J. Mol. Model. **1997**, *3*, 233 – 239

Journal of Molecular Modeling

© Springer-Verlag 1997

Three-dimensional Model Structure for the Extracellular Domains of Fibroblast Growth Factor Receptor - 1 (FGFR-1)

Sérgio Oyama Junior, Sumika Kiyota, M. Terêsa M. Miranda, Angelo G. Gambarini and Wladia Viviani*

Departamento de Bioquímica, Universidade de São Paulo, Caixa Postal 26077, 05599-970, São Paulo, SP, Brasil (wladia@quim.iq.usp.br)

Received: 21 January 1997 / Accepted: 23 June 1997 / Published: 30 June 1997

Abstract

We have developed a model for the two immunoglobulin-like extracellular domains DII and DIII of the FGF receptor 1 (FGFR-1β), giving a special attention to the determination of the appropriate Ig set. The DII domain was aligned with the C-terminal domain of myosin light chain kinase (telokin) of the I set, and the DIII domain with the variable domain of the Bence-Jones immunoglobulin of the V set. Two assemblies, corresponding to different propositions for the domains relative orientation, have been refined and compared.

Keywords: Fibroblast growth factor, receptor modeling, immunoglobulin superfamily

Introduction

The Fibroblast Growth Factors (FGFs) constitute a family of structurally related proteins involved in the control of cell growth and differentiation, embryonic development, angiogenesis and malignant transformation [1]. These proteins act upon their target cells probably through the formation of a 1:1:2 complex comprising the FGF molecule, the polysaccharide heparan sulfate and the cell surface receptor [2]. The latter is part of a family of proteins composed by two or three extracellular immunoglobulin-like (Ig-like) domains, a single transmembrane chain and two intracellular domains with tyrosine kinase activity. Upon the complex formation, the two units of FGF receptor (FGFR) suffer cross phosphorylations resulting in signals transduction toward the cellular interior which elicit the biological responses reported above. Apparently, the dimerization or oligomerization of the receptor molecules induced by the ligand binding is a general mechanism of activation for several hormones and growth factors [3].

There are at least four closely related genes encoding the FGF receptor subtypes: FGFR-1 or flg, FGFR-2 or bek, FGFR-3 or cek2 and FGFR-4 [4]. Several FGFR isoforms differ in the number of extracellular Ig-like domains. The presence of three domains (DI, DII and DIII) defines an α-type receptor, whereas those isoforms lacking the N-terminal domain (DI) are known as β-type isoforms. These isoforms are produced by RNA processing of the FGFR-1 and FGFR-2 transcripts. The FGFR-3 and FGFR-4 receptors have only been reported as three Ig-like domains isoforms [4]. Data obtained from chimerical molecules formed by different FGFR domains suggest that the DII domain is the major structure responsible for the high affinity binding of the FGFs, whereas DIII appear to be important for the ligand specificity [5]. These results indicate that DII and DIII constitute the minimal structural requirement for the binding of FGFs and heparan sulfate ligands.

* *To whom correspondence should be addressed*

Another class of FGFR isoforms results from an alternative RNA processing event that produces three different sequences for the C-terminal half of DIII [4]. The IIIa isoform corresponds to a soluble receptor of unknown function, while the IIIb and IIIc forms display different FGFs specificity [6,7].

The structural requirements for the FGFs signaling complex formation are poorly understood. Presently, there are no experimental atomic coordinates for the FGFs receptors. In the lack of these, some models have been proposed based on homology modeling. Pantoliano et al. [2] have proposed a model for the DII and DIII domains based on the automatic alignment of the C_{H2} domain (Ig Fc fragment; C1 set) and V_H domain of the HYHEL-5 Fab structures (V set), respectively. However, neither the primary sequence nor the cysteines arrangement pattern in the folded domains showed a significant target similarity for both DII and DIII. Recently, Bateman and Chothia [8] published some outlines for FGFR structures, where they proposed the use of the I set, namely the telokin structure, for both domains building. However, due to the great difference in the number of residues between telokin and DIII (93 and 105, respectively, for the aligned regions) the resulting model also presents some inconsistencies, namely about how to accommodate these extra residues.

