FGFs, stabilize extended loops. The key heparin-binding loops of FGF19 have radically different conformations and charge patterns, compared to other FGFs, correlating with the unusually low affinity of FGF19 for heparin. A model for the complex of FGF19 with FGFR4 demonstrates that unique sequences in both FGF19 and FGFR4 are key to the formation of the complex. The structure therefore offers a clear explanation for the unusual affinity of FGF19 for FGFR4 alone.

Fibroblast growth factors (FGFs)[1] are a family of polypeptide growth factors that share a common core (*1*, *2*). FGFs are found in all animals from *Caenorhabditis elegans* to vertebrates, as well as in some arthropod viruses. FGF signaling has been implicated in cell proliferation, differentiation, survival, and migration and is required for both development and maintenance of vertebrates. Recent evidence suggests that FGF signaling is also involved in altering cellular and system responses to metabolites (*3−5*), highlighting the importance of this family of growth factors.

There are 22 human FGF paralogues, ranging in size from 17 to 34 kDa. These proteins share a common core of around 140 amino acids, which has homology to the interleukin 1β (IL-1β) family of growth factors. The most recently identified members of the family have been established as FGFs by virtue of their sequence being more similar to the canonical FGFs than to IL-1β. Crystal structures have been solved for 6 of the 22 FGFs [FGF1 (*6*), FGF2 (*7*, *8*), FGF4 (*9*), FGF7 (*10*), FGF9 (*11*, *12*), and FGF10 (in complex with receptor) (*13*)]. These proteins all form a β-trefoil structure (Figure 1), an all-β structure with a pseudo-3-fold axis of symmetry. NMR results from FGF1 and FGF2 suggest that the FGF sequence corresponding to strand 11 in IL-1β does not form a genuine β-strand (*14*, *15*) and that it is more flexible than the remainder of the FGF core (*16*).

FGFs bind to two types of molecules on the cell surface. On one face of the FGF, they form a low-affinity interaction with heparan sulfate (HS). HS is a highly sulfated polysaccharide, consisting of repeating units of a disaccharide of a uronic acid and glucosamine (*17*). HS chains tend to be 100−300 saccharides long and are attached to the extracellular regions of members of the glypican and syndecan protein families. Each HS chain contains alternating regions of low and high sulfation, with an average of three highly sulfated regions of 12−14 saccharides per chain. In the highly sulfated regions, each disaccharide has three sulfate groups added, to give a total of two negative charges per monosaccharide unit. It has long been known that HS is a requirement for the cellular responses to FGF (*18*, *19*). Recently, several crystal structures have defined the nature of binding of FGFs to heparin, a soluble analogue of HS (*20−23*). Comparison of these structures has suggested that HS forms its strongest interactions to heparin via two or three sulfate groups closely positioned in space, which bind intimately into the protein backbone of the FGF (*24*). These structural results correlate with recent biochemical findings (*25*, *26*), and consensus sulfation patterns for the binding of several FGFs are being elucidated (*10*).

On a second face, FGFs bind to FGF receptors (FGFRs), their second cellular binding receptors. FGFRs are integral membrane proteins with a single transmembrane helix. The intracellular region of FGFRs consists largely of a split tyrosine kinase domain, and so FGFRs are hypothesized to act as classical tyrosine kinase receptors. The extracellular region contains three immunoglobulin-like (IgG) domains, D1−D3. The membrane proximal IgG domains D2 and D3 bind to FGFs, and the D2 domain also binds heparin and HS. Indeed, it is likely that FGFRs are permanently com-

---

[1] Abbreviations: FGF, fibroblast growth factor; FGFR, fibroblast growth factor receptor; HS, heparan sulfate; IL-1β, interleukin 1β.
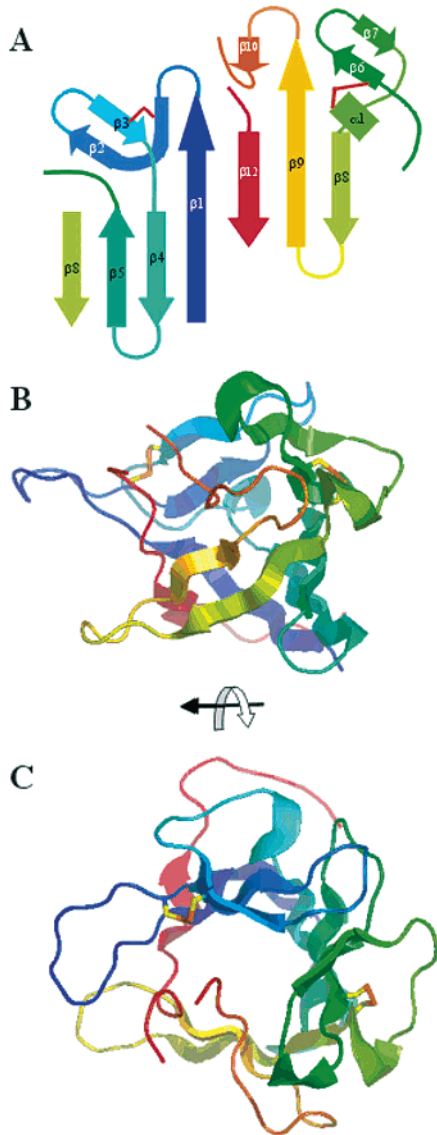
FIGURE 1: Overview of the FGF19 structure. Images show the cartoon representation of the protein backbone. The backbone is colored blue (N-terminus) to red (C-terminus). The four cystine residues are shown with Cα, Cβ, and Sγ as sticks. Sulfur is shown in orange. (A) Schematic of the secondary structure. β8 is shown twice to emphasize that strands β1, β4, β5, β8, β9, and β12 form a β-barrel. The position of disulfide bridges is shown by maroon links. (B, C) Orthogonal views of the molecule. (B) shows the β-barrel at the bottom and loops at the top. (C) View down the β-trefoil with the β-barrel at the bottom.

plexed with HS on the cell surface (*27*). FGFRs have been found in all species with FGFs, and in mammals there are five homologues (FGFR1−5). Of these, FGFR1−3 contain a major splice variation in the membrane proximal IgG domain, giving rise to two possible receptor forms of the FGFR. These lead to significant alterations in the FGF binding specificities of the receptor (*13, 28*). Therefore, there are in total eight different possible FGF-binding surfaces in mammalian FGFR ectodomains. Each FGF binds to a number of FGFRs, with differing affinities for each (*28*). Thus, the specificity of FGF binding to cellular targets will be affected by both the range of FGFRs that the cell expresses and the patterning of HS that is generated by the cellular enzymes. The FGF−FGFR−heparin system offers vertebrates a method for generating exquisitely regulated

signaling networks and gradients. It is little wonder that FGFs have been implicated in the development of every major structure and tissue type in the human body.

A number of complexes of FGFs and FGFRs have been studied by X-ray crystallography (*13, 20, 21, 29−31*). These complexes have revealed a binding interface on the surface of the FGFR that encompasses both of the IgG domains predicted by biochemical methods and the region linking these two domains. The considerable similarities in these structures, and the high sequence identity of the human FGFR ectodomain, suggest that the interaction sites observed are likely to be conserved for all FGF−FGFR pairs. However, there are some differences in the orientations of the FGFR domains relative to one another, particularly with respect to D3 and FGF (*13, 21, 30*). This suggests that shifts in the domains may be a means of obtaining optimal binding of each FGF to its receptors. A more significant difference is observed in the two complexes including FGF, FGFR, and heparin (*20, 21*). While these show very similar interactions of each molecule to each other molecule (*24*), the manner in which the FGFRs are dimerized is extremely different. This has generated a significant controversy in the field, as the two architectures for complex formation can be observed using essentially identical protein constituents. The relevance of each of these complexes for the physiological signaling complex has still not been ascertained.

The vast majority of the current literature on FGFs and FGFRs has concentrated on the two canonical FGFs, FGF1 and FGF2, and on FGFR1 and FGFR2. A large literature has surrounded these molecules, and their interactions are becoming extremely well mapped. However, FGF1 and FGF2 bear more sequence similarity to one another than to any other FGFs, and a large number of the most recently isolated FGFs are extremely different in sequence to the canonical members of the family (*2, 32*). The only structure of a noncanonical FGF (FGF10) in complex with its receptor has highlighted important differences in the binding of the receptor to the FGF (*13*). There is, therefore, a need for further data on the biology of the other members of the FGF family, not least because a number of these have been implicated in inherited diseases (*5*).

