

Inés López-Torrejón · Enrique Querol
Francesc X. Avilés · Masaharu Seno
Rafael de Llorens · Baldomero Oliva

Human betacellulin structure modeled from other members of EGF family

Received: 7 September 2001 / Accepted: 3 January 2002 / Published online: 25 April 2002
© Springer-Verlag 2002

Abstract We have modeled betacellulin (BTC) to gain insight into the structural elements that can explain its properties. The epidermal growth factor (EGF) signal transduction pathway, a significant mediator of several cell functions, is based on four closely related tyrosine kinase receptors. The ErbB receptors are transmembrane glycoproteins and signal transduction is initiated by ligand binding that induces receptor homo- or heterodimerization to form a complex containing two molecules of ligand and two molecules of receptor. The EGF family of ligands can be divided into three groups based on their ability to bind and activate distinct ErbB receptor homo- and heterodimers. Each member of the group formed by BTC, heparin binding EGF (HB-EGF) and epiregulin (EP) can interact with both the EGF receptor (EGFR) and heregulin receptors (ErbB-3 and ErbB-4), and are hence called “bispecific” ligands. BTC and EP

also present the distinctive feature that they activate all possible heterodimeric ErbB receptors. BTC has been modeled with the program MODELLER, using human EGF, human transforming growth factor alpha (hTGF α), human HB-EGF and human heregulin one alpha (hHRG-1 α) as templates. The structure of the model as well as that of the templates were optimized and a simulation of 100 ps was run for all. The main structural properties of the model and the templates were compared and in conclusion the hBTC conformation was closely similar to that of hTGF α . Electronic supplementary material to this paper can be obtained by using the Springer LINK server located at <http://dx.doi.org/10.1007/s00894-002-0072-2>.

Keywords Betacellulin · Epidermal growth factors · ErbB receptors

Abbreviations *PDB* Protein Data Bank · *RMSD* Root mean square deviation · *H-bonds* Hydrogen bonds · *MD* Molecular dynamics · *hEGF* Human epidermal growth factor · *hHB-EGF* Human heparin binding epidermal growth factor · *hTGF α* Human transforming growth factor alpha · *hHRG-1 α* Human heregulin one alpha · *hHRG-1 β* Human heregulin one beta · *hBTC* Human betacellulin · *EGF* Epidermal growth factor · *HB-EGF* Heparin binding epidermal growth factor · *TGF α* Transforming growth factor alpha · *BTC* Betacellulin · *HRG* Heregulin · *EP* Epiregulin · *AR* Amphiregulin · *EGFR* Epidermal growth factor receptor · *ErbB-[1,2,3,4]*, *HER[1,2,3,4]* Tyrosine kinase receptors related to the epidermal growth factor signal transduction pathway

Electronic supplementary material to this paper can be obtained by using the Springer LINK server located at <http://dx.doi.org/10.1007/s00894-002-0072-2>.

I. López-Torrejón · E. Querol · F.X. Avilés · B. Oliva (✉)
Institut de Biotecnologia i de Biomedicina “Vicent Villar Palasí”
and Departament de Bioquímica,
Universitat Autònoma de Barcelona,
08193 Bellaterra, Barcelona, Spain
e-mail: boliva@imim.es
Tel.: +34 93 5422933, Fax: +34 93 5422802

M. Seno
Department of Bioscience and Biotechnology,
Faculty of Engineering,
Graduate School of Natural Science and Technology,
Okayama University, 3-1-1 Tsushima-Naka,
Okayama 700-8530, Japan

R. de Llorens
Departament de Biologia, Facultat de Ciències,
Universitat de Girona, 17071 Girona, Spain

B. Oliva
Laboratori de Biologia Estructural Computacional,
Grup de Recerca d' Informàtica Biomèdica,
Departament de Ciències Experimentals i de la Salut,
Universitat Pompeu Fabra, C/ Doctor Aiguader, 80,
08003 Barcelona, Spain

Introduction

The epidermal growth factor (EGF) signal transduction pathway, a significant mediator of several cell functions, is based on four closely related tyrosine kinase receptors: ErbB-1 (also known as EGFR), ErbB-2 (or HER2/neu) for which no ligand has been described so far, ErbB-3

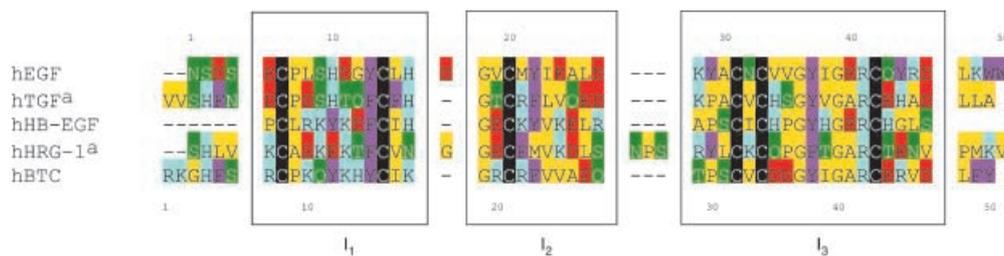


Fig. 1 Amino acid sequence alignment for hBTC and the templates used to build the model. The sequences were aligned manually using the six cysteines (in *black* background) as anchor points. [25] The amino acid type is background coloured as follows: *yellow* for apolar residues (Ala, Gly, Ile, Leu, Pro, Val and Met), *green* for polar residues (Asn, Gln, Ser and Thr), *red* for acid residues (Asp and Glu), *blue* for basic residues (Arg, His and Lys) and *pink* for aromatic residues (Phe, Trp and Tyr). In boxes I₁, I₂ and I₃ there are the equivalent C α positions used for the RMSD calculation. Two different numberings are indicated in the alignment: in the top of the alignment the numbering indicates the hEGF sequence, whilst the numbering of the hTGF α sequence is shown in the bottom. In the text, the sequence numberings of hHB-EGF and hBTC are referred to that of hTGF- α , while the sequence numbering of hHRG-1 α is referred to that of hEGF, according to Jones et al. [40]

(or HER3) which is characterized by an impaired kinase function, and ErbB-4 (or HER4). [1, 2, 3, 4, 5, 6] The ErbB receptors are transmembrane glycoproteins and signal transduction is initiated by ligand binding to the surface that induces receptor homo- or heterodimerization. [7] This dimerization activates the receptor intrinsic tyrosine kinase activity, leading to receptor transphosphorylation that induces the recruitment of intermediary effectors. [8] The multiplicity of EGF-like ligands, their receptors and their downstream substrates constitutes the basis for an interactive signalling network that is apparently aimed at increasing signal diversification.

The EGF family of ligands can be divided in three groups based on their ability to bind and activate distinct ErbB receptor homo- and heterodimers. The first group, EGF, TGF α and AR, interact with EGFR. The second, constituted by heregulins (HRGs), bind ErbB-3 and ErbB-4. The third group, formed by BTC, HB-EGF and EP, can interact with both EGFR and ErbB-4, and they are called “bispecific” ligands. This family of growth factors is characterized by six-cysteines forming a three-disulphide or cysteine-knot scaffold, so-called T-knot, [9, 10] shared by several mammalian growth factors, ω -toxins and protease inhibitors. [11] This spatial arrangement is the minimum structural motif required for full activity of all members of the EGF family.

Betacellulin (BTC) was first described from a mouse insulinoma cell line and a human breast adenocarcinoma cell line. [12, 13, 14] It is expressed in various tissues and body fluids, particularly in the pancreas. [15, 16] This expression implies that betacellulin seems to regulate the proliferation of pancreatic and breast cells. The BTC protein is synthesized as a 178-amino acid residue precursor, [13] and the released form is a 32-kDa glycoprotein including 80 amino acid residues, [15] although the 50-residue active EGF-like domain excludes the first

30 N-terminal amino acid residues. [17] The growth factors BTC and EP also have the distinctive feature that they activate all possible heterodimeric ErbB receptors. [3] These special characteristics of BTC can help to understand the structural elements involved in receptor specificity.

Up to date, four tertiary structures of the EGF family of ligands have been solved: hEGF, hTGF α , hHRG-1 α and hHB-EGF. [18, 19, 20, 21] All show a similar spatial arrangement despite primary sequence diversity. Binding of the EGF-like domain to an ErbB receptor favors dimerization of the receptor to form a complex containing two molecules of ligand and two molecules of receptor. [6] Ligand bivalency has been inferred from biophysical, [7] structure–function [22] and immunological approaches. [23] In the case of hHRG-1 β , using a recombinant chimerical ligand, [24, 25] it has been demonstrated that major determinants required for high-affinity binding to the primary ErbB receptor (ErbB-3 or ErbB-4) are contained within the N-terminus, whilst the C-terminus recruits the second ErbB partner. The second partner is preferentially ErbB-2, which has also been shown to increase the affinity of ligand binding to all ErbB receptor heterodimers. [26] According to this bivalency model, the various known ligands are not redundant, because each EGF-like domain selects its own set of preferred ErbB dimers.