Here we report an alternative model for the extracellular domains of the FGFRs. This model was built assuming that DII and DIII are sufficient for both the FGFs and heparan sulfate binding; therefore we adopted the human FGFR-1β isoform as a standard for the receptor family modeling. The atomic coordinates of telokin [9], of V_L domain of Bence-Jones immunoglobulin [10] and of the extracellular domains of human growth hormone receptor [11] were the templates used for building our model, which was further refined and studied by classical mechanics/dynamics simulations.

Results and discussion

a) Model building

The extracellular portion of the FGFRs is formed by immunoglobulin-like domains. Their main common structural feature is the presence of two β-sheets packed together in approximately hundred-residues domains; the differences in the

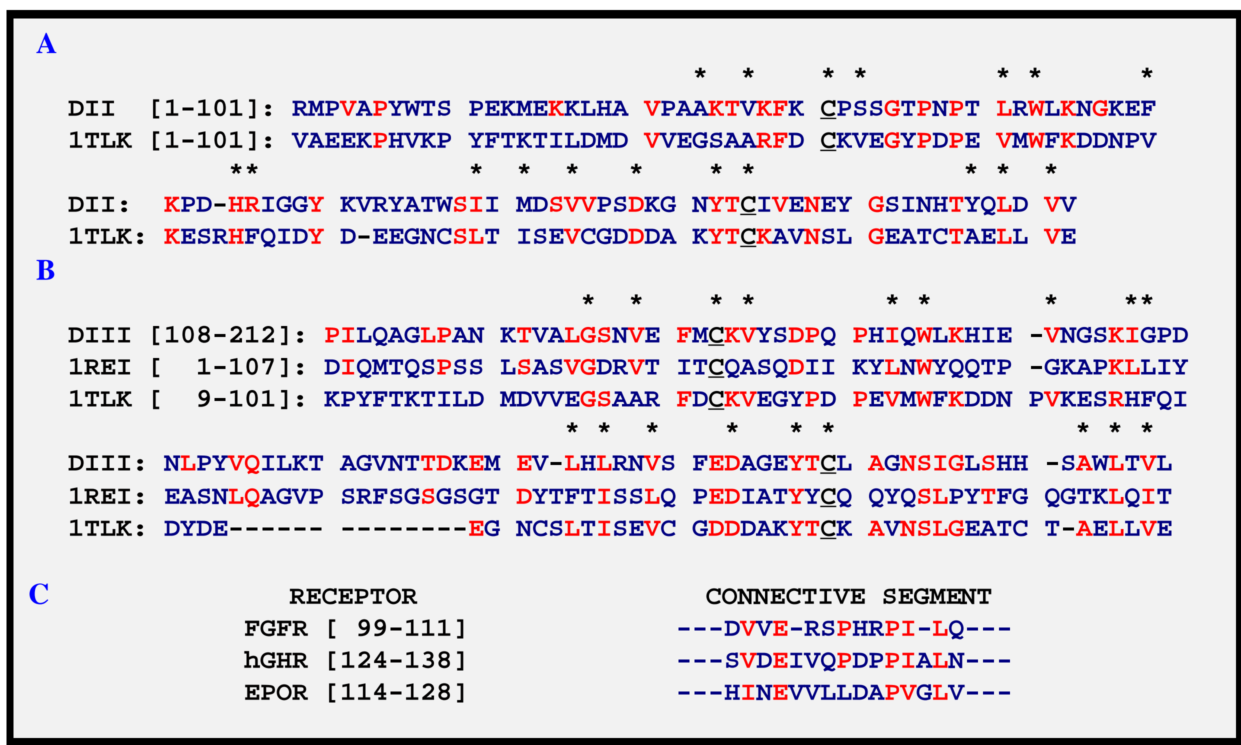
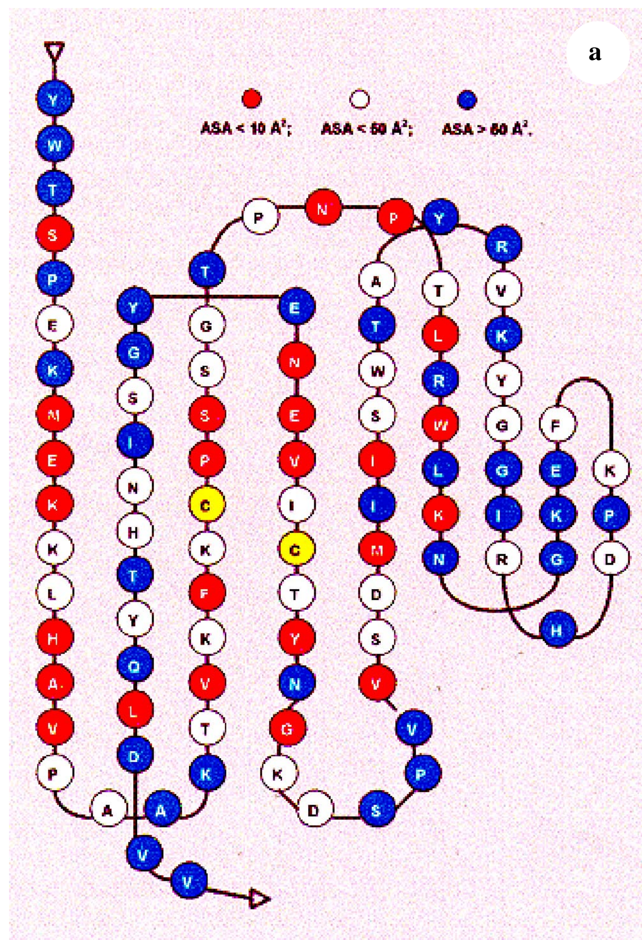


