Article No. mb982194 *J. Mol. Biol.* (1998) **284**, 541–548

JMB

COMMUNICATION

Protein Similarities Beyond Disulphide Bridge Topology

J. M. Mas[1], P. Aloy[1], M. A. Martí-Renom[1], B. Oliva[1], C. Blanco-Aparicio[2] M. A. Molina[2], R. de Llorens[2], E. Querol[1] and F. X. Avilés[1]*

[1]*Institut de Biologia Fonamental i Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193, Bellaterra, Barcelona Spain*

[2]*Unitat de Bioquímica, Departament de Biologia Facultat de Ciències Universitat de Girona 17071 Girona, Spain*

**Corresponding author*

Structural superimposition is an important procedure to analyse the relationships between proteins. A new approach and program, KNOT-MATCH, has been developed for automated structural superimposition of proteins by means of their disulphide bridge topology. As a result of the superimposition, regular secondary structures, loops and clusters of residues become correctly aligned. This fact allows us to find out important structural overlaps of residues, sometimes with functional significance, not only among proteins belonging to the same family but also between apparently non-related proteins. Different disulphide-rich protein families, such as EGF-like, defensin-like and plant protease inhibitors, have been self or cross analysed with this approach. Some amino acids that have been experimentally determined to be structural and/or functional key residues for these proteins are conserved in the three-dimensional space after superimposition by KNOT-MATCH. The program can be very useful for finding relationships among proteins that would be hidden to the current alignment methods based on sequence and on main-chain topology.

Keywords: disulphide rich proteins; disulphide bonds; protein structure; protein topology; protein superimposition

Algorithms and programs for sequence alignment of proteins are the current tools used in molecular biology and biotechnology (for a review, see Andrade & Sander, 1997). Similar developments should be addressed to the three-dimensional (3D) alignment and analysis of proteins. The actual rate at which new structures are being obtained exceeds one per day as a result of recent advances in X-ray crystallography and NMR spectroscopy. These advances demand new and faster algorithms for 3D structure comparisons. Given that the 3D structure is highly conserved in protein evolution (Rost, 1997), the comparison of 3D structures permits us to establish relationships between proteins that appear non-existent or distinct on the basis of sequence alignment. Early computer methods for such a purpose required am initial manual alignment, and were very slow or limited to close homologue proteins (Rossmann & Argos, 1976; Matthews & Rossmann, 1985). A new generation of search algorithms based upon simplified approaches has been developed in recent years, the most efficient of which allows for fully automated and rapid similarity searches through an entire database (Brenner, 1995). Structural superimposition of proteins requires the selection of certain atoms or parts of the proteins that have to be matched. There is a large number of ways in which one could do this and, frequently, such a task is based in the fitting of regular secondary structures. However, the comparison of proteins lacking regular secondary structures or being non-homologous is still very difficult, and the development of new efficient tools for the structural alignment of these proteins is necessary (Johnson *et al*., 1994; Srinivasan *et al*., 1996).

A large number of proteins with a biological relevance (growth factors, hormones, toxins, etc.) contain disulphide bridges that can determine their fold and topological tendencies (Creighton, 1992; Chang *et al*., 1994, 1995; Wu *et al*., 1998). The correct disulphide bridge topology is indispensable

Abbreviations used: 3D, three-dimensional; PCI, potato carboxypeptidase inhibitor; EGF, epidermal growth factor; PDGF, platelet derived growth factor.

E-mail address of the corresponding author: fx.aviles@blues.uab.es

0022–2836/98/480541–08 $30.00/0

for their final structure and function (Betz, 1993). In this context, Thornton (1980) and Richardson (1981) used the sequence linkage pattern between cysteine residues to establish preliminary relationships between cystine-rich proteins. More recently, structural characteristics based in the so-called β-cross (two β strands maintained by two crossed disulphide bridges) were used to make a classification of small cystine-rich proteins (Harrison & Sternberg, 1996). The development of approaches for the 3D structural alignment and topological analysis of cystine-rich small proteins, which are frequently poor in regular secondary structures, may help to assess the structural similarities among such proteins.

A simple but powerful computer-based program to study disulphide-bridge structure in proteins, called KNOT-MATCH, is presented here. The program uses exclusively the disulphide-bridge topology to superimpose cystine-rich proteins and extract potential structure/function relationships beyond the disulphide bridges. The topological analysis and the subsequent comparison of the 3D structures have been mainly performed using a non-homologous PDB database of proteins containing more than two disulphide bridges (Hobhom & Sander, 1994). The use of intramolecular geometrical relationships to describe protein structures has the advantage of being independent of the co-ordinate frame (Holm & Sander, 1993). The novel program has been designed to obtain the best superimposition of disulphide bridges between two proteins independently of their order in sequence and of any other structural characteristic than the disulphide-bridge topology. Following this procedures we made a general comparison among disulphide-rich small proteins. Similarities in the scaffold after protein superimposition have been detected in several cases and many similar chemical groups overlapping in 3D space have been identified, establishing relationships previously hidden among remotely homologous or sequentially non-homologous proteins.