Clearly, the structure differences between the members of the EGF-like family are implied in its functional behavior. Nevertheless for comparison with betacellulin, a model of its structure should be made. Additionally, betacellulin has an EGF-like domain that contributes to its biological role, [27] whilst 30 more residues at the N-terminal side of the protein may play a different role that is not the object of this study. The model of the EGF-like domain of betacellulin is obtained from the same members of the family, and the backbone chain is assumed to be very similar for all. However, the side-chain locations contribute to the protein function and differentiate it. Consequently, a model of the betacellulin EGF-like domain, hereafter indicated as hBTC model, and its comparison with the remaining members of known structure, can help in discussing the biological relevance and their biochemical roles. [27]

Methods

Figure 1 shows a multiple alignment of hBTC [25] with the sequences of hHRG-1 α , hEGF, hTGF α and hHB-EGF, whose 3D structures are known. In this

figure the numbering of the residues of these proteins is defined in order to standardize it in the text. The program MODELLER [28] was used to model build 10 conformations of hBTC. The 3D structures of hHRG-1 α , hEGF, hTGF α and hHB-EGF were used to extract the inner distances used for the modeling of hBTC according to the alignment shown in Fig. 1. All the structures on the alignment aside from hEGF were obtained from the Brookhaven Protein data Bank (PDB). [29] The 3D conformation of hEGF was taken from the Campbell et al. web site (<http://nmra.ocms.ox.ac.uk>). The average structure of the set of 10 structures of hBTC was optimized by 5000 steps of steepest descent with the program GROMOS87 and by simulated annealing from 10⁴ K to 5 K in 3 ps using a time step of 0.002 ps. The standard GROMOS with D4 parameters for the potential energy function was used for the optimization. [30] A twin-range method was used on the calculation of the non-bonding potential energies. The cut-off for the potential energy was 8 Å, without use of a switching function, whilst for the long-range interactions the cut-off was 13 Å. Bond lengths and angles were constrained by SHAKE [31] along the last 4500 steps and unconstrained in the initial 500 steps in order to accommodate the averaged structure. The energy of the optimized structure was -3.11×10^3 kJ mol⁻¹, where -2.22×10^3 kJ mol⁻¹ are for the electrostatic energy and -1.61×10^3 kJ mol⁻¹ for the Lennard-Jones interactions. The optimized structure was taken as the initial model of the 3D structure of hBTC. The initial structures of each template were obtained by optimizing the structures taken from the PDB. The initial structures are referenced by the time 0 in parenthesis.

The set formed by the template structures (hHRG-1 α , hEGF, hTGF α and hHB-EGF), plus the initial model of hBTC, was used to seed a molecular dynamics (MD) simulation for each structure. The MD simulation was run at 300 K, with a time step of 0.002 ps, using SHAKE [31] and a twin-range calculation of the non-bonding energies with cut-offs of 8 Å and 13 Å for long-range interactions. The structures were first embedded in a water box of water molecules (hHRG-1 α with 6420 water molecules, hEGF with 4522 water molecules, hTGF α with 4129 water molecules, hHB-EGF with 2906 water molecules, and the initial model of hBTC with 5009 water molecules). An optimization of each structure within the water box was run with 1000 steps of steepest descent, under similar conditions to those used for the simulation, with SHAKE and a twin-range cut-off for non-bonding energy. After optimization, a 100-ps simulation was run for each system, and the structures of each template plus that of hBTC model were extracted from the last simulation step without water molecules. The initial structures of each optimized template and the initial model of hBTC, plus the structures extracted in the last step of each simulation, were used for the analysis and comparison study. The structures here extracted are taken indicated by the time 100 in parenthesis.

Table 1 Functional residues. Amino acid residues playing an important role in the binding of hEGF, hTGF α and hHRG-1 β with their corresponding ErbB receptor. Amino acids of hHRG-1 β can be involved in the binding with ErbB-4 (in bold), ErbB-3 (in italics) or in both

hEGF ^a	hTGF- α ^b	hHRG-1 β ^c	hHB-EGF ^d	hBTC ^d
		Ser-1		Gly-3
		His-2		His-4
		Leu-3		Phe-5
		Val-4		Ser-6
		Glu-8	Arg-10	Lys-10
His-10	His-12	<i>Glu-10</i>	Tyr-12	Tyr-12
Tyr-13	Phe-15	Phe-13	Phe-15	Tyr-15
Leu-15	Phe-17	Val-15	Ile-17	Ile-17
His-16		Asn-16	His-18	Lys-18
Ile-23		Val-23	Val-24	Val-24
Leu-26			Leu-27	Glu-27
Tyr-37	Tyr-38	Phe-40	Tyr-38	Tyr-38
	Val-39		His-39	Ile-39
Glu-40		Ala-43	Glu-41	Ala-41
Arg-41	Arg-42	Arg-44	Arg-42	Arg-42
Gln-43	Glu-44	Glu-46	His-44	Glu-44
	His-45		Gly-45	Arg-45
Leu-47	Leu-48	Met-50		Leu-48
	Leu-49	Ala-51		Phe-49

^a From Hommel et al., 1992; [18] Campion and Niyogi, 1994; [37] and Groenen et al., 1994 [22]

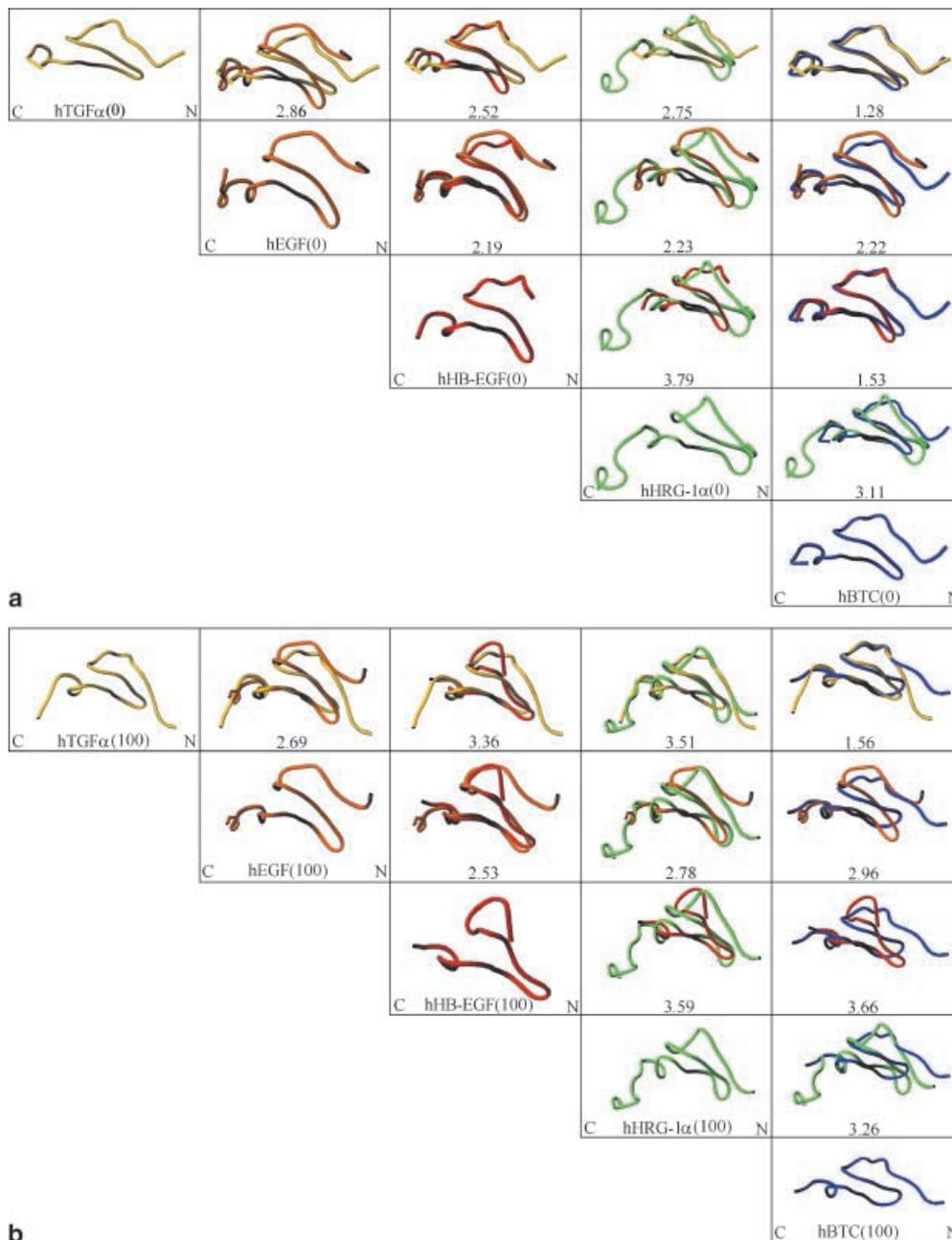
^b From McInnes and Sykes, 1997; [38] and McInnes et al., 2000 [39]

^c From Barcacci et al., 1995; [24] Jones et al., 1998 [40]; and Ballinger et al., 1998 [41]

^d Corresponding position in hHB-EGF and hBTC

The analyses of the model were done with the program XAM. [32] The structures of hHRG-1 α , [33] hEGF, [18] hTGF α [34] and hHB-EGF [21] were superimposed to the modeled structure of hBTC using the main-chain of the residues indicated within boxes I₁, I₂ and I₃ in Fig. 1. The superimposition was repeated analogously for the last step of the simulation. The superimposed structures were compared visually with the program TURBO-FRODO. [35] The comparison was done for the structures at the initial step (templates plus initial hBTC model) and for the resulting structures after 100 ps of MD simulation. The main residues involved in the interaction with their specific receptors are indicated in Table 1. The position of these residues in 3D space was carefully compared for each structure and for the model of hBTC after 100 ps of simulation. Similar positions were identified for each specific residue, and the conservation of the physico-chemical character of its side-chain was accordingly examined. The nomenclature used in the text indicates the residue by its single letter code, the residue number in the sequence, and the code of the protein (hHRG-1 α , hEGF, hTGF α , hHB-EGF, hBTC). The hydrogen bonding net of hBTC after 100 ps of simulation was calculated with XAM default values (distance of 2.9 Å between donor and acceptor, and (1450, 2150) ranges for the angle). The accessible surface per residue was calculated as the amount of the Connolly's surface per atom. [36] A probe water of

Fig. 2 Cross-superpositions presented by the trace of C_{α} atoms for the initial templates and the initial model of hBTC (a) and after 100 ps of MD simulation (b and c). **b** The comparison between the templates and the model at 100 ps. **c** The comparison by crossing the models at initial and last steps of the simulation. The RMSD value is added at the frame bottom of each superposition. C- and N-terminal orientations are from left to right. Figures were produced using the program TURBO-FRODO. [35] The colour code is: *orange* for hEGF (3D structure obtained from the web, see text), [18] *yellow* for hTGF α (PDB accession code 1yug), [34] *red* for hHB-EGF (PDB accession code 1xdt), [21] *green* for hHRG-1 α (PDB accession code 1haf) [33] and *blue* for hBTC model



1.4 Å for all the atoms of the residue was used in the calculation. The accessible surface was used to compare the structures in the initial step plus the result after 100 ps of simulation. All calculations were carried out on an Indigo Power2 (Silicon Graphics) with an R8000 processor.

