Molecular characteristics of fibroblast growth factor–fibroblast growth factor receptor–heparin-like glycosaminoglycan complex

Ganesh Venkataraman*, Rahul Raman†, V. Sasisekharan‡, and Ram Sasisekharan†§¶

*Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, ‡Department of Chemical Engineering, †Division of Bioengineering and Environmental Health, and §Center for Biomedical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

ABSTRACT Fibroblast growth factor (FGF) family plays key roles in development, wound healing, and angiogenesis. Understanding of the molecular nature of interactions of FGFs with their receptors (FGFRs) has been seriously limited by the absence of structural information on FGFR or FGF–FGFR complex. In this study, based on an exhaustive analysis of the primary sequences of the FGF family, we determined that the residues that constitute the primary receptor-binding site of FGF-2 are conserved throughout the FGF family, whereas those of the secondary receptor binding site of FGF-2 are not. We propose that the FGF–FGFR interaction mediated by the 'conserved' primary site interactions is likely to be similar if not identical for the entire FGF family, whereas the 'variable' secondary sites, on both FGF as well as FGFR mediates specificity of a given FGF to a given FGFR isoform. Furthermore, as the pro-inflammatory cytokine interleukin 1 (IL-1) and FGF-2 share the same structural scaffold, we find that the spatial orientation of the primary receptor-binding site of FGF-2 coincides structurally with the IL-1β receptor-binding site when the two molecules are superimposed. The structural similarities between the IL-1 and the FGF system provided a framework to elucidate molecular principles of FGF–FGFR interactions. In the FGF–FGFR model proposed here, the two domains of a single FGFR wrap around a single FGF-2 molecule such that one domain of FGFR binds to the primary receptor-binding site of the FGF molecule, while the second domain of the same FGFR binds to the secondary receptor-binding site of the same FGF molecule. Finally, the proposed model is able to accommodate not only heparin-like glycosaminoglycan (HLGAG) interactions with FGF and FGFR but also FGF dimerization or oligomerization mediated by HLGAG.

Fibroblast growth factors (FGFs) are important signaling molecules that play key roles during development and morphogenesis, as well as during several physiological and pathological situations such as wound healing, neovascularization, and tumor growth and progression (1–3). These growth factors induce mitogenic, chemotactic, and angiogenic activity in several different cell types (4). There are 18 FGF members in this important family, sharing between 10% and 55% sequence identity (5, 6). Despite only ≈55% sequence identity between FGF-1 and FGF-2, the nearly superimposable three-dimensional structures of FGF-1 and FGF-2 suggest that all the members of the family share a common structural fold, namely the β-trefoil scaffold (7–9).

FGFs signal though a set of four types of tyrosine kinase high-affinity receptors (FGFR-1 to -4) (10). The extracellular region of the FGFR contains three distinct Ig-like domains (I, II, and III). Alternative splicing of Ig III domain results in different spliced variants of each of the receptors (10). The

different FGFRs exhibit unique affinities and selectivity for each of the FGF family members and the signal they transduce (11–13). In addition, the activity of these growth factors is also regulated by a low-affinity receptor, which has been identified to be heparin-like glycosaminoglycan (HLGAG), the polysaccharide component of the cell surface heparan sulfate proteoglycans (HSPGs) (3, 14). The macromolecular interactions of the growth factor, HLGAGs of HSPGs, and FGFR that lead to signal transduction is key to signaling by this important class of molecules (1, 14–17). It has been suggested that HLGAGs mediate FGF dimerization or oligomerization, enabling the dimeric FGF to induce FGFR dimerization or oligomerization (18–20). The absence of structural information on the extracellular domains of the FGFRs has limited our understanding of the features and mechanisms of the interactions of FGF with FGFR and the role of HLGAGs in this process.

Earlier, Pantoliano *et al.* (21) described a model for the binding of FGF-2 to FGFR in which one molecule of FGF-2 binds to two molecules of FGFR. In this "growth hormone–like" model, the binding of one of the receptor molecules to the primary receptor binding site in FGF-2 leads to the recruitment of the second receptor molecule to occupy the second receptor binding site in the same FGF molecule. HLGAG binding to FGF-2 in this model is such that it can bridge the receptors and FGF-2, contributing to additional stability for the interaction. This model requires a 1:2 FGF-2 to FGFR ratio for signaling. However, this model is not able to account for the observation of FGF-2 dimerization or oligomerization mediated by HLGAGs and the role of this phenomenon in the formation of a ternary complex of FGF–HLGAG–FGFR.

It has been observed and reported that the FGFs and the pro-inflammatory cytokine interleukin 1 (IL-1) share the same β-trefoil fold (9). In fact, the structural scaffolds of FGF-2 and IL-1β and IL-1α are similar. In this study, upon superposition of the two structures and inspection of the primary receptor binding sites of IL-1β and FGF-2, we found that the topological positions of the two binding sites coincide. Although the specific amino acid residues that constitute the primary receptor-binding site for the FGF-2 are different from those of IL-1, these residues are entirely conserved across the FGF family. On the other hand, the amino acid sequences as well as the topological position of the secondary receptor-binding site residues are not at all conserved across the entire FGF family. The structural similarities between the IL-1 and the FGF systems enabled us here to outline molecular principles of FGF–FGFR interactions and to propose a model of the FGF–FGFR complex that satisfies many of the experimental observations, including specificity of FGF–FGR interactions,

Abbreviations: FGF, fibroblast growth factor; FGFR, FGF receptor; HLGAG, heparin-like glycosaminoglycan; IL-1, interleukin 1; IL-1R, IL-1 receptor.
¶To whom reprint requests should be addressed at: MIT, 77 Massachusetts Ave., Building 16-561, Cambridge, MA 02139. e-mail: ramnat@mit.edu.

FGF–HLGAG–FGFR interactions, and HLGAG-mediated FGF self-association leading to FGFR dimerization.