FGF19 (*33, 34*) is a member of the most distant of the seven subfamilies of the FGFs. Uniquely among the FGFs, it has affinity for only one FGFR, FGFR4 (*34*). FGFR4 is notable among the FGFRs in that it does not possess a splice variant in its extracellular domain and in that FGFR4 signaling is not capable of supporting the proliferation of cells in vitro. FGF19 expression has been observed in a range of fetal tissues (particularly the brain) and in the adult gall bladder (*33, 34*). Analysis of chicken embryos has shown that FGF19 is a key regulator of the development of the inner ear (*35*).

However, other effects of FGF19 have been hinted at by recent research. FGF19 is located on a region of chromosome 11 that has been linked to the rare disorder osteoporosis pseudoglioma syndrome, consistent with FGF19 expression in the fetal muscle and cartilage. FGF19 is upregulated in one human colon cancer cell line (*34*), and ectopic expression of FGF19 in murine skeletal muscle can lead to cancer formation (*36*). Perhaps the most intriguing observation is that the ectopic expression of FGF19 under the control of the myosin light chain promoter leads to chronic weight loss

in the mice, even in genetic backgrounds where obesity is the expected phenotype. This finding, coupled to recent evidence that FGFR4 is responsible for subtle alterations to the mammalian metabolism (*3*, *37*) and that FGF19 is involved in regulation of bile acid biosynthesis (*38*), suggests that there may be important roles for FGF19 in the adult as well as the developing mammal.

Here we report the crystal structure of FGF19, solved to 1.3 Å resolution. This structure reveals a number of features of FGF19 that have not been observed in the structures of other FGFs. These include the presence of two disulfide bonds in FGF19, where none has been seen in other FGFs; an extended loop appears to occlude the normal HS binding surface of the FGFs, but this loop may provide a means for binding HS in a different fashion; the structure also offers a rationale for the unusual FGFR binding properties of this molecule. These observations should help in determining the biology of not only FGF19 but also that of the other more poorly characterized FGF family members.

EXPERIMENTAL PROCEDURES

Chemicals. Chemicals were obtained from Sigma (St. Louis, MO) or from Melford Laboratories (Ipswich, U.K.).

Preparation of FGF19. The FGF19 cDNA clone 3849343 was ordered from the IMAGE Consortium (*39*), held at the MRC Geneservice (Babraham, U.K.). DNA corresponding to residues 39−196 of FGF19 were amplified by PCR and cloned into the pGAT2 expression vector (gift of M. Hyvonen, Cambridge, U.K.). 6His-GST-FGF19 was expressed in the Rosetta strain of *Escherichia coli* (Novagen, Madison, WI) and proved to be insoluble. Inclusion bodies were isolated, washed, and unfolded in 6 M guanidine with 50 mM Tris-HCl, pH 8.0, and 1 mM DTT. FGF19 was refolded by slow dropwise addition to 9 volumes of 0.5 M L-arginine, 0.1 M NaCl, 0.1 M Tris-HCl, and 5 mM reduced glutathione at pH 8.5.

The refolded protein was purified by affinity to nickel-NTA (Qiagen, Crawley, U.K.) and then by affinity to glutathione−Sepharose (Amersham Biosciences, Uppsala, Sweden). The sample was cleaved with thrombin to release the FGF19. FGF19 was purified from the tags by affinity to heparin−Sepharose (Amersham Biosciences), and final purification was performed by ion exchange using a Resource-Q column (Amersham Biosciences). The protein was then dialyzed into 10 mM Tris-HCl and 200 mM NaCl prior to crystallization.

Crystallization and Structure Solution. FGF19 at 500 μM was crystallized using the hanging drop technique. The optimal mother liquor was 12−14% (w/w) PEG-3350, 0.25 M MgSO$_4$, and 0.1 M potassium phosphate, pH 7.0. Diffraction-quality crystals grew over 7−14 days. Cryoprotection was achieved by the slow addition of 25% glycerol into the mother liquor. Data were collected to 1.3 Å at the Daresbury Synchrotron Radiation Source, station PX14.2, at 100 K. X-ray data were indexed and scaled using the HKL package (*40*). To provide phases, the structures of FGF1, FGF2, FGF4, FGF7, and FGF9 (PDB codes AFGF, 1BFF, 1IHK, 1QQK, and 1IJT) were used as search models using AMORE (*41*). As none of these were successful on their own, they were superimposed using the program COMPARER (*42*). The FGF19 sequence was aligned to these structures using FUGUE (*43*). The results from FUGUE were hand-edited at the C-terminus of the protein, where we considered that a better alignment could be made. All side chains that were not conserved in FGF19 were replaced by alanine in the superimposed models, and loops that were not considered to be similar in FGF19 were removed. This composite model was used as a search model for molecular replacement using AMORE. This gave a family of significant hits, which provided sufficiently good phases to start structure solution. Waters were added using Arp/Warp (*44*). Following this, the remainder of the model was built by hand, using the Xtalview viewer (*45*). Refinement was carried out using first CNS (*46*) and then Refmac (*47*) at the later stages. The model was validated using PROCHECK (*48*), WHATCHECK (*49*), and MOLPROBITY (*50*). Molecular images were created using PYMOL (*51*).

Alignment and Model Generation. For the analysis of FGF19 heparin binding, FGF19 was superimposed onto the position of FGF1 in the structure 1AXM using COMPARER.

To generate models for all of the FGFs without structures, the FGF structures 1IHK, 1QQK, 1IJT, and FGF19 were superimposed. For each FGF, an alignment of all available homologues was created using CLUSTALX, and alignments to the superimposed structures were generated using FUGUE. Models were then built from these alignments using SCORE (*52*) to build the core main chain, RAPPER (*53*, *54*) to build main chain loops, and CELIAN (*55*) to build side chains. A final alignment of all models and structures was built using COMPARER. Alignment annotation was performed using JOY (*56*). Phylogenetic trees were generated using TraceSuite II (*57*).

Computational Modeling of the FGF19−FGFR4 Interaction. A model for FGFR4 was based on the structures of FGFR2 and FGF2 complexes (1EV2 and 1FQ9). An alignment of human FGFR4 to the FGFR2 structure 1EV2 was performed using CLUSTALX (*58*). A number of other FGFR1, FGFR2, and FGFR4 sequences were also used as inputs to ensure that the optimal alignment was chosen. A model of FGFR4 was built from this as above. The FGFR4 model and FGF19 were then superimposed into the positions of FGFR1 and FGF2 in the 1FQ9 structure using COMPARER.

Optimization of the FGFR4 interacting side chains in the complex was performed by a previously described protocol for the refinement of rigid-body docking interfaces (*59*). The interaction energy was computed from potentials precalculated on a 3-D grid surrounding the 10 Å vicinity of the FGF19 molecule. Conformational sampling was performed by biased-probability Monte Carlo minimization (*60*) (as implemented in the MolSoft ICM 3.0 program; http://www.molsoft.com) of the torsional angles of all surface FGFR4 side chains in the 10 Å vicinity of the FGF19 molecule. A total number of 760000 energy evaluations (10000 per torsional angle) were achieved during the minimization, after which global energy was completely stabilized. Total computational times were around 5 h in a single 2.4 GHz Pentium-4 CPU running Linux.

Further exploration of the conformational space of the FGF19−FGFR4 refined model was achieved by a new developed protocol for hinge-bending docking, in which domains D2 and D3 in FGFR4 were treated as rigid bodies

Table 1: Sequence Identities and Backbone Atom RMSDs for the Solved FGF Structures[a]

	FGF1	FGF2	FGF4	FGF7	FGF9	FGF10	FGF19
FGF1		56	38	37	41	32	33
FGF2	0.515		44	38	40	37	34
FGF4	0.637	0.597		37	44	34	33
FGF7	0.871	0.807	0.855		42	55	38
FGF9	0.626	0.595	0.606	0.862		43	30
FGF10	0.632	0.678	0.604	0.800	0.978		35
FGF19	0.973	0.854	0.887	1.052	0.866	0.978	

[a] A structure-based sequence alignment of the FGFs with solved structures was prepared using COMPARER (*42*). The COMPARER algorithm generates RMSD values for the backbone atoms at structurally equivalent positions in each pair of molecules. An alignment of the core regions of the FGFs was generated using FUGUE (*43*) (to ensure that the missing region of FGF19 was included in the calculation). Sequence identities were calculated from this alignment using JOY (*56*). The structures used are 1JQZ (FGF1), 2FGF (FGF2), 1IJT (FGF4), 1QQK (FGF7), 1IHK (FGF9), 1NUN (FGF10), and 1PWA (FGF19). Sequence identities (%) are shown above the diagonal, and backbone atom RMSDs (Å) are shown below the diagonal.