Results and discussion

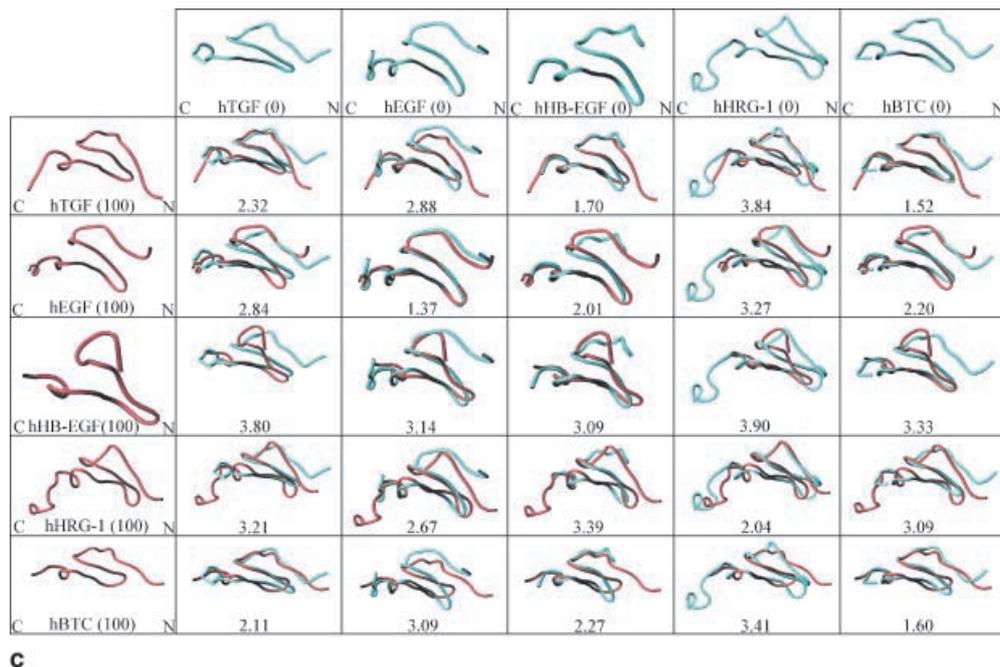
Study of the structure of the model

Root mean square calculations

The global root mean square deviation (RMSD) of the main-chain atoms has been calculated for all the pairs of

EGF-like domains, including the modeled structure of hBTC (see methods). Superpositions are shown in Fig. 2. In Fig. 2a is shown the comparison of the initial structures of the templates indicated by “(0)”, in Fig. 2b is shown a similar result for the last step of the simulation indicated by “(100)”, whilst Fig. 2c shows the cross-comparison between the initial and the final structures. The initial model of hBTC shows a small RMSD with respect to the initial structure of hTGF α (1.28 Å) and with respect to hHB-EGF (1.53 Å). With respect to the initial structure of hEGF the RMSD value is larger than these (2.22 Å), and also with respect to hHRG-1 α (3.11 Å). It seems reasonable that hBTC is more similar to hTGF α and hHB-EGF than to the hEGF or hHRG-1 α . Proteins hBTC and hHB-EGF are both able to bind both

Fig. 2 (continued)



EGFR and HRG receptors (ErbB-3 and ErbB-4), and when looking at the scaffold their structures are similar. However, not only the scaffold determines the functionality of a ligand protein, also the residue-accessibility and its chemical properties are important for its interaction. In the following paragraphs we present the results of studying the energetic features involved in the structure (namely the hydrogen bonding net) and the disposal of residues in the surface of betacellulin compared to hEGF, hTGF α , hHB-EGF and hHRG-1 α .

This is corroborated with the structures after 100 ps of simulation, where still the RMSD between the hBTC model and the hTGF α structure is about 1.56 Å, whilst the comparison between the initial hBTC model and the last structure of hBTC after 100 ps is of about 1.60 Å and 2.32 Å for a similar comparison with hTGF α . The comparison between the cross-related structures, hBTC(0) and hTGF α (100) with RMSD of about 2.51 Å, and hBTC(100) and hTGF α (0) with RMSD of about 2.51 Å, shows on average the smaller RMSD values. This also indicates that the structures have changed similarly along the simulation. On the other hand, the original small RMSD between hBTC and HB-EGF, increases to about 3.66 Å after 100 ps of simulation. This shows that hBTC is finally different to HB-EGF.

H-bonds

The hydrogen bonding net of the modeled hBTC(100) structure, and the structures of hTGF α (100), hEGF(100), hHB-EGF(100) and hHRG-1 α (100) are shown in Table 2. About 50% of the H-bonds present in hTGF α (100) are also present in hBTC(100). This corroborates the similarities found by the RMSD in these structures. However, few of the H-bonds present in the structure of hHB-EGF(100)

are also present in the modeled structure, even though its superimposition has a small RMSD value. The RMSD is a measure of the similarity/difference between the scaffold of two protein structures, and the scaffold of a protein structure defines the intra-net of H-bonds. This relation is maintained for hTGF α (100), hEGF(100) and hHRG-1 α (100) with respect to hBTC(100), but not for hHB-EGF(100).

Occasionally one of the partners of the H-bond is different in hBTC(100) than for some of the other EGF-like structures (see Table 2) although the position in the net is similarly located. For example, the side-chain-main-chain H-bond between hHB-EGF_Y12 and hHB-EGF_R10 has no equivalent H-bond in the hBTC(100) structure. Instead, the peptide bond of hBTC_Y12, in equivalent position, forms an H-bond with the main chain of hBTC_P9 (close to position 10 in the sequence).

It is noticeable that an H-bond at position hEGF_H16 in hEGF, a position that is important for the binding of this molecule with the receptor, is still present at the equivalent position hBTC_K18 in hBTC (see Table 2) but not at the hTGF α _H18 position. This location is not important for hTGF α , as seen from Table 1, and it marks a similarity between hBTC and hEGF, over the general and most common similarities found between hBTC and hTGF α . In other words, hBTC is commonly similar to hTGF α , being occasionally similar to hEGF in some residues that play some biological role for hEGF.

Structural comparison in functional residues

Chemical character

The sequence identity between the templates and the model is 48% with hTGF α , 41.5% with hHB-EGF, 36%