METHODS

Eighteen FGF sequences [human for all proteins except FGF-15 (from mouse) and FGF-10 (from chicken)] were obtained from the SwissProt and GenBank databases and aligned by using three different multiple sequence alignment programs, namely CLUST-ALW (22), MAP (Multiple Alignment Program) (23), and PIMA (Pattern Induced Multiple Sequence Alignment) (24). These three programs use different algorithms and weight matrices to perform the alignment. CLUSTALW uses a series of BLOSUMM (25) and PAM (26) weight matrices, PIMA uses its own set of weight matrices (24), and MAP uses the BLOSUMM50 (25) and PAM250 (26) weight matrices. The 18 proteins were aligned by using these sequence alignment algorithms and the default scoring matrices for these algorithms. Conserved residues in the family were identified from the aligned sequences as those residues that were identical in at least 13 of the 18 proteins at a particular position. The conserved residues were all represented in terms of the FGF-2 sequence numbers and the amino acid at that position in FGF-2. Sequence numbers for the corresponding position for any of the other FGFs can be obtained from the aligned sequences.

Calculation of the Normalized Solvent-Accessible Surface Area from the Crystal Structure. Several crystal structure coordinates of FGF-2 were used to determine the solvent-accessible surface area of each of the amino acids in the FGF-2 sequence. All the available FGF-2 crystal structures (Protein Data Bank codes: 1bas, 1fga, 2fgf, 4fgf, 1bfb, 1bfc, 1bfg, 1bla, 1bld, 2bfh, and 1bff, see www.pdb.bnl.gov for references) were used to determine solvent-accessible area of each of the residues. The solvation module of INSIGHT II (Molecular Simulations, Burlington, MA) was used to determine the solvent-accessible area of each of the residues, with a probe radius of 1.4 Å (27). The solvent-accessible areas for the different residues were almost independent of the structure used for the calculation, within small variations (less than 2%) except for some unusual substitutions. Any significant discrepancies in the accessible areas of residues in a particular crystal structure were attributable to mutations or chemical modifications in the protein.

To determine buried and exposed residues, the solvent-accessible surface areas of the residues had to be normalized to the maximum solvent-accessible area. This was accomplished by determining the maximum solvent-accessible area of all 20 amino acids in a polypeptide of the form Gly-Gly-Xaa-Gly-Gly, where Xaa stands for any amino acid. This calculation was performed for different secondary conformations of the polypeptide, including α-helix, extended, and β-strand conformations. In most cases it was found that the extended conformation gave the maximum solvent-accessible surface area. Polypeptide chains were used to minimize contributions from the free amino and carboxyl termini, and glycines were used in the polypeptide chain to remove the effect of side chains in the calculation of the area. The solvent-accessible surface area for each residue in a crystal structure of FGF-2 was divided by its corresponding maximum accessible area in the polypeptide chain to obtain the normalized solvent-accessible surface area of each residue. For the FGF-2 structure, the residue solvent accessibilities were normalized by their corresponding maximum accessible area, and an arbitrary cutoff of 15% was used to determine those residues that are surface exposed.

Homology Modeling of FGF–FGFR Complex. In this study, FGFR1 was modeled by using the coordinates of Ig domains of the IL-1 receptor (IL-1R) (see below), in contrast to other models for FGFR1 in the literature, which were based on telokins (28, 29) or BLAST sequence alignment to other proteins (30). The sequences of human FGFR1 (accession no. P11362) and IL-1R (accession no. P14778) were obtained from the SwissProt Database and aligned by using the different alignment programs described before (22–24). Alignment of the primary sequences led to about 22.7% sequence similarity between FGFR1 and IL-1R. The coordinates of the cocrystal structure of IL-1β with IL-1R and IL-1RA with IL-1R were obtained from the Protein Data Bank [1ITB.pdb (31), 1IRA.pdb (32)]. The IL-1R molecules from the two crystal structures were first compared. It was found that the structures of each of the three Ig domains were identical in the two crystal structures, although the relative orientations between them were different (31, 32). The coordinates of Ig II and Ig III from one of the structures (1ITB.pdb) were used for the homology modeling, as described below.

Using the homology module of INSIGHT II, we generated the coordinates of the Ig II and Ig III domains of FGFR1 (10) on the basis of the sequence alignments between the two proteins. Amino acid insertions or deletions in the FGFR1 sequence were modeled by generating possible chain conformations around the region of insertion or deletion and splicing it to the structure. The Ig II and Ig III structures were subject to extensive minimization, without charges first and then with charges until convergence (energy derivative less than 0.01). All the minimization runs were performed with the Discover module of the INSIGHT II modeling program, using the combined valence force field.

The Ig II domain of FGFR1 was manually docked to the primary receptor binding site of FGF-2 (33, 34), such that van der Waals contacts between the two proteins were maximized. The complex was then subject to extensive minimization to relieve any steric contact between the two proteins. Similarly, the Ig III domain was docked to the secondary receptor binding site of FGF-2 (33). The linker between the two domains was then generated and minimized. HLGAG binding to FGF was modeled based on the hexasaccharide—FGF-2 cocrystal (Protein Data Bank ID: 1bfc).

RESULTS AND DISCUSSION

Sequence Comparison of FGF Family of Molecules. As a first step in this investigation, we obtained the amino acid sequences of all 18 members of the FGF family from the database. A sequence comparison analysis of the 18 members was performed by using three different sequence alignment algorithms. Fig. 1 shows the sequence comparison of the 18 members of the FGF family using the CLUSTALW multiple sequence alignment algorithm (22). On the basis of this alignment, we chose those residue locations that are conserved as shown in Fig. 1. Use of other sequence alignment algorithms [PIMA (24) and MAP (23)], however, led to sequence alignments that are quite different, especially at the N- and C-terminal regions (data not shown). They, however, led to essentially the identical set of conserved amino acids for this family. These conserved amino acids do not have any obvious patterns in terms of their location within the primary sequence or in terms of the number of intervening amino acids.

Topological Mapping of Conserved Residues in the FGF Family. Despite limited sequence identity, several independent crystal structures of FGF-1 and FGF-2, the prototype members of the FGF family of molecules, show that the C^{α} traces of these proteins in all these crystal structures superimpose (rms deviation of less than 1 Å) (7–9). By extension, it is believed that the entire FGF family share the common "β-trefoil fold" (7–9). Therefore, we explored the topological relationship between the conserved amino acids in the FGF family by mapping them on to the three-dimensional structure of FGF-2.