Structural/functional relevance beyond disulphide-bridge topology superimposition

In order to evaluate the structural/functional relevance of disulphide-bridge topology, a few examples of proteins superimposed by such a topology have been studied in detail (Figure 1). In some cases, the overlap of their side-chains and backbones has also been compared to obtain possible functional relationships. The matches obtained in 3D space have been contrasted with the experimental results reported in the literature. The fitting made by KNOT-MATCH shows a preferential occurrence of certain disulphide-bridge topologies. This fact could be used to group proteins and to further study the relationships between them. It is noteworthy that in some cases, a correct alignment of certain secondary structures has also been found after the superimposition of the disulphide bridges between apparently unrelated non-homologous proteins.

The EGF-like family, defined by SCOP (Murzin *et al.*, 1995), presents a cystine-knot as an important structural and folding feature (Wu *et al.*, 1998), and abundant information has been collected about the structure/function relationships of the members. This fact and the large structural variability among the members prompted us to select a model for an in-depth analysis between proteins of the same family. After disulphide bridge superimposition, the members of the family have been compared in order to extract a consensus of chemical groups of side-chains matching in the 3D space: 15 amino acid locations have been found as a consensus with a percentage of appearance greater than 55% (Figure 2(a)). As expected in the superimposition within a family, the consensus is well distributed over the entire scaffold, and the substitutions are sequentially and spatially conservative. Many of these conservative residues have been previously detected as functionally important for the EGF-like family by site-directed mutagenesis and other approaches (Groenen *et al.*, 1994; Barbacci *et al.*, 1995). However, other positions in the consensus (i.e. Pro7 from EGF protein) have not been previously described as functionally or structurally important. The consensus of 3D chemical groups obtained for this family could unveil new relevant residues in its members. Similar results were obtained for the defensin family that is formed by antimicrobial peptides, with three intramolecular disulphide bonds and a clearly defined β-sheet structure. Three categories of defensins have been found on the basis of their structures, the invertebrate category (mainly from insects) and the α and β-types from vertebrates (White *et al.*, 1995; Ganzt & Lehrer, 1998). The three groups differ in the connectivity of the six conserved cysteine residues but not in the disulphide bridge topology, which is very similar, as has been shown using the KNOT-MATCH program. After the superimposition using six members of the defensin-like family: anthopleurin A (1ahl), sea anemona toxin (1atx), defensin-HNP3 (1dfnA), antihypertensin (2bds), neurotoxin I (2sh1) and β-defensin (1bnb), seven 3D conserved residues were found (data not shown) which further prove that KNOT-MATCH can be used to find structural conserved side-chain residues or regions within a family.

The superimposition by disulphide-bridge topology allows us to find structural similarities not only between clearly related proteins (i.e. belonging to the same family, as shown above), but also between proteins apparently not related in sequence or in function. We have studied in detail two apparently unrelated proteins such as potato carboxypeptidase inhibitor (PCI) and the human epidermal growth factor (EGF), belonging to the EGF-like family. PCI has been previously studied by our group and extensive structural and functional information has been obtained (Oliva *et al.*, 1991a,b, 1995; Avilés *et al.*, 1993; Molina *et al.*, 1994;