(e.g., fixed backbone and side-chain atoms) and the link between them ($E_{250}RSPHRS_{256}$) was fully flexible (all torsional angles of the linker residues E250−S256 were freely allowed to move) with loose restraints imposed to keep D2 in the vicinity of the starting conformation during the simulations. The starting conformation was prepared by randomizing the values of all torsional angles of the flexible linker. Biased-probability Monte Carlo minimization (as implemented in ICM 3.0) simultaneously explored the positional variables of the whole molecule and backbone, side-chain torsional angles of the linker residues. D3 movements were thus completely unrestricted, other than the geometrical restraints imposed by the linker conformation. The energy of the system was calculated as previously described for the interface refinement. Several minimization trajectories running from totally different random starting positions converged to the same energy values (1000000 energy evaluations per run; around 12 h in a single 2.4 GHz Pentium-4 CPU running Linux).

For comparison, similar models of the FGF19−FGFR1 and FGF19−FGFR2 low-affinity complexes were built on the basis of the FGF2−FGFR1 (1FQ9) and FGF1−FGFR2 (1DJS) complexes, respectively. Their interfaces were optimized by using the refinement protocol described above.

(For a more detailed description of the methods using in the docking experiments, see Supporting Information.)

RESULTS AND DISCUSSION

Structure Determination. Residues 39−196 of FGF19 were expressed with a 6His-GST tag in *E. coli* and refolded in vitro from inclusion bodies. This fragment contains the predicted β-trefoil core of FGF19. The FGF19 was cleaved from the tag, purified to homogeneity by ion-exchange chromatography, and crystallized. Preliminary attempts to solve the structure by molecular replacement, using individual FGFs as search models, proved unsuccessful. This was not unexpected, due to the low sequence identity between FGF19 and the other solved FGFs (Table 1). Molecular replacement was successfully performed using a model consisting of the superimposed structures of FGF1, FGF2, FGF4, FGF7, and FGF9, with nonidentical side chains

Table 2: Crystallographic Data

space group	R32
unit cell parameters	$a = 67.573, c = 193.374$
wavelength	0.975
resolution (Å)	23.33−1.3
reflections (unique)	42303
redundancy (outer shell)	12.08 (11.30)
completeness (%) (outer shell)	99.8 (100)
R_{sym} (%)[a] (outer shell)	6.2 (58.6)
$I/\sigma(I)$ (outer shell)	36.5 (4.33)
no. of non-H atoms	1277
overall R factor (%)[b]	18.1
free R factor (%)[c]	19.5
average B (Å²)	20.133
B_{11} (Å²)	−0.34
B_{22} (Å²)	−0.34
B_{33} (Å²)	0.51
B_{12} (Å²)	−0.17
B_{13} (Å²)	0.00
B_{23} (Å²)	0.00
average B, main chain (Å²)	14.274
average B, side chain (Å²)	16.644
average B, water	36.187
RMSD bonds (Å)	0.010
RMSD angles (deg)	1.44
RMSD bonded B (Å²)	1.64

[a] $R_{sym} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I_i(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. [b] R factor $= \sum_{hkl}||F_o| - |F_c||/\sum_{hkl}|F_o|$. [c] The free R factor was calculated from 5% of the data.

(approximately 70%) in the template structures replaced with alanine and loops that were considered to be dissimilar in FGF19 removed from the template structures entirely. As molecular replacement searches are conducted in Patterson space, where interatomic vectors are seen as peaks rather than atoms, the molecular replacement search requires that a significant number of correct interatomic vectors are present. The use of a greater number of models will lead to more correct vectors being present, especially the most sensitive, high-resolution vectors, and so we suspect that the composite probe, while having more inaccurate vectors, had sufficient correct vectors for a good solution to be raised above the noise.

The structure was then refined to 1.3 Å resolution using conventional methods. The refined model consists of 1 FGF19 molecule, 1 sulfate ion, 1 2-amino-2-(hydroxymethyl)-propan-1,3-diol (Tris) molecule, 1 glycerol molecule, and 267 water molecules. Only residues 41−146 and 160−176 of FGF19 could be resolved in the model. Although considerable electron density is present in the region corresponding to the missing 13 amino acids, the density was not of sufficient quality to allow the unambiguous placement of further protein in this region. This area was left unmodeled. Data collection and refinement statistics are shown in Table 2.

The FGF19 Structure in Comparison to Other FGFs. The solution of this structure by molecular replacement implies that the fold of FGF19 is very similar to that of the other FGFs, and this is indeed the case. FGF19 forms a β-trefoil-like structure (Figure 1). The backbone atom RMSD for FGF19 to the other FGFs is larger than that found between the other FGFs (Table 1). However, given that FGF19 has the lowest sequence identity of these FGFs, this increase is not surprising. The β-barrel at the base of the trefoil (strands 1, 4, 5, 8, 9, and 12) is well formed, with two of the β-strands (1 and 9) extended to form β-sheets with the exterior β-hairpins. Two features of this structure that diverge from
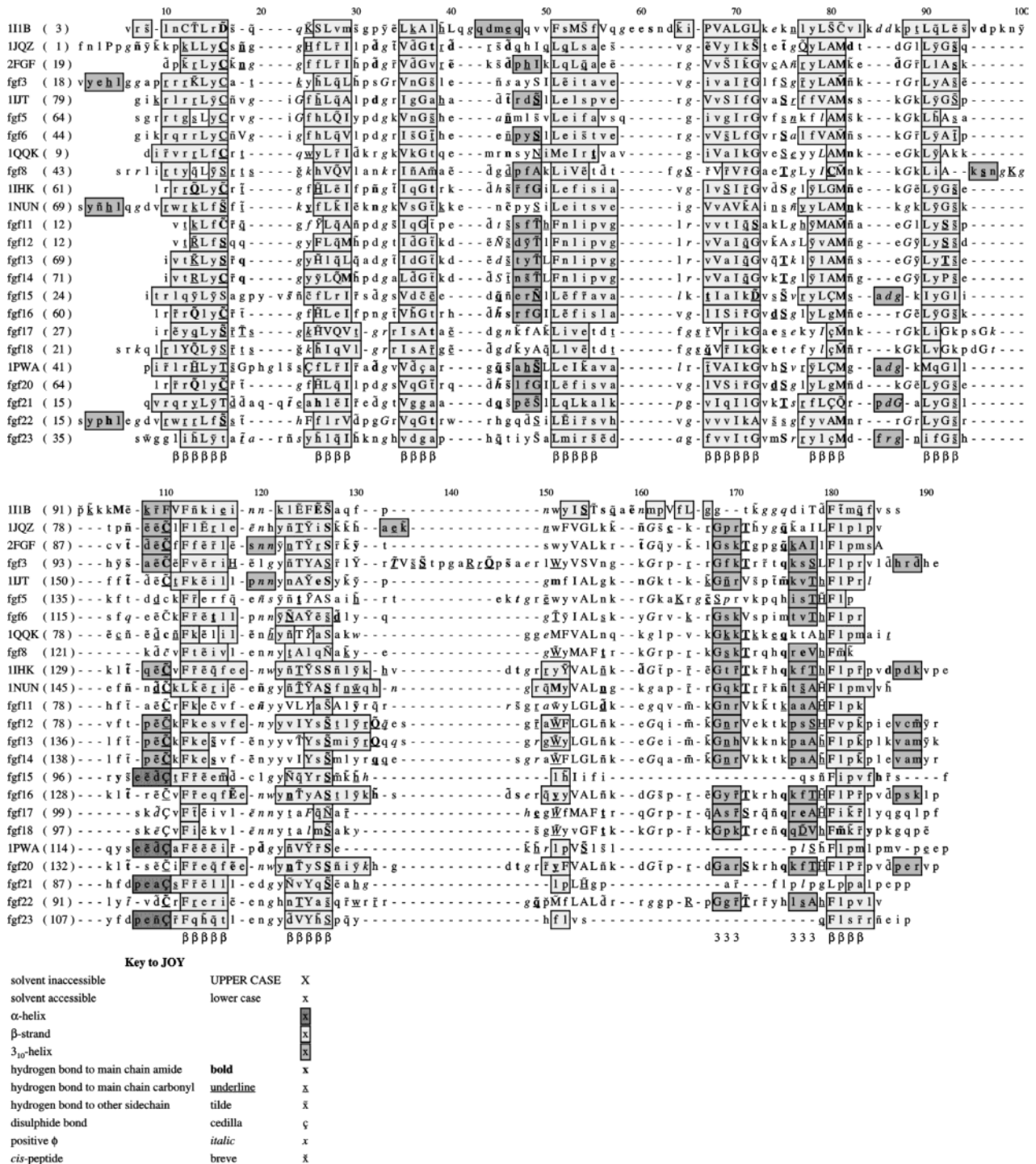
FIGURE 2: Alignment of the human FGFs. Models were built for all of the human FGFs (and mouse FGF15) using the seven solved FGF structures as a template. These models, the structures of FGF1, FGF2, FGF4, FGF7, FGF9, FGF10, and FGF19 and IL-1β, and two previously described *C. elegans* (*61*) models were superimposed to generate an optimal alignment for all of the FGFs. This alignment was annotated using JOY (*56*). PDB codes: 1I1B, IL-1β; 1JQZ, FGF1; 2FGF, FGF2; 1IJT, FGF4; 1QQK, FGF7; 1IHK, FGF9; 1NUN, FGF10; 1PWA, FGF19. For a larger, color version, see Figure 1 in Supporting Information.

the other FGFs are immediately apparent: FGF19 has two disulfide bonds (compared with none in the other FGFs), and the region corresponding to strand 11 is not resolved in the structure.