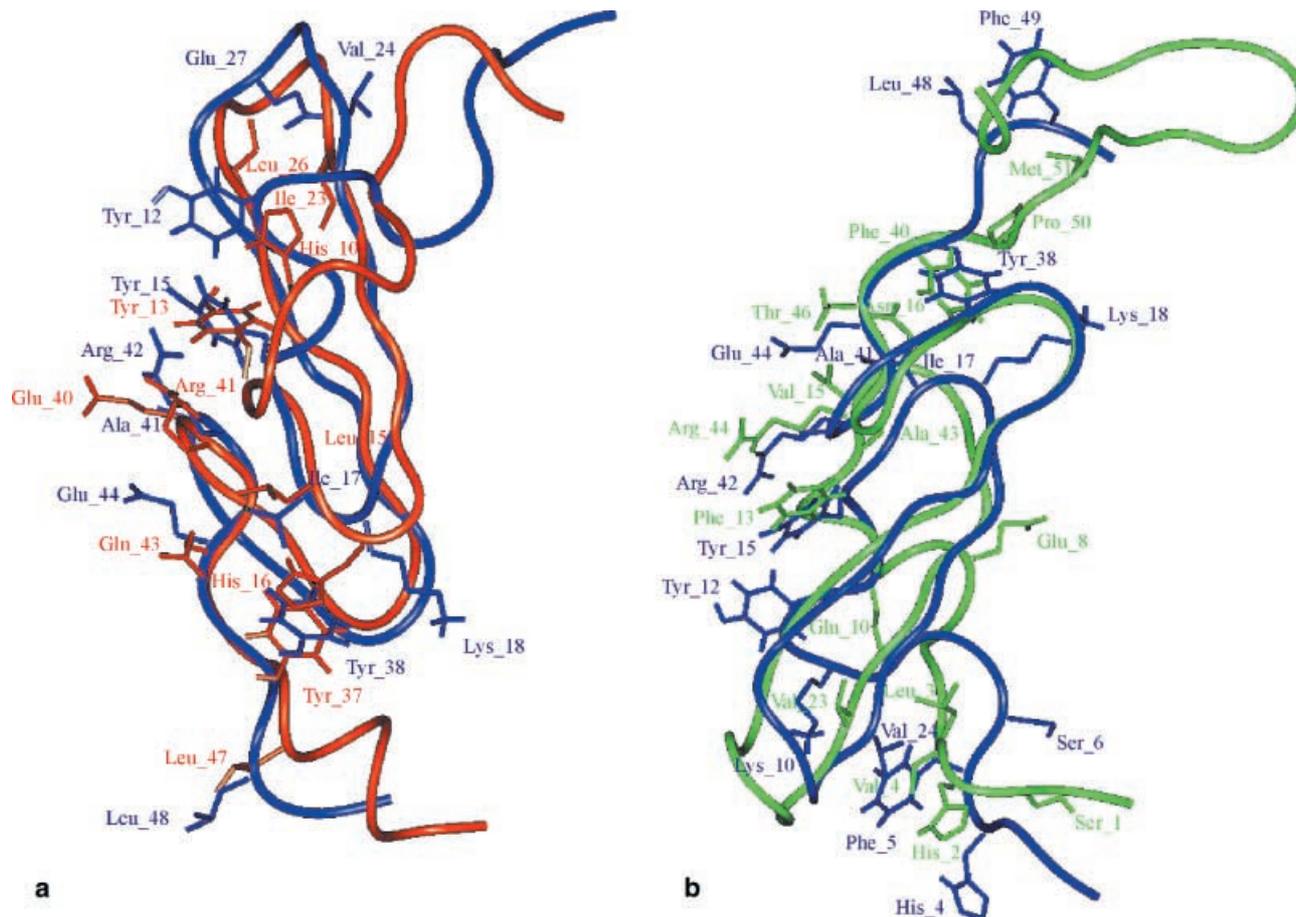


Fig. 3 Backbone traces of hBTC(100) superimposed with hTGF α (100), hEGF(100) or hHRG-1 α (100). The colour code is indicated in Fig. 2. Amino acids that are important for receptor-binding of hTGF α , hEGF and hHRG-1 α are indicated with the

correspondences in the hBTC model (**a**, **b**, and **c** respectively). **d** Residues with different chemical character located in similar positions for both hTGF α and hBTC

with hEGF and 28% with hHRG-1 α . Figure 1 shows that there is only one remarkable change of amino acid type in hBTC with respect to hTGF α in a position considered to be important for the binding of hTGF α with its receptor. This change implies a tyrosine residue (hBTC_Y12) instead of a histidine residue (hTGF α _H12) (see Fig. 3a). In hEGF there is also a histidine residue at this position, which is important for the binding of this molecule with its receptor. Therefore, the change of amino acid type in hBTC at this position with respect to hEGF is the same as with respect to hTGF α (the aromatic hBTC_Y12 instead of the basic hEGF_H10) (see Fig. 3b). For hHRG-1 α the residue in this position is also important for binding with its receptors, but the residue character is acidic (hHRG-1 α _E10) (see Fig. 3c).

Additionally, there are two more changes of amino acid type in hBTC with respect to hEGF, in both cases between apolar and acid residues (hBTC_E27 instead of hEGF_L26 and hBTC_A41 instead of hEGF_E40). The model of hBTC also changes the polarity of two residues in the N-terminal tail with respect to hHRG-1 α (hBTC_G3 instead of hHRG-1 α _S1 and hBTC_S6 instead of hHRG-1 α _V4).

It is remarkable that the N-terminal tail of hHRG-1 α has been found to contain major determinants required for its high affinity binding to ErbB-3 and ErbB-4. [24]

Because of the similar conformations found between hTGF α and the model of hBTC, their differences and similarities on the side-chain orientations of the residues located in the surface have been more thoroughly compared. The differences between the modeled structure of hBTC and the template hTGF α could be involved in a likely different biological behavior. Figure 3d shows those side-chains where the chemical character is different, as for example for the basic residue lysine 10 (hBTC_K10) in hBTC, for which the corresponding residue in hTGF α is an acidic aspartate (hTGF α _D10).

Surface accessibility

The percentages of solvent accessible surface per residue are shown in Fig. 4 for the comparison of the initial models of hBTC with hTGF α (Fig. 4a), hHB-EGF (Fig. 4b), hEGF (Fig. 4c) and hHRG-1 α (Fig. 4d) and analogously

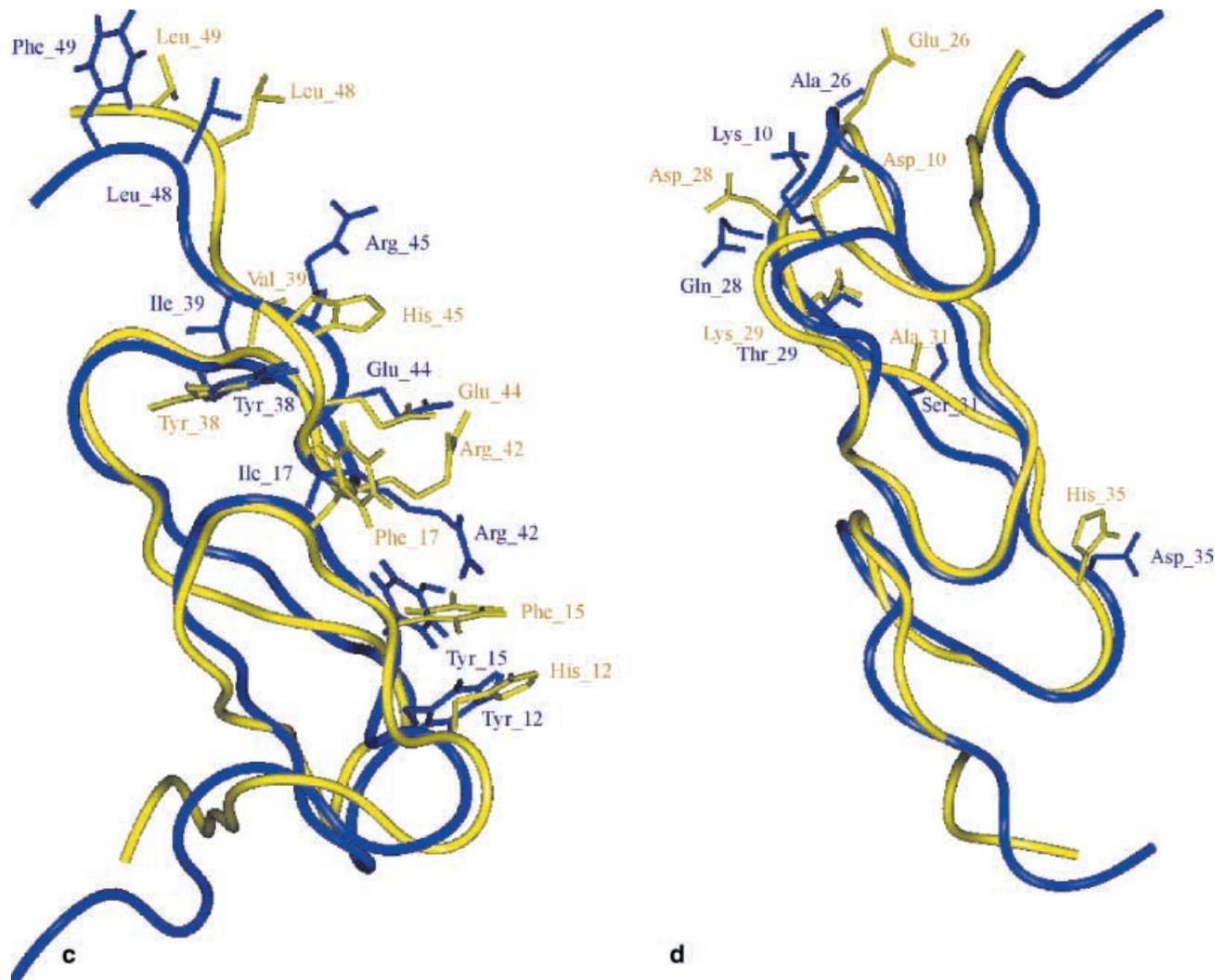


Fig. 3 (continued)

in Fig. 5 for the structures after 100 ps of simulation. A different shape of the curve around residues hTGF α _Q26 and hTGF α _K29 is observed in hBTC. The residue hTGF α _Q26 is more accessible to solvent than its equivalent residue in hBTC (hBTC_A26). This is due to the fact that the side chain of an alanine residue is apolar and shorter than the side-chain of a glutamine residue. This difference disappears after 100 ps of simulation, the two curves becoming similar in this region whilst diverging in a common arginine 22. This is due to an internal H-bond of hTGF α _R22 between the side-chain and its own carbonyl main-chain (see Table 2).

When comparing the middle region between hEGF and hBTC, it is observed that the region with large accessibility is wider in hEGF than in hBTC, and the accessibility is also larger here. The interval of this region is expanded around residues hEGF_E24 and hEGF_Y29. The reason is that there are polar residues at the boundaries of this region in hEGF (hEGF_E24 and hEGF_Y29) whereas for hBTC these residues are apolar (hBTC_V25 and hBTC_P30). Nevertheless, after 100 ps

of simulation the ranging interval becomes similar for both whilst for hBTC(100) glutamine 28 becomes more exposed than aspartate 27 of hEGF(100). It is noticeable that the accessibility of hEGF_D17 (with no equivalent C α position in hBTC) is similar to that of hBTC_R20. Also, after 100 ps of simulation, the difference becomes smoother, although a 50% accessibility of hBTC_R20 still remains. This couple of charged residues, although not aligned, could be equivalent for both proteins in terms of surface. The region mentioned above belongs to a range of three-residue insertions in hHRG-1 α with respect to hBTC. It is remarkable that the hBTC solvent accessibility is shifted two positions downstream with respect to hHRG-1 α . This way, the equivalent position for hHRG-1 α _K24 would be hBTC_E27 in the surface of hBTC, for hHRG-1 α _D25 it would be hBTC_Q28, and for hHRG-1 α _P29 it would be hBTC_T29. A similar result is found after 100 ps of simulation (see Fig. 5a–d), although differences have again become smoother.