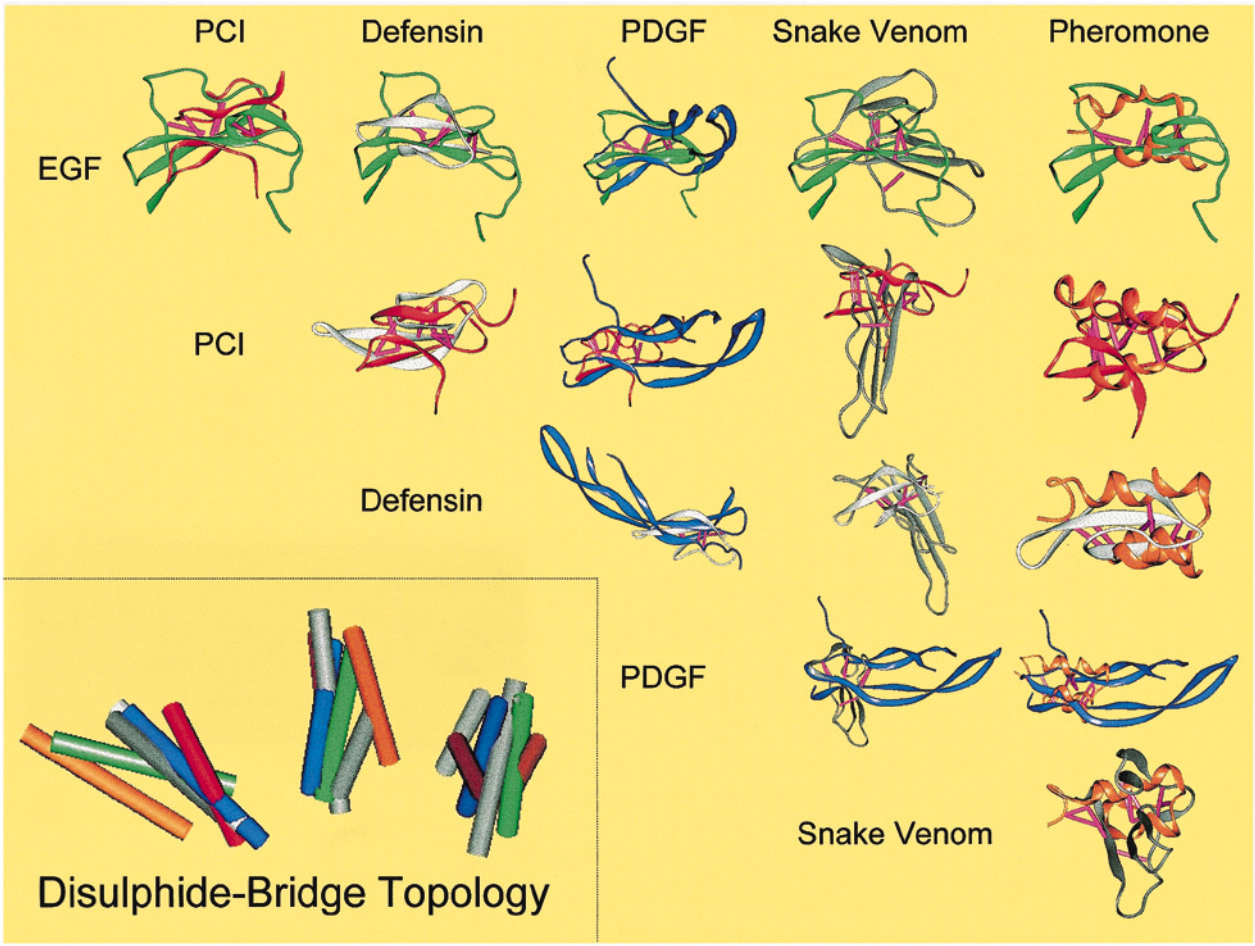

Figure 1. Superimposition of non-homologous proteins by disulphide bridge topology using KNOT-MATCH. Colour and PDB files codes: EGF in green (3egf); PCI in red (4cpaI); defensin in white (1dfnA); platelet derived growth factor (PDGF) in blue (1pdgB); snake venom toxin in grey (1coe); and pheromone in orange (1rpl). In the lower left corner of the Figure the superimposition of the vectors between C^{α} atoms of the cysteine residues bonded in the above proteins is displayed. The average and standard deviation rmsd of 2.37(±0.27) Å indicates that there is a clear match of the compared disulphide bridges. Although most of the proteins chosen for the comparison have different structures and functions, an important group of them presents some structural similarities comprising less than 20 residues. These similarities have been analysed by visual inspection using TURBO FRODO (Roussel *et al.*, 1994). The program KNOT-MATCH looks for the best superimposition of the C^{α} atoms of the cysteine residues that form a disulphide bridge in the compared proteins. Firstly, the program filters all possibilities and chooses a few possible correspondences between residues, even for an asymmetrical number of cystine residues. The program then checks the rmsd and extracts the best superimposition independently of the sequential order of the cysteine residues in both proteins. Special attention has been taken in the filtering process, this being the major problem of the computation. For example, the comparison of a protein with seven disulphide bridges (as Bowman-Birk) with respect to one with three disulphide bridges (as PCI) involves a total number of 13,440 rmsd calculations as the formula indicates:

$$\binom{7}{3} 2^{3!} = 13{,}440$$

This makes it very difficult to tackle the whole set of non-homologous disulphide-rich proteins. The use of the structural dynamic algorithm method (Taylor & Orengo, 1989) allowed a reduction in the number of calculations. The algorithm compares the angles and distances that describe the topological knot formed by disulphide bridges for each pair of proteins, and extracts 30 combinations that may yield the best rmsd. For each combination, the rmsd is calculated and only the best superimposition (with the smallest rmsd) of each pair of proteins is taken to represent its relation. The program KNOT-MATCH runs in Silicon Graphics computers and is available from the World Wide Web by ftp (ftp://luz.uab.es/pub/Knot-Match) and will be available as an on-line server in the near future through the web page http://luz.uab.es.

Blanco-Aparicio *et al.*, 1998; Martí-Renom *et al.*, 1998). PCI, a plant protease inhibitor, constitutes the only member of its protein family in the non-redundant selected PDB. These two proteins constitute a good example of very different functional proteins with a similar disulphide-bridge topology; they share the so-called T-knot structure. However, the two proteins present a different sequential