Of the 26 amino acids that are conserved, 7 are glycine/proline and 13 are hydrophobic amino acids. On mapping onto the FGF-2 structure, the conserved glycines/prolines almost always form the turns between the β-sheets. These glycines/prolines therefore appear to be part of the structural scaffold.

```
 1 HST-1  MS GPGTAAVALLPAVLLALLAPWAGRGGAAAPTAPNGTL---EAELE      RRWESLVALSLARLPVAAQP          KEAAVQSGAGDYL-LGIKRLRRL
 2 HST-2  MSRGAGRLQGTLWALVFLGILVGMVVP--SPAGTRANNTL---LDS-      RGWGTLLSRSRAGL---AG           EIAGVNWESGYL-VGIKRQRRL
 3 aFGF   --                                       MAEGEITTFT---AL         TEKFN---LPPG NYKKPKL-L
 4 bFGF   --                                       MAAGSITTLP---AL         PEDGGSGAFPPG-HFKDPKR-L
 5 GAF    -MAPLGEVGSYFGVQD-----------AVPFGNVPVLP-----VDSP-------VLLNDHLGQSEAG----GL----------PRGPAVTDLDHL-KGILRRRQL
 6 FGF16  ----MAEVGGVFASLDWDLHGF------SSSLGNVPLA-------DSP------GFLNERLGQIEG-----KL-----------QRG-SPTDFAHL-KGILRRRQL
 7 FHF-1  -- MAAAIASSLIRQKRQARESNSDRVSASKRRSSPSKDG---RSLCE-------RHVLGVFSKVRFCS---GR-----------KRPVRRRPEPQL-KGIVTR- L
 8 FHF-4  ---MAAAIASGLIRQKRQAREQHWDRPSASRRRSSPSKN----RGLCN-------GNLVDIFSKVRIFG---LK-----------KRRLRRQ-DPQL-KGIVTR--L
 9 FHF-2  ---MAAAIASSLIRQKRQARER--EKSNACKCVSSPSKG----KTSCD-------KNKLNVFSRVKLFG---SK-----------KRRRRRP-EPQL-KGIVTK--L
10 FHF-3  ----MAALASSLIRQKREVREPGGSRPVSAQRRVCPRGT----KSLCQ-------KQLLILLLSKVRLCG---GR-----------PARPDRGPEPQL-KGIVTK-~L
11 AIGF   --MGSPRSALSCLLLHLLVLCLQAQVRSAAQKRGPGAGNPADTLGQGHEDRPFGQRSRAGKNFTNPAPNYPEEGSKEQRDSVLPKVTQRHVREQSLVTDQLSRRLIRTYQL
12 FGF17  -MGAAR------LLPNLTLCLQLLILCCQTQ-----------GENH----------------PSPNFN---------------QYVRDQGAMTDQLSRRQIREYQL
13 FGF18  -MYSAPSACTCLCLHFLLLCFQVQVLVAEEN-----------------------VDFR--------------------IHVENQTRARDDVSRKQLRLYQL
14 FGF5   --MSLSFLLLLLFFSHLILSAWAHGEKRLAPKGQPGPAATDRNPIGSSSR--------QSSSSAMSSSSASSSPAASLG---------SQGSGLEQSSFQWSPSGRRTGSL
15 INT-2  ---------------------------------------------MGLIWLLLLSLLEPGWPAAGPG-----ARLRRDAGGRGGVYEHL-GGAPRRRKL
16 KGF    ---MHKWILTWILPTLLYRSCFHIICLVGTISLACNDMT------------------P--EQMATNVNCS----S-----------PER-HTRSYDYMEGGDIRVRRL
17 FGF10  ---MCKKWILTNGASAFSHLPCCCLLLLFLVSSVPVTCHDLGQDMLSP-----------EATNSSSSSSSSFPSSFSSP---------SSAGRHVRSYNHL-QGDVRKRKL
18 FGF15  ------------------MARKWNGRAVARALVLAT-------------------LWLAVSGRPLAQQ---SQ------------SVSDEDPLFLYGWGKITRLQYL
```

```
 1 HST-1  YCNVGIG--FHLQALPDGRIGGAHADT-RDSLLELSPVERG-VVSIFGVASRFFVAMSSKGKLYGSPF-FTDECTFKEILLPNNYNAYESYKYPG----------------
 2 HST-2  YCNVGIG--FHLQVLPDGRISGTHEEN-PYSLLEISTVERG-VVSLFGVRSALFVAMNSKGRLYATPS-FQEECKFRETILPNNYNAYESDLYQG-----------------
 3 aFGF   YCSNGG---HFLRILPDGTVDGTRDRSDQHIQLQLSAESVG-EVYIKSTETGQYLAMDTDGLLYGSQT-PNEECLFLERLEENHYNTYISKKHAEKN---------------
 4 bFGF   YCKNGG---FFLRIHPDGRVDGVREKSDPHIKLQLQAEERG-VVSIKGVCANRYLAMKEDGRLLASKC-VTDECFFFERLESNNYNTYRSRKYTS---------------
 5 GAF    YCRTG----FHLEIFPNGTIQGTRKDHSRFGILEFISIAVG-LVSIRGVDSGLYLGMNEKGELYGSKK-LTQECVFREQFEENWYNTYSSNLYKHVDTG-----------RR
 6 FGF16  YCRTG----FHLEIFPNGTVHGTRHDHSRFGILEFISLAVG-LISIRGVDSGLYLGMNREGELYGSKK-LTRECVFREQFEENWYNTYASTLYKHSDSE----------RQ
 7 FHF-1  FSQQG----YFLQMHPDGTIDGTKDENSDYTLFNLIPVGLR-VVAIQGVKASLYVAMNGEGYLYSSDV-FTPECKFKESVFENYYVIYSSTLYRQQESG-------RA
 8 FHF-4  YCRQG----YYLQMHPDGALDGTKDDSTNSTLFNLIPVGLR-VVAIQGVKTGLYIAMNGEGYLYPSEL-FTPECKFKESVFENYYVIYSSMLYRQQESG----------RA
 9 FHF-2  YSRQG----YHLQLQADGTIDGTKDEDSTYTLFNLIPVGLR-VVAIQGVQTKLYLAMNSEGYLYTSEL-FTPECKFKESVFENYYVLYASSMIYRQQQSG-----------RG
10 FHF-3  FCRQG----FYLQANPDGSIQGTPEDTSSFTHFNLIPVGLR-VVTIQSAKLGHYMAMNAEGLLYSSPH-FTAECRFKECVFENYVVLYASALYRQRRSG----------RA
11 AIGF   YSRTSG---KHVQVLANKRINAMAEDGDPFAKLIVETDTFGSRVRVRGAETGLYICMNKKGKLIAKSNGKGKDCVFTEIVLENNYTALQNAKYEG----------------
12 FGF17  YSRTSG---KHVQVTG-RRISATAEDGNKFAKLIVETDTFGSRVRIKGAESEKYICMNRKGKLIGKPSGKSKDCVFTEIVLENNYTAFQNARHEG----------------
13 FGF18  YSRTSG---KHIQVLG-RRISARGEDGDKYAQLLVETDTFGSQVRIKGKETEFYLCMNRKGKLVGKPDGTSKECVFIEKVLENNYTALMSAKYSG----------------
14 FGF5   YCRVGIG--FHLQTYPDGKVNGSHEAN-MLSVLEIFAVSQG-IVGIRGVFSNKFLAMSKKGKLHASAK-FTDDCKFRERFQENSYNTYASAIHRTEKTG----------RE
15 INT-2  YCATK----YHLQLHPSGRVNGSLENS-AYSILEITAVEVG-IVAIRGLFSGRYLAMNKRGRLYASEH-YSAECEFVERIHELGYNTYASRLYRTVSSTPGARRQPSAERL
16 KGF    FCRTQ----WYLRIDKRGKVKGTQEMKNNYNIMEIRTVAVG-IVAIKGVESEFYLAMNKEGKLYAKKE-CNEDCNFKELILENNYNTYASAKWTHNGGE-----------
17 FGF10  YSYNK----YFLKIEKNGKVSGTKKENCPFSILEITSVEIG-VVAVKSIKSNYYLAMNKKGKVYGSKE-FNSDCKLKERIEENGYNTYASLNWKHNGRQ-------------
18 FGF15  YSAGPYVSNCFLRIRSDGSVDCEEDQN-ERNLLEFRAVALK-TIAIKDVSSVRYLCMSADGKIYGLIRYSEEDCTFREEMDCLGYNQYRSMKHHLH---------------I
```

```
 1 HST-1  MFIALSKNGKTKKG--NRVSPTMKVTHFLPRL------------------------------------------------
 2 HST-2  TYIALSKYGRVKRG--SKVSPIMTVTHFLPRI------------------------------------------------
 3 aFGF   WFVGLKKNGSCKRG--PRTHYGQKAILFLPLPVSSD---------------------------------------------
 4 bFGF   WYVALKRTGQYKLG--SKTGPGQKAILFLPMSAKS----------------------------------------------
 5 GAF    YYVALNKDGTPREG--TRTKRHQKFTHFLPRPVDPDKVP-----ELYKDILSQS--------------------------------
 6 FGF16  YYVALNKDGSPREG--YRTKRHQKFTHFLPRPVDPSKLP-----SMSRDLFHYR---------------------------------
 7 FHF-1  WFLGLNKEGQIMKG--NRVVKKTKPSSHFVPKPIEVCMYR-----EPSLHEIGEKQGR-----SRKSSGTPTMNGGKVVNQDST------
 8 FHF-4  WFLGLNKEGQAMKG--NRVKKTKPAAHFLPKPLEVAMYR-----EPSLHDVGETVPKPGVTPSKSTSASAIMNGGKPVNKSKTT-----
 9 FHF-2  WYLGLNKEGEIMKG--NHVKKNKPAAHFLPKPLKVAMYK-----EPSLHDLTEFSRSGSGTPTKSRSVSGVLNGGKSMSHNEST-----
10 FHF-3  WYLGLDKEGQVMKG--NRVKKTKAAAHFLPKLLEVAMYQ-----EPSLHSVPEASPS-----SPPAP------------------
11 AIGF   WYMAFTRKGRPRKG--SKTRQHQREVHFMKRLPRGHHTT----EQSLRFEFLNYPPF-----TRSLRGSQRTWAPEPR-----------
12 FGF17  WFMAFTRQGRPRQA--SRSRQNQREAHFIKRLYQGQLPFPNHAEKQKQFEFVGSAPT-----RRTKRTR-R----PQPL---------
13 FGF18  WYVGFTKKGRPRKG--PKTRENQQDVHFMKRYPKGQPEL------QKPFKYTTVTKR-----SRRIRPT------HPA----------
14 FGF5   WYVALNKRGKAKRGCSPRVKPQHISTHFLPRRFKQSEQPELSFTVTVPEKKNPPSPIKSKIPLSAPRKNTNSVKYRLKFRFG----·
15 INT-2  WYVSVNGKGRPRRG- FKTRRTQKSSLFLPRVLDHRDHEMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH
16 KGF    MFVALNQKGIPVRG--KKTKKEQKTAHFLPMAIT------------------------------------------------
17 FGF10  MFVALNGRGATKRG- QKTRRKNTSAHFLPMVVMS------------------------------------------------
18 FGF15  IFIQAKPREQLQDQKPSNFIPVFHRSFFETGDQLRSKMFSLPLESDSMDPFRMVEDVDHLVKSPSFQK-----------------
```

