Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of fibroblast growth factor 9 shows a symmetric dimer with unique receptor- and heparin-binding interfaces

H. J. Hecht,[a]* R. Adar,[b,c] B. Hofmann,[a] O. Bogin,[c] H. Weich[d] and A. Yayon[b,c]

[a]Department SF, GBF – Gesellschaft für Biotechnologische Forschung, Mascheroder Weg 1, D-38124 Braunschweig, Germany, [b]Department of Molecular Cell Biology, The Weizmann Institute of Science, Rehovot 76100, Israel, [c]Prochon Biotechnology Ltd, Rehovot 76100, Israel, and [d]Department RDIF, GBF – Gesellschaft für Biotechnologische Forschung, Mascheroder Weg 1, D-38124 Braunschweig, Germany

Correspondence e-mail: hjh@gbf.de

Received 28 July 2000
Accepted 20 December 2000

PDB Reference: FGF9, 1g82.

Fibroblast growth factors (FGFs) constitute a family of at least 20 structurally related heparin-binding polypeptides active in regulating cell growth, survival, differentiation and migration. FGF9, originally discovered as a glia-activating factor, shares 30% sequence identity with other FGFs and has a unique spectrum of target-cell specificity. FGF9 crystallized in the tetragonal space group $I4_1$, with unit-cell parameters $a = b = 151.9$, $c = 117.2$ Å. The structure of the glycosylated protein has been refined to an R value of 21.0% with $R_{\rm free} = 24.8\%$) at 2.6 Å resolution. The four molecules in the asymmetric unit are arranged in two non-crystallographic dimers, with the dimer interface composed partly of residues from N- and C-terminal extensions from the FGF core structure. Most of the receptor-binding residues identified in FGF1– and FGF2–receptor complexes are buried in the dimer interface, with the $\beta8$–$\beta9$ loop stabilized in a particular conformation by an intramolecular hydrogen-bonding network. The potential heparin-binding sites are in a pattern distinct from FGF1 and FGF2. The carbohydrate moiety attached at Asn79 has no structural influence.

1. Introduction

Fibroblast growth factors (FGFs) constitute a family of at least 20 structurally related heparin-binding polypeptides which are expressed in a wide variety of cells and tissues. The biological response of cells to FGF is mediated through specific high-affinity (K_d = 20–500 pM) cell-surface receptors that possess intrinsic tyrosine kinase activity and are phosphorylated upon binding of FGF (Coughlin *et al.*, 1988). Proteoglycans provide *via* heparan sulfate moieties a second class of binding sites with lower affinity ($K_d = 2 \times 10^{-9}$ M) but large capacity (10^6 sites per cell). Heparan sulfates are ubiquitous polysaccharides of highly variable length and sulfation level composed of repeating disaccharides of glucuronate or iduronate and glucosamine residues. A unique role for these molecules is in the formation of distinct complexes essential for high-affinity binding and activation of FGF in particular and of other heparin-binding growth factors in general (Yayon *et al.*, 1991; Rapraeger *et al.*, 1991).

Ligand and receptor dimerization is a key event in the transmembrane signalling of receptor tyrosine kinases. Receptor dimerization leads to an increase in kinase activity, resulting in autophosphorylation and the induction of diverse biological responses (Schlessinger & Ullrich, 1992). Several models have been proposed for the interaction between FGF2–heparin and its receptor (Yayon *et al.*, 1991; Ruoslahti & Yamaguchi, 1991; Spivak-Kroizman *et al.*, 1994; Kan *et al.*, 1993; Guimond *et al.*, 1993; Pantoliano *et al.*, 1994). Previous

work utilizing NMR demonstrated that FGF dimers in a symmetric tetramer are formed in the presence of an active heparin decasaccharide (Moy *et al.*, 1997), suggesting that a *cis*-oriented dimer is the minimal biologically active structural unit of FGF2. Using defined heparin fragments and soluble FGF receptors further demonstrated that ligand dimerization can significantly enhance the binding of FGF2 to FGFR1, the dimerization of the receptor and the induction of downstream signal transduction pathways (Safran *et al.*, 2000). Recently, however, several studies (Plotnikov *et al.*, 1999, 2000; Stauber *et al.*, 2000) exploring the crystal structure of a complex between FGF2 and FGF1 with the extracellular domains of FGFR1 and FGFR2 have shown a 1:2 molecular ratio of ligands to receptors, with no evidence for ligand dimerization or with ligand association only through interaction with heparin (Pellegrini *et al.*, 2000).

FGFs share a homology core of around 120 amino acids in their primary sequence, including four cysteine residues, one of which is conserved in all members of the family. The core structure contains 12 antiparallel β-strands organized into a threefold internal symmetry. Equivalent folds have been observed for the soybean trypsin inhibitor and interleukin IL-1a and IL-1b. The best characterized members of the family are FGF1 (aFGF) and FGF2 (bFGF), the structures of which have been determined (Zhang *et al.*, 1991; Zhu *et al.*, 1991). Both are potent mitogens that stimulate proliferation, migration and differentiation of a large variety of cells (Folkman & Klagsbrun, 1987; Rifkin & Moscatelli, 1989). FGF9, a recently identified member of the FGF family, was originally discovered as a heparin-binding glia-activating factor (Miyamoto *et al.*, 1993). It shares a 30% overall sequence identity with other FGFs but has a unique spectrum of target-cell specificity, as it stimulates the proliferation of glia and other fibroblast-like cells but is not mitogenic for endothelial cells (Naruo *et al.*, 1993). The basis for such cell selectivity resides in its differential capacity to bind the different FGF receptors. Recombinant FGF9 binds with high affinity and in a heparin-dependent manner to FGFR3, with somewhat less affinity to FGFR2 and with considerably less affinity to FGFR1 (Hecht *et al.*, 1995).

