Molecular modelling of the sterol C-14 demethylase of Saccharomyces cerevisiae
Garrett M. Morris and W. Graham Richards
Oxford Centre for Molecular Sciences, Physical Chemistry Laboratory, South Parks Road, OX1 3QZ, U.K.

Enzymes in the cytochrome P-450 superfamily are the terminal oxidases in an NADPH-dependent mono-oxygenase system present in all eukaryotes and in some bacteria. Cytochromes P-450 are involved in drug metabolism, bioactivation of polyaromatic hydrocarbons to carcinogens, de-activation of xenobiotics, biosynthesis of bile acids and prostanolans, and steroidogenesis. In addition to these essential functions in mammalian metabolism, cytochromes P-450 are important in the catabolism of carbon sources, especially in yeasts and bacteria. Apart from their fundamental importance they are also major targets in both the pharmaceutical and agrochemical industries.

As the breadth of our experience grows in the field of protein crystallography it is becoming increasingly evident that the tertiary structures of homologous proteins are conserved more than amino acid sequence identities, and considerably more than their parent DNA sequences.

As more and more gene sequences pour into the scientific literature, the need for a rapid analysis and exploitation of these data also increases. Very often a protein sequence, readily translated from the its encoding gene, is found to have a similarity with another sequence in the currently rapidly expanding sequence databases. When a relationship is found between the new sequence and that of an evolutionarily-related protein with a known crystal structure, then a model of the three-dimensional structure of the new sequence can be built using modelling by homology. Unfortunately this sequence homology can be weak, and sequence alignments produced by the standard methods (e.g. ALIGN, based on the Needleman–Wunsch dynamic programming method, from the National Biomedical Research Foundation, Washington D.C.) frequently insert gaps into regions of known secondary structure or recognized conserved sequence motifs.

Usually when two sequences of amino acids are compared, they are treated as a matrix of pairwise similarity scores, with an arbitrary gap penalty assigned to non-aligned positions in the final alignment. These pairwise scores are derived simply from the sequence identity alone. However, when a globular protein folds in an aqueous environment, the sequence of events is thought to consist of an initial (and rapid) formation of secondary structural elements, with an overall entropic driving force compelling hydrophobic regions of this intermediate structure to aggregate to form the internal region of the protein. The importance of internal hydrogen bonding and disulphide bridges should not be overlooked in producing the final native fold. The point is that when two or more protein sequences are compared for structural similarity, a better alignment may be found by comparing their profiles not just their amino acid identities. These profiles portray important amino acid properties, such as hydrophobicity, antigenicity, secondary structure preference, B-value flexibility, residue bulk, refractivity, molecular mass, etc. A comparison matrix can therefore be built up from a subtraction of the profile values at every possible matrix position. Small values in the matrix represent very similar profile values, and thus runs of amino acids with similar profiles would show up as diagonals with near-zero elements. Thus a sequence alignment based on the comparison of information profiles can be deduced.

Abbreviations used: P-450 cam, camphor 5-exo-hydroxylase; P-450 sec, side-chain cleavage; P-450 l4sh, sterol C-14 demethylase.
If a known structure exists for one of the compared sequences, then penalties may be assigned for gap insertion within elements of secondary structure in that protein. This allows alignments to form which do not destroy the known fold by inserting into important elements of the fold.

Of the 160 or so P-450 isozymes currently sequenced, only one has been solved crystallographically: cytochrome P-450cam. The tertiary structure consists of 12 helical segments accounting for \( \sim 45\% \) of the residues; four antiparallel \( \beta \)-pairs but neither parallel \( \beta \)-structure nor extended sheet structure. Such a predominance of helical structure is a characteristic shared by most haem proteins for which X-ray structures are available. The eukaryotic P-450 isoenzymes are, however, membrane bound and being lipophilic are understandably difficult to crystallize for X-ray determination. Amphiphilic detergents recently assisted in crystallizing bovine cytochrome P-450resc, although its three-dimensional structure remains to be solved. Here, atomic co-ordinates from the high resolution, 1.63\( \AA \), crystal structure of cytochrome P-450 camphor 5-exo-hydroxylase of the soil bacterium \( \text{Pseudomonas putida} \) (P-450cam) due to Poulos and co-workers\(^1\) have been used to build a three-dimensional model of the cytochrome P-450 lanosterol 14a-demethylease from \( \text{Saccharomyces cerevisiae} \) (P-45014DM) using the amino acid sequence of Kalb et al.\(^2\).

The procedure followed is a synthesis of secondary structure prediction, global and local sequence alignment plus distance constraints, using attribute profiles of important amino acid physico-chemical properties, such as hydrophobicity, secondary structure propensities, molecular mass, bulk, refractive index, etc., to assist in detecting weak homology. Although the sequence identity is relatively low, use of the program CAMELEON (Oxford Molecular Ltd., Terrapin House, South Parks Road, Oxford OX1 3UB, U.K.) has enabled sufficient homology to be identified for a sequence alignment to be produced and a model to be built. The attribute profiles used included Kyte and Doolittle hydropathy\(^3\) and secondary structure propensities due to Chou and Fasman\(^4\) and Levitt\(^5\) although a choice of some 25 properties is available in CAMELEON.