Figure 1. *Alignment of: (A) the FGFR-1β DII domain sequence with telokin (1TLK) [9]; (B) the DIII domain sequence with the V_L domain of Bence-Jones immunoglobulin (1REI) [10] and telokin; and (C) the connective segments between the Ig-like domains of human FGFR-1β, the human Growth Hormone Receptor (hGHR) [11] and the human Erythropoietin Receptor (EPOR) [17]. The conserved residues are colored in red. Mutations Lys/Arg; Asp/Glu; Leu/Ile/Val and Thr/Ser were scored as conservatives. The asterisks indicate the key residues implicated in the maintenance of the I and V sets as described by Harpaz and Chothia [14].*



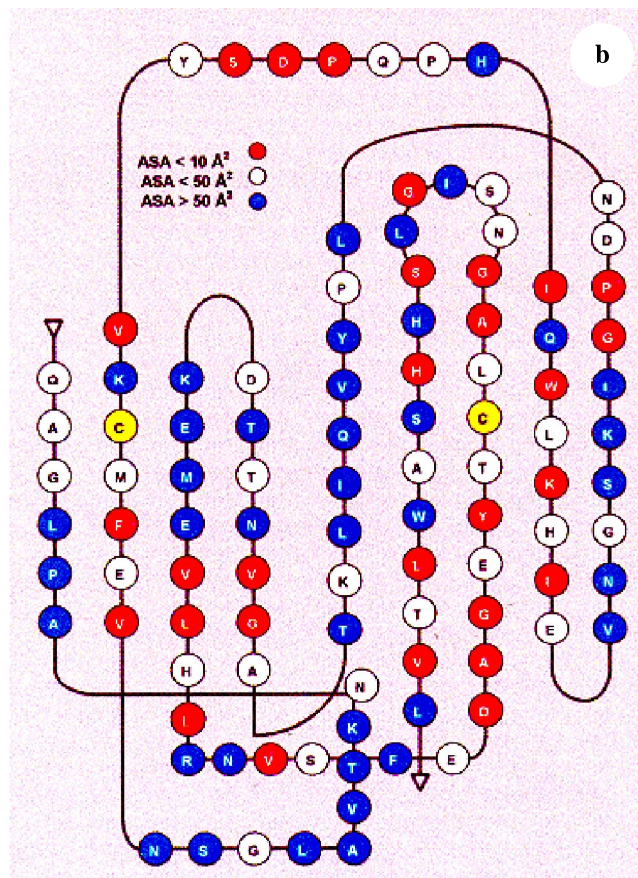


Figure 2. *Proposed residues arrangement for (A) DII domain of FGFR-1β corresponding to the typical I set folding pattern and (B) DIII domain model structure based on the V set arrangement. In both domains, the cysteine groups location in space allow the formation of a disulfide bond. Accessible surface area (ASA, in Å²) for each residue in both models is represented accordingly to the color scheme.*

number and size of their β-strands led to the classification into structural subgroups referred to as "sets" [12].

The building of an Ig-like domain by homology is usually based on the alignment of its primary structure with a similar template sequence. However, for sequences that present a low similarity level with their analogues, as is the case of our domains, a crude sequence homology alignment may lead to a doubtful result. For this reason, special attention has to be given to the template folding pattern that defines the set it belongs to, which has to be compatible with the characteristics of the target primary sequence. The choice of the set to which a new structure will be attributed is of critical importance for the correct prediction of its folding and the solvent accessibility of the different side chains. On the other hand, an incorrect choice of the subset, that might be the result of an alignment based on mere sequence homology, would compromise the reliability of the model itself.