In order to further understand the molecular basis for this specificity, we have explored the three-dimensional structure of FGF9.

2. Materials and methods

The full-length coding region for human FGF9 (Miyamoto *et al.*, 1993) cDNA was isolated as a *Bam*HI/blunt fragment from pET vector (Kuriyama *et al.*, 1995) and was subcloned into the vector pBacPAK9 digested with *Bgl*II and *Sma*I. The FGF9 cDNA was used without changes in the coding region. Plasmids contaning the cDNA species in the proper orientation were isolated from bacteria used for transfection into Sf9 cells with purified linearized baculovirus DNA. Screening for recombinant viruses, cloning and propagation of recombinant viruses were performed as described by Fiebich *et al.* (1993). Insect cells at a density of 1×10^6 cells ml^{-1} were used for baculovirus infection. For purification of FGF9 protein from the insect cell serum-free supernatant, it was adjusted to 0.6 *M* NaCl and purified over a 5 ml HiTrap heparin column (Pharmacia Amersham). FGF9-containing samples were pooled, diluted 1:3 with 20 m*M* Tris–HCl pH 7.4 and applied to a 5 ml TSK heparin-affinity FPLC column (TosoHaas). Bound proteins were eluted with a 20 ml gradient of 0.4–1.5 *M* NaCl in buffer *A* (20 m*M* Tris–HCl pH 7.4). Aliquots of a 1 ml fraction containing FGF9 were used for SDS–PAGE and for silver staining of the gel. Total recovery of FGF9 was 4–6 mg l^{-1} insect cell supernatant.

The protein concentration was measured with a standard assay (BCA, Pierce). For amino-terminal sequencing of glycosylated FGF9, 20 mg protein from the biologically active fractions (estimated with Balbc-3T3 cells, not shown) was loaded onto an Applied Biosystems 473A gas-phase protein sequencer. 20 rounds of Edman degradation were carried out using standard protocols and chemicals supplied by Applied Biosystems (*ca* 50% pos. 19 and 50% pos. 34 of the coding region).

Crystals were grown with the sitting-drop method to typical dimensions of $0.2 \times 0.2 \times 0.2$ mm from solutions containing FGF9 at a concentration of 2.1 mg ml^{-1} and 2.0 *M* ammonium sulfate buffered at pH 5.2 with 0.1 *M* MES/Tris buffer. The statistics of the native data set collected at the MPG-GBF beamline BW6 of the DESY synchrotron from a shock-cooled crystal to a resolution of 2.6 Å are given in Table 1. Indexing and scaling the data set with *MOSFLM* (Collaborative Computational Project, Number 4, 1994) and *SCALA* (Collaborative Computational Project, Number 4, 1994) showed the space group to be tetragonal $I4_1$, with unit-cell parameters $a = b = 151.9$, $c = 117.2$ Å. The asymmetric unit contains four molecules showing clear twofold symmetry in a pseudo-$I4_122$ arrangement and in addition a pseudo-cubic threefold axis in the self-rotation function calculated with *GLRF* (Tong & Rossmann, 1997). The structure was solved by molecular replacement. The successful run of *EPMR* (Kissinger *et al.*, 1999) used the coordinates of FGF1 (PDB code 2afg), modified by replacement of all non-glycine residues by alanine, and identified clearly three of the four molecules in the asymmetric unit with a correlation factor of 0.296. The fourth molecule (*B*) was placed manually so that it formed a dimer with the third molecule (*C*) identical to the *AD* dimer. Introduction of the fourth molecule reduced the *R* value from 46.0 to 39.7% and $R_{\rm free}$ from 47.4 to 41.3% in an initial refinement cycle using *CNS* (Brunger *et al.*, 1998) which consisted of rigid-body refinement, molecular-dynamics calculation using the slow-cool protocol and restrained refinement. These calculations were carried out after introduction of the correct sequence using *O* (Jones *et al.*, 1991) and included fourfold non-crystallographic symmetry restraints. Further refinement proceeded by iteration of manual adjustment of the structure using *O* (Jones *et al.*, 1991) and restrained refinement with non-crystallographic symmetry restraints at 2.6 Å resolution. Water molecules were added using *ARPP*/*REFMAC* (Collaborative Computational Project, Number 4, 1994) until there was no further decrease

Table 1
Data set and refinement statistics.

Values in parentheses are for the highest resolution shell, 2.74–2.6 Å.