In general, for most simulations the differences found on the initial structure of hBTC compared with each template, become smoother after 100 ps of simulation, it being difficult to differentiate between them. However, when analyzing the curve of the surface accessibility of

Table 2 Hydrogen bonds. Observed hydrogen bonds in hBTC(100) and in templates hTGF α (100), hEGF(100), hHB-EGF(100) and hHRG-1 α (100). They are classified according to the origin of the atoms involved in the hydrogen bond, belonging to the main-chain

(top rows) or to the side-chain (bottom rows). Structurally equivalent hydrogen bonds in different structures are all in the same row in this table. Residues implicated in binding are in bold (see also Table 1)

hBTC		hTGF- α		hHB-EGF		hEGF		hHRG-1 α	
-	-	-	-	-	-	-	-	-	-
-	-	6 Asn H	23 Phe O	-	-	-	-	-	-
-	-	-	-	-	-	-	-	5 Lys H	22 Met O
-	-	-	-	-	-	-	-	6 Cys H	20 Cys O
12 Tyr H	9 Pro O	-	-	12 Tyr H	10 Arg O	10 His H	7 Pro O	-	-
-	-	-	-	-	-	13 Tyr H	10 His O	-	-
13 Lys H	10 Lys O	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	14 Cys H	11 Lys O
17 Ile H	42 Arg O	17 Phe H	42 Arg O	-	-	15 Leu H	41 Arg O	15 Val H	44 Arg O
-	-	-	-	-	-	-	-	16 Asn H	45 Cys O
-	-	-	-	19 Gly H	33 Ile O	-	-	-	-
20 Arg H	33 Val O	20 Thr H	33 Val O	-	-	18 Gly H	16 His O	-	-
22 Arg H	31 Ser O	22 Arg H	31 Ala O	-	-	19 Val H	32 Asn O	19 Glu H	35 Lys O
-	-	-	-	22 Lys H	33 Ile N	21 Met H	30 Ala O	21 Phe H	33 Leu O
-	-	-	-	-	-	-	-	-	-
23 Phe H	7 Arg O	23 Phe H	7 Asp O	-	-	-	-	22 Met H	5 Lys N
-	-	24 Leu H	29 Lys O	-	-	23 Ile H	28 Lys O	23 Val H	31 Arg O
-	-	25 Val H	4 His O	-	-	-	-	-	-
27 Glu H	24 Val O	-	-	27 Leu H	24 Val O	-	-	-	-
-	-	-	-	-	-	27 Asp H	23 Ile O	-	-
-	-	29 Lys H	24 Leu O	-	-	28 Lys H	23 Ile O	31 Arg H	23 Val O
-	-	29 Lys H	27 Glu O	-	-	-	-	-	-
31 Ser H	22 Arg O	-	-	-	-	30 Ala H	21 Met O	33 Leu H	21 Phe O
33 Val H	20 Arg O	33 Val H	20 Thr O	33 Ile H	20 Glu O	32 Asn H	19 Val O	35 Lys H	19 Glu O
35 Asp H	18 Lys O	35 His H	18 His O	35 His H	18 His O	-	-	-	-
-	-	-	-	-	-	34 Val H	37 Tyr O	-	-
38 Tyr H	35 Asp O	38 Tyr H	35 His O	-	-	37 Tyr H	34 Val O	40 Phe H	37 Gln O
38 Tyr H	36 Glu O	-	-	-	-	-	-	-	-
39 Ile H	45 Arg O	39 Val H	45 His O	-	-	38 Ile H	44 Tyr O	41 Thr H	47 Glu O
-	-	-	-	39 His H	44 His O	-	-	-	-
-	-	-	-	40 Gly H	42 Arg O	-	-	-	-
43 Cys H	40 Gly O	-	-	43 Cys H	15 Phe O	42 Cys H	39 Gly O	-	-
-	-	44 Glu H	39 Val O	-	-	-	-	-	-
45 Arg H	39 Ile O	-	-	-	-	44 Tyr H	38 Ile O	-	-
47 Asp H	37 Gly O	47 Asp H	37 Gly O	47 Ser H	37 Gly O	46 Asp H	36 Gly O	49 Val H	39 Gly O
-	-	-	-	-	-	50 Trp H	46 Asp O	-	-
-	-	-	-	-	-	51 Glu H	47 Leu O	-	-
-	-	-	-	-	-	52 Leu H	51 Glu N	-	-
-	-	-	-	-	-	2 Ser H	5 Glu OE ₁	-	-
-	-	-	-	8 Cys H	12 Tyr OH	-	-	-	-
11 Gln H	11 Gln OE ₁	10 Asp H	10 Asp OD ₁	-	-	-	-	-	-
-	-	-	-	14 Asp H	14 Asp OD ₁	-	-	-	-
-	-	-	-	-	-	13 Tyr HH	29 Tyr O	12 Thr HG ₁	10 Glu O
-	-	-	-	-	-	-	-	13 Phe H	12 Thr OG ₁
18 Lys HZ3	36 Glu O	-	-	-	-	16 His HD₁	42 Cys O	16 Asn HD₂₂	45 Cys O
-	-	-	-	-	-	-	-	-	-
-	-	22 Arg HH ₁₂	22 Arg O	-	-	-	-	-	-
-	-	-	-	23 Tyr H	31 Ser OG	-	-	-	-
-	-	28 Asp H	28 Asp OD ₁	-	-	-	-	42 Gly H	41 Thr OG ₁
42 Arg HH12	15 Tyr O	-	-	-	-	-	-	-	-
-	-	-	-	42 Arg HH ₁₁	14 Asp O	41 Arg HE	12 Gly O	-	-
-	-	-	-	-	-	41 Arg HH₁₂	12 Gly O	-	-
-	-	48 Leu H	47 Asp OD ₁	44 His HD ₁	42 Arg O	-	-	-	-
49 Phe H	47 Asp OD ₁	-	-	-	-	-	-	-	-
-	-	-	-	-	-	49 Trp H	46 Asp OD ₁	-	-
-	-	-	-	-	-	-	-	54 Gln HE ₂₁	55 Asn O
-	-	-	-	-	-	-	-	58 Lys H	54 Gln OE ₁
-	-	-	-	-	-	2 Ser HG	5 Glu OE ₁	-	-
-	-	-	-	-	-	4 Ser HG	24 Glu OE ₁	-	-
-	-	-	-	28 Arg HE	23 Tyr OH	-	-	-	-
-	-	29 Lys HZ ₂	27 Glu OE ₂	-	-	-	-	-	-
-	-	42 Arg HE	44 Glu OE₁	-	-	-	-	-	-
-	-	-	-	-	-	-	-	46 Thr HG ₁	47 Glu OE ₁
-	-	-	-	-	-	-	-	46 Thr HG ₁	47 Glu OE ₂
50 Tyr HH	36 Glu OE ₁	-	-	-	-	45 Arg HH ₁₂	37 Tyr OH	-	-