On the basis of these considerations we decided that our main objective, when searching for our template structure, would be the attribution of the human FGFR-1β domains [13] to the most suitable immunoglobulin set, rather than the primary sequence similarity which we have considered of secondary importance. For this purpose, we adopted the interval between cysteine residues in the primary sequence. It is known that this parameter has been reported to be very variable within the Ig superfamily: differences in this interval size have been reported; besides, the number of cysteine residues itself is not a constant [12]. However, we found some consistencies which qualifies the cysteine separation as a suitable parameter for discriminating among the various Ig sets: (a) the number of cysteines residues and/or the disulfide bridges among the elements of the same set is relatively conserved; (b) variations longer than 10 residues are seldom found within a set; these should be attributed to the existence of a further strand and therefore to a folding pattern variation from one set to another. Therefore, guided by the Cys-Cys interval size, we proceeded to the selection of the best suited Ig sets.

This parameter itself was sufficient to indicate, without ambiguity, that the DII domain of FGFR-1β should be attributed to the recently described structural set called the "I" set

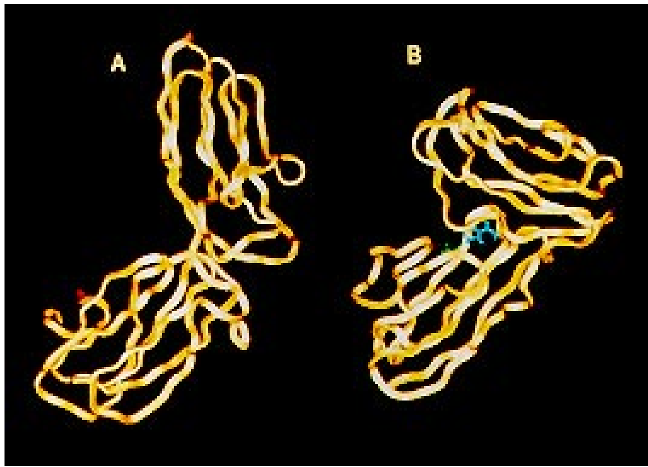


Figure 3. *Proposed relative orientation for the FGFR-1β assembly of domains DII and DIII based on (A) the orientation for the vascular cell adhesion molecule domains (VCAM) [16] and (B) the human growth hormone receptor extracellular domains (hGHR) [11]. The presence of a salt bridge between residues Lys17 (DII, blue) and Asp173 (DIII, green) is shown in model B.*