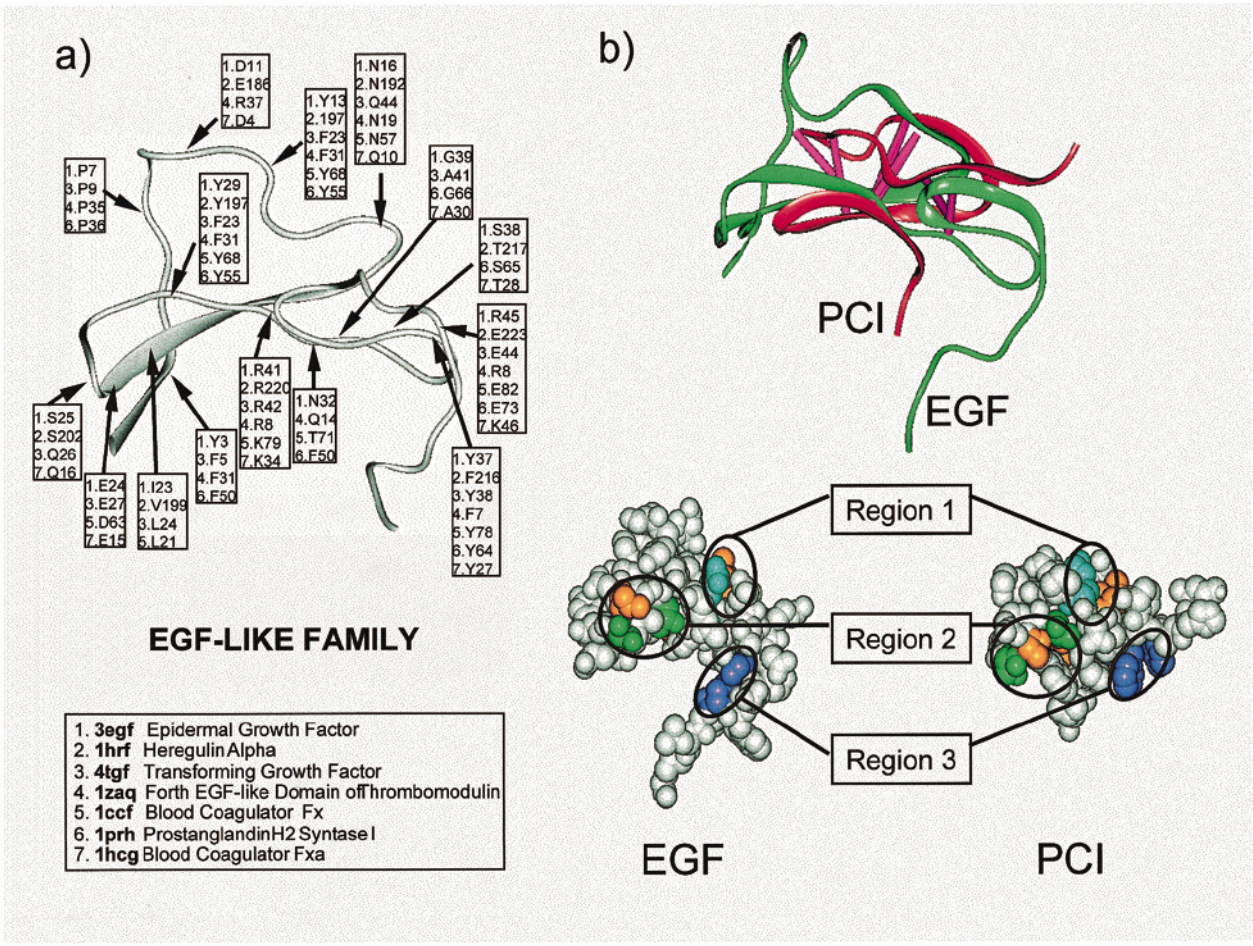

Figure 2. Similarities within family members and between apparently non-related proteins detected by KNOT-MATCH. (a) Imageplate of the EGF (3egf) scaffold. The members of the EGF-like family have been overlapped by their disulphide-bridge topologies. A set of positions in the space with conserved physicochemical properties has been found for several members of the family. These positions are indicated in the Figure with the number of the involved residue in each protein. Some of these residues have been reported as structurally or functionally important for these proteins. (b) Structural comparison between EGF and PCI structures. They are superimposed by their disulphide-bridge topology and displayed using van der Waals surfaces and ribbon-like representations. In the van der Waals surface representation, several residues have been coloured according to their physicochemical properties using the following code: magenta for aromatic, green for polar, blue for basic and orange for non-polar residues. The ribbon-like representations superimpose the α-carbon chains of the PCI (red) and EGF (green). Three regions showing conservative similarities have been encircled: region 1 includes residues L15 and R41 for EGF and I7 and K10 for PCI; region 2 includes residues V19, L26, S25 and T30 for EGF and A26, S30, A31 and T33 for PCI; region 3 includes the aromatic residues Y37 and W50 for EGF and W22 and F23 for PCI. The separation between equivalent residues in both structures is less than 0.7 Å, on average.

pattern of linked cysteine residues (1-3, 2-4, 5-6 for the EGF and 1-4, 2-5, 3-6 for the PCI). The different characteristics of these proteins makes it very difficult to perform a comparison between them with the current methodologies. Here, the use of KNOT-MATCH for superimposing both structures has been very useful in extracting their structural similarities. A large number of matches in 3D space for important functional residues previously identified in the EGF-like family were also found in this superimposition (Figure 2(b)). A total of ten locations conserving the physico-chemical properties of the involved residues were found. These specific locations have been described as structurally or functionally important, either for PCI or for EGF (Rees & Lipscomb, 1982; Brown *et al.*, 1989; Brown & Wüthrich, 1992; Ullner *et al*, 1992; Groenen *et al.*, 1994; Picot *et al.*, 1994; Molina *et al.*, 1994; Barbacci *et al.*, 1995; Jacobsen *et al.*, 1996; Blanco-Aparicio *et al.*, 1998). It is worth mentioning that four of the 15 consensus positions previously found within the EGF family have strong correlation with PCI, increasing the significance of the overlaps. These structural relationships found between PCI and EGF-like proteins could explain why PCI acts as a growth factor antagonist through its binding to the EGF receptor, as recently shown by our group (Blanco-Aparicio *et al.*, 1998).