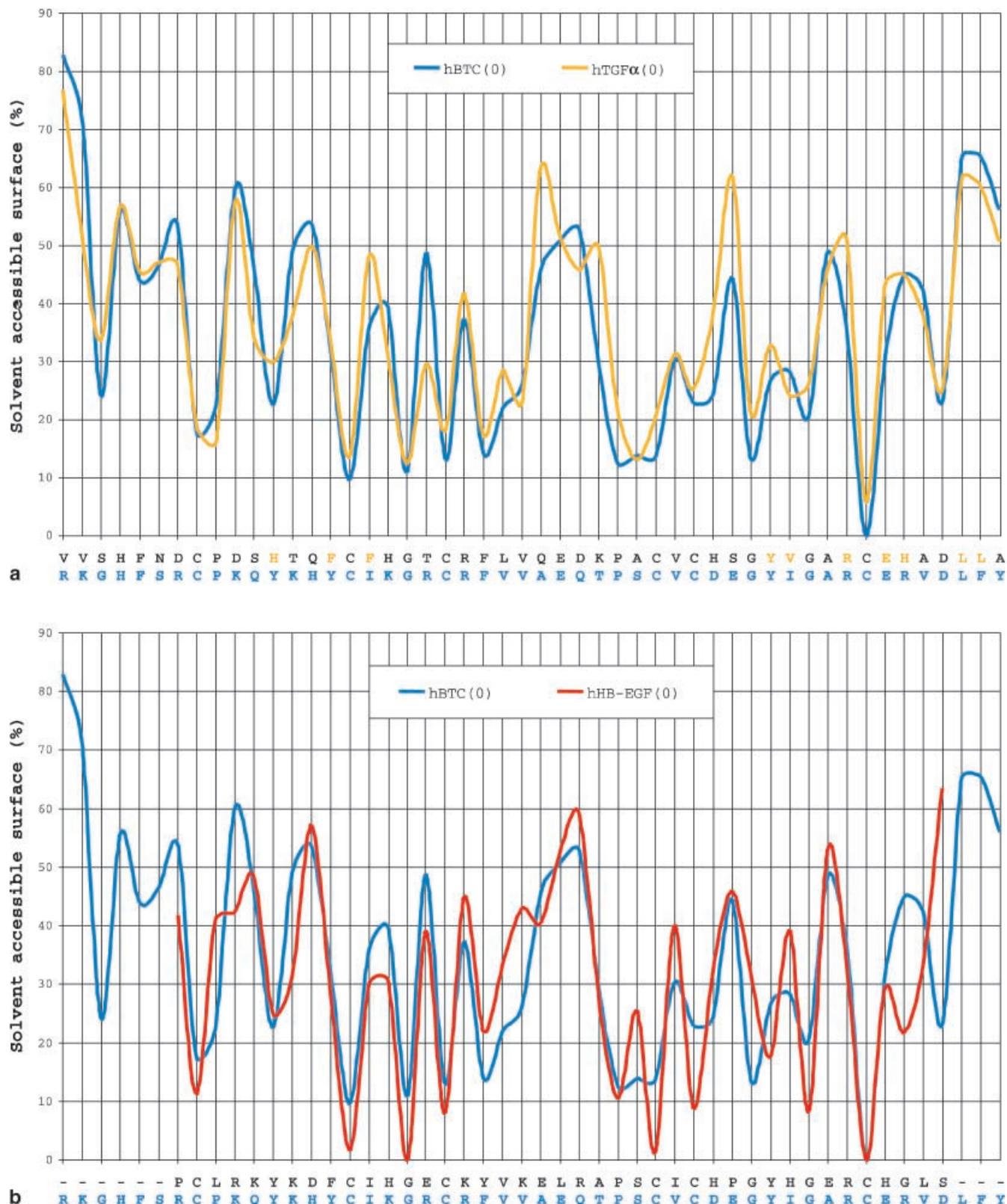


Fig. 4 Percentage of solvent accessible surface per residue for the initial models of hBTC compared with the initial structures of hTGF α (a), hHB-EGF (b), hEGF (c) and hHRG-1 α (d). On the x-axis is shown the alignment of hBTC with respect to the template sequence (hTGF α , hEGF, hHB-EGF and hHRG-1 α). Template residues involved in receptor binding (Table 1) are coloured

as in Fig. 2. For hHRG-1 α different colours are used according to different binding receptors: *green* is used for those residues affecting the hHRG-1 β binding to ErbB-3 and ErbB-4, *pink* for those mainly affecting the hHRG-1 β binding to ErbB-3 and *yellow* for those mainly affecting the hHRG-1 β binding to ErbB-4

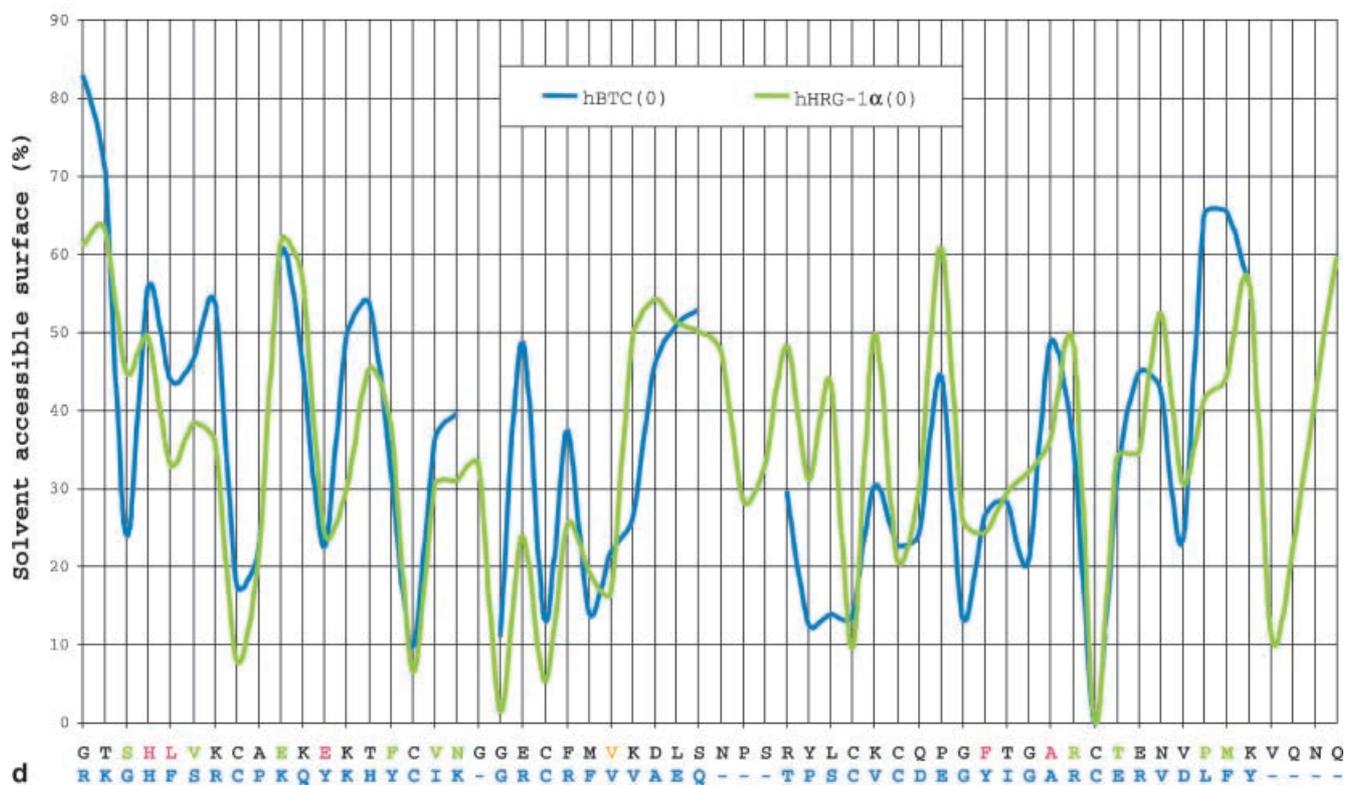
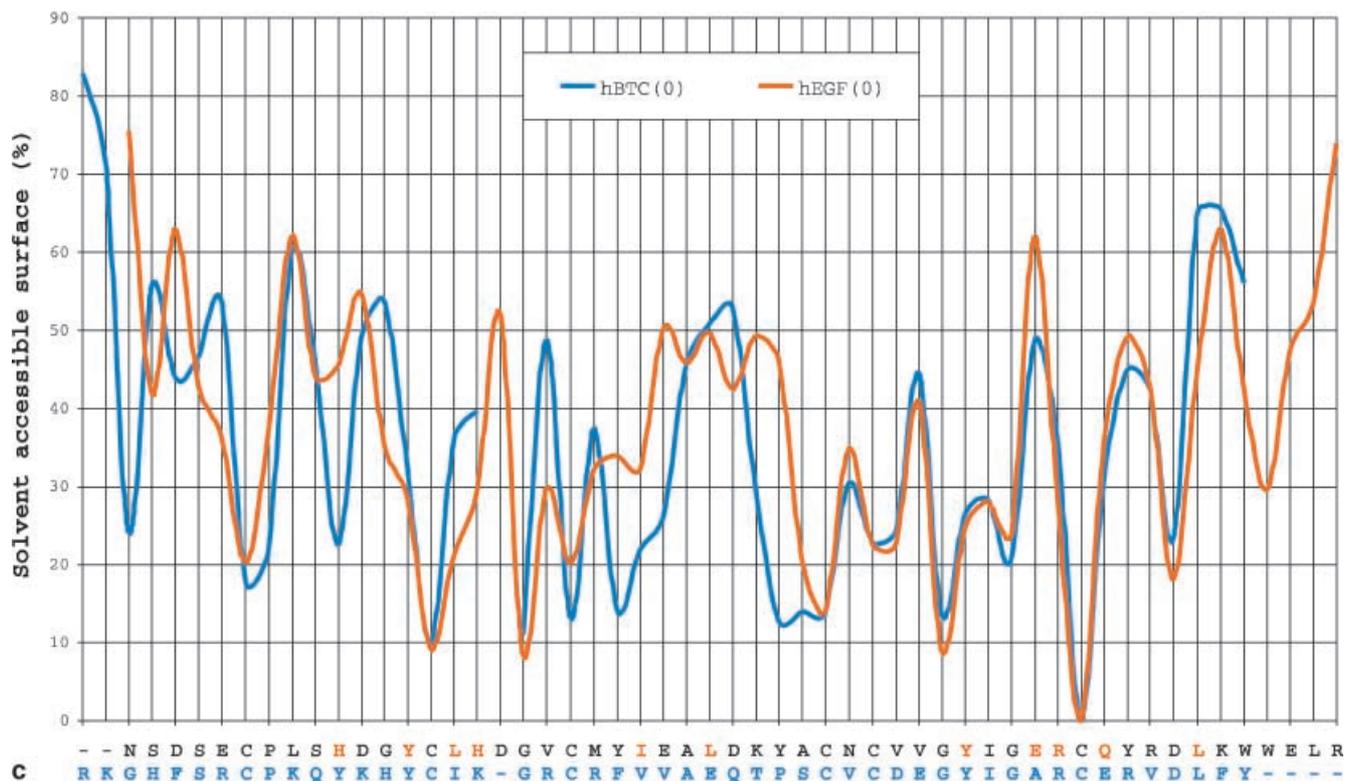


Fig. 4 (continued)