[14]. The I set is similar to the classical V set found in the variable domains of immunoglobulins, the main differences being the absence of C" strand and the presence of a shorter C' strand in the I set; consequently, the distances between cysteines in the former are shorter than those found in the latter. Among the I set three-dimensional structures available in PDB, we selected telokin (1TLK) [9], the C-terminal domain of myosin light chain kinase, as the best candidate for the matching sequence to the FGFR domain II. Figure 1a shows that in the alignment of the sequences of telokin and DII extracellular domain, the intervals between cysteines are very similar, differing by only one residue. Despite the relatively low homology between the two proteins, many of the key residues suggested by Harpaz and Chothia [14] as being responsible for the maintenance of the I set conformation could be aligned, supporting the reliability of the model.

This final result (Figure 2a) is quite different from that described by Pantoliano et al. [2], since they have used the C1 set for building the DII domain. It should be noted that this work was published after the telokin coordinates were determined by X-ray crystallography [9], but prior to the definition of the I set [14]. On the other hand, our model perfectly matches the recently published outline structure proposed by Bateman and Chothia [8], which is based on the conserved location of specific key residues in telokin and FGFRs primary sequences.

For the third extracellular domain of FGFR-1β (DIII) our findings, still based on the primary sequence interval between cysteines and key residues conservation, do not support any of the previously published models. As it can be seen in Figure 1b, the interval between cysteines is 14 residues greater in DIII than that found in telokin. By using the latter as template, the model proposed by Bateman and Chothia [8] resulted in quite a large insertion in the original folding pattern. This problem would have been avoided by accommodating the additional residues in an extra strand, providing a folding closely related to that found in members of the V set (Figure 2b). Our choice for this domain contemplates the V set as the best folding pattern; within this set, we selected the variable domain of Bence-Jones immunoglobulin (1REI) [10] as the template structure for building the DIII domain. By doing that, the separation between cysteines showed the best match; also, the residues proposed as responsible for the folding stability [8] were as well conserved as they would have been in the telokin alignment (Figure 1b). A further element of reliability is provided by the exact location of the cysteines which happen to be in the same β-strands for both the template and its target (Figure 2b).

The merge of the DII and DIII models deserved some special attention, namely the conformation of their connective segment. This segment, currently known as "hinge" in the immunoglobulin context, starts on the last b-strand of DII and extends along the first β-strand of DIII, being quite hydrophilic in its most exposed portion. It has been proposed by Wilkie et al. [15] that the relative orientation of these two domains could present some correspondence with that of the vascular cell adhesion molecule (VCAM) [16] domains, due to their similarity in both size and sequence of the two hinges (Figure 1c). We used this hypothesis to build a first structure for the FGF receptor domains assembly, which we called model A (Figure 3a). However, we found that the connective segment of FGF receptor presents a better similarity, in terms of primary structure, with the corresponding region of the extracellular domains of two cytokine class 1 receptors, the human growth hormone receptor (hGHR) [11] and the erythropoietin receptor (EPOR) [17] (Figure 1c). Another similarity between FGF, hGH and EPO receptors should be noticed. The relative orientation between the two Ig-like domains in hGHR is further stabilized by the presence of a salt bridge between the residues Arg39, located at the end of the first β-strand of domain I, and Asp132 which is at the beginning of the first β-strand of domain II [11]. A similar feature can be observed in the recently determined structure of the extracellular domains of EPOR [17], where a salt bridge between residues Arg32 and Asp122 participates in the stabilization of the two domains native orientation. As a consequence, it can be observed that the angle between the two Ig-like domains is maintained at the value of approximately 90° for both hGH and EPO receptors, despite the differences in the domains folding pattern. Of course, for the comprehension of this conformational preference, the similarity of the connective segment of the two proteins should not be neglected as well.