The significance of these features is emphasized by an alignment of models of the FGF family (Figure 2; for a larger, color version, see Figure 1 in Supporting Information). With the additional information provided by the FGF19 structure, equivalent disulfide bonds to the second observed in FGF19 are modeled in FGF8, FGF15, FGF17, FGF18, FGF21, and FGF23. In addition, the two *C. elegans* homologues show cysteines in the equivalent positions, and so these are likely also to form a disulfide. In contrast, only FGF15 shows an equivalent disulfide to the first disulfide seen in FGF19. In the cases of FGF15, FGF21, and FGF23, the modeling software did not build a model for strand 11. This is in part due to the high similarity of the rest of these models for FGF19 and in part due to the low similarity to the equivalent loops in the other FGFs. It is therefore clear that these features are not a peculiarity of FGF19 but are
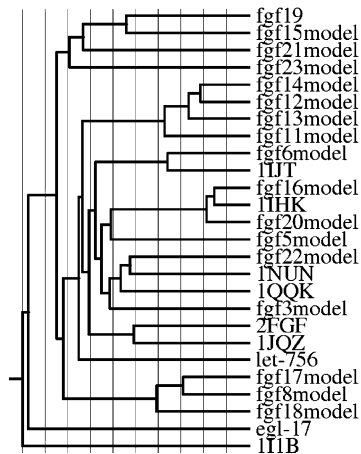
FIGURE 3: Phylogenetic tree of the FGFs. The alignment generated above was used, altered to include the missing loops in FGF15, FGF19, FGF21, and FGF23. The new alignment was submitted to TraceSuite II (*57*). PDB codes as in Figure 2.

common to a broader subsection of the FGF family. A phylogenetic tree, derived from this structure-based sequence alignment, separates the mammalian FGFs into seven subfamilies (Figure 3). It is notable that the disulfide bond is limited to the two most divergent (FGF19 and FGF8) subfamilies [and to the *C. elegans* and *Drosophila* (cf. ref *61*) homologues], while the loss of the strand 11 region is limited to the most divergent (FGF19) subfamilies. As all of the previously solved structures belong to the least divergent families, it is clear that an examination of these features offers the potential to teach new insights into FGFs.

The disulfides are both in positions where they can stabilize two of the FGF β-hairpins. The first disulfide connects C58 at the start of strand 2 with C70 at the end of strand 3, while the second links C102 in strand 6 with C120 in the short α-helix between strands 7 and 8. One of the cysteine positions is conserved throughout the FGF family (Figure 2). Two other cysteines are substituted for amino acids of similar size, and only the cysteine in strand 2 substitutes for a significantly larger side chain. Accordingly, the disulfides form at positions where the backbone is similarly positioned to the equivalent backbone in the other FGFs. The first disulfide, which is conserved only in FGF15 (the rodent orthologue of FGF19), is between two strands that are essentially identical in conformation in FGF19 and the other FGFs (Figure 4A). However, the loop between strands 1 and 2 (which C58 stands at the end of) is considerably extended in FGF19 compared to the other FGFs (Figures 2 and 4A) and adopts an extended conformation. The importance of the disulfide may lie in maintaining the conformation of the loop, which is functionally important (see below). Although FGF21 (a member of the FGF19 subfamily) has a loop of seven residues (cf. eight in FGF19 and seven in FGF15) without a disulfide, the FGF19 loop contains one proline and two glycine residues and so is likely to be far more conformationally flexible. This suggests that the presence of a disulfide here is likely to be required for stabilizing the long and potentially flexible loop of FGF19.

The second disulfide bond is associated with a change in the protein backbone. While strand 6 is in a very similar conformation in FGF19, the loop between strands 7 and 8 (where C120 lies) shows an extension of one residue. This leads to a change from a short 3_{10} helix to an α-helix in

FGF19 (Figure 4B). The importance to the local structure of this disulfide is supported by the alignment with other FGFs (Figure 2). In this case, the disulfide is maintained across two subfamilies of the FGFs (FGF8 and FGF19 subfamilies), as well as in the *C. elegans* and *Drosophila* homologues. In all 10 of these FGFs, the β7−β8 loop is extended by one amino acid beyond the length of the loop in the canonical FGFs. There is a perfect correlation between the disulfide and the extended helix. As this side of the molecule is not implicated in the binding of either the receptor or heparin, there is no obvious requirement for this loop extension in FGF function. One speculation is that the conformation imposed on the helix in this loop may be required for the correct folding of the FGF and for enhancing FGF stability. This is clearly an issue in the case of FGF23. FGF23 is required in the circulation for the maintenance of normal vitamin D_3 circulation and calcium homeostasis (*62*). Therefore, unlike the canonical FGFs, whose actions are mainly local to the cell producing the FGF, FGF23 requires the stability to survive in the bloodstream. This may also be the case for FGF19, whose overexpression in the mouse muscle leads to a phenotype of increased energy expenditure, with effects seen in tissues distant to the site of FGF production (*4*). The segregation of disulfides to these two divergent subfamilies of FGFs (and to primitive species) has a further implication. There is considerable evidence for functions for FGFs in the cytosol. In particular, FGF1, FGF2, FGF9, FGF11−FGF14, FGF16, and FGF20 have no ER signal sequences and are synthesized in the cytosol (*1*). It has been hypothesized that FGF11−FGF14 are not growth factors but cytosolic scaffold proteins (*63*). Additionally, there is good evidence for the release of FGF1 and FGF2 into the cytosol upon endocytosis of FGF−FGFR complexes (*64*), and direct functions in regulating intracellular signaling have been proposed (e.g., ref *65*). All of these functions are clearly less available to a disulfide-bonded protein, as the stability of such a protein will be significantly compromised in the heavily reducing environment of the cytosol. We would therefore expect that FGF8, FGF15, FGF17−FGF19, FGF21, and FGF23 would not have intracellular functions in the same manner as the less divergent FGFs.

The discovery of a disulfide bond involving C120 is intriguing as this cysteine is the only amino acid that is absolutely conserved across the entire FGF family, across all species. To date, there has been no rationale found for the extreme conservation of this amino acid, which is located near to neither the FGFR nor the heparin-binding interfaces. Furthermore, mutation to serine causes no obvious alteration in either the stability or the structure of FGF2 (*8*). In other FGF structures, the conserved cysteine shows a backbone buried within the hydrophobic interior of the protein and the side chain sulfur partly buried and partly exposed to the solvent. The side chain makes a hydrogen bond to a main-chain amide in most of the solved structures (Figure 2), which is conserved in FGF19 (main-chain amide of S116). One further possibility is that this face of the molecule may be involved in interactions with the recently identified FGF binding protein (*66*), whose interaction site with FGFs has not been clearly determined.

