Figure Legends

Figure 1

An example of a useful 'loose' similarity between 3D structure detected using the MAP method and a secondary structure prediction. a) The alignment found by the method between the predicted pattern for HNF–3 and the helical DNA binding motif within phage 434 repressor. Boxed, bold-faced, upper-case regions indicate aligned predicted and experimental secondary structures. Sec denotes the PHD prediction for HNF–3, and a 3-state DSSP secondary structure assignment for 434 repressor. Bur shows predicted and experimental states of burial for HNF–3 and 434 repressor: b = buried, e = exposed; u = intermediate/unknown. b) The equivalent alignment found using the STAMP (Russell & Barton, 1992) structure comparison algorithm. Boxed, bold-faced, upper-case regions indicate structural equivalences. Sec denotes DSSP 3–state secondary structures for both proteins. c) and d) show the crystallographic structures of the matched regions of HNF–3 and 434 repressor, with structurally equivalent residues shown in ribbon/coil format, and unequivalent regions shown as C_α trace. The N- and C- termini of the structures are labelled.

Figure 2

Search pattern for the von-Willebrand factor type A domain (derived from Edwards & Perkins, 1995) as discussed in the text. α helices are indicated by cylinders, β strands by arrows. The range of numbers given beside each secondary structure or loop are the range of predicted lengths. Bullets (•) show those secondary structure that are required for any possible map (i.e. those involved in distance restraints). Two distance restraints, one from a putative disulphide bond (9.5 Å)and the other from knowledge of two residues thought to be involve in metal coordination (15 Å) are shown to the left of the figure.

Figure 3

Maps from the top six scoring folds found during a search with the PID pattern. Details are given in the text.

Table 6

Results of running MAP using secondary structure assignments (I) and PHD secondary structure predictions (II) shown beside THREADER results (III) for eleven protein structures having type B and C similarities (Russell & Barton 1994) within the domain database. The first column for each method shows the top ten scoring domains, which are denoted by a PDB four letter code (Bernstein *et al.*, 1977), a chain identifier as the fifth character (if any), followed by an underscore and a Roman numeral denoting the domain (if any). Bold inverted text denotes a correct match using the strict classification, grey backgrounds show loose classifications (see text). The second column shows the score for each domain, the third the protein structure class, and the fourth the name of the fold/structure. Upper case denotes fold families under the strict definitions. Upper case names in parentheses (if present) denote the name of the loose family classification. The globins 1HBG, 1MYGA and 1ECA and the cupredoxin 1PAZ are sequence similar to the query so are not shown inverted and are not included in the evaluation statistics (see text).

Strict fold classifications: 4HB-1= Up-down-up-down four helix bundle (4HB); 4HB-2= up-up-down-down (interleukin-4 type) 4HB; GLOBIN= globin-type folds; W-HTH= winged helix-turn-helix (HTH) folds; EF-HAND= calcium binding EF hands; CYTOC= cytochromes C; THIO= thioredoxin-like folds; FLAVO= flavodoxin-like folds; ROSS= Rossman folds; PBL= periplasmic binding protein-like folds; ACTIN-ATPASE= actin/HSC-70/hexokinse like folds; G-PROT= G-protein (ras) like folds; FAD-BIND= FAD/NAD binding protein-like folds; $\alpha\ \beta$ -BARREL= $\alpha\ \beta$ (TIM) barrels; β -GRASP= β -grasp (ferredoxin) like folds; IG= Immunoglobulin superfamily; CUP= Cuppredoxins (plastocyanin-like); β -TREFOIL= β -trefoils (interleukin-1-β -like); OB-FOLD= oligonucleotide/oligosaccharride binding folds.

Loose fold classifications: 4HB= 4HB-1, 4HB-2, ferritin; HTH= W-HTH, λ-rep., trp-rep.; DWAB (doubly-wound-$\alpha\ \beta$)= ROSS, FLAVO, THIO, PBL, G-PROT, sugar phosphatase, pfk, pgk, dhfr; GKBS (greek key β sandwich)= IG, CUP, α -amylase inhibitor, sod, macromycin, prealbumin.

Other abbreviations: sod: superoxide dismutase; pfk= phosphrofuctokinase; pgk= phosphoglcerate kinase; dhfr= dihdrofolate reducatse; ldh= lactate dehydrogenase; ser-prot= serine proteinase; asp-prot= aspartic proteinase; inh.= inhibitor; rep.=repressor; glut.=glutathione; red.=reductase; thym. phosph.=thymidine phosphorylase; ribo.=ribonuclease; glyc.=glycoprotein; P-glucomutase= phosphoglucomutase; glyc. ribo trans.= glycinamide ribotransferase.

Table 7

Summary of fold recognition success rates. Strict and Loose refer to the critera for structural similarity discussed in the text. Class refers to structural class success as discussed in the text. (1st) refers to success measured as a correct fold at rank 1, (Top 10) as a correct fold in the top 10 ranked structures.