We have also found other examples where two unrelated proteins show regions that overlap after

the disulphide-bridge topology superimposition, e.g. PCI superimposed with other small disulphide-rich proteins. The region C27-C34 of PCI, fits with the loop G26-H34 of scorpion toxin agitoxin II, and with the loop K27-R34 of charibdotoxin protein with rmsd of around 0.7 Å. Coagulation factor X (PDB code, lccf) has a loop between D63 and T70 that also superimposes with C27-C34 PCI loop, showing an rmsd of 1.8 Å. This PCI loop is also similar to loop H22-Y29 of EGF (rmsd of 0.7 Å). Other loops from PCI have also been studied in detail by such a procedure, the C18-C24 loop being comparable to snake venom toxin (1coe) K47-C54 loop, presenting a rmsd of 0.6 Å. A similar result has been observed between the same PCI loop and the K47-C54 loop from a member of the EGF-like family (the coagulation factor X) with a rmsd of 0.5 Å. It is remarkable that all these loop similarities have been clearly detected only when the superimposition is based in the disulphide-bridge topology.

Figure 3 shows in detail the disulphide bridge superimposition of three non-homologous proteins, PCI (4cpaI), scorpion venom toxin (1agt) and defensin-HNP3. Although the superimposition of these three proteins gives a relatively large rmsd between C^{α} atoms of the disulphide bridges (above 2 Å), visual inspection of the superimposed structures reveals clear similarities in certain regions of the protein scaffolds. Some of the 3D overlapped loops between these three proteins (Figure 3) are defined in the literature as structurally or functionally important. For example, loop Q25-G35 of PCI, which has been proposed to be involved in the binding to EGF receptor (Blanco-Aparicio *et al.*, 1998) overlaps with the Y17-F29 defensin loop. This latter loop has been reported to contain residues which are related to the dimeric assembly of