FGF19 Loop Conformations and Heparin Binding. Although the structure of FGF19 is in general similar to that of the other FGFs, there are a number of regions at the
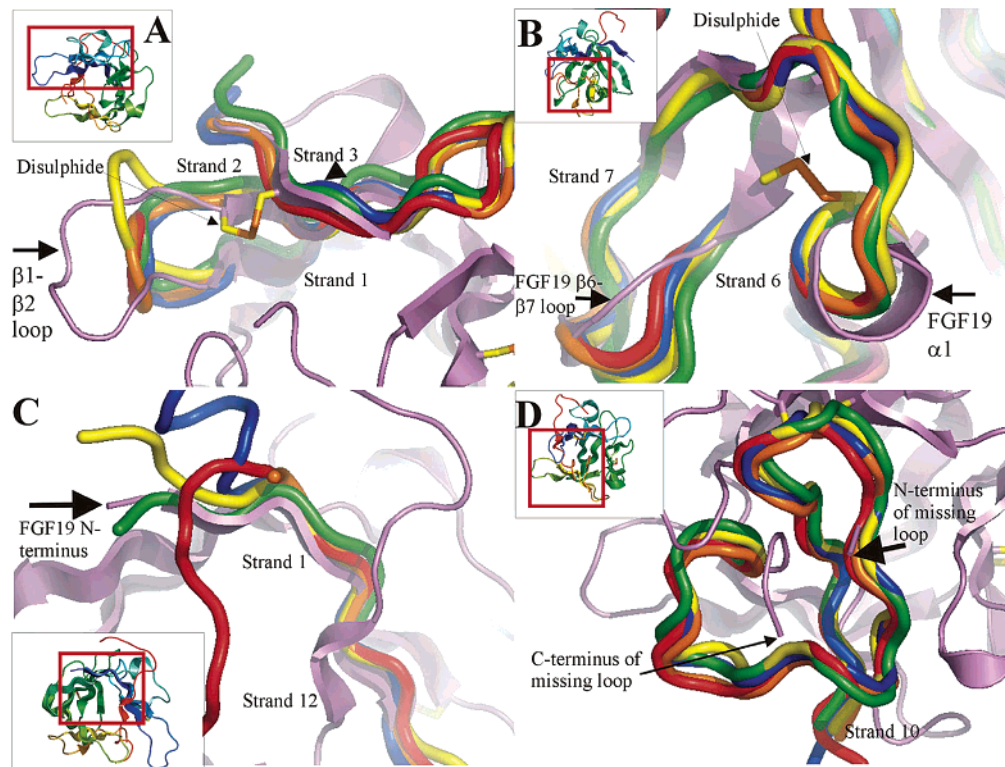
FIGURE 4: Comparison of solved FGF structures. FGFs were superimposed using COMPARER (*42*). FGF19 is shown in cartoon form. Other FGFs are shown as a backbone ribbon (in the region of interest only). Colors: FGF1, red; FGF2, orange; FGF4, yellow; FGF7, green; FGF9, blue; FGF19, violet. Where shown, the FGF19 cystine side chains are shown as sticks. Insets display FGF19 in cartoon form in an identical orientation to the figure, colored as in Figure 1. The red box shows the region displayed in the main figure. (A) FGF19 $\beta1-\beta2$ loop (black arrow) is extended considerably compared with other FGFs. The extended conformation of this loop is anchored at the C-terminal end by a disulfide bond. (B) The helix between strands 7 and 8 (black arrow) is extended in FGF19 from a 3_{10}-helical loop to an α-helical loop. This extension is again anchored by a disulfide forming from one end of the helix to a neighboring strand. The $\beta6-\beta7$ loop (black arrowhead) also shows a considerable divergence from the conformation seen in other FGFs. (C) The FGF19 N-terminal strand (black arrow) is extended in comparison to the other FGFs, in a manner similar to FGF7. (D) The heparin-binding loop of FGF19 is missing from the structure (black arrow, N-terminal end; black arrowhead, C-terminal end). The equivalent loop in all other FGFs is well conserved in the structure, but the FGF19 structure diverges from this at each end of the missing loop, suggesting that the conformation of this loop will be very different in FGF19.

periphery of the molecule that show considerable divergence from the other FGFs (Figure 4). The first of these regions is at the extreme N-terminus of FGF19 (Figure 4C). Here, strand 1 is extended toward the N-terminus, compared to the equivalent strands in FGF1 and FGF2. This strand extension is observed in FGF7 and FGF10 and has been postulated to provide a key interaction with the "b" splice form of FGFR2 (*13*). The presence of this extended strand in FGF19 is intriguing, as FGF19 has been shown to have no affinity for FGFR2b (*34*). The N-terminal extension then follows a similar path to FGF7 and FGF10.

FGF19 shows a subtle alteration in the loop linking strands 6 and 7 (Figure 4B). In FGF19, a short 3_{10} helix is formed here, with two consecutive positive ϕ backbone angles. This surface of the FGF is involved in binding neither heparin nor FGFR, so this alteration in conformation is likely to be of little significance to FGF function.

FGF19 shows a considerable extension of the loop connecting strand 1 to strand 2 (Figure 4A). In FGF19, this loop is eight residues long, compared with between three and five in other FGFs with a solved structure. An alternative conformation for this loop is not unprecedented: The FGF4 structure also shows divergence from the loop structure of the canonical FGFs, although the extent of this bulge is not nearly as great as that of FGF19. The conformation of this

loop in FGF19 is stabilized by the presence of a sulfate ion (Figure 4A), which forms hydrogen bonds with amide groups of H53, L55, S56, and possibly G51 (the G51 amide can also form a less favorable H-bond with the side chain of S50). As this loop is at the heparin-binding site (see below), the binding of sulfate to this position is likely to be functionally relevant. So, even though this loop is at a crystal contact, it is likely that this conformation is physiological.

Perhaps the most significant feature of this structure is the region corresponding to strand 11. In this crystal structure, this region is disordered, suggesting that it does not form a β-structure in this molecule. The path of the molecule at either end of the missing loop diverges from that of the other FGFs (Figure 4D). There is considerable electron density in the missing region, which we could not build into an acceptable protein conformation. However, this density clearly does not permit a loop conformation of any similarity to the loop seen in the other FGFs (data not shown). The alternative conformation of the FGF19 heparin-binding loop is confirmed by the conformation of the protein backbone on either side of the missing loop (Figure 4D). The loop conformation is well conserved in the other solved FGF structures. Examination of the backbone conformation of FGF19 shows that the backbone diverges from that of the other FGFs in the residues preceding and following the
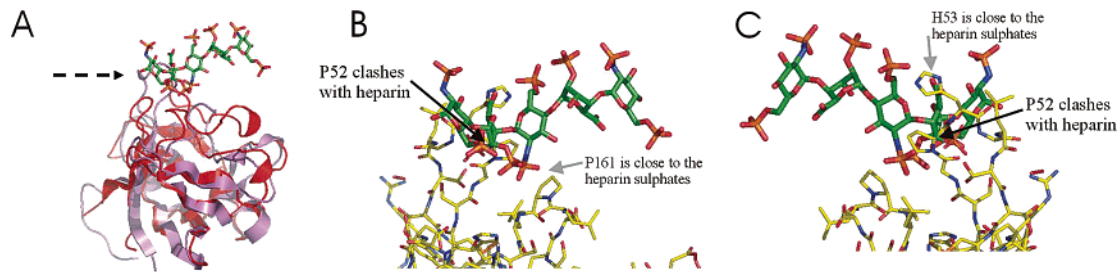
FIGURE 5: The FGF19 structure was superimposed onto the FGF1−heparin structure (*23*) using COMPARER (*42*). (A) Comparison of FGF1 and FGF19 structures in relation to heparin. Proteins are shown in cartoon representation, with FGF1 in red and FGF19 in violet. The FGF19 $\beta1-\beta2$ loop (dashed arrow) extends considerably beyond the FGF1 loop and clashes with the heparin path. (B) Close-up of the FGF19−heparin superimposition. FGF19 and heparin are shown as all-atom representation. Colors: FGF19 carbon, yellow; heparin carbon, green; nitrogen, blue; oxygen, red; sulfur, orange. P52 in the $\beta1-\beta2$ loop (bold arrow) shows the strongest clash with the heparin saccharide backbone. The gray arrow shows the close contact between the side chain of P161 and two sulfate groups. (C) Reverse view of the FGF19−heparin superposition. Colors as in (B). The black arrow shows the clash between P52 and heparin. The gray arrow shows the close contact between the side chain of H53 and an iduronic acid residue.

missing loop and that the FGF19 backbone at the C-terminal region passes through the path of the loop in other FGF structures. This suggests that the missing loop is likely to be forming a radically different structure from that observed in the other FGF homologues. This is further corroborated by the similarities in sequence of the other FGFs. Several residues are well conserved in this loop, in particular the positive ϕ glycine at the start of the loop (Figure 2). This glycine is conserved even in the *C. elegans* and *Drosophila* FGF homologues (*61*). However, the sequences of FGF15, FGF19, FGF21, and FGF23 in this region are considerably diverged from the other FGFs. This evidence suggests that the structure of the missing loop of FGF19, and the equivalent loops in its close homologues, is likely to be considerably divergent from those seen in other FGFs.