Based on these considerations, we built a second model for the extracellular domains of FGFR, which we will refer to as model B (Figure 3b), where the domains are set to a relative orientation based on that found for hGH and EPO

Table 1. *Energies obtained during the models refinement steps.*

Molecule	Energy minimized (kcal/mol)	Simulated Annealing (kcal/mol)	ΔE (kcal/mol)
DII	1031.46	1014.04	17.42
DIII	886.72	857.51	29.21
Model A	1701.02	1664.39	36.63
Model B	1747.65	1656.86	90.79

receptors. The salt bridge residues themselves are not conserved in none of their primary structures; an alternative salt bridge between Lys17 (DII) and Asp173 (DIII) could be observed (Figure 3b) in a spatial location correspondent to that found in hGH and EPO receptors. The relative stability of model B and the previously proposed model A, is compared and discussed in the next section.

b) Model Refinement

The stability of the proposed folded structures for DII and DIII was checked separately by observing the spontaneous intramolecular motions during a 100 ps simulation (after a 10 ps equilibration step) at 600 K, the temperature we found to be well suited for checking out the viability of a proposed Ig-like folded structure. All over this time, both domains demonstrated their tendency to keep the proposed structure.

We proceeded then to the structure refinement, carried out independently for the two separated domains DII and DIII. This refinement step consisted of a simulated annealing of the average structure assumed by the model during the previous 600 K simulated step: heating to 600 K for 50 ps, then successively cooling by 100 K each 20 ps down to 200 K, with an equilibration step of 5 ps at every temperature modification. The resulting structure was submitted to energy minimization using the Steepest Descent algorithm followed by the Conjugated Gradient algorithm until an energy variation up to 0.1 kcal/mol·Å and 0.05 kcal/mol·Å, respectively, was achieved. This resulted in a potential energy lowering of 17.42 kcal/mol from the minimized homology built to the refined structure of DII; similarly, DIII potential energy was lowered by 29.21 kcal/mol (Table 1).

The refined model structures of DII and DIII were then merged according to the process described in the previous section yielding two distinct assemblies (models A and B). These two models, whose difference consists in the bulk domains orientation, could have been tested and compared by a long step of molecular dynamics simulation in order to monitor their relative tendency to keep the initial conformation or undergo bulk domain motions. However, since bulk domains motion is a phenomenon which demands a long time scale (10^{-9} to 1 sec) for being observed, we decided to avoid the fastidious and time consuming molecular dynamics simulation by setting an alternative strategy by which the two structures would be firstly energetically refined and secondly compared to each one of them in order to check for the best solution. In consequence, we refined both models A and B by simulated annealing, starting the calculation at 300 K for 20 ps; the structures were cooled in 50 K intervals down to 50 K, ending by a 25 K and, finally, by a 10 K step. Every cooling step was preceded by equilibration and lasted for 5 ps, and cross terms were included in the two last steps. The resulting structures energy was minimized by using the same procedure as described above. The simulated annealing of model A resulted in a potential energy lowering of 36.63 kcal/mol; similarly, model B potential energy was lowered by 90.79 kcal/mol (Table 1).

Models comparison and concluding remarks

The energy difference between these structures, ΔE(a,b) = -7.53 kcal/mol, indicates that model B is the most stable. Of course part of this situation should be attributed to the existence of an ionic bond between Lys 17 and Asp 173 in model B. Let us remind that an analogue salt bond is believed to stabilize the experimentally determined conformations of hGH and EPO receptors.

It should however be noted that the salt bridge (Lys17, Asp173) contribution to the potential energies, that we estimated by driving apart the side chains involved in the bond, by a virtually infinite distance, has a calculated value of ΔE = –2.77 kcal/mol (Table 1) in a medium with dielectric constant $\varepsilon = 80$. This value can be subtracted from the calculated potential energy for B, showing that: a) even without this ionic contribution, the DII/DIII relative orientation in structure B keeps being ΔE(a,b) = –4.77 kcal/mol more stable than the arrangement presented in model A; b) the ionic bond