Space group	$I4_1$
Unit-cell parameters	
a (Å)	151.9
b (Å)	151.9
c (Å)	117.2
Resolution range (Å)	39.5–2.6
Unique reflections	40985
Data redundancy	6.3
Completeness (%)	99.9 (99.9)
$I/\sigma(I)$	8.5 (3.3)
R_{merge}† (%)	5.7 (24.2)
R_{all}‡ (%)	21.0
R_{free}§ (%)	24.8
No. of residues	624
No. of sugars	10
No. of sulfates	12
No. of waters	147
Coordinate error¶ (Å)	0.25
Ramachandran core region†† (%)	88.4
NCS r.m.s.‡‡ (Å)	0.55
R.m.s. bond length (Å)	0.034
R.m.s. bond angle (°)	2.7
R.m.s. *B* values (Å^2)	
Chain *A*	43.4
Chain *B*	79.9
Chain *C*	51.7
Chain *D*	44.6

† $R_{merge} = [\sum I_i(hkl) - \langle I(hkl)\rangle]/\sum I_i(hkl)$. ‡ $R_{all} = [\sum|F_o(hkl) - F_c(hkl)|]/\sum F_o(hkl)$. § R_{free} data set consisted of 5% of the unique reflections. ¶ Coordinate error calculated with *REFMAC*5 (Murshudov *et al.*, 1997; Collaborative Computational Project, Number 4, 1994). †† Ramachandran plot calculated with *PROCHECK* (Laskowski *et al.*, 1993). ‡‡ R.m.s. deviation of Cα protein atoms related by non-crystallographic symmetry calculated with *LSQMAN* (Kleywegt & Jones, 1997).

in the free *R* factor. In the last stages of the refinement, positional restraints for the non-crystallographic symmetry were dropped; however, owing to the limited resolution, only grouped temperature factors for main-chain and side-chain atoms were refined. In a final refinement cycle, individual atom temperature factors were introduced and refined using *REFMAC*5 (Murshudov *et al.*, 1997; Collaborative Computational Project, Number 4, 1994); this reduced the *R* value from 22.3 to 21.0%, while R_{free} remained constant at 24.8%.

N-terminal sequencing and MALDI mass spectrometry indicated heterogeneity of the crystallized protein, with the major components starting at residues 19, 34, 38 and 42 (SWISSPROT ID FGF9_HUMAN). The glycoconjugate is, according to MALDI mass spectrometry (M. Nimtz, unpublished work), of the three-mannosyl insect type with two *N*-acetylglucosamines, three mannose and one fucose moiety, a minor component having two fucose molecules, as expected from the expression system. In the N79 glycosylation site in three of the four molecules, density for the two *N*-acetylglucosamines together with one fucose molecule is clearly shown; the rest of the carbohydrate is disordered. In the fourth molecule only the first *N*-acetylglucosamine is visible. In the crystal, all four molecules of the asymmetric unit show flexibility of the N-terminal and, to a lesser extent, the C-terminal residues.

The first residue visible in the electron density is Pro49 in one molecule and Val51 or Thr52 in the others; C-terminal residues are visible to Ser208, the native C-terminus, in one molecule, to Gln207 and Ser206 in two others, and to Asp203 in the last molecule. The average r.m.s.d. between all C^α atoms common to the four molecules in the asymmetric unit is 0.5 Å and is 0.3 Å for residues 62–193. The final refinement statistics for the model consisting of 624 amino-acid residues, ten carbohydrate molecules, 147 water molecules and 12 sulfate molecules are given in Table 1.

3. Results and discussion

3.1. Overall structure description

The core unit of the FGF9 structure (Fig. 1) is formed by residues 62–193 and is very similar to the structures of FGF1 and FGF2, as expected from the sequence alignment (Fig. 2). The r.m.s. difference between C^α atoms of these residues and the corresponding C^α atoms of FGF1 (PDB code 2afg) and FGF2 (PDB code 2fgf) is 1.0 Å. Major differences (r.m.s.d. > 1.5 Å) occur at Thr70–Gly71, where FGF1 and FGF2 have an additional glycine, at the loop Asp88–Ser90, which may be correlated with the C-terminal extension in FGF9, and at Tyr153–Arg161, where in FGF9 an insertion of three (relative to FGF1) or five (relative to FGF2) residues occurs. The loop Glu141–Asn146 already shows some variability in FGF1 and FGF2. Compared with FGF1, the largest difference in C^α positions in this loop is 4.0 Å (at Ala142), while the largest difference is 2.0 Å (at Ala142) compared with FGF2. In FGF9, the loop containing the glycosylation site at Asn79 is identical to that of FGF1 and FGF2. N-terminal sequencing and

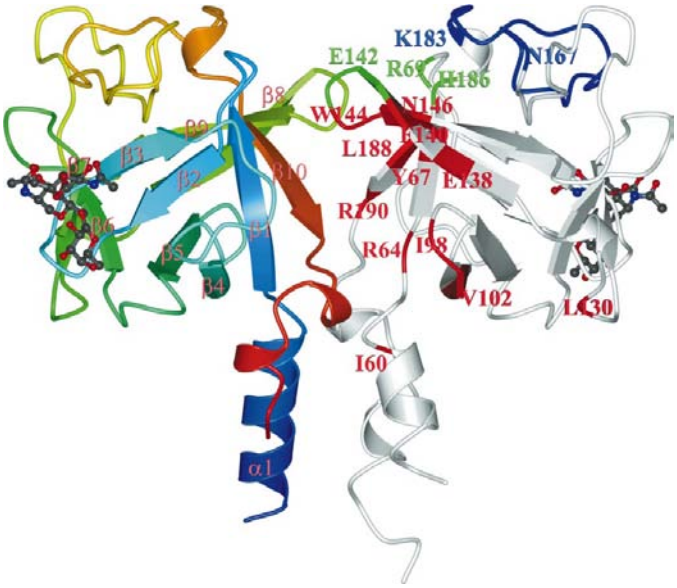

Figure 1
Ribbon representation of the FGF9 dimer composed of chains *A* and *D* showing the carbohydrate moiety. Chain *D* is sequentially colour coded from blue (N-terminus) to red (C-terminus); secondary-structure elements are labelled corresponding to the sequence alignment. Primary (red) and secondary (green) receptor-binding residues according to Plotnikov *et al.* (1999) are indicated on chain *A*, with residue labels for the first and last residue of each region. The heparin-binding loop is shown in blue. [Drawn with *Molscript* (Kraulis, 1991) and rendered with *gl_render* (L. Esser, unpublished program) and *PovRay*.]