This loop in the other FGFs contains strand 11 and is the major heparin-binding region of the FGF (*20, 21, 23*). The disordered nature of this loop suggests that, in FGF19, strand 11 does not form. It is likely that this lack of strand formation is biological, as P142 in strand 10 will not offer a hydrogen bond, reducing the possibility for the formation of a strand 11. The loss of strand 11 is consistent with data from other FGFs, which suggest that strand 11 is rather weak in the FGF family. NMR data from FGF1 and FGF2 are inconsistent with a strand 11 (*14, 15*). The *sstruc* algorithm (*67*) used in JOY, which uses hydrogen bonds to assign secondary structure, does not predict a strand 11 for any of the FGFs, while such a strand is predicted for IL-1β (Figure 2). The SEGNO algorithm, by contrast, uses the backbone dihedral angles to assign secondary structure. This algorithm assigns no secondary structure in this region for FGF4 and FGF7 and assigns a polyproline strand (a form of secondary structure that does not allow regular β-sheet hydrogen bonding) for strand 11 for the other FGFs (data not shown). As both of these algorithms tend to overassign secondary structure, and neither assigns a regular β-strand for strand 11, we conclude that this loop does not form a regular secondary structure in the FGF family. This is in contrast to the IL-1β family, where there are 12 regular β-strands forming the protein core.

The altered conformation of this loop is of greatest significance because of the consequences of this for heparin binding. The role of HS in the activity and regulation of FGFs has been established beyond doubt by the inability of cells without HS to respond to FGF signaling (*18, 19*).

Additionally, the binding of FGF19 to FGFR4 increases with the concentration of heparin in a dose-dependent fashion (*34*).

However, during the purification of FGF19, it became apparent that the affinity of this protein for heparin is far lower than that of FGF1 or FGF2. The concentration of NaCl required to release FGF19 from a heparin−Sepharose support is less than 400 mM (data not shown). This compares very unfavorably with the 1.3 M NaCl required to liberate FGF1 or greater than 1.5 M NaCl required to liberate FGF2-GST. Comparison with the crystal structures of FGF1 and FGF2 in complex with heparin suggests why this might be. The strongest interactions between FGF1 and FGF2, and heparin, are formed between sulfate groups and the protein backbone, in the region that is disordered in this structure (*24*). As the conformation of the loop is likely to be considerably different to that of the other FGFs, this loop is unlikely to bind to heparin in the same manner as other FGFs. It may be that the increased flexibility of this region is a hindrance to heparin binding in FGF19 or that this region will only adopt a stable conformation in the presence of heparin. In terms of possible side-chain interactions with heparin, the disordered region contains two arginines, one lysine, two glutamines, and one asparagine side chain. This compares favorably with other FGFs in terms of side chains that could form hydrogen bonds with heparin. We therefore cannot predict whether this region of FGF19 will be able to bind to heparin. However, the deviation of the structure from the key heparin-binding surface of other FGFs does suggest that the FGF binding affinity is likely to be considerably reduced.

Superimposition of the FGF19 structure onto the structures of an FGF1−heparin complex (*23*) demonstrates that the extended loop $\beta1-\beta2$ is likely to have a role in heparin binding (Figure 5). In the superimposition, this loop occupies the same space as the extension of the heparin along the FGF. This suggests that this loop may significantly alter the manner in which this FGF binds to heparin. In the structure, the loop is bound to a sulfate ion, which forms three or four hydrogen bonds to main-chain amides. As the strongest bonds between the canonical FGFs and heparin are sulfate groups interacting with main-chain amides (*24*), these interactions are likely to be representative of an interaction of this loop with heparin. Although this loop is also at the FGFR binding site of the FGF (*20, 21*), modeling of the complex with FGFR4 (see below) demonstrates that this conformation of the loop is consistent with binding to the

receptor. An extension in the $\beta1-\beta2$ loop is found in all of the members of the FGF19 subfamily, which are the FGFs that have an abnormal heparin-binding loop. These two features may be correlated, suggesting that there is a significant alteration in the heparin binding of FGF19. The ends of the strand 11 region in FGF19 are closer to the heparin than the equivalent residues in FGF1, suggesting that the true conformation of the missing loop is likely to be more extended from the FGF core than that of FGF1. Therefore, the best hypothesis for the binding of FGF19 to heparin is that the heparin molecule is located further from the core of the FGF than in FGF1 and FGF2. This would overcome the clash with the $\beta1-\beta2$ loop and account for the radically different structure of the "$\beta11$ strand" in FGF19. Binding to a more conformationally flexible molecule would also be expected to reduce the affinity for heparin. The costs of reduced entropy in the loop and of losing favorable interactions in the flexible state to form the heparin-bound state would both reduce the energetic gain from forming hydrogen bonds to the main-chain carbonyls and to the positively charged side chains in the missing loop.

Computational Modeling of the FGF19−FGFR4 Interaction. One of the intriguing features of FGF19 is that it binds only to FGFR4 (*34*). Of the previously characterized FGFs, only the FGF7 subfamily has shown affinity to just one receptor (*28*). Given the similar overall structure of FGF19 to the canonical FGFs, the rationale for this specificity is quite unclear. We therefore attempted to generate a model of the FGF19−FGFR4 complex to further probe this specificity.

As it is thought that the interactions with domain 3 are important for the specificity of FGF−FGFR complexes, we chose the structures of FGF2−FGFR complexes as a basis for our model. FGF1 forms very limited interactions with domain 3 of the receptor and has high affinity for all FGFRs. The structures 1EV2 and 1FQ9 were therefore chosen to build the FGFR4 model.

The FGFR4 model and the solved structure of FGF19 were then built into the positions of FGFR1 and FGF2, respectively, in the structure of the homologous FGF2−FGFR1 complex (1FQ9). To remove major clashes presented in some of the FGFR4 interacting side chains, we applied a refinement protocol that had already proved to be highly effective for relaxing clashing interfaces after rigid-body docking simulations (*59*). After optimization of the conformation of the interacting FGFR4 side chains, all major clashes were removed.

The geometry of the FGFR-D2/FGF interface is well conserved among the different currently available structures, so we can be confident about this part of the model. However, the FGFR-D3/FGF interface is hypothesized to provide most of the specificity of the FGF receptor interactions and, perhaps because of that, different conformations have been reported for that interface (*20, 21*). Therefore, considering the possibility of alternative binding modes, we have explored the conformational space of the FGFR4-D3/FGF19 interface using a hinge-bending docking protocol (described in Experimental Procedures). After the simulations (Figure 6B), we found a well-defined global energy minimum in which the lowest energy conformation had an RMSD of 1.0 Å with respect to the starting conformation (RMSD calculated for all FGFR Cα atoms after superimposing the

FGF structures). These results suggest that the proposed model for the FGF19−FGFR4 interaction represents an energetically favorable conformation. As a control of this technique, we built a model of the FGF2−FGFR4 interaction based on the FGF2−FGFR2 complex (1EV2). As expected from this high-affinity interaction (*28*), the lowest energy conformation is found at RMSD 0.39 Å from the initial model (data not shown).

We performed a similar energy and conformational analysis of a putative model for the low-affinity interaction between FGF19 and FGFR1 based on the structure of the homologous complex FGF2−FGFR1 (1FQ9). In this case, the lowest energy solution obtained after hinge-bending docking was far from the initial model (RMSD 24.0 Å; Figure 6B). This suggests that the putative model for the FGF19−FGFR1 interaction based on the homologous FGF2−FGFR1 structure does not seem to have an energetically favorable behavior, as expected from its experimentally determined low affinity. It seems that the FGFR1 domain 3 $\beta C'-\beta E$ loop conformation (as in the FGF2−FGFR1 complex) is clearly clashing with FGF19 (Figure 6A) and thus contributing to the poor energetics of this model. It has been reported that this loop is quite flexible in FGFR1 and FGFR2 in their interaction with different FGFs (*30*). The domain 3 $\beta C'-\beta E$ loop in the FGF1−FGFR2 complex (1DJS) has a different conformation to that of the FGF2−FGFR1 complex (1FQ9). To analyze the role of this different loop conformation, we built a putative model for the low-affinity interaction FGF19−FGFR2 based on the homologous complex FGF1−FGFR2 (1DJS). The lowest energy conformation obtained after simulations was also far from the starting model (RMSD 25.8 Å). This suggests that, regardless of the different conformation of the domain 3 $\beta C'-\beta E$ loop, interaction between FGF19 and FGFR1 or FGFR2 according to the proposed models is energetically unlikely, which is consistent with the detected low affinity of these interactions (*34*).