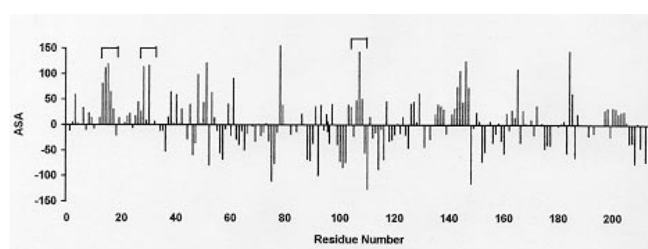


Figure 4. *Comparison of FGFR-1β model A and model B accessible surface area for individual residues. The profile was calculated as [ASAa - ASAb] (Å²) ; the residues indicated by brackets (Lys13, Met14, Glu15, Lys16, Lys28, Lys30 and Arg107) are those located at the DII/DIII interface and whose accessibility were lowered in the model B proposed orientation.*

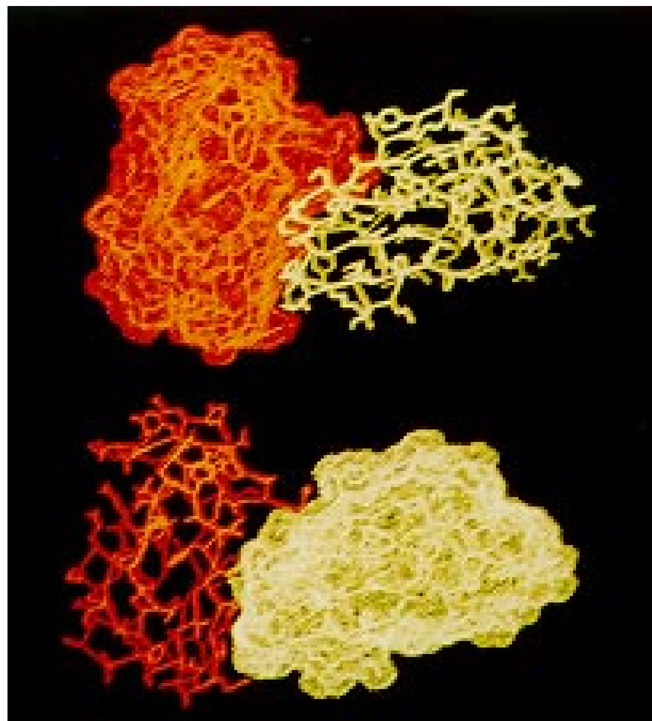


Figure 5. *Spatial complementary render of the DII (orange) and DIII (yellow) domains in model B, after refinement. For sake of visual clarity, the accessible surface is shown for one domain at a time.*

itself, by its 36.8% contribution in the stabilization of orientation B (when compared to A) is a crucial element for this conformation preference for FGFR extracellular domains.

Considering the spatial arrangement in B another interesting feature concerns the length between the Ig domains, the hinge region. The particular residue alignment attributed to this segment in B (Figure 1c) resulted in the bulk domains approximation of 2.1 Å when compared to the VCAM-based (A) hinge segment as reference. Of course that approximation in model B corresponds to a more packed structure; the total accessible surface area (ASA) of the molecule is lowered by 6.25%. The subtraction of individual ASA values for each residue leads to the profile of Figure 4. One can see how the new buried residues correspond to the ones located in the contact area between the domains, indicated by the brackets in Figure 4. An arbitrary domains approximation would have resulted in a very unfavorable intramolecular steric hindrance, illustrated by a large number of atom bumps.

This situation was definitely not the case for B, since no atomic distance lesser than 1.72 Å was found between the DII and the DIII domains. It is interesting to see how the molecular shape of these domains (Figure 5) present a spatial complementary in the contact region, to which the absence of atomic crashes should be imputed.

In the light of the above remarks, the structure of model B, inspired on the hGH and EPO receptors structures, seems much more suitable for the FGFR-1β than the VCAM-based one, as proposed by Wilkie et al. [15] and tested here with the name of model A. The same propositions can be extended to the other FGFR isoforms, that show 55 to 72% homology to each other [4] and whose structures can be easily built by using the model we proposed here as template.