MALDI mass spectrometry indicated heterogeneity of the crystallized protein, with the major components starting at residues 19, 34, 38 and 42. In the structure residues become visible in one of the molecules (*D*) at residue 49, forming a helix between residues 49–62. In the other three molecules of the asymmetric unit the first residues visible in the electron density are 51 (*B*) or 52 (*A*,*C*), with the helix starting at residue 52 or 54. The C-terminal residues starting from residue 193 form an irregular helix which shows some variability in the four molecules of the asymmetric unit. Together, these N- and C-terminal parts form an extension which is clearly separate from the core structure.

3.2. Quaternary structure

There is increasing evidence for the capacity of FGFs to undergo either spontaneous or heparin-induced oligomerization, although the relation of such dimers and higher order oligomers to receptor binding and activation is still unclear. For FGF1 a heparin-linked dimeric structure has been reported (DiGabriele *et al.*, 1998), while for FGF2 in the presence of heparin both monomer and dimer structures were observed (Faham *et al.*, 1996). Moreover, chemical cross-linking, ultracentrifugation experiments (Herr *et al.*, 1997) and mass-spectrometric techniques (Davis *et al.*, 1999) provided evidence of self-oligomerization for FGF2 in the presence and in the absence of heparin. Nevertheless, in the structures of the FGF2–receptor complex (Plotnikov *et al.*, 1999) and the FGF1–receptor complex (Stauber *et al.*, 2000) both FGF molecules are separate and are only linked *via* the receptor molecules. In these structures, heparin is postulated to bind into a positively charged groove created in the receptor dimer, with the two termini bound to the heparin-binding domains of the FGF2 molecules (Plotnikov *et al.*, 1999; Stauber *et al.*, 2000).

FGF9 readily dimerizes under physiological conditions, probably more easily than other FGFs, and dimers of FGF9 are frequently observed by immunoblotting lysates of RCJ3.1C5.18 mesenchymal cells (Garofalo *et al.*, 1999) and L-8 myoblasts (unpublished observation). Accordingly, the FGF9 structure crystallized in the absence of heparin shows the four molecules of the asymmetric unit organized in two dimers related by non-crystallographic symmetry.

The solvent-accessible surface area calculated with *GRASP* (Nicholls *et al.*, 1991) varies between 8709 and 9008 Å^2 for the individual molecules depending on the length of the extensions. The surface areas of the dimeric molecules, chains *AD* and *BC*, are 15 292 and 15 311 Å^2, respectively, yielding buried surface areas of 2395 and 2637 Å^2, respectively, or approximately 1200 Å^2 per chain, well above the cutoff value of 400 Å^2 per chain used as one of the classification criteria by the Protein Quaternary Structure server PQS (http://pqs.ebi.ac.uk/pqs-doc/pqs-doc.shtml). A large part of the buried surface of the dimer is contributed by the N- and C-terminal extensions, as the buried area per dimer is reduced to 1551 and 1594 Å^2, respectively, when only the residues 62–193 of the FGF core structure are used in the calculation.