FIG. 1.   Multiple sequence alignment of FGF-1 to FGF-18 by the CLUSTALW program (22). Boxes mark regions of highly conserved amino acid residues as described in *Methods*. FGF-1 to FGF-9 and FHF-11 to FGF-14 were obtained from the SwissProt database. FGF-10 and FGF-15 to FGF-18 were obtained from the GenBank database. The accession numbers for the FGFs are as follows: FGF-1 (acidic FGF, aFGF), P05230; FGF-2 (basic FGF, bFGF), P09038; FGF-3 (INT-2, protooncogene protein precursor), P11487; FGF-4 (KS3 or heparin secretory transforming protein 1, HST-1), P08620; FGF-5, P12034; FGF-6 (heparin secretory transforming protein 2, HST-2), P10767; FGF-7 (keratinocyte growth factor, KGF), P21781; FGF-8 (androgen-induced growth factor, AIGF), P55075; FGF-9 (glia-activating factor, GAF), P54130; FGF-10, 2911146; FGF-11 (FGF homologous factor 3, FHF-3), Q92914; FGF-12 (FGF homologous factor 1, FHF-1), Q92912; FGF-13 (FGF homologous factor 2, FHF-2), Q92913; FGF-14 (FGF homologous factor 4, FHF-4), Q92915; FGF-15, 2257959; FGF-16, 2911170; FGF-17, 3041790; and FGF-18, 3355904.