The analysis of the models shows that the conformation of the $\beta C'-\beta E$ loop in FGFR4 favors interaction with FGF19 in larger extent than what the other loop conformations found in the FGFR1−FGF2 (1FQ9) or FGFR2−FGF1 (1DJS) complexes do. Indeed, a sequence comparison of FGFR4 with the different FGFRs shows differences in the region corresponding to the mentioned loop (Figure 6D). First, there is a deletion of two amino acids (aspartic acid and lysine) that shorten the loop in FGFR4 with respect to FGFR1c and FGFR2c. Second, an aspartic acid appears in FGFR4 instead of the highly conserved glycine. This suggests that this FGFR4 loop could play an essential role in the high specificity of FGF19 for this receptor. In addition, the FGF19 loop that interacts with the FGFR4 $\beta C'-\beta E$ loop shows a highly strained conformation. This $\beta4-\beta5$ loop shows a tight (two-residue) β-turn, with a positive ϕ leucine residue (L87). This positive ϕ is not seen at the same position in the other solved FGF structures: most other FGFs have a glycine in the second position in the turn (Figure 2; cf. R88 in FGF19). The result is that the β-turn in FGF19 is type 1, compared with type 2 in the other FGFs. Only FGF11−FGF15 also lack a glycine. Binding of this loop to FGFR4 is therefore likely to be a feature of the FGF15−FGF19 pair [as FGF11−FGF14 are thought to be intracellular (*63*)]. Thus we can predict that other members of the FGF19 subfamily (i.e., FGF21 and FGF23) will not have the specific affinity to
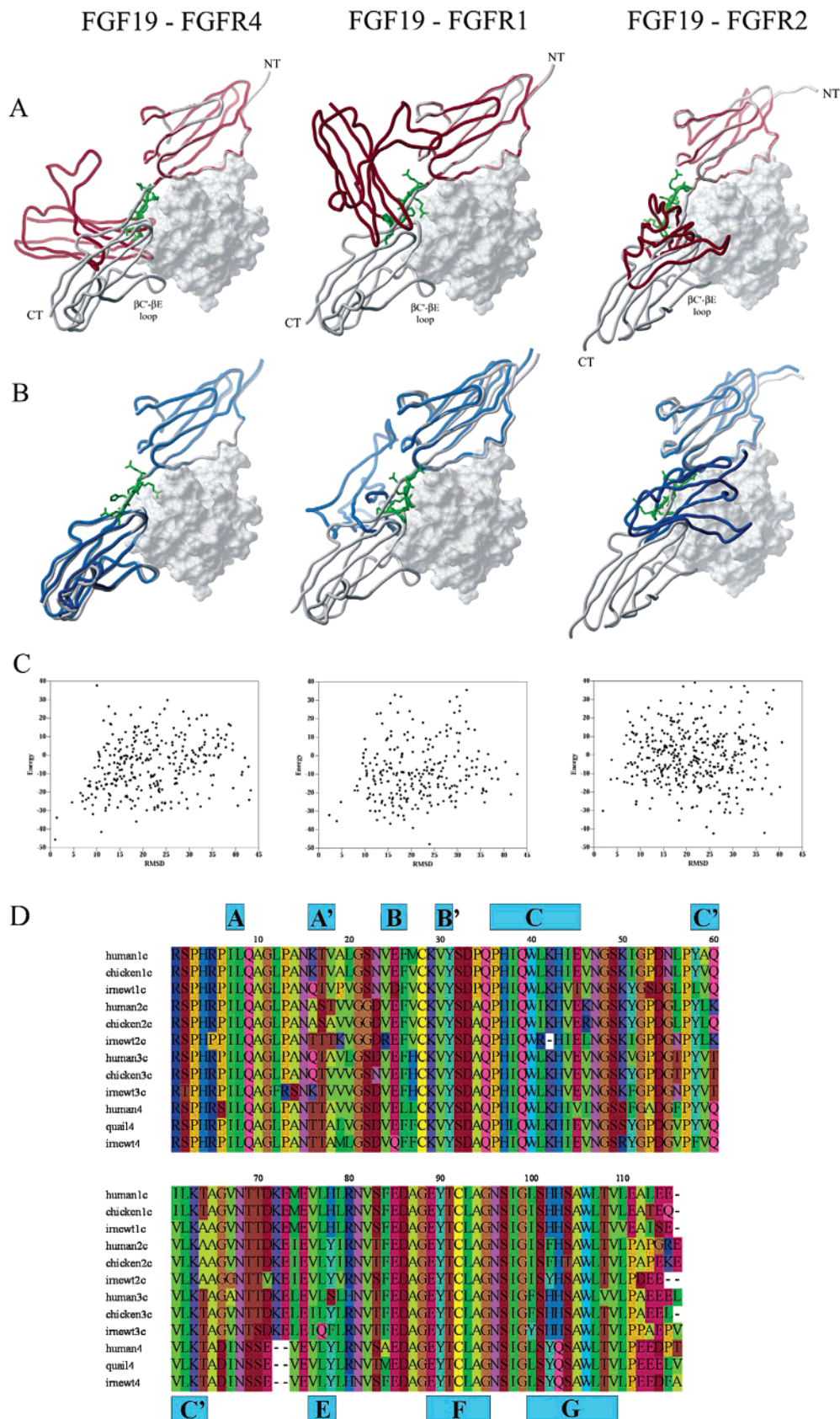
FIGURE 6: Docking of FGF19 to 3-D models of FGF receptors. (A) The starting conformation created by randomizing the torsion angles of the flexible linker. The initial model created by homology and further refinement of the interface is shown as a white ribbon (FGFR) and white solid surface (FGF19), with the starting conformation for docking shown in red ribbon (linker in green). (B) The lowest energy solution obtained after hinge-bending docking (represented in blue). (C) Energy−RMSD distribution of the docking solutions obtained from the hinge-bending docking simulations. (D) Alignment of FGFR immunoglobulin domain 3 from human, chicken/quail, and Iberian ribbed newt (*Pleurodeles waltl*). Molecules are shown as species and then FGFR number. The secondary structure follows that of Plotnikov et al. (*30*).

FGFR4. The sequence and structure of FGF19 therefore also support an interaction with FGFR4 that is specific for both partners.

The structure of FGF19, presented here, both explains several novel features of FGF19 and gives further insight into the biology of other members of the FGF family. FGF19 shows two disulfide bonds, one of which is conserved among several members of the family: this suggests that these FGFs are likely to have greater stability and a longer range of activity from the FGF secreting cell. It also suggests that the intracellular functions associated with some FGFs will not be present in these members of the family. FGF19 additionally shows a significantly reduced affinity for heparin, which is likely to be due to the expansion of one heparin-binding loop and a dramatic alteration in conformation of the other (in comparison to the other FGFs). Furthermore, we propose a model for the interaction of FGF19 with FGFR4, which explains the unusually specific affinity of FGF19 for this receptor. Our model proposes that this affinity is due to the specific structure of FGFR4 and FGF19 and represents a specific determinant of each molecule for the other.