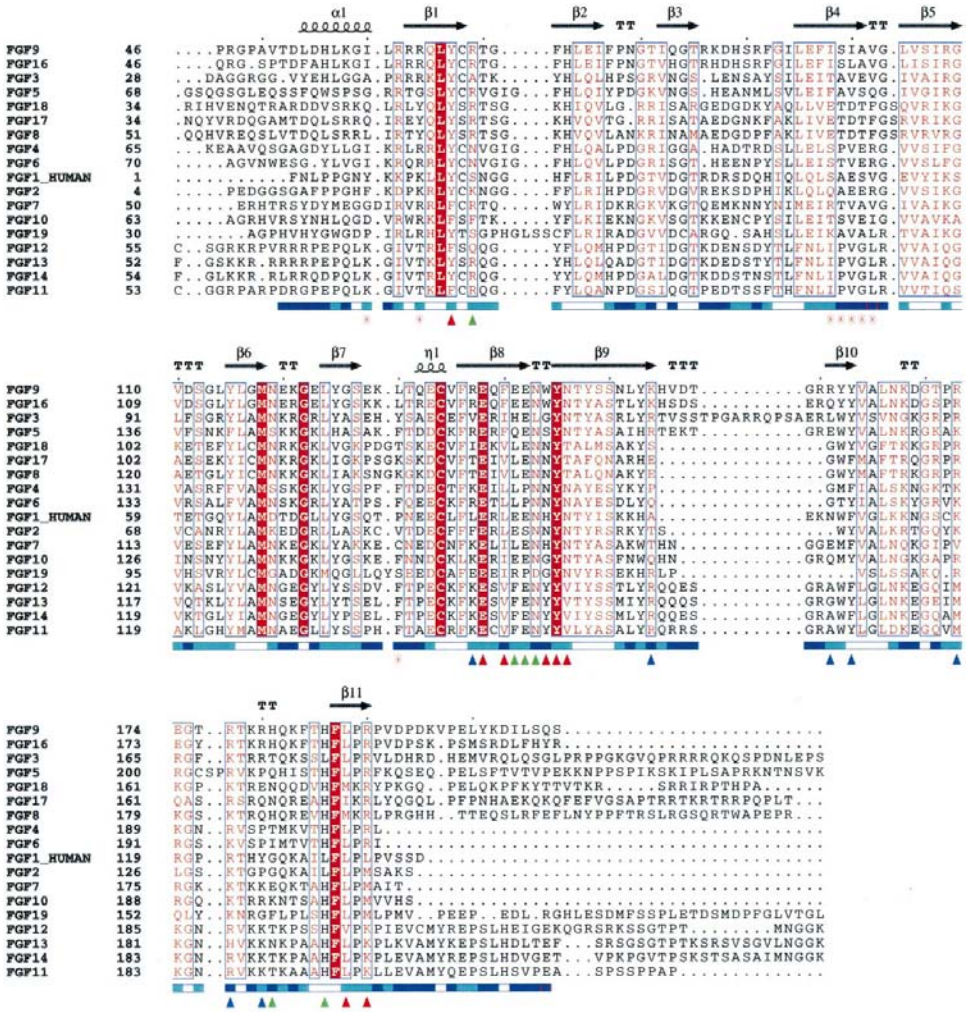

Figure 2
Sequence alignment of the human fibroblast growth factors showing schematically the secondary structure together with the primary (red triangle, contacts to receptor domain D2 and linker region; red star, contacts to receptor domain D3) and secondary (green) receptor binding sites according to Plotnikov *et al.* (1999). Blue triangles indicate potential heparin-binding sites in FGF9. The bar below each sequence block indicates high (blue) to low (white) solvent accessibility of residues in a monomer molecule.

The lack of these structured terminal extensions may therefore be one of the reasons why similar dimer formation has not been observed in the absence of heparin in the FGF1 and FGF2 structures. For FGF2, the crystal structure with the highest resolution (PDB code 1bgf) showed disorder for the first 19–20 N-terminal residues (Ago *et al.*, 1991), confirmed by NMR studies of complete FGF2 (PDB code 1rml), which showed disorder for the N-terminal 28 residues (Moy *et al.*, 1996). For FGF1 the crystal structure with the highest resolution (PDB code 2afg) showed disorder for the N-terminal 9–10 residues (Blaber *et al.*, 1996); for the NMR structure (PDB code 1rml) a molecule N-terminally truncated at residue 25 was used (Lozano *et al.*, 1998).

3.3. Dimer interface

The dimer interface in FGF9 consists mainly of hydrophobic contacts but includes four hydrogen bonds and two salt bridges related by non-crystallographic twofold symmetry. The hydrogen bonds connect the side chain of Tyr67 with the side chain of Asn143 and the side chain of Arg64 with the backbone carbonyl group of Val192 where the C-terminal extension starts, while the salt bridges connect Arg62 with Asp193, also at the start of the C-terminal extension. The hydrophobic contacts are concentrated in a prominent hydrophobic cluster of the residues Leu54, Leu57, Ile60, Leu61, Pro194, Val197 and Leu200 at the base of the structure, close to where the terminal extensions join the core. At the centre and top of the core structure, Pro191 and Leu188 together with Pro189 and the hydrophobic parts of the side chains of Arg190, Trp144 and Tyr145 form an additional, though less pronounced, hydrophobic patch.

A potentially important structural difference between FGF9 and FGF1 and FGF2 occurs in the dimer interface with the noticeable shift of the β-turn linking β8 and β9 (residues 139–146, corresponding to 96–104 in FGF2). In FGF9, the loop conformation is fixed by a hydrogen-bond network involving residues His181, His186, Glu141 and Glu142 (Fig. 3). The arrangement is stabilized further by a salt bridge between Glu142 and Arg69. Residues from this loop have been implicated in receptor binding (Venkataraman *et al.*, 1999); in the experimental FGF–receptor complexes, where residues from this loop make extensive contacts to the receptor, the loop has been found to undergo some conformational changes upon receptor binding (Plotnikov *et al.*, 1999; Stauber *et al.*, 2000). This conformational adaptation is likely to be much reduced in FGF9 because of the hydrogen-bonding network. Stabilization of this loop in a particular conformation by residues not directly involved in receptor binding, as in FGF9, could therefore have significant implications for receptor affinity. In the structure of FGF7 (Ye *et al.*, unpublished work; PDB code 1qqk), where Glu142 and Arg69 are conserved, the loop is in a conformation similar to that in FGF1 and FGF2 but lacks the salt bridge. It is most likely that the loop conformation in FGF9 is influenced by the hydrogen bond between Glu141 and His181, which is unique to FGF9 and FGF16. Similar interactions could occur in FGF5, which has two glutamines in these positions, and in FGF10, which has glutamic acid and lysine.

With the exception of residues from the terminal extensions, most of the residues (Fig. 1) involved in the dimer interface in FGF9 correspond to residues identified as belonging to the major receptor-binding sites in FGF2 (Venkataraman *et al.*, 1999; Plotnikov *et al.*, 1999, 2000; Stauber *et al.*, 2000). This is particularly true for residues Tyr67, Tyr145, Leu188, Ile60 and His186, corresponding in FGF2 to Tyr24, Tyr103, Leu140, Phe17 and Leu138 and in FGF1 to Tyr15, Tyr94, Leu133, Tyr8 and Leu131, which were found by Plotnikov *et al.* (1999) and Stauber *et al.* (2000) to be in contact with the receptor. These residues are almost completely buried (less than 10 Å^2 solvent-accessible surface calculated with *AREAIMOL*; Collaborative Computational Project, Number 4, 1994) in the FGF9 dimer interface and in order for them to become accessible to the receptor, molecule dissociation of these preformed dimers has to occur, at least in FGF9.