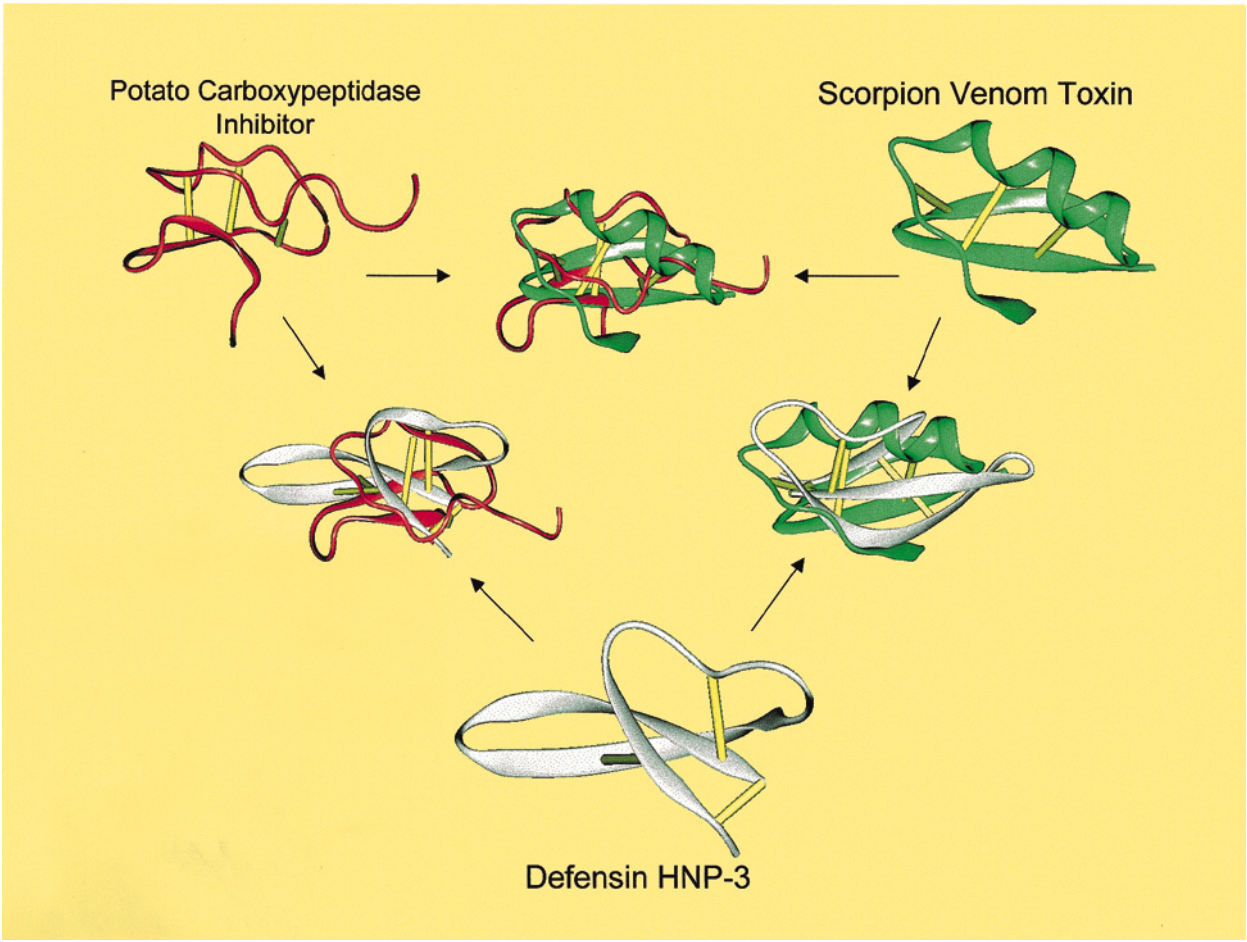

Figure 3. In several cases the superimposition by disulphide-bridge topology allows for good alignment of secondary structures even for proteins of different families. The Figure shows such a superimposition for three apparently non-related proteins (PCI, defensin-HNP3 and scorpion venom toxin). The overlaps of PCI and scorpion venom toxin are found in loops: Q25-G35 of PCI *versus* G26-H34 of scorpion venom toxin and P11-S19 of PCI with S7-I15 of scorpion venom toxin. The superimposition of PCI and defensin-HNP3 also shows two overlapped loops: C8-C18 of PCI with D2-I11 of defensin-HNP3 and Q25-G35 of PCI with Y17-F29 of defensin-HNP3. Finally, two more loops overlap after superimposition of scorpion venom toxin and defensin-HNP3: G1-C10 of scorpion venom toxin *versus* I21-R13 of defensin-HNP3 and D2-P8 of defensin-HNP3 with C28-D20 of scorpion venom toxin. The possible significance of these overlaps is still unknown, although they could be used for establishing further relationships between these proteins or for biotechnological purposes.

the protein that is essential to bind and permeabilize the lipid bilayer (Hill *et al*., 1991; White *et al*., 1995). On the other hand, the D2-P8 defensin loop involves some residues also defined as important in protein dimerization (Hill *et al*., 1991) and overlaps with scorpion toxin D20-C28 loop, which participates in the interaction with K^+ channels (MacKinnon *et al*., 1998). It has been suggested that scorpion toxins share a common folding motif (CSαβ motif), based on a cystine stabilized αβ core, with other small cystine-rich proteins such as insect defensins (Cornet *et al*., 1995). These analogies were not found for mammalian defensins, such as defensin-HNP3 analysed here. However, all these proteins show a similar disulphide-bridge topology and spatial arrangement of β strands when superimposed by KNOT-MATCH. This superimposition also allows aligning the N-terminal loop and CSαβ motif of scorpion venom toxin with equivalent structural elements of PCI, particularly regions S7-I15 and P11-S19 of each one (see Figure 3). Further analysis, both experimental and theoretical, should be done to define and rationalise the relationships between these proteins with different functions but with 3D structure similarities.

Biological relevance of the procedure

The capability of superimposing two unrelated proteins by their disulphide-bridge topology, independently of the sequential connectivity pattern, allows a versatile detection of similar spatial positioning of residues or secondary structures and has several implications in the area of structural and evolutionary comparison of proteins.

As shown above for the EGF-like family, the disulphide-bridge topology can be used to determine structurally and functionally relevant residues within a family of proteins. Sometimes, these relevant residues cannot be assessed by a sequential alignment because they become clearly overlapped only in 3D space. One example where the application of the program could help in the definition of a structural family is the so-called T-knot proteins (Isaacs, 1995). Two different current classifications of T-knot proteins have been defined. Lin & Nussinov (1995) defined the T-knot motive as an intrinsic linked pattern between cystine residues in a given protein sequence (1-4, 2-5 and 3-6). On the other hand, Harrison & Sternberg (1996) describe this topology as based on the disulphide β-cross topology, which is composed of two disulphide bridges and a β-sheet. Our approach is also capable of distinguishing between the T-knot proteins by exclusively using the topology of three disulphide bridges, without taking into account either the linkage pattern or the presence of a β-sheet. This feature can complement the definitions given by Lin & Nussinov (1995) and clarify relationships. For example, PCI and EGF do not have a common pattern of linked cystine residues, and only EGF contains a clearly defined β-sheet. Nevertheless, some loops of both proteins are similar and present several locations in the space with conserved residues, besides a common knot of disulphide bridges in the 3D space.

The case of PCI further shows that one of the potential merits of the proposed superimposition is that we can find unexpected relationships between non-homologous proteins. Thus, the disclosed overlaps in the 3D space suggested that PCI could have a biological function related to growth factors, which has further been demonstrated experimentally (Blanco-Aparicio *et al*., 1998). In fact, PCI is the first example of a human EGF antagonist found, and the superimposition of both molecules using KNOT-MATCH, clarified the mechanism of its antitumoral properties, based on its binding to the EGF receptor (Blanco-Aparicio *et al*., 1998). In general, our results also suggest that PCI could play a peptide hormone/growth factor-like role in plants, as do other small disulphide-rich proteins such as a plant insulin-like 38-residue peptide (Watanabe *et al*., 1994). The potential convergent/divergent evolution of the above molecules to acquire common structural determinants to bind cellular receptors is an open issue.

In general, the lack of sequence similarity between the proteins compared here (except for the comparison between the members of the EGF-like family) does not allow us to establish a clear evolutionary relationship between them. Most of the comparisons were done between remote homologues and analogues (Russell *et al*., 1997), finding structural similarities between them involving a small number of residues (not more than 10 or 20) or small regions. These results are in agreement with the analysis performed by Rost (1997), which shows that only 3-4% of all residues can be considered crucial for protein structure and function (anchor residues). It has been suggested that some of the disulphide-bridge topologies may have diverged from a common ancestor, such as the small and large α and β scorpion toxin fold (Harrison & Sternberg, 1996). However, many of the similarities in disulphide-bridge topology between proteins may be possibly explained by universal structural rules, preferences for close cysteine packing, or kinetics effects in the folding process (Harrison & Sternberg, 1996).