Most of the remaining conserved amino acids map to the core region of the FGF-2 structure, whereas a select few of the amino acids map to the protein surface as shown in Fig. 2. A quantitative measure of the surface-exposed versus buried nature of the residues can be obtained by using the normalized solvent-accessibility area of the different residues as described in *Methods* (Table 1). The solvent-accessibility data show that 18 of the conserved amino acids are buried with little or no solvent contact and presumably contribute to the conserved structural scaffold within the family. Among the solvent-accessible residues, seven of the amino acids are located in spatial proximity to each other on the FGF-2 structure. These seven amino acids (Y24, E96, S100, N101, Y103, L140, and P141) form a contiguous van der Waals surface patch on the FGF-2 structure as shown in Fig. 3. Five of the above seven amino acids (Y24, E96, N101, Y103, and L140) have been implicated to be important for receptor binding on the basis of

site-directed mutagenesis studies (33, 35). These amino acids form the primary FGFR binding site, as discussed below. It remains to be seen whether the remaining two amino acids in the van der Waals surface patch in Fig. 3 play a role in receptor binding.

FGF–FGFR Interactions: Primary and Secondary Receptor Binding Sites. Two different receptor-binding sites have been identified in the FGF-2 (33–35). These sites have been termed the high- and low-affinity sites, based on their relative affinities for FGFR binding. As stated above, the primary receptor binding site is composed of a set of discrete amino acids forming a hydrophobic cluster on the FGF-2 surface, with amino acids Y24, E96, N101, Y103, L140, and M142. The amino acid residues K110, Y111, and W114 within the linear stretch between 106–115 in FGF-2 constitute the secondary binding site of FGF-2 (33, 34). The primary and the secondary receptor binding sites of FGF-2
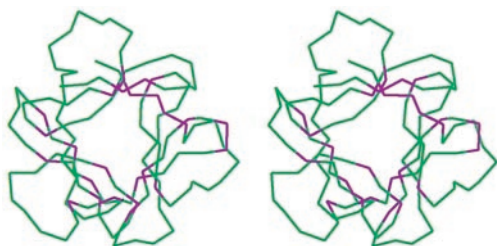
Fig. 2. Stereo representation of the C$^\alpha$ trace of the FGF-2 molecule. The conserved amino acids were mapped onto the FGF-2 crystal structure (Protein Data Bank ID: 1fga). The conserved amino acids shown in Fig. 1 are colored purple, while the rest of the C$^\alpha$ atoms are colored green. Most of the conserved amino acids fall within the core region of the FGF-2 structure, whereas few of the amino acids are close to the FGF-2 surface.

Fig. 3. The conserved solvent-accessible residues on the surface of the FGF-2 structure. These residues form a van der Waals surface patch. A Connolly surface rendering of the FGF-2 generated by using INSIGHT II with the solvent-accessible conserved amino acid residues colored in purple is shown. This site coincides with the primary receptor binding site identified on the basis of site-directed mutagenesis studies (33).

map onto diametrically opposite sides of the molecule (Fig. 4).

As discussed in the previous section, the primary receptor-binding site is conserved in all the members of the FGF family. On the other hand, the residues corresponding to the secondary receptor-binding site of FGF-2 are not conserved in the FGF family. At the location of the secondary receptor binding site of FGF-2, the residues are highly variable, with amino acid deletions and/or insertions in the other members (Fig. 1). The facts that the primary receptor-binding site contributes to a significant binding interaction with the FGFR (33) and that this site is conserved in the FGF family suggest that the dominant interaction between all the FGF molecules and FGFR is likely to be similar if not identical. However, it is known that the FGF family members are able to bind and activate the FGFR differently (36). This observation suggests that the "specificity" role in FGFR binding and activation is contributed by the secondary binding site. For the other members of the FGF-2 family, the binding affinity contributed by the secondary site could be weak or strong compared with FGF-2.

Comparison of IL-1 Receptor-Binding Site with FGF-2 Receptor-Binding Site. Although sequence similarity between FGF and the IL-1 family is very low, it has been observed by several investigators (7–9) that the three-dimensional structure of IL-1 is very similar to that of FGF, in that both have the β-trefoil fold. Furthermore, the cocrystals of IL-1β with IL-1R and also IL-1R-IL-1 RA (IL-1 receptor antagonist that binds to IL-1R, but does not signal) clearly point to the amino acid residues on IL-1β (or IL-1 RA) that interact with the receptor

(31, 32). Similar to FGF-2, IL-1 also binds to IL-1R through a primary receptor-binding site and a secondary receptor-binding site (31). The interaction of IL-1 RA with IL-1R is the same at the primary binding site, but the secondary binding site interaction with IL-1R is different, leading to their functional differences (32).

Structural superposition of C$^\alpha$ trace of FGF-2 and IL-1β crystal structures results in overall superposition of 50 amino acids (rms deviation of 0.59 Å). Fig. 5 compares the spatial relationship of the residues involved in receptor binding for FGF-2 and IL-1β. Note that the primary receptor binding van der Waals surface patch of FGF-2 shown in Fig. 3 forms part of the corresponding receptor-binding van der Waals patch of IL-1β. Thus, the spatial orientation of the primary receptor-binding site of FGF-2 coincides structurally with the IL-1β receptor-binding site when the two proteins are superimposed. This coincidence is significant, considering that the two proteins do not have any similarity functionally. Furthermore, similar to IL-1R, the extracellular domain of FGFR is also composed of three Ig-like domains and is believed to adopt a fold similar to IL-1R (32). Thus, the similarities between FGF and IL-1— namely, (*i*) the presence of primary and secondary receptor binding sites, (*ii*) the coincidence in the spatial orientation of the primary receptor-binding site, and (*iii*) the structures of the extracellular domains of the receptors—taken together, provided a framework to develop a model for the FGF system. However, other factors, including HLGAG interactions, need to be incorporated into the model.