In the experimental FGF–receptor complexes (Plotnikov *et al.*, 1999; Stauber *et al.*, 2000; Pellegrini *et al.*, 2000) both the FGF ligand molecules are separate and linked only by heparin, either directly or *via* the receptor molecules. A complete separation of the FGF9 dimer requires the separation of the extensive hydrophobic interactions at the N- and C-terminal extensions. As it seems unlikely that these hydrophobic residues remain exposed to solvent, at least three alternative scenarios can be proposed.

(i) At present, there is no experimental evidence that residues outside of the core FGF structure participate in receptor binding, although in the FGF1–FGFR2 complex both FGF termini are in the vicinity of the receptor. In addition, preliminary results suggest that a complete deletion of both

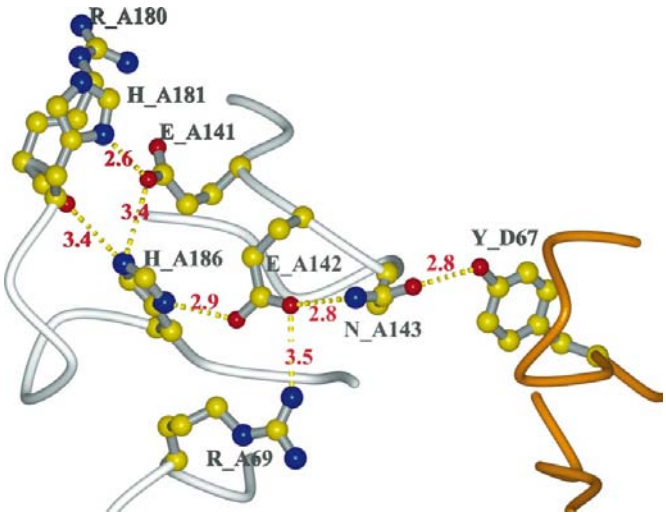

Figure 3
Hydrogen-bond network stabilizing the β8–β9 loop. Molecule *D* is represented with a grey chain trace and molecule *A* with an orange chain trace; hydrogen-bond distances (Å) are shown in red. [Drawn with *Molscript* (Kraulis, 1991) and rendered with *gl_render* (L. Esser, unpublished program) and *PovRay*.]

termini may have no apparent functional implications as evidenced by the capacity of a truncated form to both bind receptor and induce cell proliferation (Adar *et al.*, in preparation). The function of these terminal residues could therefore be to provide stability to the unliganded FGF molecules, probably correlated with the function as a non-cleaved secretion signal attributed to the 60 N-terminal residues by Revest *et al.* (1999), but to become redundant and flexible on receptor binding. It is intriguing to suggest that the observed heparin-independent self-association of FGF9 could have physiological significance by rendering the non-receptor-bound FGF in a protected inactive form by utilizing the same residues defined for receptor binding for a homotypic dimer interface.

(ii) These residues remain as a connecting region between the FGF molecules after a conformational change that exposes the buried receptor-binding residues. Preliminary modelling suggests that this could be possible with hinge regions probably in the region of residues 62 and 190–192. In this case, the terminal extensions could connect adjacent ligand–receptor complexes to form multimeric assemblies.

(iii) In the experimental receptor–ligand complexes (Plotnikov *et al.*, 1999; Stauber *et al.*, 2000) the secondary receptor-binding sites are different from the sites identified by site-directed mutagenesis as influencing receptor binding (Springer *et al.*, 1994; Zhu *et al.*, 1997, 1998). This discrepancy is presently not clear and may point to the involvement of other determinants in FGF receptor binding and activation.

At least some of the FGFs, especially FGF3 and FGF16, show a similar pattern of hydrophobic and hydrophilic residues in these terminal extensions in the sequence alignment. However, owing to the high sequence diversity and the structural flexibility, more structural investigation of these homologues is still required.