There is a striking correlation between the inverse of the hydropathy profiles (averaging window length, \( 12 \leq L \leq 19 \)) of both P-450s, and the so-called 'distance profile'. This is calculated from the known tertiary stucture of cytochrome P-450cam, by defining its origin at the centre of the active site (here the C-5 carbon in camphor) and then calculating the distance to the centre of each of the side-chains in the fold. Each distance is plotted above the appropriate residue on-screen, scaled such that the range of ordinates for all residues is scaled between the sequence axis and the upper-y limit of the graphics window. This striking correlation implies that three-dimensional information appears to be encoded in hydropathy, hydrophobicity and hydrophilicity profiles, which also indicate a similar fold in both cytochromes P-450cam and P-45014DM. This reflects the notion that hydrophobic residues prefer to pack in the interior of a protein, while polar and charged residues generally prefer surface locations.

The hydropathy profile also reveals a putative transmembrane-spanning segment within the first 50 or so residues of P-45014DM which would help to explain why the fungal P-450 is membrane bound and why it is difficult to crystallize.

The model of \( \text{S. cerevisiae} \) P-45014DM was built using the following method. A format-conversion program was used to make the CAMELEON homology file readable by QUANTA, the general, all-purpose molecular modelling system. This homology file contains the amino acid sequence alignment of P-450cam and P-45014DM in particular the locations of deletions and insertions, and new amino acid labels. The homology file was used in QUANTA to prepare automatically the completely mutated model of 14DM, taking the backbone atomic co-ordinates of Poulos 2cpp.pdb crystal structure was conserved from P-450cam. The side-chains of P-450 are automatically replaced by those of the P-45014DM segments aligned with it. QUANTA also deletes appropriate sections and indicates points of insertion. This method of modelling by homology improves on an earlier methodology which involved single residue mutations carried out one-by-one, since (i) it is significantly faster in doing mutations; (ii) it accounts explicitly for any insertions or deletions; (iii) it generates a model of the entire target protein, not just the active site.

The deletions needed in the P-450cam structure were as follows I (79-87), II (127-157), and III (299-300). The insertions were: I (P-45014DM:1-88 before P-450cam:1), II (P-45014DM:277-294 between P-450cam:228 and P-450cam:229), III (P-45014DM:394-442 between P-450cam:329 and P-450cam:330), and IV (P-45014DM:528-530 after P-450cam:414). Of the insertions, II is predicted using secondary structure prediction to be a helix-extended strand unit which would prefer to
pack against the core of the protein: this is how it has been built; it does not contact the active site. The major insertion III consists of two extended strands and two helices, the strands being notably hydrophobic. There is evidence to suggest that when the part of the gene encoding for the putative transmembrane anchor is excised, the resulting mutant \( P-450_{14\alpha M} \) appears to be still membrane bound. Therefore it seems that the first two strands of insertion III are membrane spanning. The three-dimensional positions of insertions I and III are such that the substrate access channel would be pointing towards the membrane, with the proximal basin (which is encircled in \( P-450_{cam} \) by helices B, C, J, K, and L; an analogue of each of these helices exists in \( P-450_{14\alpha M} \) pointing in the opposite direction, accessible for electron transfer.

The \( P-450_{14\alpha M} \) model thus produced had bad contacts between certain amino acids within the structure, so it was necessary to carry out a molecular mechanics minimization using CHARMM [6], the force field used in the QUANTA package. The haem-iron was PATCHed into the porphyrin of the active site, thus allowing molecular mechanics minimization of the true active site to be carried out, not just a dihydrogenated porphyrin. Polar hydrogens were modelled explicitly.

\( P-450_{14\alpha M} \) shows significant homology in the C-terminal helical domain of the sequences, which is focused on the functionally important conserved amino acid motifs. The distal \( \alpha \)-helix I, \( \alpha \)-helices J and K known in \( P-450_{cam} \) show strong homology to a similarly hydrophobic region in \( P-450_{14\alpha M} \) predicted to be \( \alpha \)-helical. The distal helix houses a small binding pocket, with conserved motif GXXT. The all-important haem-binding domain with conserved motif FXXGXXXXCXG is also detected, within a region of conserved anti-parallel \( \beta \)-bulge and the long proximal \( \alpha \)-helix L.

Two deletions, I and III, far apart in the sequence of \( P-450_{cam} \), but adjacent in the tertiary structure, when regularized form a wide, largely hydrophobic access channel in \( P-450_{14\alpha M} \). There are two hydroxyl-containing residues and one acidic residue which fall in a line down one side of the access channel. These residues appear to be responsible for lining up the lanosterol-hydroxyl end of the substrate before binding in the correct orientation within the binding pocket.

The model of \( P-450_{14\alpha M} \) has a largely hydrophobic binding pocket which is indeed able to accommodate the substrate lanosterol, and benzo-\( [\alpha] \)-pyrene, which can also fit through the access channel. The presence of the hydrophobic access channel has provided an explanation for how ketoconazole, itraconazole and other similarly large demethylase inhibitors are able to bind to cytochrome \( P-450_{14\alpha M} \). Only the nitrogen-heterocyclic head of the demethylase inhibitor binds to the haem-iron, with most of the rest of the molecule occupying the access channel, and the end exposed to the solvent.

Current work is involved in automatically docking the following Public Domain demethylase inhibitors into the model of the active site of \( P-450_{14\alpha M} \): pachlobutrazol, triadimefon, diniconazole, propiconazole, fluotrimazole, flutriafol, triarimol, pyriflumizole. Including enantiomers, this amounts to 30 molecules. It is hoped that the model will be able to reproduce known trends in activity and enantiomeric differentiation.

Supported as a CASE award with I.C.I. Agrochemicals, Jealotts Hill, Berkshire. I would like to thank Dr Keith Heritage and Dr Brian Baldwin for their excellent assistance and advice.


Received 11 April 1991