A Model for FGF–FGFR Interaction. In FGFR, it has been shown that Ig domain I is sometimes missing in the functional

Table 1. Amino acid residues that are conserved in the FGF family

Conserved residue	SA score	Conserved residue	SA score	Conserved residue	SA score
L23	0.0015	M76	0.0016	Y106	0.0033
Y24	**0.2775**	G80	0.0004	S108	0.0053
L32	0.0015	L82	0.0115	G122	0.0409
G38	0.0753	C92	0.0052	G127	0.0015
G42	0.0884	F94	0.0005	L138	0.0991
V63	0.0516	**E96**	**0.3037**	F139	0.0007
I65	0.0022	**S100**	**0.9254**	**L140**	**0.2997**
G67	0.0093	**N101**	**0.4518**	**P141**	**0.2745**
Y73	**0.1729**	**Y103**	**0.1788**		

The conserved amino acid residues were identified by aligning the sequences of all 18 FGF family members by using three different multiple sequence alignment programs (22–24). The numbering of the amino acids corespond to their position in the FGF-2 sequence. All three of the sequence alignment programs resulted in identifying the identical set of conserved residues, although the alignment (not shown in the table) was quite different, especially at the N and the C termini. The normalized solvent-accessible surface areas (SA scores) of the conserved amino acids in the FGF-2 crystal structure are tabulated. Surface-accessible residues selected as described in *Methods* are shown in boldface type. These normalized solvent-accessible surface area values were computed by using coordinates of the FGF-2 crystal structure (Protein Data Bank ID: 1 fga) as described in *Methods*. Note that amino acids R22 and M142 were not included in the analysis, although they have been included in the site-directed mutagenesis studies (33). These amino acids correspond to the N- and the C-terminal regions of the crystal structures and hence did not provide reliable solvent-accessibility values.
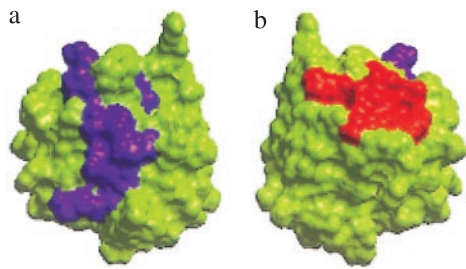
FIG. 4.   Primary and secondary receptor binding sites of FGF-2. The binding sites are identified on the basis of site-directed mutagenesis studies (33–35). (*a*) On the left is the Connolly surface rendering of the FGF-2 molecule with the residues corresponding to the primary site in purple and those corresponding to the secondary site in red. Notice that the secondary site (red) is not visible in this view as it is on the diametrically opposite end of the molecule compared with the primary site. (*b*) Same molecule as in *a* but rotated about a vertical axis in the plane of the paper by 180° so as to view the molecule from the other side.

forms of FGFR and the binding of different FGFs to FGFR lies primarily in domains II and III (1). The Ig II domain is conserved in sequence among the various FGFRs, and variations in the C-terminal half of Ig III contributes to the specificity of FGF binding (10, 37). All the above observations taken together lead us to propose that the binding of FGFR to FGF–2 is such that one of the receptor Ig domains (Ig II) selectively recognizes and binds to the primary receptor-binding site on FGF-2, whereas a second Ig domain (Ig III) on the same receptor molecule binds to the secondary binding site on FGF-2. In the FGF–FGFR model proposed here (Fig. 6*a*), the receptor sites are such that the two domains of the FGFR wrap around the FGF-2 molecule to result in a 1:1 FGF-2-to-FGFR binding stoichiometry. The HLGAG chains bind to the FGF-2 molecule such that they come in close proximity to both the FGFR domains and the linker between the receptor domains.

Extension of this model to other members of the FGF family is easily achieved (Fig. 6*a*). The sequence of the primary binding site of the FGF family molecules is conserved (Table 1), and hence it seems logical that this site interacts with the Ig II domain of the receptor, contributing to a common binding. The specificity of the FGF–FGFR interaction in the FGF family arises, therefore, from the binding of the variable nonconserved secondary site on FGF with the Ig III domain of FGFR. In this model, the relative spatial orientations of the Ig II and the Ig III domains and consequent receptor oligomerization are provided by the flexibility of the linker between Ig II and Ig III. Therefore, the FGF–FGFR binding seems to be such that the Ig III domain and the linker between Ig II and Ig III contribute to the specificity by recognizing specific secondary sites on the FGF molecule. On the other hand, the interaction at the primary binding site of FGF to the Ig II domain of FGFR seems conserved across the family.

In our model, the heparin-binding site and the receptor sites are oriented in such a way that a HLGAG bound to FGF would interact with the receptor as well (see below). Specifically, the C-terminal region of the Ig II domain, the linker between Ig II and Ig III, and parts of the Ig III domain come in close proximity with the HLGAG chain. The presence of a basic cluster of amino acids [the K18K region of Ig II and the linker region (38, 39)] is consistent with such an interaction between receptor and HLGAG. Such a trimolecular complex between HLGAG, FGF, and FGFR is shown in Fig. 6*b*.