Testing whether the relationships found by our approach are due to topological restrictions, kinetics effects or protein convergent evolution, is a major challenge. It is possible that the detected superimpositions encode more structural and functional relationships that escape our analysis and interpretation, and that are likely to be further disclosed when improved algorithms and experimental data become available.

It is worth mentioning that many of the proteins used in this analysis (growth factors, PCI, defensins, toxins, pheromones, etc.) are targets for biotechnological applications. The knowledge of the common determinants of their 3D structure and function can be of great help in facilitating

their rational redesign. This is the case of PCI and EGF, where the engineering of PCI-like EGF antagonists with improved properties has a clear clinical interest. The superimposition by the disulphide-bridge topology using KNOT-MATCH can be a quite general approach to understand weak relationships between cystine-rich proteins and, consequently, help in the redesign and minimisation of them.

Acknowledgements

The authors gratefully acknowledge Dr T. Blundell and Dr R. Sowdhamini (University of Cambridge) and Dr M. J. E. Sternberg (Imperial Cancer Research Fund, London) for helpful suggestions and advice about the present research. This work has been supported by grants BIO95-0848, BIO97-0511, and SAF94-0989 from the CICYT (Ministerio de Educación y Ciencia, Spain) and by the CERBA (Centre de Referència en Biotecnología) de la Generalitat de Catalunya. J.M.M. is a pre-doctoral fellowship recipient of the CIRIT-CERBA (Generalitat de Catalunya).

References

Andrade, M. A. & Sander, C. (1997). Bioinformatics: from genome data to biological knowledge. *Curr. Opin. Biotechnol.* **8**, 675–683.

Avilés, F. X., Vendrell, J., Guash, A., Coll, M. & Huber, J. (1993). Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and the activation process. *Eur. J. Biochem.* **211**, 381–389.

Barbacci, E. G., Guarino, B. C., Stroh, J. S., Singleton, D. H., Rosnack, J. K. & Moyer, J. D. (1995). The structural basis for the specificity of epidermal growth factor and heregulin binding. *J. Biol. Chem.* **16**, 9585–9589.

Betz, S. F. (1993). Disulfide bonds and the stability of globular proteins. *Protein Sci.* **2**, 1551–1558.

Blanco-Aparicio, C., Molina, M. A., Fernandez-Salas, E., Fraizer, M. L., Mas, J. M., Querol, E., Avilés, F. X. & de Llorens, R. (1998). Potato carboxypeptidase inhibitor, a T-knot protein, is an epidermal growth factor antagonist that inhibits tumor cell growth. *J. Biol. Chem.* **273**, 12370–12377.

Brenner, S. E. (1995). BLAST, BLITZ, BLOCKS and BEAUTY: sequence comparison on the net. *Trends Genet.* **11**, 330–331.

Brown, L. R. & Wüthrich, K. (1992). Nuclear magnetic resonance solution structure of the α-neurotoxin from Black Mamba. *J. Mol. Biol.* **227**, 1118–1135.

Brown, S. C., Mueller, L. & Jeffs, P. W. (1989). ^{1}H assignment and secondary structure of human transforming growth factor α. *Biochemistry,* **28**, 593–599.

Chang, J., Canals, F., Shindler, P., Querol, E. & Avilés, F. X. (1994). The disulfide folding pathway of potato carboxypeptidase inhibitor. *J. Biol. Chem.* **269**, 33087–33094.

Chang, J., Shindler, P., Ramseier, U. & Lai, P. (1995). The disulfide folding pathway of human epidermal growth factor. *J. Biol. Chem.* **270**, 9207–9216.

Cornet, B., Bonmatin, J. M., Hetru, C., Hoffmann, J. A., Ptak, M. & Vovelle, F. (1995). Refined three-dimensional structure of insect defensin A. *Structure,* **3**, 435–448.

Creighton, T. E. (1992). Folding pathways determined using disulphide bonds. In *Protein Folding* (Creighton, T. E., ed.), pp. 301–351, W. H. Freeman & Company, New York, USA.

Ganzt, T. & Lehrer, R. I. (1998). Antimicrobial peptides of vertebrates. *Curr. Opin. Immun.* **10**, 41–44.

Groenen, L. G., Nice, E. C. & Burgess, A. W. (1994). Structure-fuction relationships for the EGFT/TGF-α family of mitogens. *Growth Factors,* **11**, 235–257.

Harrison, P. & Sternberg, M. J. E. (1996). The disulphide β-cross: from cysteine geometry and clustering to classification of small disulphide-rich protein folds. *J. Mol. Biol.* **264**, 603–623.

Hill, C. P., Yee, J., Selsted, M. E. & Eisenberg, D. (1991). Crystal structure of Defensin HNP-3, an amphiphilic dimer: mechanism of membrane permeabilization. *Science,* **251**, 1481–1485.

Hobohm, U. & Sander, C. (1994). Enlarged representative set of protein structures. *Protein Sci.* **3**, 522–524.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **223**, 123–138.