3.4. Potential heparin-binding sites

Heparin-binding sites have been structurally identified in the heparin-linked FGF1 dimer (DiGabriele *et al.*, 1998; PDB code 2axm) and in heparin complexes with FGF2 monomers (Faham *et al.*, 1996; 1bfb), where prominent interactions of basic residues with the sugars sulfate or carboxylate groups are involved. The surface of FGF9 contains three clusters of basic residues potentially suitable for heparin binding (Fig. 4). At least some of these sites contain a bound sulfate molecule, although the discrimination between bound water and sulfate is not completely certain owing to the limited resolution. The first site is in a pocket created by the insertion at Tyr153/Arg161 and the sulfate ion is bound to Arg180, Tyr163 and the backbone N atom of Arg161. This pocket is at a distance of approximately 14 Å from the nearest heparin-binding site in FGF1 and FGF2, but could occur also in FGF16, FGF13 and FGF11 which have a highly homologous insertion and identical or homologous residues in position 163 and 180. The second site, where Arg137, Lys154 and Arg161 form a cluster highly suggestive of sulfate binding, is even further away from the FGF1 and FGF2 homologous sites and is located almost on the opposite side of the molecule. A similar arrangement could also occur in FGF16, where Arg161 is replaced by a glutamine. The third site is formed by Arg173 and Arg177, which correspond to Lys118 and Arg122 in the heparin-binding loop in FGF1 (DiGabriele *et al.*, 1998) and to Lys125 and Lys129 in FGF2 (Faham *et al.*, 1996). Fitting the heparin structures observed in FGF1 (DiGabriele *et al.*, 1998; PDB code 2axm) and FGF2 (Faham *et al.*, 1996; PDB code 1bfb) to FGF9, however, shows that the high-affinity heparin-binding site described by the residues Asn28, Lys126 and Gln135 in FGF2 (Faham *et al.*, 1996) is partially blocked in FGF9 by the side chain of Phe184, which makes the backbone N atoms less accessible for sulfate binding, as observed for FGF2 and FGF1. In the experimental FGF2–FGFR1 complex, this site contains a bound sulfate ion and is proposed to bind the terminal part of heparin (Plotnikov *et al.*, 1999). Sulfate ions visible in the experimental FGF1–FGFR2 complex (Stauber *et al.*, 2000; PDB code 1djs), however, seem to correspond well with the potential heparin-binding sites on FGF9. In this complex, three sulfate ions are bound to FGF1 Lys128, Lys118 and Arg122, probably with contributions by Lys112 and

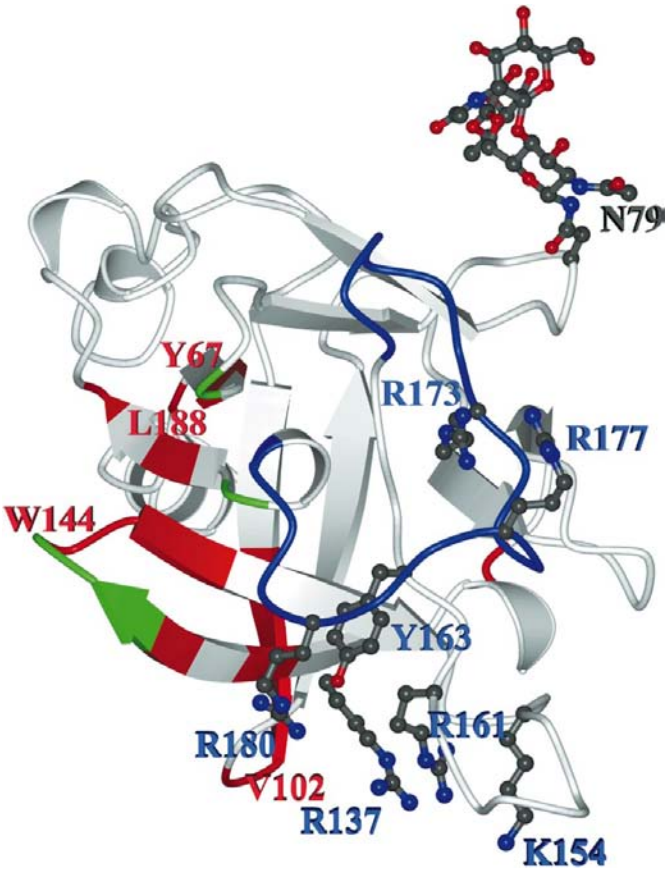

Figure 4
Potential heparin-binding sites in FGF9. Chain *A* is shown in an orientation perpendicular to Fig. 1 as a ribbon representation; the colour code is identical to Fig. 1. The sulfate-binding residues are shown in ball-and-stick representation and are identified by blue labels. Red labels for primary receptor-binding sites and the glycosylation site (black labels) have been included for reference to Fig. 1. [Drawn with *Molscript* (Kraulis, 1991) and rendered with *gl_render* (L. Esser, unpublished program) and *PovRay*.]

Arg119. In FGF9, Lys183 corresponds to FGF1 Lys128 and, in addition, Arg69 is directed very close to the sulfate bound to FGF1 Lys128. FGF9 Arg173 and Arg177 correspond to FGF1 Lys118 and Arg122, respectively, and only a small adjustment owing to Phe184 would be necessary for similar sulfate binding to the complex. These fine adjustments in the spatial organization of the heparin-binding residues in FGF9 may well coordinate with the distinct structural variants of sulfated domains on heparan sulfates required for binding and activation of different members of the FGF family as well as of other heparin-binding growth factors (Ornitz, 2000).

References

Ago, H., Kitagawa, Y., Fujishima, A., Matsuura, Y. & Katsube, Y. (1991). *J. Biochem (Tokyo)*, **110**, 360–363.

Blaber, M., DiSalvo, J. & Thomas, K. A. (1996). *Biochemistry*, **35**, 2086–2094.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Coughlin, S. R., Barr, P. J., Cousens, L. S., Fretto, L. J. & Williams, L. T. (1988). *J. Biol. Chem.* **263**, 988–993.

Davis, J. C., Venkataraman, G., Shriver, Z., Raj, P. A. & Sasisekharan, R. (1999). *Biochem. J.* **341**, 613–620.

DiGabriele, A. D., Lax, I., Chen, D. I., Svahn, C. M., Jaye, M., Schlessinger, J. & Hendrickson, W. A. (1998). *Nature (London)*, **393**, 812–817.

Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J. & Rees, D. C. (1996). *Science*, **271**, 1116–1120.

Fiebich, B. L., Jaeger, B., Schoellmann, C., Weindel, K., Wilting, J., Kochs, G., Marm, D., Hug, H. & Weich, H. A. (1993). *Eur. J. Biochem.* **211**, 19–26.