FGFR Oligomerization. A ternary complex of FGF–HLGAG–FGFR leading to FGFR clustering has been proposed to be critical for signal transduction (3, 20). It has been proposed that FGFR clustering is promoted by HLGAG-mediated FGF-2 dimers and oligomers. We have earlier

proposed that FGF-2 forms oriented dimers and oligomers by preferentially self-associating in a side-by-side fashion (40). In the model proposed here, the clustering of the receptor molecules can be achieved by the oligomerization of FGF–FGFR complex (Fig. 6) by translation as in a beads on a string as shown in Fig. 7*a*. Here, the extracellular domains of the receptors bound to neighboring FGF molecules interact with significant van der Waals contact, providing additional stabilization. The recently solved cocrystal of FGF-1 with HLGAG
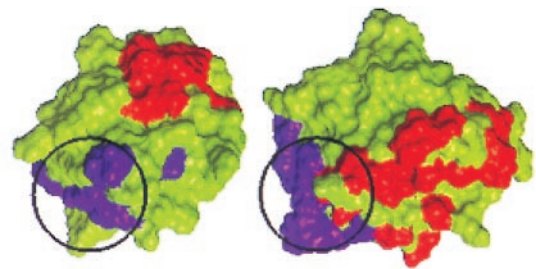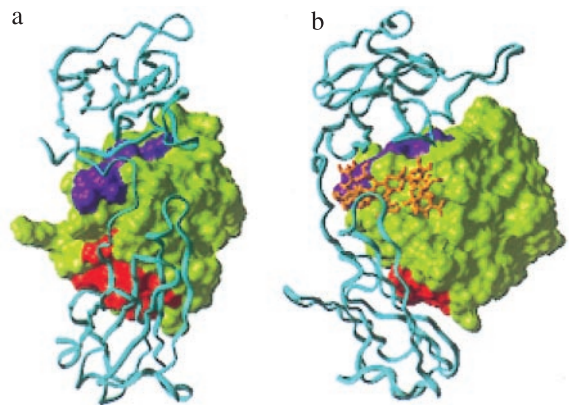

FIG. 5.   Comparison of the spatial orientation of the primary (purple) and secondary (red) receptor-binding sites of FGF-2 and the primary (purple) and secondary (red) receptor-binding sites of IL-1β (31). The Cα traces of the FGF-2 molecule and the IL-1β molecule were superimposed. The FGF-2 molecule was then translated horizontally for comparison. The primary binding site (highlighted with a black circle) of FGF-2 forms a part of the primary binding site of IL-1β [identified on the basis of the interaction with receptor domains I and II in the cocrystal structure (31)]. The relative spatial orientations of the two primary binding sites are similar (when superimposed they are at the bottom left of the molecules). The number of residues constituting the IL-1β primary binding site is much higher than in FGF-2. Although the three-dimensional orientations of the set of noncontiguous amino acids that confer receptor binding are very similar in the two proteins, there is no one-to-one correspondence in the amino acid composition at the receptor-binding sites. Note also that the secondary binding sites (red), which are believed to provide functional specificity, are spatially located differently in FGF-2 and IL-1β.

FIG. 6.   (*a*) Molecular model of the FGF–FGFR complex. FGF-2 molecule (Connolly surface rendering, colored in green) interacts with the Ig II and Ig III domains of the receptors. The Ig III domain of the receptor (bottom of the figure) interacts with the secondary binding site of FGF-2 (red) whereas the Ig II domain of FGFR interacts with the primary binding site of FGF-2 (purple). The Ig I domain of the receptor is not shown in this model. Notice that the linker between Ig II and Ig III is long and flexible to position the two Ig domains at diametrically opposite ends of the FGF-2 molecules. Extension of this model to the other members of the FGF family would suggest that the interaction at the primary site (purple and Ig II) would remain the same, whereas the interaction of the secondary site (Ig III and red) can vary to topologically other regions of the molecule. (*b*) A model of FGF-2–HLGAG–FGFR interaction. The HLGAG-binding site of FGF-2 is sandwiched between the primary and the secondary receptor-binding sites. HLGAG can bind to FGF-2 such that it interacts with both the receptor domains and the linker between the domains.
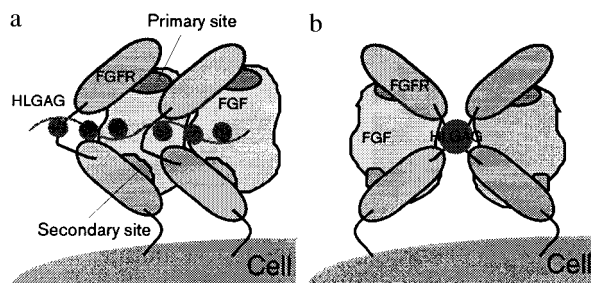
FIG. 7. (*a*) Schematic representation of the receptor clustering and oligomerization resulting from a side-by-side oligomerization of FGF as in a "beads-on-a-string" model. (*b*) Schematic representation of the receptor dimerization induced by a FGF-1–HLGAG dimer as observed in the cocrystal structure of FGF-1–decasaccharide (41). This model is schematically similar to the model suggested by DiGabrielle *et al.* (41).

decasaccharide (41) shows a symmetry-related dimer bridged by HLGAG with no protein–protein contact. In such a dimer, the receptor-binding surfaces of the two FGF molecules are still well exposed (41). The symmetry-related molecules are such that the two primary sites of the FGF-1 molecules in the dimer orient along the same direction. The clustering of the receptors promoted by such dimers is shown in Fig. 7*b*, where the receptors are now related by a twofold symmetry and the location of the heparin chain with regard to the receptor is different.

In summary, we show that the residues that constitute the primary receptor-binding site in FGF-2 are a set of surface-exposed amino acids that not only form a van der Waals surface patch in FGF-2 but also are conserved in the entire FGF family of molecules. We find this van der Waals surface patch in FGF-2 to structurally coincide with the primary receptor-binding site of IL-1, suggesting a common theme for the mode of receptor interaction for the two systems. We have proposed that the "conserved" primary site interactions mediated by Ig II is likely to be similar if not identical for the entire FGF family, whereas the "variable" secondary site interactions, mediated by the spliced variants of Ig III, govern specificity of a given FGF to a given FGFR isoform. Finally, the proposed model provides a framework for HLGAG interactions with FGF and FGFR as well as plausible modes of FGF dimerization or oligomerization mediated by HLGAG and how this impinges on FGFR clustering.

Note Added in Proof. A recent report by McKeehan's group (42), published following the submission of this manuscript, demonstrates that not only is IgII (the primary site) of FGFR R1 essential for binding FGF-1, FGF-2, and FGF-7, but this domain binds to each of the FGFs with equal affinity. Importantly, alteration in the heparin-binding region (conserved regions of IgII–linker–IgIII of FGFR1) and site-specific alteration of IgIII (secondary site) of FGFR1 modify FGF affinity and specificity for FGFR, respectively.