Methods

The sequence of human Fibroblast Growth Factor Receptor 1 [13] was extracted from SWISS-PROT data bank [18]. Crystallographic coordinates of the template structures used in this work were obtained from the Brookhaven Data Bank [19].

Model building was performed on a IBM-RS6000 (3AT) workstation by using the HOMOLOGY and BIOPOLYMER modules of INSIGHT II package (Biosym/Molecular Simulation Inc., San Diego, CA). Refinement of the model structure and other calculation steps were carried in vacuum with the DISCOVER module of INSIGHT II, using the Consistent Valence Force Field (CVFF) and a dielectric constant $\varepsilon = 80$ in order to have a rough, implicit simulation of the aqueous solvent around the protein.

Acknowledgments: We wish to thank FAPESP, PADCT II/ FINEP and CNPq for financial support.

Supplementary material: 3D coordinated of FGFR-1 in PDB-format.

References

1. Burgess, W.H.; Maciag, T. *Annu. Rev. Biochem.* **1989**, *58*, 575.
2. Pantoliano, M.W.; Horlick, R.A.; Springer, B.A.; VanDyk, D.E.; Tobery, T.; Wetmore, D.R.; Lear, J.D.; Nahapetian, A.J.; Bradley, J.D.; Sisk, W.P. *Biochemistry* **1994**, *33*, 10229.
3. Ullrich, A.; Schlessinger, J. *Cell* **1990**, *61*, 203.
4. Johnson, D.E.; Williams, L.T. *Advances in Cancer Research* **1993**, *60*, 1.
5. Wang, F.; Kan, M.; Xu, J.; Yan, G.; McKeehan, W. *J. Biol. Chem.* **1995**, *270*, 10222.
6. Partanen, J.; Makela, T.P.; Eerola, E.; Korhonen, J.; Hirvonen, H.; Claesson-Welsh, L.; Alitalo, K. *EMBO J.* **1991**, *10*, 1347.
7. Fernig, D.G.; Gallagher, J.T. *Prog. Growth Factor Res.* **1994**, *5*, 353.
8. Bateman, A.; Chothia, C. *Nature Struct. Biol.* **1995**, *2*, 1068.
9. Holden, H.M.; Ito, M.; Hartshorne, D.J.; Rayment, I. *J. Mol. Biol.* **1992**, *227*, 840.
10. Epp, O.; Lattman, E.E.; Schiffer, M.; Huber, R.; Palm, W. *Biochemistry* **1975**, *14*, 4943.

11. De Vos, A.M.; Ultsch, M.; Kossiakoff, A.A. *Science* **1992**, *255*, 306.
12. Williams, A.F.; Barclay, A.N. *Annu. Rev. Immunol.* **1988**, *6*, 381.
13. Dionne, C.A.; Crumley, G.; Bellot, F.; Kaplow, J.M.; Searfross, G. ; Ruta, M.; Burgess, W.H.; Jaye, M.; Schlessinger, J. *EMBO J.* **1990**, *9*, 2685.
14. Harpaz, Y.; Chothia, C. *J. Mol. Biol.* **1994**, *238*, 528.
15. Wilkie, A.O.M.; Morriss-Kay, G.M.; Jones, E.Y.; Heath, J.K. *Curr. Biol.* **1995**, *5*, 500.
16. Jones, E.Y.; Harlos, K.; Bottomley, M.J.; Robinson, R.C.; Driscoll, P.C.; Edwards, R.M.; Clements, J.M.; Dudgeon, T.J.; Stuart, D.I. *Nature* **1995**, *373*, 539.
17. Livnah, O.; Stura, E.A.; Johnson, D.L.; Middleton, S.A.; Mulcahy, L.S.; Wrighton, N.C.; Dower, W.J.; Jolliffe, L.K.; Wilson, I.A. *Science* **1996**, *273*, 464.
18. Bairoch, A.; Boeckmann, B. *Nucl. Acids. Res.* **1991**, *19*, Supp., 2247.
19. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Meyer, E.F.Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Schimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.