Folkman, J. & Klagsbrun, M. (1987). *Nature (London)*, **329**, 671–672.

Garofalo, S., Kliger-Spatz, M., Cooke, J. L., Wolstin, O., Lunstrum, G. P., Moshkovitz, S. M., Horton, W. A. & Yayon, A. (1999). *J. Bone Miner. Res.* **14**, 1909–1915.

Guimond, S., Maccarana, M., Olwin, B. B., Lindahl, U. & Rapraeger, A. C. (1993). *J. Biol. Chem.* **268**, 23906–23914.

Hecht, D., Zimmerman, N., Bedford, M., Avivi, A. & Yayon, A. (1995). *Growth Factors*, **12**, 223–233.

Herr, A. B., Ornitz, D. M., Sasisekharan, R., Venkataraman, G. & Waksman, G. (1997). *J. Biol. Chem.* **272**, 16382–16389.

Jones, T. A, Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kan, M., Wang, F., Xu, J., Crabb, J. W., Hou, J. & McKeehan, W. L. (1993). *Science*, **259**, 1918–1921.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.

Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **276**, 525–546.

Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kuriyama, M., Sakamoto, J. I., Nakatu, M., Kurokawa, T. & Sawada, H. (1995). *Ferment. Bioeng.* **80**, 327–333.

Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lozano, R. M., Jimenez, M., Santoro, J., Rico, M. & Gimenez-Gallego, G. (1998). *J. Mol. Biol.* **281**, 899–915.

Miyamoto, M., Naruo, K., Seko, C., Matsumoto, S., Kondo, T. & Kurokawa, T. (1993). *Mol. Cell Biol.* **13**, 4251–4259.

Moy, F. J., Safran, M., Seddon, A. P., Kitchen, D., Bohlen, P., Aviezer, D., Yayon, A. & Powers, R. (1997). *Biochemistry*, **36**, 4782–4791.

Moy, F. J., Seddon, A. P., Bohlen, P. & Powers, R. (1996). *Biochemistry*, **35**, 13552–13561.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Naruo, K., Seko, C., Kuroshima, K., Matsutani, E., Sasada, R., Kondo, T. & Kurokawa, T. (1993). *J. Biol. Chem.* **268**, 2857–2864.

Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281.

Ornitz, D. M. (2000). *BioEssays*, **22**, 108–112.

Pantoliano, M. W., Horlick, R. A., Springer, B. A., Van Dyk, D. E., Tobery, T., Wetmore, D. R., Lear, J. D., Nahapetian, A. T., Bradley, J. D. & Sisk, W. P. (1994). *Biochemistry*, **33**, 10229–10248.

Pellegrini, L., Burke, B. F., von Delft, F., Mulloy, B. & Blundell, T. (2000). *Nature (London)*, **407**, 1029–1034.

Plotnikov, A. N., Hubbard, S. R., Schlessinger, J. & Mohammadi, M. (2000). *Cell*, **101**, 413–424.

Plotnikov, A. N., Schlessinger, J., Hubbard, S. R. & Mohammadi, M. (1999). *Cell*, **98**, 641–650.

Rapraeger, A. C., Krufka, A. & Olwin, B. B. (1991). *Science,* **252**, 1705–1708.

Revest, J. M., DeMoerlooze, L. & Dickson, C. (1999). *J. Biol. Chem.* **275**, 8083–0890.

Rifkin, D. B. & Moscatelli, D. (1989). *J. Cell Biol.* **109**, 1–6.

Ruoslahti, E. & Yamaguchi, Y. (1991). *Cell*, **64**, 867–869.

Safran, M., Eisenstein, M., Aviezer, D. & Yayon, A. (2000). *Biochem. J.* **345**, 107–113.

Schlessinger, J. & Ullrich, A. (1992). *Neuron*, **9**, 383–391.

Spivak-Kroizman, T., Lemmon, M. A., Dikic, I., Ladbury, J. E., Pinchasi, D., Huang, J., Jaye, M., Crumley, G., Schlessinger, J. & Lax, I. (1994). *Cell*, **79**, 1015–1024.

Springer, B. A., Pantoliano, M. W., Barbera, F. A., Gunyuzlu, P. L., Thompson, L. D., Herblin, W. F., Rosenfeld, S. A. & Book, G. W. (1994). *J. Biol. Chem.* **269**, 26879–26884.

Stauber, D., DiGabriele, A. & Hendrickson, W. A. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 49–54.

Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**, 594–611.

Venkataraman, G., Raman, R., Sasisekharan, V. & Sasisekharan, R. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 3658–3663.

Yayon, A., Klagsbrun, M., Esko, J. D., Leder, P. & Ornitz, D. M. (1991). *Cell*, **64**, 841–848.

Zhang, J. D., Cousens, L. S., Barr, P. J. & Sprang, S. R. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 3446–3450.

Zhu, H., Anchin, J., Ramnarayan, K., Zheng, J., Kawai, T., Mong, S. & Wolff, M. E. (1997). *Protein Eng.* **10**, 417–421.

Zhu, H., Ramnarayan, K., Menzel, P., Miao, Y., Zheng, J. & Mong, S. (1998). *Protein Eng.* **11**, 937–940.

Zhu, X., Komiya, H., Chirino, A., Faham, S., Fox, G. M., Arakawa, T., Hsu, B. T. & Rees, D. C. (1991). *Science*, **251**, 90–93.