1. Conrad, E. (1998) *Heparin Binding Proteins* (Academic, New York), pp. 301–348.
2. Basilico, C. & Moscatelli, D. (1992) *Adv. Cancer Res.* **59**, 115–165.
3. Sasisekharan, R., Ernst, S. & Venkataraman, G. (1997) *Angiogenesis* **1**, 45–54.
4. Friesel, R. B. & Maciag, T. (1995) *FASEB J.* **9**, 919–925.
5. Smallwood, P. M., Munozsanjuan, I., Tong, P., Macke, J. P., Hendry, S. H. C., Gilbert, D. J., Copeland, N. G., Jenkins, N. A. & Nathans, J. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 9850–9857.
6. Naski, M. C. & Ornitz, D. M. (1998) *Front. Biol.* **3**, D781–D794.
7. Zhu, X., Komiya, H., Chirino, A., Faham, S., Fox, G. M., Arakawa, T., Hsu, B. T. & Rees, D. C. (1991) *Science* **251**, 90–93.
8. Erickson, A. E., Cousens, L. S. & Mathews, B. W. (1993) *Protein Sci.* **2**, 1274–1284.
9. Zhang, J., Cousens, L. S., Barr, P. J. & Sprang, S. R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3446–3550.
10. Givol, D. & Yayon, A. (1992) *FASEB J.* **6**, 3362–3369.
11. Ornitz, D. M., Xu, J., Colvin, J. S., McEwen, D. G., MacArthur, C. A., Coulier, F., Gao, G. & Goldfarb, M. (1996) *J. Biol. Chem.* **271**, 15292–15297.
12. Wang, F., Kan, M., Xu, J., Yan, G. & McKeehan, W. L. (1995) *J. Mol. Biol.* **270**, 10222–10230.
13. Wang, F., Kan, M., Yan, G., Xu, J. & McKeehan, W. L. (1995) *J. Mol. Biol.* **270**, 10231–10235.
14. Rapraeger, A. C. (1993) *Curr. Opin. Cell Biol.* **5**, 844–853.
15. Nugent, M. A. & Edelman, E. R. (1992) *Biochemistry* **31**, 8876–8883.
16. Schlessinger, J., Lax, I. & Lemmon, M., (1995) *Cell* **83**, 357–360.
17. Faham, S., Linhardt, R. J. & Rees, D. C. (1998) *Curr. Opin. Struct. Biol.* **8**, 578–586.
18. Ornitz, D. M., Yayon, A., Flanagan, J. G., Svahn, C. M., Levi, E. & Leder, P. (1992) *Mol. Cell. Biol.* **12**, 240–247.
19. Mascarelli, F., Fuhrmann, G. & Courtois, Y. (1993) *Growth Factors* **8**, 211–233.
20. 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.
21. 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.
22. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
23. Huang, X. (1994) *Comput. Appl. Biosci.* **10**, 227–235.
24. Smith, R. F. & Smith, T. F. (1992) *Protein Eng.* **5**, 35–41.
25. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
26. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure,* ed., Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), 5th Ed., pp. 345–352.
27. Lee, B. & Richards, F. M. (1988) *J. Mol. Biol.* **55**, 379–400.
28. Bateman, A. & Cothia, C. (1995) *Nat. Struct. Biol.* **2**, 1068–1074.
29. Pellegrini, L., Commander, P., Mulloy, B., Marin-Martinez, M., Blundell, T. L. & Burke, D. F. (1998) *Biochem. Soc. Trans.* **26**, 545–549.
30. Lam, K., Rao, V. S. R. & Qasba, P. K. (1998) *J. Biomol. Struct. Dyn.* **15**, 1009–1027.
31. Vigers, G. P. A, Anderson, L. J, Caffes, P. & Randhuber, B. J. (1997) *Nature (London)* **386**, 190–194.
32. Schreuder, H. Chantal, T., Trump-Kallmeyer, S., Soffientini, A., Sarrubi, E., Akeson, A., Bowlin, T., Yanofsky, S. & Barrett, R. W. (1997) *Nature (London)* **386**, 194–198.
33. Springer, B. A., Pantoliano, M. W., Barbera, F. A., Gunyulzu, P. L., Thompson, L. D., Herblin, W. F., Rosenfeld, S. A. & Book, G. W. (1994) *J. Biol. Chem.* **269**, 26879–26884.
34. Baird, A., Schubert, D., Ling, N. & Guillemin, R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2324–2328.
35. Zhu, H., Ramnarayan, K., Anchin, J., Miao, W. Y., Sereno, A., Millman, L., Zheng, J., Balaji, V. N. & Wolff, M. E. (1995) *J. Biol. Chem.* **270**, 21869–21874.
36. Ornitz, D. M., Xu, J., Colvin, J. S., McEwen, D. G., MacArthur, C. A., Coulier, F., Gao, G. & Goldfarb, M. (1996) *J. Biol. Chem.* **271**, 15292–15297.
37. Cheon, H.-G., Larochelle, W. J., Bottaro, D. P., Burgess, W. H. & Aaronson, S. A. (1989) *Proc. Natl. Acad. Sci. USA* **91**, 989–993.
38. Kan, M., Wang, F., Kan, M., To, B. Gabriel, J. L. & McKeehan, W. L. (1996) *J. Biol. Chem.* **271**, 26143–26148.
39. Kan, M., Wang, F., Xu, J., Crabb, J. W., Hou, J. & McKeehan, W. L. (1993) *Science* **259**, 1918–1921.
40. Venkataraman, G., Sasisekharan, V., Herr, A. B., Ornitz, D. M., Waksman, G., Cooney, C. L., Langer, R. & Sasisekharan, R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 845–850.
41. DiGabrielle, A. D., Lax, I., Chen, D. I., Svahn, C. M., Jaye, M., Schlessinger, J. & Hendrickson, W. A. (1998) *Nature (London)* **393**, 812–817.
42. Wang, F., Lu, W., McKeehan, K., Mohamedali, K., Gabriel, J., Kan, M. & McKeehan, W. L. (1999) *Biochemistry* 38, 160–171.