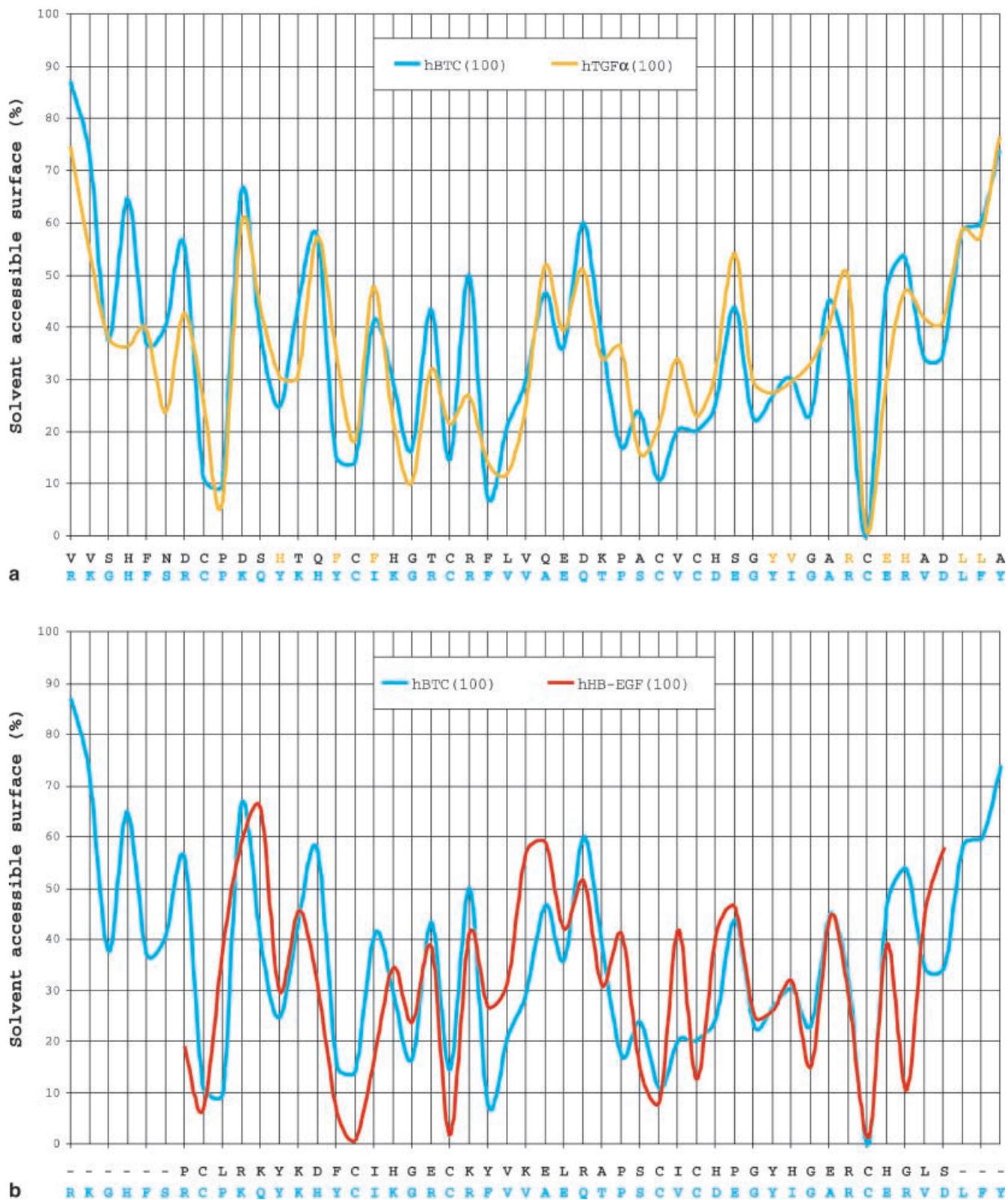


Fig. 5 Percentage of solvent accessible surface per residue for the model hBTC(100) compared with those of hTGF α (a), hHB-EGF (b), hEGF (c) and hHRG-1 α (d) at the last step of the simulation (100 ps). See legend to Fig. 4 for codes

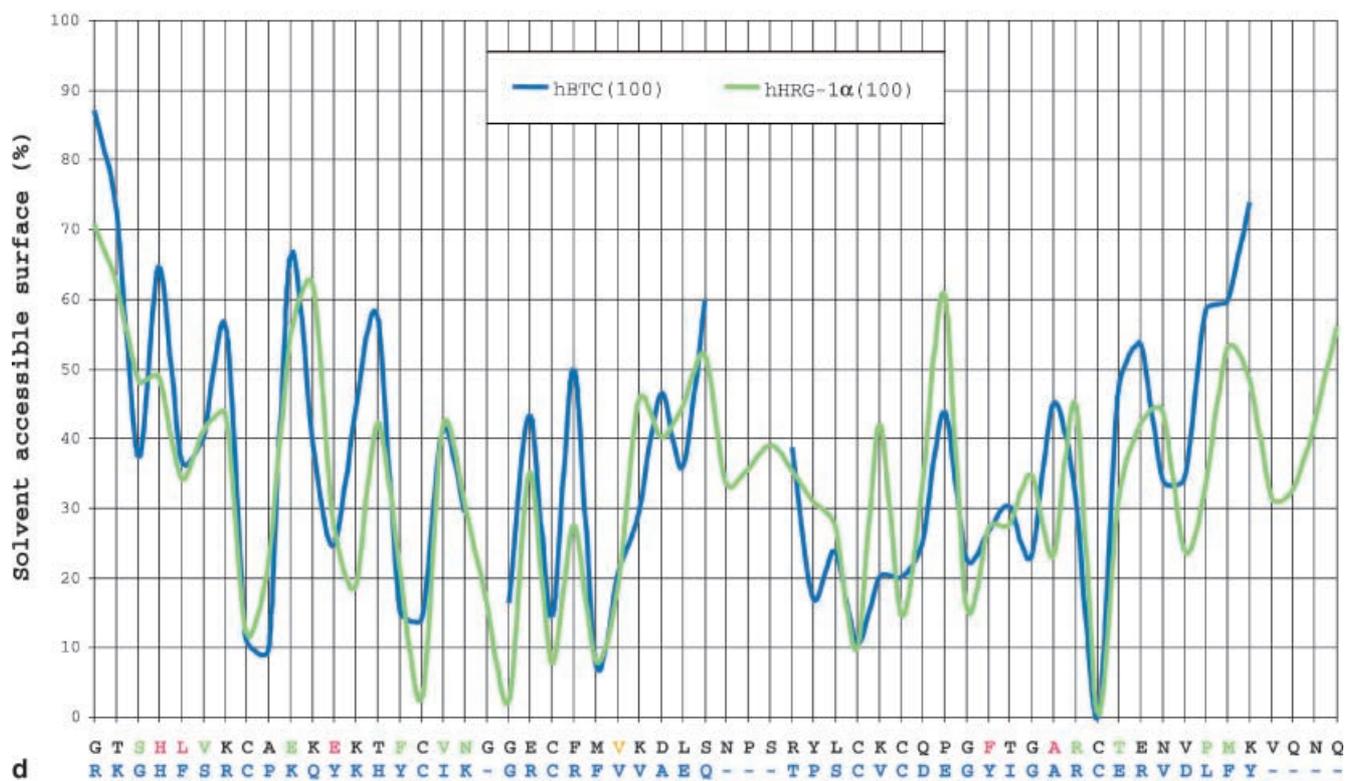
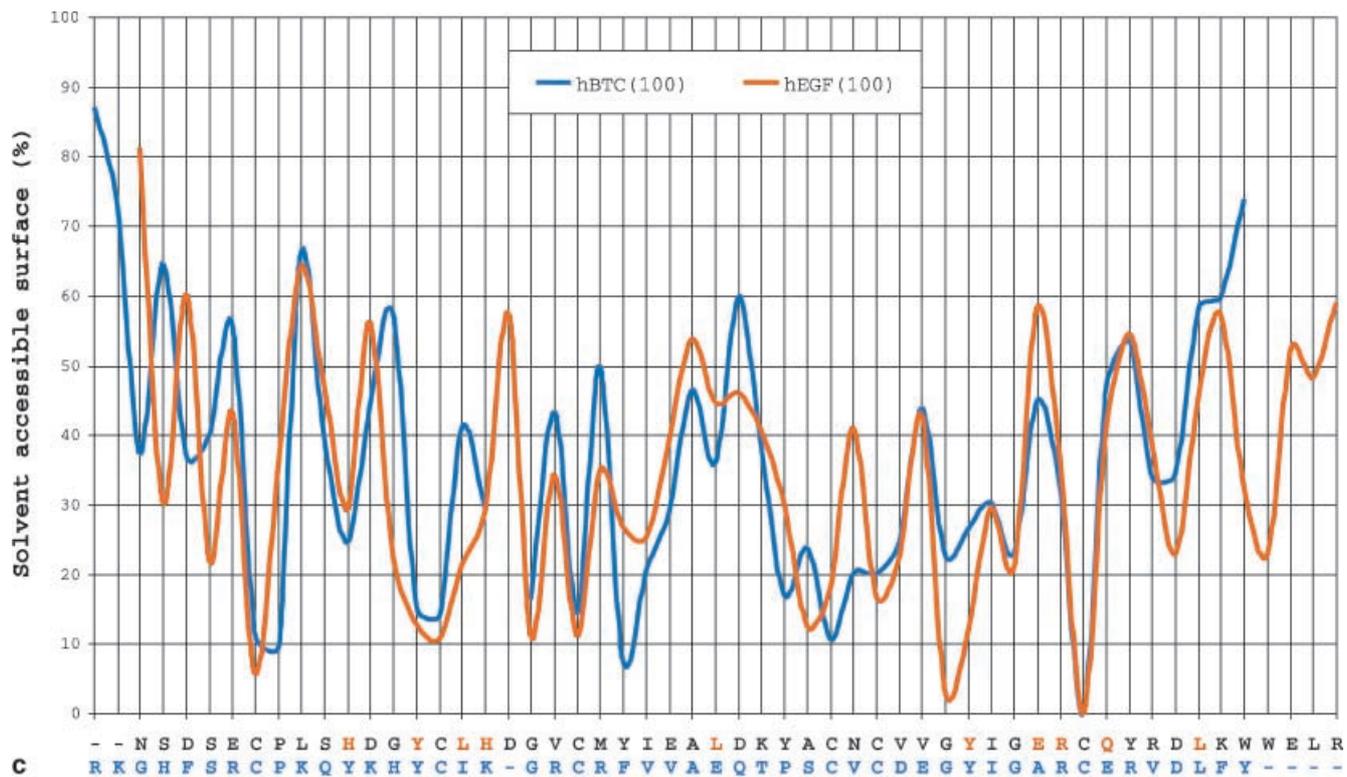


Fig. 5 (continued)

the region ranging between hBTC_V24 and hBTC_S31 compared with the same region in hHB-EGF the similar surfaces found between both at the initial step are conserved at the end of the simulation. This indicates that the surface is similar and besides it behaves similarly for both. It is remarkable that these two proteins can bind both EGFR and the HRG receptors (ErbB-3 and ErbB-4); therefore the common surface could be involved in the binding specificity and their interaction with each corresponding receptor.