Isaacs, N. W. (1995). Cysteine knots. *Curr. Opin. Struct. Biol.* **5**, 391–395.

Jacobsen, N. E., Abadi, N., Sliwkowski, M. X., Reilly, D., Skelton, N. J. & Fairbrother, W. J. (1996). High-resolution structure of the EGF-like domain of heregulin-α. *Biochemistry,* **35**, 3402–3417.

Johnson, M. S., Srinivasan, N., Swdhanini, R. & Blundell, T. L. (1994). Knowledge-based protein modeling. *Crit. Rev. Biochem. Mol. Biol.* **29**, 1–68.

Lin, S. L. & Nussinov, R. (1995). A disulphide-reinforced structural scaffold shared by small proteins with diverse functions. *Nature Struct. Biol.* **2**, 835–837.

MacKinnon, R., Cohen, S. T., Kuo, A., Lee, A. & Chait, B. T. (1998). Structural conservation in prokaryotic and eukaryotic potassium channels. *Science,* **280**, 106–109.

Martí-Renom, M. A., Stote, R. H., Querol, E., Avilés, F. X. & Karplus, M. (1998). Refolding of potato carboxypeptidase inhibitor by molecular dynamics simulations with disulphide bond constraints. *J. Mol. Biol.* **284**, 145–172.

Matthews, B. W. & Rossmann, M. G. (1985). Comparison of protein structures. *Methods Enzymol.* **115**, 397–420.

Molina, M. A., Marino, C., Oliva, B., Avilés, F. X. & Querol, E. (1994). The C-tail valine is a key residue for the stabilization of the complex between potato inhibitor and carboxypeptidase A. *J. Biol. Chem.* **269**, 21467–21472.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.

Oliva, B., Wastlund, M., Cardenas, R., Querol, E., Avilés, F. X. & Tapia, O. (1991a). On the stability and fluctuations of the potato carboxypeptidase inhibitor fold: a molecular dynamics study. *Biochem. Biophys. Res. Commun.* **176**, 616–621.

Oliva, B., Wastlund, M. ,Nilssonn O., Cardenas, R., Avilés, F. X., Querol, E. & Tapia, O. (1991b). Aspects of model building applied to carboxypeptidase inhibitor protein: a molecular dynamics study

of a putative Pro36/Gly mutant. *Biochem. Biophys. Res. Commun.* **176**, 627–637.
Oliva, B., Daura, X., Nilsson, O., Querol, E., Avilés, F. X. & Tapia, O. (1995). Structure and fluctuation patterns of potato carboxypeptidase A inhibitor protein in aqueous solution. A molecular dynamics study. *Eur. Biophys. J.* **24**, 1–12.
Picot, D., Loll, P. J. & Garavito, R. M. (1994). The X-ray crystal structure of the membrane protein prostanglandin H2 synthase-1. *Nature,* **367**, 243–249.
Rees, D. C. & Lipscomb, W. N. (1982). Refined crystal structure of the potato carboxypeptidase inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J. Mol. Biol.* **160**, 475–498.
Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.
Rossmann, M. G. & Argos, P. (1976). Exploring structural homology of proteins. *J. Mol. Biol.* **105**, 75–95.
Rost, B. (1997). Protein structures sustain evolutionary drift. *Folding Design,* **2**, S19–20.
Roussel, A., Inisan, A. & Knoops-Mouthy, E. (1994). *TURBO FORBO (version 5.0a) Manual, BIOGRAPHICS*, Technopole de Chateaux-Gombert, Marseille, France.
Russell, R. B., Saqi, M. A. S., Sayle, R. A., Bates, P. A. & Sternberg, M. J. E. (1997). Recognition of analogous and homologous protein folds: analysis of sequence and structure conservation. *J. Mol. Biol.* **269**, 423–439.
Srinivasan, N., Bax, B., Blundell, T. L. & Parker, P. J. (1996). Structural aspect of the functional modules in C-alpha deduced from comparative analysis. *Proteins: Struct. Funct. Genet.* **26**, 217–235.
Taylor, W. R. & Orengo, C. (1989). Protein structure alignment. *J. Mol. Biol.* **208**, 1–22.
Thornton, J. M. (1981). Disulphide bridges in globular proteins. *J. Mol. Biol.* **151**, 261–287.
Ullner, M., Selander, M., Persson, E., Stenflo, J., Drakenberg, T. & Teleman, O. (1992). Three-dimensional structure of the apo form of the N-terminal EGF-Like module of blood coagulation factor X as determined by NMR spectrocopy an simulated folding. *Biochemistry,* **31**, 5974–5983.
Watanabe, Y., Barbashov, S. F., Komatsu, S., Hemmings, A. M., Miyagi, M., Tsunasawa, S. & Hirano, H. (1994). A peptide that stimulates phosphorylation of the plant insulin-binding protein. Isolation, primary structure and cDNA cloning. *Eur. J. Biochem.* **224**, 167–172.
White, S. H., Wimley, W. C. & Selsted, M. H. (1995). Structure, function, and membrane integration of defensins. *Curr. Opin. Struct. Biol.* **5**, 521–527.
Wu, J., Yang, Y. & Watson, J. T. (1998). Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry. *Protein Sci.* **7**, 1017–1028.

Edited by J. Thornton

(Received 9 February 1998; received in revised form 4 August 1998; accepted 19 August 1998)