ACKNOWLEDGMENT

SUPPORTING INFORMATION AVAILABLE

A fuller description of the methods used to generate the docking simulations and a larger, color version of the FGF family alignment (Figure 2). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Powers, C., McLeskey, S., and Wellstein, A. (2000) *Endocr. Relat. Cancer 7*, 165−197.
2. Ornitz, D., and Itoh, N. (2001) *Genome Biol. 2*, reviews 3005.1−3005.12.
3. Yu, C., Wang, F., Jin, C., Wu, X., Chan, W.-K., and McKeehan, W. (2002) *Am. J. Pathol. 161*, 2003−2010.
4. Tomlinson, E., Fu, L., John, L., Hultgren, B., Huang, X., Renz, M., Stephan, J., Tsai, S., Powell-Braxton, L., French, D., and Stewart, T. (2002) *Endocrinology 143*, 1741−1747.
5. The ADHR Consortium (2000) *Nat. Genet. 26*, 345−348.
6. Eriksson, A., Cousens, L., Weaver, L., and Matthews, B. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3441−3445.
7. Zhang, J., Cousens, L., Barr, P., and Sprang, S. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3446−3450.
8. Zhu, X., Komiya, H., Chirino, A., Faham, S., Fox, G., Arakawa, T., Hsu, B., and Rees, D. (1991) *Science 251*, 90−93.
9. Bellosta, P., Iwahori, A., Plotnikov, A., Eliseenkova, A., Basilico, C., and Mohammadi, M. (2001) *Mol. Cell. Biol. 21*, 5946−5957.
10. Ye, S., Luo, Y., Lu, W., Jones, R., Linhardt, R., Capila, I., Toida, T., Kan, M., Pelletier, H., and McKeehan, W. (2001) *Biochemistry 40*, 14429−14439.
11. Hecht, H., Adar, R., Hofmann, B., Bogin, O., Weich, H., and Yayon, A. (2001) *Acta Crystallogr., D: Biol. Crystallogr. 57*, 378−384.
12. Plotnikov, A., Eliseenkova, A., Ibrahimi, O., Shriver, Z., Sasisekharan, R., Lemmon, M., and Mohammadi, M. (2001) *J. Biol. Chem. 276*, 4322−4329.
13. Yeh, B., Igarashi, M., Eliseenkova, A., Plotnikov, A., Sher, I., Ron, D., Aaronson, S., and Mohammadi, M. (2003) *Proc. Natl. Acad. Sci. U.S.A. 100*, 2266−2271.
14. Ogura, K., Nagata, K., Hatanaka, H., Habuchi, H., Kimata, K., Tate, S.-I., Ravera, M., Jaye, M., Schlessinger, J., and Inagaki, K. (1999) *J. Biomol. NMR 13*, 11−24.
15. Moy, F., Seddon, A., Campell, E., Bohlen, P., and Powers, R. (1995) *J. Biomol. NMR 6*, 245−254.
16. Chi, Y.-H., Kumar, T., Kathir, K., Lin, D.-H., Zhu, G., Chiu, I.-M., and Yu, C. (2002) *Biochemistry 41*, 15350−15359.
17. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. (1999) *Essentials of Glycobiology*, 1st ed., Cold Spring Harbor Laboratory Press, New York.
18. Ornitz, D., Yayon, A., Flanagan, J., Svahn, C., Levi, E., and Leder, P. (1992) *Mol. Cell. Biol. 12*, 240−247.
19. Rapraeger, A., Krufka, A., and Olwin, B. (1991) *Science 252*, 1705−1708.
20. Schlessinger, J., Plotnikov, A., Ibrahimi, O., Eliseenkova, A., Yeh, B., Yayon, A., Linhardt, R., and Mohammadi, M. (2000) *Mol. Cell 6*, 743−750.
21. Pellegrini, L., Burke, D., von Delft, F., Mulloy, B., and Blundell, T. (2000) *Nature 407*, 1029−1034.
22. Faham, S., Hileman, R., Fromm, J., Linhardt, R., and Rees, D. (1996) *Science 271*, 1116−1120.
23. DiGabriele, A., Lax, I., Chen, D., Svahn, C., Jaye, M., Schlessinger, J., and Hendrickson, W. (1998) *Nature 393*, 812−817.
24. Pellegrini, L. (2001) *Curr. Opin. Struct. Biol. 2001*, 629−634.
25. Kreuger, J., Salmivirta, M., Sturiale, L., Giménez-Gallego, G., and Lindahl, U. (2001) *J. Biol. Chem. 276*, 30744−30754.
26. Ostrovsky, O., Berman, B., Gallagher, J., Mulloy, B., Fernig, D., Delehedde, M., and Ron, D. (2002) *J. Biol. Chem. 277*, 2444−2453.
27. Powell, A., Fernig, D., and Turnbull, J. (2002) *J. Biol. Chem. 277*, 28554−28563.
28. Ornitz, D., Xu, J., Colvin, J., McEwan, D., MacArthur, C., Coulier, F., Gao, G., and Goldfarb, M. (1996) *J. Biol. Chem. 271*, 15292−15297.
29. Stauber, D., DiGabriele, A., and Hendrickson, W. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 49−54.
30. Plotnikov, A., Hubbard, S., Schlessinger, J., and Mohammadi, M. (2000) *Cell 101*, 413−424.
31. Plotnikov, A., Schlessinger, J., Hubbard, S., and Mohammadi, M. (1999) *Cell 98*, 641−650.
32. Coulier, F., Pontarotti, P., Roubin, R., Hartung, H., Goldfarb, M., and Birnbaum, D. (1997) *J. Mol. Evol. 44*, 43−56.
33. Nishimura, T., Utsunomiya, Y., Hoshikawa, M., Ohuchi, H., and Itoh, N. (1999) *Biochim. Biophys. Acta 1444*, 148−151.
34. Xie, M., Holcomb, I., Deuel, B., Dowd, P., Huang, A., Vagts, A., Foster, J., Liang, J., Brush, J., Gu, Q., Hillan, K., Goddard, A., and Gurney, A. (1999) *Cytokine 11*, 729−735.
35. Ladher, R., Anakwe, K., Gurney, A., Schoenwolf, G., and Francis-West, P. (2000) *Science 290*, 1965−1967.
36. Nicholes, K., Guillet, S., Tomlinson, E., Hillan, K., Wright, B., Frantz, G., Pham, T., Dillard-Telm, L., Tsai, S., Stephan, J., Stinson, J., Stewart, T., and French, D. (2002) *Am. J. Pathol. 160*, 2295−2307.
37. Yu, C., Wang, F., Kan, M., Jin, C., Jones, R., Weinstein, M., Deng, C.-X., and McKeehan, W. (2000) *J. Biol. Chem. 275*, 15482−15489.
38. Holt, J., Luo, G., Billin, A., Bisi, J., McNeill, Y., Kozarsky, K., Donahee, M., Wang, D., Mansfield, T., Kliewer, S., Goodwin, B., and Jones, S. (2003) *Genes Dev. 17*, 1581−1591.
39. Lennon, G., Auffrey, C., Polymeropoulos, M., and Soares, M. (1996) *Genomics 33*, 151−152.
40. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C., Jr., and Sweet, R., Eds.) pp A307−A326, Academic Press, New York.
41. Navaza, J. (1994) *Acta Crystallogr. A 50*, 157−163.
42. Sali, A., and Blundell, T. (1990) *J. Mol. Biol. 212*, 403−428.
43. Shi, J., Blundell, T., and Mizuguchi, K. (2001) *J. Mol. Biol. 310*, 243−257.
44. Lamzin, V., and Wilson, K. (1993) *Acta Crystallogr., D: Biol. Crystallogr. 49*, 129−149.
45. McRee, D. (1999) *Practical Protein Crystallography*, 2nd ed., Academic Press, San Diego, CA.

46. Brünger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T., and Warren, G. (1998) *Acta Crystallogr., D: Biol. Crystallogr. 54*, 905−921.

47. Murshudov, G., Vagin, A., and Dodson, E. (1997) *Acta Crystallogr. D 53*, 240−255.

48. Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. (1993) *J. Appl. Crystallogr. 26*, 283−291.

49. Hooft, R., Vriend, G., Sander, C., and Abola, E. (1996) *Nature 381*, 272.

50. Lovell, S., Davis, I., Arendall, W., III, de Bakker, P., Word, J., Prisant, M., Richardson, J., and Richardson, D. (2003) *Proteins 50*, 437−450.

51. DeLano, W. (2002) The Pymol Molecular Graphics System, DeLano Scientific, San Carlos, CA.

52. Deane, C., Kaas, Q., and Blundell, T. (2001) *Bioinformatics 17*, 541−550.

53. de Bakker, P., DePristo, M., Burke, D., and Blundell, T. (2003) *Proteins 51*, 21−40.

54. DePristo, M., de Bakker, P., Lovell, S., and Blundell, T. (2003) *Proteins 51*, 41−55.

55. L. Chen, S. Lovell, R. Smith, and T. Blundell, unpublished results.

56. Mizuguchi, K., Deane, C., Blundell, T., Johnson, M., and Overington, J. (1998) *Bioinformatics 14*, 617−623.

57. Innis, C., Shi, J., and Blundell, T. (2000) *Protein Eng. 13*, 839−847.

58. Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F., and Higgins, D. (1997) *Nucleic Acids Res. 24*, 4876−4882.

59. Fernandez-Recio, J., Totrov, M., and Abagyan, R. (2002) *Protein Sci. 11*, 280−291.

60. Abagyan, R., and Totrov, M. (1994) *J. Mol. Biol. 235*, 983−1002.

61. Nagendra, H., Harrington, A., Harmer, N., Pellegrini, L., Blundell, T., and Burke, D. (2001) *FEBS Lett. 501*, 51−58.

62. Bai, X.-Y., Miao, D., Goltzman, D., and Karaplis, A. (2003) *J. Biol. Chem. 278*, 9843−9849.

63. Schoorlemmer, J., and Goldfarb, M. (2001) *Curr. Biol. 11*, 793−797.

64. Prudovsky, I., Bagala, C., Tarantini, F., Mandinova, A., Soldi, R., Bellum, S., and Maciag, T. (2002) *J. Cell Biol. 158*, 201−208.

65. Skjerpen, C., Nilsen, T., Wesche, J., and Olsnes, S. (2002) *EMBO J. 21*, 4058−4069.

66. Tassi, E., Al-Attar, A., Aigner, A., Swift, M., McDonnell, K., Karavanov, A., and Wellstein, A. (2001) *J. Biol. Chem. 276*, 40247−40253.

67. Smith, D. (1989) SSTRUC A Computer Program, Department of Crystallography, Birkbeck College, University of London.