Conclusions

The initial model of betacellulin (hBTC) has several structural features similar to hTGF α and hHB-EGF. Nevertheless, after 100 ps of simulation the two structures become very different, although the surface still keeps some of its original shape in the region between residues 24 and 31 of hBTC. In general, the scaffold and the hydrogen bonding net are best maintained between hBTC and hTGF α . The surface of hBTC is equivalent to the surface of both molecules, although the chemical character of its residues is more conserved if compared to that of hTGF α . However, hTGF α binds only EGFR, whilst hHB-EGF (like hBTC) can bind both EGFR and HRG receptors. This could also be explained by considering the different features found in the residues shown in Fig. 3d if more experimental data were available. Additional information could be extracted from binding experiments that should be performed with these bifunctional growth factors to determine the distinctive interaction of these molecules with EGFR and with ErbB-3 or ErbB-4.

There are few structural similarities between hBTC and hHRG-1 α , the only known EGF-like domain that can bind to ErbB-3 and ErbB-4 but not to EGFR. The N-terminal tail of HRGs has been found to be essential to bind ErbB-3 or ErbB-4. [24] Barbacci et al. [24] constructed an EGF/HRG chimera in which the N-terminal tail of EGF was replaced by the N-terminal tail of HRG, and they found that this chimera (called "biregulin") was able to bind both ErbB-3 and ErbB-4 without avoiding binding to EGFR. Looking at the structure of our model of hBTC, its N-terminal tail is structurally more similar to hTGF α 's than to hHRG-1 α 's. Perhaps a new chimera of EGF with the N-terminal tail of hBTC (or hTGF- α) would also be able to bind heregulin receptors. The N-terminal tail of EGF is probably designed in such a way that avoids EGF binding to another ErbB receptor but EGFR.

The surface differences found between the template structures and the hBTC model have evolved after 100 ps of simulation until becoming more similar, whilst the original similarities found between hBTC and HB-EGF surfaces have remained. A thorough analysis of the exposed residues indicates differences that could be involved in the biological role of each structure, characterizing the likely interaction with different receptors.

The presence of the positive residue hBTC_R7, being lysine in hHRG-1 α (hHRG-1 α _K5), and a negative charge for hEGF and hTGF α (hEGF_E5 and hTGF α _D7 respectively) may help to avoid the interaction with any ErbB receptor other than EGFR, both positive and negative residues being similarly exposed. Also the location of hBTC_R42 may be important in the interaction with the receptor, because the guanidine group seems to make a direct chemical bond with ErbB receptors. [22] The spatial orientation of this residue is probably determined by the different H-bonds present at this position, as shown in Table 2.

Finally, the simulation of the whole set of structures has shown that the hBTC modeled structure is similar to that of hTGF α , and also the final structures have comparable changes after 100 ps of simulation. The hydrogen bonding net has been reduced after 100 ps of simulation (results not shown), but the main features involved on residues that are functionally important on the role of each structure have remained.

Supplementary material

Two coordinates of the hBTC model are given in PDB format. These coordinates are taken from the seeding step of the MD simulation, named as "hBTC(0).pdb", and at the final step after 100 ps of simulation, named as "hBTC(100).pdb". The coordinates in "hBTC(0).pdb" were obtained from comparative modeling and further energy optimization, as indicated in the text.

Acknowledgments This research has been supported by grants BIO98-0362 and 2FD97-0458 from the CICYT (Ministerio de Educación y Ciencia, Spain) and by the Centre de Referència en Biotecnologia (CERBA) of the Generalitat de Catalunya. I. L.-T. is a predoctoral fellowship recipient of the CIRIT-CERBA (Generalitat de Catalunya).

References

1. Alroy I, Yarden Y (1997) FEBS Lett 410:83
2. Riese DJ, Stern D (1998) Bioessays 20:41
3. Tzahar E, Yarden Y (1998) Biochim Biophys Acta 1377:M25
4. Hackel PO, Zwick E, Prenzel N, Ullrich A (1999) Curr Opin Cell Biol 11:184
5. Olayioye M, Neve RM, Lane HA, Hynes NE (2000) EMBO J 19:3159
6. Bogdan S, Klämbt C (2001) Curr Biol 10:R292
7. Lemmon MA, Bu Z, Ladbury JE, Zhou M, Pinchasi D, Lax I, Engelman DM, Schlessinger J (1997) EMBO J 16:281
8. Ullrich A, Schlessinger J (1990) Cell 61:203
9. McDonald NQ, Hendrickson WA (1993) Cell 73:421
10. Sun PD (1995) Annu Rev Biophys Biomol Struct 24:269
11. Mas JM, Aloy P, Marti-Renom MA, Oliva B, Blanco-Aparicio C, Molina MA, De Llorens R, Querol E, Avilés FX (1998) J Mol Biol 284:541
12. Shing Y, Christofori G, Ono Y, Sasada R, Igarashi K, Folkman J (1993) Science 259:1604
13. Sasada R, Ono Y, Taniyama Y, Shing Y, Folkman J, Igarashi I (1993) Biochem Biophys Res Commun 190:1173
14. Dunbar AJ, Goddard C (2000) Int J Biochem Cell Biol 32:805
15. Seno M, Tada H, Kosaka M, Sasada R, Igarashi K, Shing Y, Folkman J, Ueda M, Yamada H (1996) Growth Factors 13:181

16. Miyagawa JI, Hanafusa T, Sasada R, Yamamoto K, Igarashi K, Yamamori K, Seno, M, Tada H, Nammo T, Li M, Yamagata K, Nakajima H, Namba M, Kuwajima M, Matsuzawa Y (1999) *Endocrine J* 46:755
17. Watanabe T, Shintani A, Nakata M, Shing Y, Folkman J, Igarashi K, Sasada R (1994) *J Biol Chem* 269:9966
18. Hommel U, Harvey TS, Driscoll PC, Campbell ID (1992) *J Mol Biol* 227:271
19. Harvey TS, Wilkinson AJ, Tappin MJ, Cooke RM, Campbell ID (1991) *Eur J Biochem* 198:555
20. Nagata K, Kohda D, Hatanaka H, Ichikawa S, Matsuda S, Yamamoto T, Suzuki A, Inagaki F (1994) *EMBO J* 13:3517
21. Louie GV, Yang W, Bowman ME, Choe S (1997) *Mol Cell* 1:67
22. Groenen LC, Nice EC, Burgess AW (1994) *Growth Factors* 11:235
23. Katsuura M, Tanaka S (1989) *J Biochem* 106:87
24. Barbacci EG, Guarino BC, Stroh JG, Singleton DH, Rosnack KJ, Moyer JD, Andrews GC (1995) *J Biol Chem* 270:9585
25. Tzahar E, Pinkas-Kramarski R, Moyer JD, Klapper LN, Alroy I, Levkowitz G, Shelly M, Henis S, Eisentein M, Ratzkin BJ, Sela M, Andrews GC, Yarden Y (1997) *EMBO J* 16:4938
26. Jones JT, Akita RW, Sliwkowski MX (1999) *FEBS Lett* 447:227
27. Dunbar AJ, Goddard C (2000) *Int J Biochem Cell Biol* 32:805
28. Sali A, Blundell TL (1993) *J Mol Biol* 234:779
29. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) *J Mol Biol* 112:535
30. van Gunsteren WF, Berendsen HJC (1987) Groningen molecular simulation (GROMOS) library manual. BIOMOS, Groningen
31. Ryckaert JP, Ciccotti G Berendsen HJC (1977) *J Comput Phys* 23:327
32. Xia TH (1992) Software for determination and visual display of NMR structures of proteins: the distance geometry program DGLAY and the computer graphics programs CONFOR and XAM. Dissertation, ETH Zurich, Nr 9831
33. Jacobsen NE, Abadi N, Sliwkowski MX, Reilly D, Skelton NJ, Fairbrother WJ (1996) *Biochemistry* 35:3413
34. Moy FJ, Li YC, Rauenbuehler P, Winkler ME, Scheraga HA, Montelione GT (1993) *Biochemistry* 32:7334
35. Roussel A, Inisan AG, Knoop-Mouthy E (1991) Turbo-Frodo Manual v 5.0. Biographics, Technopole de Chateaux-Gombert, Marseille, France
36. Richmond TJ (1984) *J Mol Biol* 178:63
37. Campion SR, Niyogi SK (1994) *Prog Nucleic Acids Res* 49:353
38. McInnes, C Sykes BD (1997) *Biopolymers* 43:339
39. McInnes C, Grothe S, O'Connor-McCourt M, Sykes BD (2000) *Prot Eng* 13:143
40. Jones JT, Ballinger MD, Pisacane PI, Lofgren JA, Fitzpatrick VD, Fairbrother WJ, Wells JA, Sliwkowski MX (1998) *J Biol Chem* 273:11667
41. Ballinger MD, Jones JT, Lofgren JA, Fairbrother WJ, Akita RW, Sliwkowski MX, Wells JA (1998) *J Biol Chem* 273:11675